

Lexical Coverage in L1 and L2 Listening Comprehension: The Same or Different from Reading Comprehension?

*HILDE VAN ZEELAND and **NORBERT SCHMITT

University of Nottingham, Nottingham, UK

*E-mail: aexhv@nottingham.ac.uk

**E-mail: Norbert.Schmitt@nottingham.ac.uk

Most research on L2 reading comprehension has indicated that 98 per cent lexical coverage provides adequate comprehension of written text. This figure has been transferred to listening comprehension and has been used to set vocabulary size targets for L2 learners. This study directly investigates coverage in regard to listening comprehension, to determine whether such transfer is reasonable. The coverage of four spoken informal narrative passages was manipulated, and participants' (36 native and 40 non-native speakers) listening comprehension of factual information was measured. Results showed that most native and non-native participants could adequately comprehend the spoken texts with only 90 per cent coverage, although the non-natives showed considerable variation at this level. At 95 per cent coverage, non-native participants also demonstrated relatively good comprehension, but with much less variation. Based on a 95 per cent coverage figure, language users would need to know between 2,000 and 3,000 word families for adequate listening comprehension, compared with Nation's (2006) calculation of 6,000–7,000 families based on a 98 per cent figure.

INTRODUCTION

Research has shown that it takes a relatively small group of frequent words in a language to cover a large proportion of the words in written and spoken discourse (Nation 2001). Of particular interest to the field of second language pedagogy is the percentage of words in written or spoken discourse which enables successful comprehension. This percentage of known words in a piece of discourse is referred to as *lexical coverage* (Adolphs and Schmitt 2003) or sometimes *text coverage* (Nation and Waring 1997). Lexical coverage is an essential measure, for it allows the calculation of estimates of the vocabulary size necessary for comprehension of written and spoken texts.

Most research on lexical coverage in relation to L2 comprehension has been conducted on reading. We have a fairly good idea of the percentage of vocabulary that needs to be known to allow comprehension of written text, as well as how many word families are required to reach this percentage. Unfortunately, very few studies have focused on vocabulary knowledge and listening

comprehension (in either the L1 or the L2), and so we lack such information about spoken language. Researchers have therefore applied the results from reading studies to listening, with the underlying assumption that the relationship between lexical coverage and comprehension is similar in the two modalities. However, a number of studies have indicated that listening comprehension involves different processes than reading comprehension (e.g. Lund 1991), and that vocabulary knowledge plays a different role in the two modalities (e.g. Mecartty 2000; Stæhr 2008). Such findings raise the question of whether coverage figures derived from reading research can indeed be transferred to listening. This study seeks to resolve this issue by investigating the relationship between lexical coverage and both L1 and L2 listening comprehension and determining what implications this has for vocabulary size learning targets for spoken discourse.

BACKGROUND

Lexical coverage in L1 reading comprehension

Research has established that vocabulary knowledge and L1 reading comprehension are closely related (Anderson and Freebody 1981; Stahl 1990). Unfortunately, there is virtually no direct research on the relationship between lexical coverage and comprehension in L1 contexts. The only relevant study we found is Carver (1994). He explored the relationship between the relative difficulty and the amount of unknown vocabulary in written text. Relative difficulty was calculated as a text's difficulty level (as ascertained by two readability measures) minus the reader's ability level (as determined by a 100-word vocabulary test), in grade equivalent units. The amount of unknown vocabulary was obtained by readers underlining words they had not seen before or did not know the meaning of. The study involved two different text types (fictional versus factual, sampled from library books and school curriculum materials) and two participant groups of different proficiency levels. It was found that the percentage of unknown words decreases as the relative easiness of the text increases. Generally, easy texts contained around 0 per cent unknown words, difficult texts around 2 per cent or more unknown words, and texts that were of a difficulty level appropriate to the learner around 1 per cent unknown words. Although Carver does not directly relate these findings to lexical coverage, he suggests that 99 per cent is the appropriate coverage level for written text. However, it should be emphasized that Carver relates word knowledge to text difficulty and not to text comprehension.

Lexical coverage in L2 reading comprehension

In contrast, there has been much more research in the L2 context. A strong relationship has been demonstrated between receptive vocabulary knowledge and L2 reading comprehension, with correlations ranging from 0.40 to 0.85

(Koda 1989; Qian 1999, 2002; Albrechtsen *et al.* 2004; Stæhr 2008; Laufer and Ravenhorst-Kalovski 2010; Schmitt *et al.* 2011). The relationship between lexical coverage and reading comprehension has also been addressed, with the earliest research by Laufer (1989) suggesting that 95 per cent lexical coverage was needed to allow adequate L2 reading comprehension. A score of 55 per cent on her reading test was considered adequate comprehension, and at the 95 per cent coverage level, there were significantly more participants with a score above than below 55 per cent.

In a later study, Hu and Nation (2000) suggested that readers needed knowledge of 98 per cent of the words in a fiction text to achieve successful comprehension. In this study, the coverage of a fiction text was manipulated by means of replacing words with non-words. Four different coverage degrees were created and L2 readers' comprehension of these texts was tested. At the 80 per cent coverage level, no learner achieved adequate comprehension (as defined by a score of 70 out of the maximum of 124), and at 100 per cent the majority did. Relatively few gained adequate comprehension of the texts with 90 and 95 per cent coverage. Hu and Nation thus concluded that the required coverage was somewhere between 95 and 100 per cent and suggested a figure of 98 per cent.

The two above studies attempted to identify a coverage 'threshold' at which adequate comprehension of written texts can be achieved. Schmitt *et al.* (2011) took a different approach, investigating each percentage point of coverage between 90 and 100 per cent, in an attempt to describe the overall relationship 'curve' between coverage and comprehension. This revealed a linear relationship between the two, which suggests that the coverage level required depends on the degree of comprehension aimed for. Based on the data, if 60 per cent comprehension is the goal, 98 per cent lexical coverage is needed.

Laufer and Ravenhorst-Kalovski (2010) support this argument of basing the required coverage level on the comprehension wished for. The authors suggest that two lexical coverage thresholds can be identified, depending on the definition of 'adequate' comprehension: 98 per cent as the optimal and 95 per cent as the minimal. Based on the performance of their Israeli participants, the authors conclude that 95 per cent coverage enables acceptable comprehension and that 98 per cent coverage leads to successful comprehension by most learners.

These lexical coverage figures are crucial, because they allow a calculation of the vocabulary size which speakers need in order to use language. For example, Nation (2006) calculated that it takes a vocabulary size of 8,000–9,000 word families to reach a 98 per cent coverage level in written text. Similarly, Laufer and Ravenhorst-Kalovski (2010) calculated that it takes around 8,000 word families to yield a coverage of 98 per cent (including proper nouns), but only about 4,000–5,000 to reach a 95 per cent level.

Lexical coverage in L2 listening comprehension

Although the lexical coverage and vocabulary size requirements seem relatively established for L2 reading, this is far from the case for L2 listening.

We could find only two published studies focusing on this area. Bonk (2000) was the first to address, albeit indirectly, the issue of lexical coverage in relation to L2 listening comprehension. In his study, participants ($N=59$) listened to four short passages with increasing percentages of low-frequency words. Their lexical knowledge was measured with a dictation test and their comprehension with a written free-recall test. It was found that higher dictation scores generally meant better listening comprehension, with a modest correlation of 0.45 being reported. A lexical threshold was not found, as there was a considerable amount of variation in comprehension scores. In general, however, learners with lexical coverage scores below 90 per cent showed poor comprehension and those with lexical coverage scores above 95 per cent showed good comprehension.¹

However, there are issues with Bonk's instruments that call his findings into question. First, knowledge of words in the listening passages was measured by a dictation test, and it is unclear to what extent such an 'integrative' assessment method (drawing on many facets of linguistic knowledge) directly measures vocabulary knowledge. Secondly, to measure comprehension, the participants were asked to write down everything they recalled from the listening passages, and these accounts were then rated on a four-point scale. This introduced a level of subjectivity to the marking. More importantly, the scale was a relatively blunt measure, especially as 'the construction of the rating scale was guided by a desire to, as much as possible, separate those with ratings of 1–2 (Inferior comprehension) from those of 3–4 (Good comprehension)' (p. 21). These issues make it difficult to draw any strong conclusions about the precise relationship between lexical coverage and listening comprehension from Bonk's study.

A recent study by Stæhr (2009) looked at the effects of vocabulary size and depth on listening comprehension. Participants' ($N=115$) comprehension was assessed with a standardized test from the Cambridge Certificate of Proficiency in English. Their vocabulary size was measured with the Vocabulary Levels Test (VLT), and their depth of vocabulary knowledge with a slightly adapted version of the Word Associates Test. Stæhr found that both size and depth correlated significantly with listening comprehension (0.70 and 0.65, respectively). In addition, he indirectly explored the lexical coverage/listening comprehension relationship. Although he did not have a lexical coverage measure per se, he estimated it by submitting his listening passages to a *Vocabulary Profiler* analysis on the *Compleat Lexical Tutor* website (Cobb n.d.), and then matching the frequency profiles to the participants' VLT scores. For example, participants who mastered² the VLT 3,000 level were assumed to have the vocabulary to achieve about 94 per cent lexical coverage of the listening passage (i.e. the 94% coverage consisted of words from the first three 1,000 levels), and those participants produced a mean of nearly 60 per cent on the listening test. Participants who mastered the 5,000 level (with an assumed 98 per cent lexical coverage) achieved a mean listening score of 73 per cent. Although indicative, Stæhr's indirect method of measuring lexical coverage

makes it hard to come to any strong conclusions concerning the lexical coverage/listening comprehension relationship.

THE ROLE OF VOCABULARY KNOWLEDGE IN L2 READING AND LISTENING COMPREHENSION

It is clear that the two studies discussed above have not provided an empirical basis on which to establish a lexical coverage percentage which allows adequate listening comprehension. This is problematic, because in the absence of an empirically derived figure, the tendency has been to simply use the lexical coverage figures from reading research and assume that they also work for listening comprehension. However, there is research which indicates that the influence of vocabulary knowledge differs in the two modalities, at least in the L2. Mecartty (2000), for example, studied lexical and grammatical knowledge in both reading and listening comprehension and found that vocabulary knowledge explained a larger amount of variance in reading than in listening. Similarly, Stæhr (2008) found that vocabulary size correlated more strongly with reading and writing abilities (0.83 and 0.73) than the skill of listening (0.69). A recent study by Mehrpour and Rahimi (2010) investigated the effects of both general and text-specific vocabulary knowledge on reading and listening comprehension. It appeared that learners' general vocabulary knowledge influenced their reading comprehension, but not their listening comprehension. In contrast, familiarity with text-specific vocabulary (tested by giving half of the participants a glossary with keyword definitions) had a significant effect on both reading and listening comprehension scores, but with a stronger effect on reading performance.

Why would the role of vocabulary knowledge be smaller in L2 listening than L2 reading? It may be related to differences between processing spoken and written text. Researchers have found, for example, that readers rely more on linguistic information from the text (i.e. vocabulary) than listeners do (Reves and Levine 1988), while listeners rely more on top-down processing than readers (Lund 1991; Park 2004). This suggests that listening comprehension may be largely based on factors such as world knowledge and topic familiarity. Such top-down information is believed to be compensatory in use, in the sense that it is employed strategically by listeners to compensate for inadequate knowledge of the L2 or an inability to recognize words in continuous speech (Field 2004; Vandergrift 2011). In addition, listening achievement can be partly explained by learners' metacognitive knowledge (i.e. their awareness of listening processes and ability to regulate them) (Vandergrift *et al.* 2006). In-depth discussions of the various factors known to affect listening can be found in Field (2008), Rost (2002), and Vandergrift (2011).

Another factor, related to the processing differences, may be the characteristics of spoken versus written language. Listeners may focus less on vocabulary because of the fleeting nature of spoken discourse: whereas written text

has a permanent character which allows the reader to refer back to lexical items to decode the message, this cannot be done in listening (Reves and Levine 1988). In fact, this requirement of quick, online processing is what makes listening a great challenge to many learners. Online word segmentation has been identified by learners as the main obstacle in L2 listening comprehension (Goh 2000). This is due not only to the speed of spoken discourse but also to linguistic features such as elision, reduction, assimilation, and cliticization (Field 2008), as well as to learners' use of L1 phonotactic conventions (Weber and Cutler 2006). In addition, learners often experience difficulties recognizing individual lexical forms correctly (Cutler 2005).

However, contrary to these inhibiting factors, there are a number of characteristics of spoken language that may compensate for gaps in linguistic knowledge. One such characteristic is prosody. Features such as sentence stress and intonation affect listeners' interpretation of texts and simultaneously help them in guessing the meaning of unknown words (Buck 2001). Another characteristic is non-verbal information. Communication devices such as gestures and facial expressions aid L2 listening comprehension (Harris 2003). Still another aspect of spoken language is that it tends to be lexically less dense than written text (McCarthy and Carter 1997), containing many fillers, interactive markers, and repetitions (Rost 2002).

The various facilitating factors involved in listening, such as top-down processing, prosody, and (where provided) visual input, suggest that a lower lexical coverage might be needed for learners to comprehend spoken discourse than to comprehend written discourse (Adolphs and Schmitt 2003). This line of thinking is in agreement with the research finding that the influence of vocabulary knowledge is smaller in listening than in reading comprehension. However, the time constraints involved in speech processing and the difficulties with spoken word recognition suggest that a higher degree of lexical coverage might be required. Clearly, this tension makes it difficult to predict the relationship between lexical coverage and listening comprehension (particularly in the L2).

Vocabulary size targets for L2 listening

Considering the differences between reading and listening discussed above, it would be imprudent to assume that the relationship between vocabulary and listening comprehension can be derived in a straightforward manner from research exploring the relationship between vocabulary and written comprehension.

Nevertheless, this is exactly what has been done. In an influential paper, Nation (2006) used Hu and Nation's (2000) 98 per cent figure based on reading comprehension to calculate the vocabulary size requirements for operating in a spoken environment. He used the Wellington Corpus of Spoken English (WCSE) and BNC word frequency lists to analyze the vocabulary of spoken English, finding that it took 6,000–7,000 word families plus proper nouns to

reach nearly 98 per cent coverage. (For comparison, he found that it takes a vocabulary size of 8,000–9,000 word families to reach a 98 per cent coverage level in written text.)

Furthermore, two recent studies by Webb and Rodgers (2009a) set vocabulary size targets based on 95 and 98 per cent coverage figures which have come from reading research. They analyzed the language of 88 television programs and found that knowledge of the most frequent 3,000 word families (plus proper nouns and marginal words) provided 95.45 per cent coverage. (This ranged from 2,000 to 4,000 word families in different TV genres). Knowledge of the most frequent 7,000 word families provided 98.27 per cent coverage (range 5,000–9,000 word families). They also analyzed 318 film scripts (Webb and Rodgers 2009b) with the use of the BNC frequency lists and found that the most frequent 3,000 word families provided 95.76 per cent coverage (the range was 3,000–4,000 word families depending on the movie genre). Knowledge of the most frequent 6,000 word families provided 98.15 per cent coverage (range 5,000–10,000 word families). But surely listeners use the visual support to interpret word meaning from context when watching television or movies, suggesting that knowledge of a smaller number of words may be required for comprehension of these media.

Are Nation's and Webb and Rodgers' vocabulary size requirements valid? As they are directly contingent upon the lexical coverage figure for their calculation, it is impossible to know unless one can be sure about the underlying coverage figure. Given the frequency distribution of vocabulary, vastly different vocabulary size requirements accrue from what might appear to be relatively trivial lexical coverage differences. However, the size requirements are anything but trivial; the difference between teaching and learning 3,000 word families (based on Webb and Rodgers' 95 per cent findings) and 6,000–7,000 word families (based on Nation's 98 per cent calculations) is vast and may require quite different teaching approaches and investments of time and effort. This makes it essential to come to a firmer determination of the lexical coverage required for listening comprehension, in order to better establish appropriate vocabulary size targets for operating in spoken contexts. This study seeks to be a first step in establishing the lexical coverage requirements empirically by directly measuring and comparing lexical coverage and listening comprehension of one type of oral discourse (i.e. informal narratives).

METHODOLOGY

Aim

This research follows the methodology used by Hu and Nation (2000), involving the insertion of non-words into listening passages to achieve various levels of lexical coverage, and measuring participants' comprehension of these passages. The study involves two separate participant groups: one of

native speakers and one of non-native speakers. The following two research questions will be addressed:

- 1 What is the relationship between lexical coverage and listening comprehension among both native and non-native speakers of English?
- 2 Is the degree of lexical coverage necessary for L2 listening comprehension the same as that necessary for L2 reading comprehension?

Implications for vocabulary size targets will then be discussed in light of the answers to the above questions.

Participants

In total, 76 individuals participated in this study: 36 native speakers and 40 non-native speakers of English. The majority of native participants ($N=32$) were first-year undergraduate students at a British university, and four others were postgraduate students. Almost all participants were female ($N=34$, 94 per cent) and the average age was 20.25 years ($SD=3.19$). Of the non-native participants, the majority was female ($N=31$, 77.5 per cent) and just over half had Dutch as their L1 ($N=21$). The rest of the participants came from a variety of L1 backgrounds. The average age was 32 years, but ages varied considerably ($SD=12.77$). Most of them ($N=28$) were advanced learners of English who studied a postgraduate university course, the majority in England ($N=18$) and the remainder in the Netherlands ($N=10$). The courses of the latter group involved regular exposure to English in the classroom and study materials. Those participants who were not students ($N=12$) were generally of high-intermediate level who were exposed to English on a regular basis in their work and through the media. Although not all participants in this group were advanced, they were very well capable of using the English language. All non-native participants are therefore estimated to be of proficiency levels B2 or C1 on the Common European Framework of Reference for Languages scale. They were also asked how many hours a week they had contact with English, in which 37.5 per cent estimated 1–10 h a week, 22.5 per cent 11–20 h a week, another 22.5 per cent 21–30 h a week, and 17.5 per cent over 30 h a week. This indicates that most of the participants, even those who did not study a postgraduate course, had considerable engagement with English on a weekly basis. Clearly, the subject pool was quite varied overall, but was considered acceptable, as the aim of this study was to relate lexical coverage to general L2 listening comprehension without focusing on any specific L2 learner group or L2 proficiency level.

Listening passages

In comparing lexical coverage in reading and listening comprehension, one approach would be to use the same texts and comprehension questions in both modalities. In this study, for example, texts used in previous reading research

might be recorded in spoken form and presented to L2 listeners with their original comprehension questions. In this way, the texts and comprehension tests would be exactly the same and any differences found in comprehension related to coverage would be related to the modality (reading versus listening) without any interference of textual qualities. Yet this approach would overlook the inherent differences between written versus spoken language as previously discussed and would lead to questionable results. We, therefore, decided to employ authentic spoken passages in our study.

The four passages used were short stories told in the first person found on the internet. They were all anecdotes of people getting into unusual situations and thus similar, and also thought to be interesting for participants to listen to. These passages of storytelling were chosen because they reflect a common type of spoken discourse that natives or non-natives often listen to. All stories were told as naturally as possible, in the way that a speaker would tell a story to a friend. Online Supplementary Material, Appendix 1 provides one story as an example. It shows the passages contain the usual connected speech patterns (*I'd, we're*), as well as other spoken language features such as fillers and interactive markers (*you know, well*), repetition, and redundancy (*... I stayed incredibly calm. I was probably frozen with fear but I just remained calm*), additive ordering (most sentences start with *And...*, or *So...*), and topic-comment structures (*I saw what looked like a smoke stack. You know, one of the towers looked like that*) (Rost 2002). The speaker also used prosodic devices to express his attitude (*I thought that was INCREDIBLY sweet*) and indicate salient parts (*I later found out SHE was doing the EXACT SAME THING that I was*). Participants listened to the story without any source of visual input.

As the purpose of this study was to explore the degree to which lexical coverage affects comprehension, it was important that the four stories were of the same level of difficulty so that difficulty would not be a variable. In order to ensure this, 10 native speakers of English were asked to read all four stories and rate the degree of difficulty of each story on a scale from 1 to 10. In rating 'difficulty', they were asked to consider the content of the passages, that is the story topic, amount of background knowledge necessary, and amount of detail. Scale number 1 was defined as 'completely incomprehensible; I cannot understand the overall story at all and I couldn't catch any of the details either', and number 10 as 'completely comprehensible; I can understand the overall story and all of the details without any problems'. The native speakers were also asked if they felt some stories were more difficult than others, and if so, if they could rank the stories in order of difficulty. In the rating task, the means of the four stories' difficulty ratings were all within 0.4 (8.6, 8.6, 8.8, and 9.0). In the ranking task, 7 out of 10 participants chose not to rank the stories because they considered them of equal difficulty. There was no consistent pattern of difficulty in the rankings of the three participants who did the ranking task. In order to ensure that the difficulty of the four stories was also equal in the listening mode, five native speakers of English were asked to listen to recordings of the stories and to do the same rating and

ranking task. Again, mean scores of the difficulty level were very similar (8.7, 8.5, 8.5, and 8.8), and none of the five native speakers chose to rank the stories. Although it can never be guaranteed that four separate texts are absolutely identical in difficulty, the survey results demonstrated that the four stories were very similar. The linguistic difficulty of the four texts was also compared with the Flesch–Kincaid readability formula, which found the four stories had a reading ease of 78.2, 83.4, 89.6, and 94.2, and a grade level of 6.4, 5.6, 4.3, and 3.2. This shows that all four texts were easy to comprehend, with a slight difference in linguistic difficulty. Clearly, text difficulty is much more than just linguistic complexity, and we believe that the content-based survey with the native speakers provides a better idea of the actual difficulty as experienced by the participants. The issue of text difficulty was thus largely controlled for.

Next, the vocabulary of the four stories was manipulated by means of replacing a number of words with non-words, so that each story had a particular degree of lexical coverage: 100, 98, 95, or 90 per cent (i.e. one story was 100 and another was 98 per cent). The frequency levels of the words in the four passages were identified using the *Vocabulary Profiler (BNC-20 version)*, available on the *Compleat Lexical Tutor* website (Cobb n.d.). Words not within the 2,000 word frequency level were replaced by non-words. In the 90 per cent coverage story, 10 per cent of all word tokens were replaced by non-words; in the 95 per cent coverage story, this was 5 per cent; and in the 98 per cent coverage story, this was 2 per cent. The 100 per cent story contained no non-words. All the passages, including the 100 per cent story, were then checked to ensure that all words were within the 2,000 frequency level.³ The type-token ratios indicate similar lexical variety in the four texts (Table 1), which also points to their similarity.

The *ARC Nonword Database* was used to create the non-words (Rastle *et al.* 2002). All non-words entered into the passages conformed to the phonologic and orthographic constraints of English and were five to eight phonemes long. The non-words sounded like the word class they replaced, so that participants knew what part of speech a non-word was. This is illustrated by the following extracts from the 95 per cent story: ‘we were both on *spacks* [flights] at New York airport’, ‘some people of course *splaffed* [panicked] and were *thwarfing* [crawling] around’, and ‘as fast as my *twacsy* [chubby] legs could run out of there’.

One short sentence was added at the beginning of each story to introduce the narrator. This would make it clear to the participants who were referred to in the comprehension questions. These sentences were short and did not change the content or coverage levels of the four stories. To control for personal speaker characteristics such as pronunciation, accent, and speaking style, the four stories were told by the same speaker (a native British English male). The final stories were all between 470 and 485 words long and the recordings were all between 1.58 and 2.08 min long (Table 1).

Table 1: The number of words and non-words in the four stories

Lexical coverage (%)	Number of words in text	Number of non-words in text	Type-token ratio	Recording time (min.s)
100	484	0	36	1.59
98	475	9	38	1.58
95	472	23	43	2.02
90	471	47	36	2.08

Listening comprehension test

Participants' comprehension of each story was measured with a multiple choice test. In order to decide which information from the stories should be covered by the tests, three native speakers of English were asked to read the four stories (without the non-word substitutions) and underline idea units that they thought were important for comprehension. For each story, six multiple choice items were then written based on idea units that had been underlined by all three native speakers. This made sure that the multiple choice tests covered the most important idea units. Another nine questions were written for each passage, which focused more on the stories' details. This made a total of 15 questions for each story. Each item had four options, as well as an 'I don't know' option. All items on the tests were written using vocabulary within the 2,000 frequency level.⁴

After the 15 items for each story were written, they were piloted in a listening test with 10 native and 10 non-native speakers. The purpose of the piloting was to ensure that the items worked well in the spoken mode and were answerable given the information in the listening passages. Participants therefore listened to recordings of the stories without non-words. The 10 best items for each test were retained, considering both the pilot results and the range of story content covered.⁵ Participants' scores on these 10 items showed that the total number of questions answered incorrectly was very low: for all native speakers together, in each of the four tests, the number of incorrect answers was not higher than 2 (10 participants $\times$ 10 questions = 100 answers, of which only 2 incorrect). For all non-native speakers together, in each of the four tests, the number of incorrect answers was not higher than 5 (10 participants $\times$ 10 questions = 100 answers, of which only 5 were incorrect). This shows that the final 40 items match the four stories: if participants have full knowledge of the words in the passages, they should be able to answer the questions correctly. As an example, the listening comprehension test of the 95% coverage story can be found in online Supplementary Material, Appendix 2.

Vocabulary test

As all words in the stories, except the non-words, fell within the 2,000 frequency band, it was crucial that participants mastered the 2,000 level so that the four coverage levels were accurate. The revised version of Nation's (1983) VLT was used to test this (Schmitt *et al.* 2001). The complete VLT consists of five parts, but in this study only knowledge of the 2,000 vocabulary level was tested.

In the VLT, participants were asked to match 30 words within the 2,000 level to their correct definition. Participants who did not reach a score of at least 29 out of 30 were excluded from the study. Of the 40 participants, 65 per cent achieved the maximum score of 30, and 35 per cent gained a score of 29. Having one mistake on the whole VLT test suggests a participant could be unfamiliar with up to a maximum of 3.3 per cent of the words in the 2,000 level (i.e. one error from 30 target words). If we had accepted a higher error rate, participants might have been unfamiliar with many more words at the 2,000 level. A possible effect of this would have been that our participants may not have known the percentage of words in the listening passages which we claim at each of our coverage levels. It is interesting to note that previous studies have not been as rigorous in this regard. In the study by Hu and Nation (2000), a minimum score for the participants to be included was 14 out of 18 in the 2,000 test and 10 out of 18 in the 3,000 test. This leaves potentially 22.2 per cent of the words in the 2,000 level unknown, meaning that the participants in this study were unlikely to have actually reached the reported coverage levels. The same flaw makes Stæhr's (2009) coverage levels somewhat questionable: 27 out of 30 on the VLT was the minimum for participants to show they mastered a specific vocabulary level. The high VLT minimum used in our study made it likely that the participants actually knew the reported coverage levels in the stories.

Procedure

Participants were told that the study was about the effect of known and unknown words on listening comprehension and were forewarned that some of the stories contained non-words. If no warning was given, the non-words would likely distract participants while listening and interrupt their processing of the stories. Participants then completed the VLT test, followed by the listening comprehension test. The whole experiment took approximately 45 min.

In the listening test, all participants listened to all four stories. Each story was played twice. Listening only once would likely lead to incorrect answers due to memory effects rather than to poor comprehension. As the aim of this study was to measure the effect of lexical coverage and not memory, listening twice was considered more appropriate. When the stories were played for the first time, participants were not allowed to look at the comprehension questions yet. As soon as the recording had been played once, participants were asked to

read and answer the comprehension questions. When participants listened to the same story for a second time, they could change or add any answers in the test. This procedure was the same for all four stories. The stories were counter-balanced among different participants to avoid any order, fatigue, and/or practice effects. There were two order groups. To make sure the counterbalancing worked, Mann–Whitney U tests were performed to compare the results on each comprehension test (i.e. on each of the four coverage levels) between the two order groups. In the native speaker group, scores on all four tests were comparable for the two groups. In the non-native speaker group, scores on only the 95 per cent story were found to differ between the two groups ($U=121.500$, $z=-2.193$, and $p=0.028$). This shows that, overall, the order in which participants listened to the four stories did not affect their comprehension.

RESULTS

Native speakers

An overview of the comprehension scores achieved by the native speaker participants can be found in Table 2. As can be seen, with 100 per cent coverage almost all participants (33 out of 36) achieved the maximum score of 10. Three participants scored 9, leading to a final mean of 9.92. We can thus conclude that native speakers who know all the words in a short spoken narrative are generally able to fully comprehend the factual information in it. With 2 per cent unknown words, one-third of the participants reached the maximum comprehension score of 10 (13/36), and about three-quarters achieved a score of either 9 or 10 (28/36). Overall, the comprehension remained relatively high, although the mean comprehension score dropped to 8.97. An increase in standard deviation is visible here (as well as in the 95 and 90 per cent levels), indicating a larger dispersion than in the 100 per cent story test. With 95 per cent coverage, the mean score is 8.19, and with 90 per cent, this is 8.53. Many of the scores at these coverage levels dip down into the lower figures (6–8). However, in absolute terms, the scores are still high.

In order to fully understand the relationship between coverage and comprehension, it is important to also analyze the relative comprehension at the different coverage levels. As Shapiro–Wilk tests showed that the data were not normally distributed, non-parametric tests were used. An independent-samples Kruskal–Wallis test showed that there was a significant difference in comprehension of the four stories, [$\chi^2(3)=57.0$ and $p<0.01$]. Wilcoxon signed ranks tests revealed that the significant differences existed between the 100 per cent story and the other three stories ($p<0.001$) as well as between the 98 and 95 per cent story ($p<0.01$) (Figure 1). Therefore, 100 per cent coverage leads to the best listening comprehension, then 98 per cent, but there was no difference in comprehension between 95 and 90 per cent coverage.

Table 2: Comprehension scores on the four tests by native speakers (N=36)

Comprehension score	Coverage			
	100%	98%	95%	90%
5	0	1	0	0
6	0	0	1	4
7	0	3	9	2
8	0	4	10	10
9	3	15	14	11
10	33	13	2	9
Mean score	9.92	8.97	8.19	8.53
Standard deviation	0.28	1.13	0.98	1.25

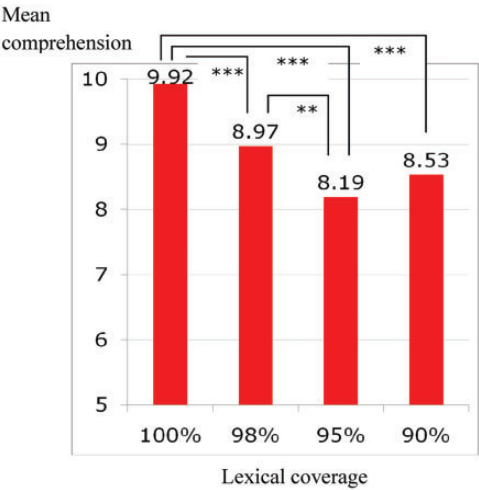


Figure 1: Mean comprehension scores by native speakers on the four coverage levels. *** $p < 0.001$; ** $p < 0.01$

Non-native speakers

Table 3 provides an overview of the numbers of non-native participants reaching various comprehension scores on the four tests. In the 100 per cent coverage story, the majority (29/40, 73 per cent) of the participants achieved a maximum score of 10. It can be seen that the rest of the participants also gained high scores (mean=9.62). This indicates that 100 per cent lexical knowledge enabled most participants to fully comprehend the factual

Table 3: Comprehension scores on the four tests by non-native speakers (N = 40)

Comprehension score	Coverage			
	100%	98%	95%	90%
3	0	0	0	5
4	0	0	0	0
5	0	1	1	2
6	0	3	9	3
7	1	7	6	9
8	2	12	14	6
9	8	9	7	10
10	29	8	3	5
Mean score	9.62	8.22	7.65	7.35
Standard deviation	0.70	1.31	1.29	2.13

information in the passage, yet it did not guarantee this. Then, for the 98 per cent coverage story, scores were lower: only 8 out of 40 participants (20 per cent) were able to reach the maximum score of 10, and the mean comprehension score was 8.22. In addition, we see more variation in comprehension scores, suggesting that the comprehension ability of individual L2 listeners is affected differently by a decrease in coverage. Interestingly, among the native participants, there was no such increase in variation as coverage decreased: the standard deviations on the different coverage levels were quite similar. Next, the stories with 95 and 90 per cent coverage show a further decrease in average comprehension scores to 7.65 and 7.35, respectively. Although lower than the native scores, this suggests that the L2 participants could still comprehend the passages quite well if as much as 5 or even 10 per cent of the words were unknown. For example, if one accepts that a score of 7 out of 10 reflects adequate comprehension, 75 per cent of the L2 participants could still adequately comprehend the passage with 95 or even 90 per cent lexical coverage. With a stricter criterion of 8 out of 10, more than 50 per cent still obtained adequate comprehension at both coverage levels.

The Kruskal–Wallis test indicated that the difference in comprehension of the four stories was significant [$\chi^2(3) = 52.4$ and $p < 0.01$]. Wilcoxon signed ranks tests showed that comprehension of the 100 per cent story was significantly better than that of the stories with 98, 95, and 90 per cent coverage ($p < 0.001$). Comprehension of the 98 per cent coverage story was significantly better than that of the 95 and 90 per cent coverage stories ($p < 0.01$), but there was no significant difference between comprehension of the 95 and 90 per cent stories (Figure 2).

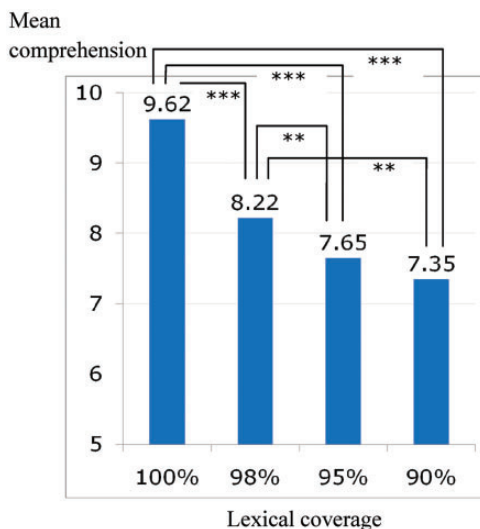


Figure 2: Mean comprehension scores by non-native speakers on the four coverage levels. *** $p < 0.001$; ** $p < 0.01$

Although the difference in comprehension of the 95 and 90 per cent stories is not significant, the two levels of coverage seem to affect comprehension differently. As can be seen in Table 3 and Figure 3, the standard deviation of the scores on the 90 per cent story is larger than that of the 98 per cent story. This can also be seen in the larger range of scores: at the 95 per cent coverage level, the scores range from 5 to 10, while at the 90 per cent level, the scores range from 3 to 10. This indicates that there was more variation in comprehension at the 90 per cent coverage level. It seems that individual L2 listeners vary greatly in how their listening ability is affected by an increase in unknown vocabulary. We should thus be careful not to conclude from the relatively high mean comprehension score that 90 per cent coverage necessarily enables adequate L2 listening comprehension.

DISCUSSION

Lexical coverage and listening comprehension

The results clearly demonstrate that lexical knowledge contributes to listening comprehension both in the L1 and in the L2. In both studies, comprehension of the spoken passage with 100 per cent coverage was significantly better than that of the passages with lower coverage levels. Although both native and non-native speakers could generally understand the stories quite well at the three lower coverage levels, optimal comprehension required 100 per cent coverage; below this level, the comprehension dropped significantly.

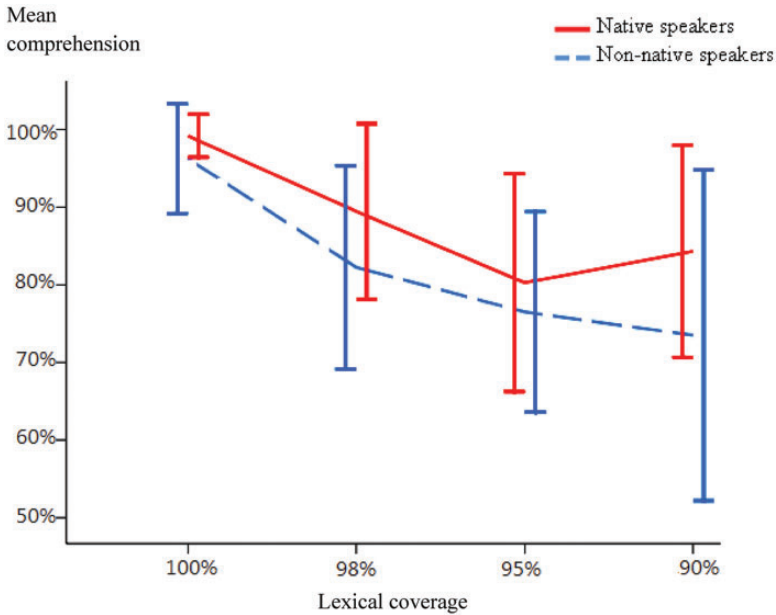


Figure 3: Mean comprehension scores by both native and non-native speakers. Error bars show the mean $\pm$ 1 standard deviation

Although there was no significant difference between comprehension of the stories with 90 and 95 per cent coverage in either the native or non-native group, these two coverage levels clearly had different effects on the comprehension of non-natives. The higher standard deviation and range of the 90 per cent story, relative to that of the 95 per cent story, indicate that there was more variation in comprehension of this story (Table 3 and Figure 3). Whereas a high number of participants could still comprehend the passage very well with 90 per cent coverage, there were also a number who produced rather low scores. This variety in scores indicates that some L2 listeners seem to cope better with unknown vocabulary than others, particularly when there is a lot of it (i.e. 90 per cent level). This is reminiscent of findings by Bonk (2000) and Stæhr (2008, 2009), who similarly found considerable variation in listening comprehension scores.

There might be various reasons for the fact that non-native speakers showed more variation in their comprehension than native speakers. As all native participants were university students while the non-native participant pool was quite varied, there may have been more variance in listener attributes in the L2 group. For example, although all were high intermediate and advanced learners, there was likely some variation in their general L2 proficiency and their skilfulness in online word segmentation and automaticity, which may have influenced their listening ability (Goh 2000). Related to

this is the possibility of different levels of depth of vocabulary knowledge: a greater depth of knowledge may aid learners in integrating information across discourse and thereby contribute to their overall comprehension in listening (Stæhr 2009) and reading (Qian 1999). Additionally, as there was variance in the non-natives' L1s, the listening process of individuals may have suffered from L1 intrusion to different extents (e.g. Cutler 2005). Individual listeners may also have used somewhat different processing approaches: some may have relied relatively more on a bottom-up approach, whereas others used a more global and context-driven approach. Moreover, it is not impossible that the higher scorers had better metacognitive control of the comprehension process using effective combinations of different cognitive processes such as inferring, elaborating, monitoring, evaluating, and predicting. There are also many other factors contributing to listening comprehension (as discussed above) which may have led to the large amount of variation.

Lexical coverage in reading versus listening

The high mean comprehension scores for the 95 and 90 per cent stories suggest that the 98 per cent coverage figure set in reading research may not necessarily be the most appropriate for spoken discourse. It therefore appears that listening comprehension (at least of factual information in informal narratives) requires lower lexical coverage than reading comprehension does, which is in line with findings of earlier research comparing the role of vocabulary in reading and listening, as discussed in the literature review. It also appears that there is less variation among L2 readers than listeners in how strongly comprehension depends on vocabulary knowledge. With regard to reading, Schmitt *et al.* (2011) found that the variation of participants' comprehension did not change much along coverage levels between 90 and 100 per cent, that is at all coverage levels, the standard deviation was found to be similar. Results of our study suggest that this is not the case in listening.

Lexical coverage and vocabulary size targets

So how much lexical coverage is necessary for listening comprehension? This depends not on crossing a defined threshold but on the degree of comprehension desired. Hundred per cent coverage led to almost full comprehension and is statistically better than any lower coverage level. However, 98 per cent coverage also led to very high comprehension rates, and had a statistical advantage over lower coverage levels for both native and non-native speakers. If only very high comprehension will do, then 98 per cent is probably a good coverage target. But if less stringent comprehension rates are acceptable, then 90 and 95 per cent seem to enable this. It seems that 95 per cent may be the best lexical coverage target for L2 listening comprehension of informal narratives: it still leads to relatively high comprehension rates, but avoids the large variation which occurs at the 90 per cent level. It might be noted that our

conclusions for spoken language are similar to those of Laufer and Ravenhorst-Kalovski (2010) for reading lexical coverage: 98 per cent coverage for optimal comprehension and 95 per cent for 'adequate' comprehension, although in contrast, we also found that 90 per cent coverage could lead to successful comprehension by L2 listeners, while Laufer and Ravenhorst-Kalovski could not recommend this level of coverage for reading.

Based on our results, vocabulary size targets set for listening comprehension based on the 98 per cent figure from reading research (e.g. Nation 2006), may be unnecessarily inflated, except for cases where full comprehension is required. For situations where good, but not necessarily complete, comprehension is desired, 95 per cent coverage might be sufficient in most cases. This would reduce estimates of the vocabulary size necessary for listening comprehension from 6,000 to 7,000 word families (based on 98 per cent coverage) to around 2,000–3,000 families (based on 95 per cent coverage) (Adolphs and Schmitt 2003; Nation 2006⁶; Webb and Rodgers 2009a, 2009b). This is obviously a much more manageable prospect pedagogically.

Furthermore, it appears that listeners can largely comprehend informal narratives with lexical coverage as low as 90 per cent. Adolphs and Schmitt (2003) calculated that 750 word families or 1,500 individual words was necessary to reach 90 per cent coverage of the CANCODE, while Nation (2006) calculated that it would take a little over 2,000 word families to reach the same coverage figure for the WCSE. Thus, it seems that even with a vocabulary size of between 750 and 2,000 word families, L2 learners should be able to engage with a considerable amount of spoken discourse.

Limitations and suggestions for further research

One limitation of the study is that participants' knowledge of the 2,000 vocabulary level was measured by means of a written test. As we know, orthographic vocabulary knowledge does not necessarily entail phonologic word knowledge (Milton and Hopkins 2006), and so it would have been better to use a spoken vocabulary test rather than a written one.

After the results had been analyzed, another issue in the study's design appeared. The test of the 95 per cent coverage story contained one question which was relatively difficult to answer. Especially the native speakers seemed to find the item problematic. This is why the results show a rather low mean comprehension score of this story (Table 2). The item was likely difficult because it involved knowledge of a non-word. Although this was true for more items in the various tests, this item was particularly difficult because there was no possibility for listeners to decipher the meaning of the word elsewhere in the text or by using background knowledge. The piloting process did not highlight this issue, most likely because the piloting was only done when the stories did not have non-words inserted. However, we do not see this as a limitation but rather a reflection of the study's ecological validity. In authentic texts, unknown words can be relatively more or less essential for comprehension.

In this case, the non-word related to an idea unit our native-speaking respondents indicated was important for comprehension (see the 'listening comprehension test' section), and if the lower frequency word in the original passage was unknown, comprehension would suffer, just as our test showed. Nevertheless, future research should carefully monitor the relationship between word replacement and comprehension questions, and pilot the test both with and without non-words. With regard to the results of this study, without this item the score on the 95 per cent story might have been a little bit higher, but it is unlikely that this would have led to a significant difference in comprehension of the stories with 90 and 95 per cent coverage.

One decision that had to be made in the process of designing this study was whether to let participants listen to each story once or twice. As the purpose of the study was to investigate the effect of lexical coverage and not memory, it was decided that participants should listen to the stories twice. However, if participants were given only one chance to listen to the stories, the comprehension scores would certainly be lower. This is important to take into account, especially because input will be heard only once in most natural listening contexts, and so the results from this study might be viewed as 'best-case' performance. Similarly, the fact that participants were given the written questions may have cued their listening process and helped them answer the questions correctly (Sherman 1997). With this in mind, the reported comprehension scores should be considered optimal.

Another factor which should be considered in interpreting the results of this study is that the comprehension scores are based only on the texts used in this study. The stories used here were all short stories, told in an informal manner. As narrative has been found to be most comprehensible to listeners (Rubin 1994), other genres (e.g. academic lectures) could have led to slightly lower comprehension results. The higher coverage figure of 98 per cent needed for listening comprehension found by Stæhr (2009) may be due to the fact that the listening material included discourse types other than informal speech. As lexical coverage studies have generally been limited to one text type (e.g. Laufer and Ravenhorst-Kalovski 2010), or analyzed a range of types together (e.g. Stæhr 2009), it is difficult to predict the effect of this variable. Some insight is provided by Schmitt *et al.* (2011), who used only academic texts but controlled for the degree of background knowledge participants would bring to the reading process. This revealed that background knowledge aids reading comprehension, but only at higher (94 per cent and above) coverage levels. Listening research should similarly look at how comprehension is affected by lexical coverage as related to discourse type and background knowledge, for it is likely to differ (Nation and Webb 2011).

Clearly, another factor affecting the results is the comprehension tests. The test items required the extraction of factual information from the four passages. Of course listening comprehension involves various other skills and resources, such as inferencing and interpreting (Buck 2001), but these

were not assessed. Participants' ability to answer the questions correctly was therefore strongly related to their knowledge and use of the linguistic information provided in the texts. This is important in interpreting the results, for if more global, inferential questions had been used, the lexical coverage effect may have been weaker. However, without further research, this is only speculation.

The importance of discourse type and comprehension measure should also be taken into account in comparing our results to those found in reading. They are perhaps best compared with the results of Hu and Nation (2000), who used a similar text type (simple, short narrative written in the first person), and a comparable comprehension measure (one of their tests was a multiple-choice format concerning key information from the text). Based on this comparison, listening comprehension may well require lower vocabulary coverage than reading. However, we see our study as only the first step toward addressing this issue, and believe that many factors will need to be explored (e.g. discourse type, type of comprehension required, and individual differences of the listener) before a final set of answers can be established.

SUPPLEMENTARY DATA

Supplementary material is available at *Applied Linguistics* online.

NOTES

- 1 Bonk related his coverage scores to only a limited number of lexical target words in the texts. Subsequently, Schmitt (2008) calculated the lexical coverage figures which included all content and function words in the texts, and it is his figures which we report here.
- 2 Mastery was established as 27 correct out of a possible 30 items.
- 3 If a story had more low-frequency words than the percentage of non-words required, a number of these low-frequency words were replaced by words within the 2,000 level. If a story needed a percentage of non-words but did not reach this percentage of low-frequency words, a number of high-frequency words were replaced by low-frequency words with the same meaning, and these were then replaced by non-words. Before manipulation, the four texts all contained around 5 per cent post-2,000 level words, which is another indication of the similarity of the different passages.
- 4 Exceptions were proper nouns and a few words for which no 2,000 level synonym was known but which were not expected to be difficult for participants: *restaurant*, *housemate*, *wheelchair*, *flight*, and *screaming*.
- 5 It should be noted that the test items focused on the factual information in the texts, and not on other possible aspects of listening comprehension, such as inferencing or developing a personal interpretation of what was said.
- 6 Although Nation (2006) based his main analysis around a 98 per cent lexical coverage requirement, he also shows figures for 95 per cent coverage, which required 3,000 word families + proper nouns.

REFERENCES

- Adolphs, S.** and **N. Schmitt.** 2003. 'Lexical coverage and spoken discourse,' *Applied Linguistics* 24/4: 425–38.
- Albrechtsen, D., K. Haastrup,** and **B. Henriksen.** 2004. 'The relationship between vocabulary size and reading comprehension in the L2' in D. Albrechtsen, K. Haastrup, and B. Henriksen (eds): *Angles on the English-Speaking World, Vol. 4: Writing and Vocabulary in Foreign Language Acquisition.* Museum Tusculanum Press, pp. 129–47.
- Anderson, R. C.** and **P. Freebody.** 1981. 'Vocabulary knowledge' in J. Guthrie (ed.): *Comprehension and Teaching: Research Reviews.* International Reading Association, pp. 77–117.
- Bonk, W. J.** 2000. 'Second language lexical knowledge and listening comprehension,' *International Journal of Listening* 14/1: 14–31.
- Buck, G.** 2001. *Assessing Listening.* Cambridge University Press.
- 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.
- Cobb, T.** n.d. 'Web Vocabprofile,' available at
- Cutler, A.** 2005. 'The lexical statistics of word recognition problems caused by L2 phonetic confusion,' *Interspeech* 413–6.
- Field, J.** 2004. 'An insight into listeners' problems: too much bottom-up or too much top-down?,' *System* 32/3: 363–77.
- Field, J.** 2008. *Listening in the Language Classroom.* Cambridge University Press.
- Goh, C. C. M.** 2000. 'A cognitive perspective on language learners' listening comprehension problems,' *System* 28/1: 55–75.
- Harris, T.** 2003. 'Listening with your eyes: The importance of speech-related gestures in the language classroom,' *Foreign Language Annals* 36/2: 180–7.
- Heatley, A.** and **P. Nation.** 1994. 'Range,' *Victoria University of Wellington, NZ.* Computer program, available at <http://www.vuw.ac.nz/lals/>, accessed 1 April 2012.
- Hu, M.** and **P. Nation.** 2000. 'Unknown vocabulary density and reading comprehension,' *Reading in a Foreign Language* 13/1: 403–30.
- Koda, K.** 1989. 'The effects of transferred vocabulary knowledge on the development of L2 reading proficiency,' *Foreign Language Annals* 22/6: 529–40.
- Laufer, B.** 1989. 'What percentage of lexis is essential for comprehension' in C. Lauren and M. Nordman (eds): *Special Language: From human Thinking to Thinking Machines.* Multilingual Matters, pp. 316–23.
- Laufer, B.** and **G. C. Ravenhorst-Kalovski.** 2010. 'Lexical threshold revisited: Lexical coverage, learners' vocabulary size and reading comprehension,' *Reading in a Foreign Language* 22/1: 15–30.
- Lund, R. J.** 1991. 'A comparison of second language listening and reading comprehension,' *The Modern Language Journal* 75/2: 196–204.
- McCarthy, M.** and **R. Carter.** 1997. 'Written and spoken vocabulary' in N. Schmitt and M. McCarthy (eds): *Vocabulary: Description, Acquisition and Pedagogy.* Cambridge University Press, pp. 20–39.
- Mecartty, F. H.** 2000. 'Lexical and grammatical knowledge in reading and listening comprehension by foreign language learners of Spanish,' *Applied Language Learning* 11/2: 323–48.
- Mehrpour, S.** and **M. Rahimi.** 2010. 'The impact of general and specific vocabulary knowledge on reading and listening comprehension: A case of Iranian EFL learners,' *System* 38/2: 292–300.
- Milton, J.** and **N. Hopkins.** 2006. 'Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners?,' *Canadian Modern Language Review* 63/1: 127–47.
- Nation, I. S. P.** 1983. 'Testing and teaching vocabulary,' *Guidelines* 5/1: 12–24.
- Nation, I. S. P.** 2001. *Learning Vocabulary in Another Language.* Cambridge University Press.
- Nation, I. S. P.** 2006. 'How large a vocabulary is needed for reading and listening?,' *The Canadian Modern Language Review* 63/1: 59–81.
- Nation, I. S. 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 University Press, pp. 6–19.
- Nation, I. S. P.** and **S. Webb.** 2011. *Researching and Analyzing Vocabulary.* Heinle.

- Park, G.** 2004. 'Comparison of L2 listening and reading comprehension by university students learning English in Korea,' *Foreign Language Annals* 37/3: 448–58.
- Qian, D. D.** 1999. 'Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension,' *Canadian Modern Language Review* 56/2: 282–307.
- Qian, D. D.** 2002. 'Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective,' *Language Learning* 52/3: 513–36.
- Rastle, K., J. Harrington, and M. Coltheart.** 2002. '358,534 nonwords: The ARC Nonword Database,' *Quarterly Journal of Experimental Psychology* 55/4: 1339–62.
- Reves, T. and A. Levine.** 1988. 'The FL receptive skills: Same or different?,' *System* 16/3: 327–36.
- Rost, M.** 2002. *Teaching and Researching Listening*. Longman.
- Rubin, J.** 1994. 'A review of second language listening comprehension research,' *The Modern Language Journal* 78/2: 199–221.
- Schmitt, N.** 2008. 'Instructed second language vocabulary learning,' *Language Teaching Research* 12/3: 329–63.
- Schmitt, N., X. Jiang, and W. Grabe.** 2011. 'The percentage of words known in a text and reading comprehension,' *The Modern Language Journal* 95/1: 26–43.
- Schmitt, N., D. Schmitt, and C. Clapham.** 2001. 'Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test,' *Language Testing* 18/1: 55–88.
- Sherman, J.** 1997. 'The effect of question preview in listening comprehension tests,' *Language Testing* 14/2: 185–213.
- Stæhr, L. S.** 2008. 'Vocabulary size and the skills of listening, reading and writing,' *Language Learning Journal* 36/2: 139–52.
- Stæhr, L. S.** 2009. 'Vocabulary knowledge and advanced listening comprehension in English as a foreign language,' *Studies in Second Language Acquisition* 31/4: 577–607.
- Stahl, S. A.** 1990. 'Beyond the instrumental hypothesis: Some relationships between word meanings and comprehension,' Technical Report no. 505. University of Illinois at Urbana Champaign, Center of the Study of Reading.
- Vandergrift, L.** 2011. 'L2 listening: Presage, process, product and pedagogy' in E. Hinkel (ed.): *Handbook of Research in Second Language Teaching and Learning*. Vol. II. Routledge, pp. 455–471.
- Vandergrift, L., C. Goh, C. Mareschal, and M. H. Tafaghodtari.** 2006. 'The metacognitive awareness listening questionnaire (MALQ): Development and validation,' *Language Learning* 56/3: 431–62.
- Webb, S. and M. P. H. Rodgers.** 2009a. 'Vocabulary demands of television programs,' *Language Learning* 59/2: 335–66.
- Webb, S. and M. P. H. Rodgers.** 2009b. 'The lexical coverage of movies,' *Applied Linguistics* 30/3: 407–27.
- Weber, A. and A. Cutler.** 2006. 'First-language phonotactics in second-language listening,' *Journal of the Acoustical Society of America* 119/1: 597–607.