

How well does teacher talk support incidental vocabulary acquisition?

Marlise Horst
Concordia University in Montreal
Canada

Abstract

Opportunities for incidental vocabulary acquisition were explored in a 121,000-word corpus of teacher talk addressed to advanced adult learners of English as a second language (ESL) in a communicatively-oriented conversation class. In contrast to previous studies that relied on short excerpts, the corpus contained all of the teacher speech the learners were exposed to during a 9-week session. Lexical frequency profiling indicated that with knowledge of 4,000 frequent words, learners would be able to understand 98% of the tokens in the input. The speech contained hundreds of words likely to have been unfamiliar to the learners, but far fewer were recycled the numbers of times research shows are needed for lasting retention. The study concludes that attending to teacher speech is an inefficient method for acquiring knowledge of the many frequent words learners need to know, especially since many words used frequently in writing are unlikely to be encountered at all.

Keywords: incidental vocabulary acquisition, L2 vocabulary, ESL teacher speech, spoken corpus, lexical frequency profiling, frequency list, coverage

Paul Nation's contribution to language teaching research and practice is both remarkable and remarkably well known. If a teacher of English asks "What is the most important grammar to teach?" I think many applied linguists would dismiss her question as naive. Some might point to corpus work by Biber, Johansson, Leech, Conrad, and Finegan (1999) that has recently begun to provide an answer. But were she to ask the same question about vocabulary, she would immediately be referred to work by Nation. Over the course of four decades, his landmark *Vocabulary Levels Test* (Nation, 1990), continually updated frequency lists, and host of publications on teaching and learning vocabulary have provided practical answers to the questions of a generation of language teachers, and, along with the work of the other Paul (Meara), changed the way the field thinks about lexis.

I count myself among the many teachers who found answers to pressing questions in Nation's work. In the early 1990s I was working in Oman with university students who needed to be able to comprehend academic texts and lectures delivered in English and hoped to be able to do so right away—even though their previous schooling had hardly prepared them for anything resembling this. Teachers at Sultan Qaboos University tackled the challenge following two strategies advocated by Nation. One involved intensive direct teaching of high frequency

vocabulary in the computer lab (see Cobb, 1997, 1999 for descriptions of these projects). The second involved implementing a program of extensive reading. We wondered how effective this approach would be. What would the students gain?

Thanks to the work of Nation and his colleagues, there was an experimental methodology in place for answering this question. Following Saragi, Nation, and Meister's study (1978) of acquiring invented *nadsat* words through reading Anthony Burgess' novel *A Clockwork Orange*, my co-researchers and I identified words in a graded reader that were largely unknown to the Omani students, administered an unexpected test of these words once the reading had been completed, and identified a modest though convincing amount of vocabulary uptake, especially of items that were repeated often (Horst, Cobb, & Meara, 1998). The idea of investigating whole books and more recently, large corpora of language input for their potential to support incidental vocabulary acquisition is now well established. Genres that have been investigated in terms of coverage—the percentages of words likely to be known (and unknown) to learners at various levels of vocabulary knowledge—include simplified readers (Wodinsky & Nation, 1988), newspapers (Hwang & Nation, 1989), teen novels (Hirsh & Nation, 1992), academic texts (Coxhead, 2000; Sutarsyah, Nation, & Kennedy, 1994), native speaker conversations (Adolphs & Schmitt, 2004), film (Nation, 2006), and television (Webb & Rodgers, 2009). The study I will report continues in this tradition by exploring the following question: What are the opportunities for learning new vocabulary through exposure to that most fundamental type of input, the speech language learners hear in class?

Evidence that new word knowledge can be acquired incidentally through exposure to spoken input is well established. Twenty years ago, Elley (1989) reported that children retained knowledge of new words they heard in stories read aloud. Since then other studies have shown that learners of a second language (L2) can achieve small but significant vocabulary gains through comprehension-focused listening. Activities that have been investigated include self-directed exploration of a video disk (Brown, 1993), attending to a video-taped dialogue in class (Duquette & Painchaud, 1996), following audio-taped instructions to complete a classroom task (Ellis & He, 1999), watching video both with and without captions (d'Ydewalle & Van de Poel, 1999; Markham, 1999), and listening to stories from graded readers read aloud (Brown, Waring, & Donkaewbua, 2008). In their carefully controlled study, Brown et al. found a repetition effect; as had been found in studies of L2 reading (e.g., Rott, 1999; Zahar, Cobb, & Spada, 2001), words met more often were more likely to be retained. But the main purpose of their study was to compare incidental vocabulary gains when the same stories were read in three exposure conditions: reading only, reading while listening to a text, and listening only. Performance on measures of word knowledge showed the listening condition to be the least effective; gains proved to be very small and susceptible to decay over time; the authors conclude that in order for knowledge acquired through comprehension-focused listening to be lasting, learners may need to hear new words as many as 30 times or more (Brown et al., 2008, p. 18). The extent to which vocabulary is repeated in the spoken input of the language classroom is clearly important, but it has been difficult to investigate because researchers have had to rely on samples of teacher talk that are short and discontinuous. In the corpus study reported here, all of the teacher talk that a group of learners were exposed to in an entire English-as-a-second-language (ESL) course was explored to determine the extent to which the teacher used words that were likely to be new and the extent to which they were repeated.

The studies mentioned above explored a variety of listening activities, many of which are typically used in language classrooms. But findings are necessarily limited to the particular kind of listening treatment that was investigated; the studies cannot make claims about the incidental vocabulary learning opportunities available to learners in whole courses of study. An early attempt to explore teacher talk on a larger scale is Meara's investigation of a series of English lessons broadcast on the BBC (1993). He assumed that listeners would probably know words on West's (1953) General Service List (GSL) of the 2,000 most frequent English word families, and that the lexical challenge would increase, with later lessons containing more unknown word types than earlier ones. The research showed that the word learning opportunities did not change much over time, and that in fact a different kind of spoken text, a *Tintin* comic, offered better opportunities to meet words that were likely to be unfamiliar. The study stands out for its innovative use of lexical frequency profiling, a methodology used again in an investigation of classroom teacher speech by Meara, Lightbown, and Halter (1997). The researchers identified numbers of "off-list" types in short excerpts of transcribed speech addressed to young French-speaking learners of English in intensive classes in Quebec, the working assumption being that words most likely to be new and learnable in this population would be words not found on the basic GSL 2,000 list or on the University Word List (Xue & Nation, 1984)—the latter being "school" words that would probably be familiar. Results based on 10 teachers ranged from 0 to 6 off-list words per 500-word sample, initially suggesting that the spoken input in the ESL classrooms was lexically impoverished. But extrapolation of the findings to the full 5-hour school day indicated that the young learners in the intensive classes were probably exposed to as many as 50 off-list types per day. It was also recognized that not only the off-list words were likely to be new. A replication study (Horst, 2009) conducted in a comparable Quebec context examined a much larger 104,000-word corpus of teacher talk and confirmed the findings of the research by Meara, Lightbown, and Halter (1997): In fact, the teacher speech appeared to offer young Quebec ESL learners at least a hundred opportunities to hear new words in use every day of class. But since the corpus consisted of speech produced by several different teachers on different days, little could be said about opportunities for learners to hear the words repeatedly over time. It was simply not possible to know whether a "learnable" word that occurred in a teacher's speech on a given day was being used for the first time or the fortieth.

The research reported here investigates the word families that occurred in a corpus that consists of all the spoken language addressed to a group of advanced adult immigrant learners during a 9-week ESL conversation course. The study considers the comprehensibility of the teacher speech, the occurrence of words that were likely to be new, and importantly, the extent to which potentially new words were repeated. The data were also examined for evidence of increasing intervals of time between repetitions, following learning research summarized in Mondria and Mondria-de Vries (1994) that shows a retention advantage for learning that is distributed in this way. In addition, the study explores the possibility that particular types of spoken input may provide better opportunities to meet new words than others. For instance, since written language typically has higher proportions of content words than spoken (O'Keeffe, McCarthy, & Carter, 2007), scripted speech (e.g., songs and textbook passages read aloud) might be expected to offer more opportunities to encounter unfamiliar words than speech used to give instructions for activities. Finally, there is the question of potentially important and learnable words that may

never be heard in class; the study also explores this possibility. The research questions are as follows:

1. Was the teacher talk comprehensible? How many words did the learners need to know in order to be able to understand it?
2. To what extent did the teacher use words that were likely to be new?
3. How often were these words repeated? Did repetitions occur at increasingly expanding intervals (regardless of whether or not this was planned)?
4. Do particular genres within the teacher talk vary? Was there a particular type of talk that provided more opportunities for learning new words?
5. What kinds of words were never used?

Method

The Corpus

The corpus used to answer the research questions consists of teacher talk addressed to a class of 20 high-intermediate and advanced ESL students recruited through a community centre in Montreal. The students were placed in the course on the basis of an integrated skills test administered by the centre. They were all recent immigrants to Canada; first languages in the group were Arabic, Chinese, Farsi, Korean, Spanish, Rumanian, and Russian. Many of the students also knew French. The teacher was a native speaker of English and a graduate student in Applied Linguistics with training in communicative language teaching; she had spent 7 years teaching English in Canada and abroad. She was unaware of the goals of the research. The classes, which focused mainly on developing speaking skills through communicative activities, were about 2 hours long and met twice a week for 9 weeks in the spring of 2003. The speaking and listening activities were adapted from the *Canadian Concepts, Level 5* textbook and supplemented with group activities designed by the teacher to give additional opportunities for conversational interaction (see Springer & Collins, 2008, for details). There were no tests on linguistic material covered in class, so word knowledge the learners gained in the course can be assumed to have been acquired incidentally (Hulstijn, 2003), though it is possible that students noted and studied some of the vocabulary explained in the course. The classes were held in a classroom research facility at a Montreal university. All 18 classes were audio- and videotaped and transcribed; the teacher wore a microphone to ensure that the quality of the recorded speech was high. The recordings were collected for a research project led by Laura Collins (Springer & Collins, 2008). Collins made the machine readable transcripts available for the study reported here. The transcripts, which represent 32 hours of class time in total, had also been colour-coded to identify five types of teacher speech: (a) classroom and activity management, (b) language focused talk, (c) text-based input, (d) discussion of text-based input, and (e) personal anecdotes.

The corpus contains all of the natural speech produced by the teacher during the 18 classes; it also includes some scripted speech: a song, some lines of textbook read aloud, and a dictation. Since one of the purposes of the research for which the corpus was originally created was to examine native speaker input, native speech from other sources was included as well. This additional material consists of a few remarks made by research assistants and four audio-taped listening passages. The corpus does not include student productions even though some of their talk was transcribed. The reasons for excluding the student speech are practical: Sometimes a student's response to a teacher query was inaudible, and group and paired activities meant that many students were often speaking at once. As a result, the transcriptions of student talk are of uneven quality and incomplete. It is recognized that by focusing on teacher talk alone, the study takes into account an important part of the spoken input that listeners were exposed to but not all of it.

A computer spellchecker was used to identify unconventional spellings in the corpus and make the following changes: French words used by the teacher in a handful of cases were deleted and variant spellings of speech fillers (*ehm, uhmmm, uhh*, etc.) were regularized to *um* or *uh*. In addition, a few contracted forms such as *hafta* and *sorta* were regularized to *have to* and *sort of*. This was necessary because the frequency software used to analyze the corpus (described below) categorizes spellings it cannot recognize (e.g., *hafta*) as very rare English words. How non-native listeners process reduced forms such as *hafta* is unclear, but work by Jenkins (2002) gives reason to think that the advanced learners in the study were likely to have understood the form as being composed of *have* and *to*. She identifies use of schwa in unstressed *to* as “non-core” in terms of intelligibility (p. 98), and concludes that pronouncing the full vowel sound may actually hinder rather than help comprehension. The fact that such chunks are frequent in spoken English (O’Keeffe, McCarthy, & Carter, 2007), gives further reason to suppose that they were readily recognized in their reduced forms. Regularizing *hafta* as *have to* meant that it was recognized by the software as belonging to the high frequency *have* family. This is consistent with Nation’s classification of *gonna*, *kinda*, and *dunno* as members of the *go*, *kind* and *know* families respectively, in his experimental frequency lists based on the British National Corpus (BNC) of written and spoken English (Nation, 2006).

Files representing the 2-hour classes range in length from 5,817 to 8,544 words (tokens) of teacher talk. With the 18 classes taken together, the total length of the entire corpus is 120,553 words.

Analysis

The teacher talk corpus was analyzed using the BNC-based frequency lists developed by Nation (2006) and corpus tools available at Cobb’s (2009) Lextutor website. Answering the first question about the comprehensibility of the teacher speech involved use of the experimental Vocabprofile BNC-20 program—an online version of Range (Heatley, Nation, & Coxhead, 2002)—to determine the levels of coverage offered by each of 20 frequency lists. This approach assumes that the learners needed to be able to understand the teacher speech well enough to work out the meanings of the unfamiliar words they heard. Research into known word densities that support adequate reading comprehension (Hu & Nation, 2000; Laufer, 1989) have identified 95% as a minimum coverage requirement; that is, with 95% or more of the words in a text

known, L2 readers are able to comprehend it well enough to answer comprehension questions successfully. Recent studies of spoken input such as film (Nation, 2006) and television (Webb & Rodgers, 2009) have set a higher 98% known word coverage criterion. So far, the coverage needed for successful listening comprehension has not been determined experimentally. It is possible that a higher level of known word support is needed for listening than for reading because the processing of spoken input occurs rapidly in real time with little opportunity to reconsider contexts surrounding new words. It can also be argued that understanding classroom speech may need less support due to the availability of visual support for meaning. In this study, I determined the numbers of words learners would need to know to meet both the 95% coverage criterion set by Hu and Nation (2000) and Laufer (1989) and the 98% criterion used by Nation (2006) and Webb and Rodgers (2009).

In order to answer the question about potentially learnable words in the input, the Vocabprofile BNC-20 output was examined to identify words that could be assumed to be unfamiliar to the learners. First, any words that were not on the lists that provided known word coverage at the 95% level were defined as unfamiliar for the purposes of the study; a second analysis identified words not on the lists that provided coverage at the 98% level as unfamiliar. The software was adjusted slightly by the author of Vocabprofile BNC-20 so that filler words used frequently in the teacher speech would not be identified as unfamiliar. Twelve speech fillers and interjections that had been originally categorized as off-list (i.e., not among the 20,000 frequent words of English) were reclassified as a 1,000-level family. These were *ah, aw, eh, ha, hmm, huh, mm, oh, sshh, uh, um, and wow*. Some proper names such as *Korea, Christian, and Saturday* occur in the corpus and were categorized according to their frequency on the BNC lists, but names of students in the class were not included in the analyses. Their absence is due to the fact that the students' names were transcribed as initials, and the online lexical profiler automatically deletes single capitals other than the pronoun *I*. As a result, a teacher utterance such as *So this is Nargis and she is a new student* was processed as *So this is and she is a new student* in the analyses reported here. Had they been written out fully, the profiling software would have categorized the student names as unfamiliar words (even though they are almost certain to have been easily understood). A manual analysis of five randomly chosen files identified 177 uses of student names in 18,320 words of speech. Extrapolation of these figures to the entire corpus points to an estimated 1,165 instances of student names in the entire corpus, just under 1% of the total. Thus the deletion of the single capitals means that the proportions of words that were likely to have been understood are slightly higher than reported in the analyses below.

To answer the question about repetitions, I used text-based Range software (also available at Lextutor) to identify unfamiliar words that were recycled six times or more—in the corpus as a whole and over the 18 sub-corpora that represented days of teaching. The minimum of six repetitions is based on studies of the numbers of encounters required for reliable retention. Figures vary from study to study, with figures ranging from 6 to 15 (see Zahar, Cobb, & Spada, 2001, for an overview). The following method was used to determine the extent to which there were increasing intervals of time between repetitions: The classes met on Wednesdays and Thursdays each week, which meant that it was possible that an unfamiliar word was used on a Wednesday and again the next day. A word that met this condition and was used again the following week (or in any subsequent class), was considered to have met the basic conditions of distributed learning.

Delineating the word learning opportunities available in the different types of teacher talk involved compiling sub-corpora of comparable size and tallying numbers of unfamiliar word families that occurred in each. It proved possible to assemble three 6,000-word sub-corpora of the following types of teacher talk: (a) text-based input, (b) language focused speech, and (c) classroom and activity management. Two other genres, text-based discussion and anecdotes, were found in very small amounts (about 1% of the entire corpus) and were not included in the comparison. The text-based input included a song, audio taped radio broadcasts, a dictation, and textbook passages read aloud. Language focused input was speech that explicitly drew attention to points of grammar, pronunciation, vocabulary and spelling, including corrections of errors. This kind of talk is often referred to as focus on form (e.g., Laufer, 2006). The management-related talk included announcements, course procedures and instructions for classroom tasks. Proportions of the corpus that are accounted for by the three types of speech were determined by analyzing three randomly chosen transcripts (9,500 words total). By extrapolation to the entire corpus, the findings indicated that over half of all of the teacher talk (53%) is devoted to classroom and activity management, 41% is language-focused speech, and 5% is text-based.

Findings

Comprehensibility

The lexical frequency profile of the entire corpus is shown in Table 1. The first row labelled K1 shows the extent to which words on the BNC list of the 1,000 most frequent English word families are found in the speech, the K2 row shows the data for the next 1,000 most frequent BNC families, and so on. There is a general pattern of decreasing frequency such that more infrequent words tended to be used in smaller numbers. The two K20 families at the low end of the frequency scale were *swizzle* and *vermicelli*. The next to last row labelled “off-list” reflects the presence of words in the corpus that are not on any of the 20 BNC frequency lists. As can be seen in the rightmost column, the overwhelming proportion of the teacher talk consists of very basic words, with almost 93% of it accounted for by the K1 list of the 1,000 most frequent BNC families. How much vocabulary knowledge is needed to be able to understand the teacher talk? If, as discussed above, students need to recognize the meanings of 95% of the words in the input they hear, the cumulative percentages in this column show that this level of coverage is achieved with knowledge of the words on the K1 and K2 lists; in fact, knowledge of these 2,000 basic word families appears to provide nearly 96% known word coverage. To meet the higher 98% coverage level, the figures indicate that over 4,000 families (lists K1, K2, K3, K4, and part of K5) would need to be known.

The learners in the course can be expected to have known many high frequency words—most if not all of the K1 and K2 lists and many on the K3, K4, and K5 lists. No vocabulary size data is available for the group but given their advanced placement, they may be comparable to the English majors at a Chinese university who were reported to have a mean vocabulary size of 4,000 high frequency words (Laufer, 2000). Thus it seems reasonable to conclude that the spoken input was indeed comprehensible such that meanings of new words met in the classroom talk were well supported. Certainly in reading the transcripts, one has the impression of a lively,

interactive classroom where students and teacher understood each other well. Interestingly, there is evidence that on the first day of class, the teacher simplified her speech more than usual. The lexical frequency profile of her speech for that day show that the K1 and K2 lists alone account for almost 98% of all the words she used (K1 + K2 = 97.68%). In other words, only 2% (or 1 in 50) of the words students heard that day were not on the lists of the 2,000 most frequent BNC families. The profiles of each of the 17 subsequent days show lower K1 + K2 coverages that equal or approximate the 96% figure identified for the corpus as a whole. The 96% figure means that on average, about 4% of the words the students heard (or one in 25) were not on these basic lists.

Table 1. *Lexical frequency profile of the teacher talk corpus*

Frequency level	Family	Type	Token	Coverage (%)	Cumulative (%)
K1 words	881	1,896	112,926	92.59	92.59
K2 words	583	900	3,866	3.17	95.76
K3 words	309	406	1,538	1.26	97.02
K4 words	184	227	944	0.77	97.79
K5 words	126	160	593	0.49	98.28
K6 words	79	90	323	0.26	98.54
K7 words	65	80	328	0.27	98.81
K8 words	40	50	197	0.16	98.97
K9 words	33	39	136	0.11	99.08
K10 words	27	32	141	0.12	99.20
K11 words	24	28	92	0.08	99.28
K12 words	14	14	47	0.04	99.32
K13 words	12	13	49	0.04	99.36
K14 words	4	5	56	0.05	99.41
K15 words	12	12	35	0.03	99.44
K16 words	10	12	26	0.02	99.46
K17 words	2	2	6	0.00	99.46
K18 words	2	2	5	0.00	99.46
K19 words	4	6	19	0.02	99.48
K20 words	2	2	13	0.01	99.49
Off-list	?	180	627	0.51	100.00
Total	2,413 + ? ^a	4,156	121,967 ^b	100.00	100.00

Note. ^aThe Vocabprofile BNC-20 software groups words on the 20 frequency lists into families, such that occurrences of *happy*, *unhappy*, *happily*, and *happier* register as a single family in the analysis. However, it is not able to do this for words not found on the lists. Hence the question mark that appears for numbers of off-list families.

^bThis figure differs from the 120,553 total given earlier for the number of tokens in the corpus due to the way Vocabprofile BNC-20 processes contracted forms such as *he's* and *don't*. These are each counted as two words, *he is* and *do not*.

Opportunities to Meet Unfamiliar Words

In terms of opportunities for learning new words, Table 1 shows that the teacher used many word families that qualify as unfamiliar according to the definitions discussed above. That is, if the comprehensibility criterion is set at 95% (only words on the K1 and K2 lists are assumed to be known), the number of unfamiliar word families in the corpus amounts to 949, or an average of about 53 families per 2-hour class. This count is based on words that occur on the K3–K20 lists only; if off-list words are added, the opportunities for hearing new words in use are even greater. When the stricter 98% comprehensibility criterion is applied such that words on the K1–K4 lists are considered to be known, then the number of unfamiliar BNC families is reduced to 456 or about 25 per class. Since no measures of vocabulary size were administered, it is not possible to verify which set of figures better reflects the learning opportunities actually available to these students. As Cobb (this issue) points out, vocabulary testing often identifies mixed learner profiles that show unexpected mastery of unusual words and surprising gaps in learners' knowledge of more frequent ones. In any case, it is unlikely that the learners came to the class with knowledge of the entire lists of the 2,000 (or 4,000) most frequent words of English in place but knew none of the words in subsequent lists. In sum, even though it is difficult to quantify learning opportunities in exact numbers, it is reasonable to conclude that the learners were exposed to dozens of words they had not met before each class through listening to their teacher.

Repetitions

The extent to which learners encountered unfamiliar words repeatedly in the teacher talk is shown in Table 2.

Table 2. *Numbers and percentages of encounters with unfamiliar words*

Occurrence	K3–K20 (raw count)	K3–K20 (%)	K5–K20 (raw count)	K5–K20 (%)
1–2	438	46	222	49
3–5	266	28	124	27
6–9	133	14	59	13
10+	112	12	51	11
Total	949	100	456	100

The columns on the left half of the table show repetitions in raw counts and percentages for the K3–K20 families (i.e., words that qualify as unfamiliar when the known word coverage is assumed to be 95%). The right half of Table 2 shows repetitions of families on the K5–K20 lists—words considered unfamiliar according to the 98% coverage criterion. Strictly speaking, knowledge of some K5 words is needed to achieve this level of coverage. As Table 1 shows, the coverage of K1–K4 lists is 97.79%, which approaches but does not quite reach the full 98% figure. The word the teacher used most frequently was *lingerie*, a K14 word that occurred 48 times in the corpus. Other families repeated more than 40 times were *steal* (K3), *vocabulary* (K4), *pants* (K4), and *dialogue* (K5).¹ Only 14 families were repeated the 30 or more times mentioned by Brown et al. (2008). As the first row shows, large proportions—almost half—of the families were used just once or twice, regardless of how unfamiliar is defined. The third and fourth rows, which show the figures for the target zone of six repetitions or more, indicate that only a quarter of the families meet the learnability criterion. The total for K3–K20 words used six times or more is 245 (133 + 112); if only K5–K20 families are considered unfamiliar, the

total is reduced to 110 (59 + 51). This amounts to an average of only six new families per 2-hour class. It appears that without deliberate attention to systematic review, repetition in the amounts needed to support acquisition does not naturally occur, at least in the classroom context investigated here.

If the families are considered in terms of their distributions over different days, a similar picture emerges: The classroom speech did not recycle many unfamiliar families from one class to the next. The findings for the 14 families used in six or more classes are shown in Table 3.

Table 3. *Families repeated in 6 or more of 18 classes*

Word	Number of occurrences	No. of classes	BNC frequency
vocabulary*	46	13	K4
verb*	35	13	K5
review*	32	9	K4
grammar*	29	9	K3
pronunciation*	24	8	K6
dialogue*	47	7	K5
newcomer	22	7	K7
translate*	22	6	K3
personality	20	6	K3
thief	20	6	K3
adjective*	14	6	K7
metro	13	6	K4
noun*	13	6	K6
angry	9	6	K3

Note. * indicates technical vocabulary.

As the table shows, most of the words that meet this level of recycling are words that are typical of language classrooms such as *vocabulary* and *verb*. These terms can be seen as belonging to a domain-specific specialist or technical vocabulary (Chung & Nation, 2004) that learners are likely to know through previous language study. Technical vocabulary that was probably familiar to the students includes terms for parts of speech like *verb* and *adjective*, units of language like *sentence* and *paragraph*, domains of study like *grammar* and *pronunciation*, and classroom activities like *review* and *translate*. Some students had taken French courses that are available to immigrants to Quebec and would have encountered a similar technical vocabulary in these classes. French terms used to talk about language have readily recognizable English cognate equivalents (e.g., *grammaire*, *dialogue*, *verbe*). The nine technical terms in Table 3 are highlighted with asterisks. The five non-technical words that remain are *newcomer*, *personality*, *thief*, *metro*, and *angry*. All of these except *newcomer* are on the lists of K3 and K4 families; that is, they are words that would not be new to students who already know 4,000 frequent English words. In sum, few of the words that were recycled often seem likely to have been new.

In answer to the question about the distribution of exposures, the analyses showed evidence of a pattern of increasing intervals between exposures for just 35 families. These were unfamiliar words that were used at least six times in the speech corpus overall and had at some point been used on two subsequent days followed by use again after a longer lapse. These families are listed

in Table 4 according to their BNC frequencies. As in the case of words recycled frequently across six or more classes (Table 3), there is a preponderance of relatively frequent words (most are K3 or K4) and a number of technical items like *verb* and *dialogue* that may not really be new.

Table 4. K3–20 families used on two subsequent days and again later

BNC frequency	Families	Number (35 total)
K3	cassette, custom, grammar*, habit, jeans, jewellery, lazy, leisure, liquid, mystery, nervous, personality, polite, questionnaire, silence, steal, thief, translate*	18
K4	jail, native, phrase*, professor, review*, trend, vocabulary*, zip	8
K5	dialogue*, lace, verb*	3
K6	pronunciation*	1
K7	adjective*, immigrate, newcomer, trait	4
K10	shirk	1

Note. * indicates technical vocabulary.

Speech Genres

In considering the different kinds of teacher talk, it was hypothesized that the text-based speech would prove to be richer in terms of word learning opportunities than the two other unscripted genres, management-related talk and language-focused speech. However, comparison of 6,000-word samples of the three genres reveals a learning advantage for language-focused speech. There were 149 unfamiliar families in the language-focused speech, 122 in the text-based talk, and perhaps not surprisingly, only 54 in the classroom management speech. These results are shown in Figure 1.

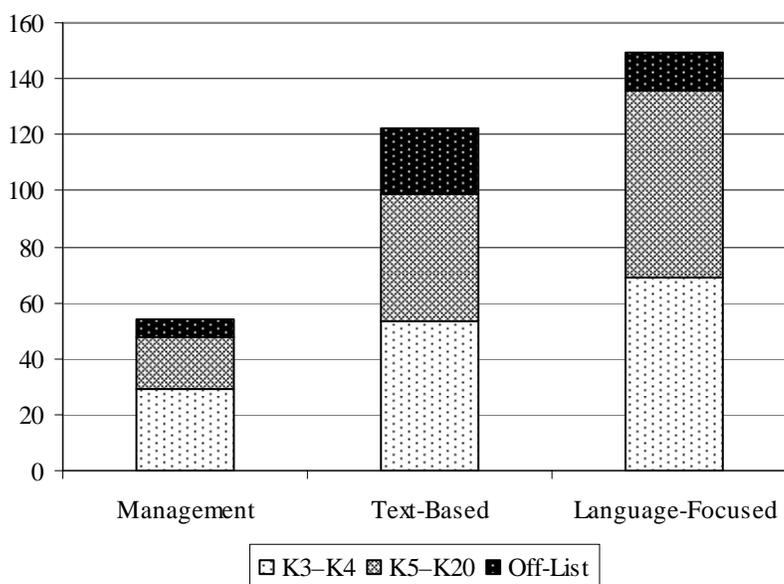


Figure 1. Unusual types in three kinds of teacher talk.

The following excerpts with unfamiliar (K3–20) families in the teacher speech underlined illustrates these differences. The language-focused excerpt contains nine families defined as unfamiliar, several of which are used repeatedly, while the text-based excerpt of a similar length contains six. There is just one unfamiliar family in the management-related excerpt, the K3 word *wander*.

Language-focused talk about a picture (104 tokens, student query excluded):

Teacher: Nylon. You know nylon? We also call pantyhose nylons. Stockings? No no. Ah! Yeah thinner. If you can see the skin. Yeah. Not, not, not transparent, but translucent. Transparent means you can see through it, but not easily. Transparent? This is transparent. Translucent. I don't know if you would call it translucent. You wouldn't . . . Opaque. Opaque is eh, you can't see through it. Translucent is . . . , maybe you would say that M's scarf is translucent. Maybe. Translucent is something that's colour and you can sort of see through it, but not clearly. O-p-a-q-u-e.

Student: [unintelligible] this one?

Teacher: Uh. Use sports shoes, running shoes. You'd say hightops or basketball shoes.

Text-based—audio taped listening passage (111 tokens):

Speaker 1: For men, the lingerie shop is a mysterious and exclusively female domain. Much like a women's washroom, you don't really understand what goes on in there, and we'd rather not go in to find out.

Speaker 2: Guys, they come in and they're like I don't know where to look. Do I look at the sales clerks, do I look at the tables, do I look at the lingerie, do I look at the pictures? Where do I look at? There isn't really a safe place to look.

Speaker 1: Men have a love-hate relationship with lingerie. Men like it when women wear lingerie, of course, but when it comes to choosing chiffon or chenille

Classroom and activity management (102 tokens):

Teacher: Has everyone got one or two words? Yeah? What I'd like you to do now, if you want you can take your piece of paper. Uh, if you don't want to you don't have to, is to stand up and walk around and wander around and tap someone on the shoulder and tell them your word. Okay? If they don't know what your word means, you can explain it to them. Okay, so you need to know what this word means Alright? You will also be hearing words that other people are telling you. So, if you don't know what it means

As Figure 1 shows, text-based talk appears to be richer in off-list words than the other two genres. Off-list words clearly add colour to the classroom discourse; among the 23 off-list words in the text-based talk are items that featured in a radio broadcast on Valentine gifts (shown in part above) such as *bikini*, *bustier*, *chemise*, *pleather*, and *underwire*. Questions might be raised about the usefulness of knowing these. There is also the problem that the text-based materials, which offered relatively good opportunities to learn unfamiliar words, represent a very small proportion (5%) of the corpus; only 6,000 words were found. In fact, most (53%) of the speech the learners were exposed to in the class consisted of management-related talk, the speech where

opportunities to hear unfamiliar words in use were the fewest. However, the proportion of language-focused talk, the type that offered the most opportunities to hear unusual words, is substantial, accounting for about 43% of the entire corpus. This would seem to be good news for vocabulary acquisition, though there is some reason to doubt that all of the words identified as unusual were truly unfamiliar. Figure 1 shows that a large proportion of the words are from the K3 and K4 frequency lists. As we have seen, these lists include technical words like *translate* and *phrase* that are unlikely to have been entirely new.

Words Not in the Speech

Finally, there is the question of words that the teacher never used. Obviously, there are thousands—too many to investigate across all 20 BNC lists. However, it was feasible to consider the K1 category. The data show that the teacher used 881 K1 families, almost the entire set. These can be easily “subtracted” from the full K1 list of 1,004 families using the VocabProfile-Negative program at Lextutor. It is possible to identify patterns in the 123 families that remain. For instance, about 20 of the never-used words are specific to business topics (*budget, contract, pension, client*). Another dozen pertain to government (*council, district, county, king*), and about 10 more to the physical world (*farm, field, boat, tree*). Given the many topics that might be raised in a speaking class, it is to be expected that some were not touched upon, yet it is interesting to note that similar gaps were found in a previous corpus study of communicative language teaching (Horst, 2009). The remaining unused words share no apparent semantic theme but many of them seem more characteristic of writing than speech (*origin, develop, presume, previous, regard*). The absence of words that are more typical of text than speech in teacher talk is hardly surprising given that the BNC lists reflect a corpus that consists largely of written texts (Nation, 2006). The written language bias in the lists may also explain why lexis related to business, government, and the physical world were among the unused K1 words. Teachers can be encouraged to address this shortfall by adjusting the topics discussed in communicative language classrooms—students may be more willing to talk about money and politics than we realize—but the problem of the lexis of literacy (*origin, develop, presume, etc.*) remains. It seems clear that classroom speech alone cannot be expected to familiarize learners with many words of written English that are important to know.

Discussion

The exploration of the corpus showed that as might be expected, spoken ESL teacher talk is a simplified genre—at least in the case of native speaker of English investigated here. Lexical frequency profiling indicated that nearly 98% of the 120,553 tokens in the corpus were accounted for by the 4,000 most frequent BNC families. On the first day of class, the teacher talk was even more simplified with 98% of tokens accounted for by 2,000 frequent BNC families, indicating that the teacher was able to adjust the lexical difficulty of her speech. These findings stand in marked contrast to the much greater amounts of vocabulary knowledge needed to achieve 98% coverage of authentic spoken materials intended for native speakers. For instance, Nation’s (2006) study of the children’s movie *Shrek* found that knowledge of 6,000 frequent English word families would be required, a vocabulary size that many learners do not attain even

after many years of study (Laufer, 2000). Webb and Rodgers (2009) identified a higher figure of 7,000 families as the vocabulary size needed to reach 98% coverage of television shows.

Initially, the research suggested that opportunities for incidental vocabulary acquisition in the course were substantial. If it is assumed that the spoken classroom input was indeed comprehensible to the learners, and that any words not on the lists of the 4,000 most frequent families were new and potentially learnable, nearly 500 families were found to meet this criterion in the corpus. The figure approaches 1,000 if words not on lists of the 2,000 most frequent families are also considered new. Unfortunately, little evidence of repetition at the levels that support long-term retention was found, and this changes this picture dramatically. Only 110 families from the K5 list and beyond were repeated six times or more in the entire corpus, an average of about six each class. If the more generous criterion that includes words on the K3 and K4 lists is applied, the total amounts to 245 families or about 14 per class; but this is almost certainly an overestimate since words like *angry* (K3), *bicycle* (K3), *pants* (K4), and *uncle* (K4) must then be seen as unfamiliar. The investigation of two other repetition patterns—exposures over six or more different classes and repetitions at increased intervals of time—also resulted in small numbers, 16 and 35 families, respectively. Both analyses included technical classroom terms like *verb* and *review* that were unlikely to have been truly new. Given the limitations of lexical profiling methodology, the numbers of learnable words reported here are estimates at best; given differences in prior word knowledge among the students, learning opportunities available in the input must have varied considerably in the group. But if—as the study suggests—there were dozens of learnable words in the speech, retention rates reported in the study by Brown et al. (2008) do not give reason for much confidence in the effectiveness of attending to teacher talk as a method for acquiring new vocabulary. They found that less than one word in 28 acquired through listening was retained after 1 week; after 3 months evidence of ability to provide translations for words met through listening had virtually disappeared—even for words that had been repeated as many as 15 times (p. 151).

The analysis of genres within the teacher talk showed that learners were most likely to hear the teacher use unfamiliar words when she was talking about features of language. The proportion accounted for by this kind of talk is substantial: Nearly half of the entire corpus (43%) is made up of language-focused speech. A closer look at the transcripts of this speech reveals an irony: Even though the study as a whole found that opportunities for new word learning were very few, a great deal of the teacher's language-focused speech appears to be devoted to that very purpose. Dozens of words appear to have been explained every day, often using an elicitation approach as in the following exchange. Transcribed student speech that is not part of the teacher talk corpus is included here to make the character of the exchange clear.

Teacher: A clasp. Do you remember a clasp?

Student: Clasp is something [unintelligible]

Teacher: Uh-huh, you can use it as a verb, to clasp someone. Okay, but a clasp, what is a clasp as a noun?

Student: Like this [demonstrates].

Teacher: Right. Something, something that ah attaches, something that hooks together. Okay, if you clasp someone's arm, you'll go like this [demonstrates]. But a clasp is a, a hook or something that attaches. In the ah, the bag for the camera, okay, there are clasps. Okay? That that hook together.

Students: Oh! Okay.

Student : [unintelligible]

Teacher: Yeah. Very very hard. Something that keeps two things together.

Student: Like cleeps?

Teacher: Like a clip.

Student: Yeah

Teacher: Kind of, yeah, kind of like a clip. There are, there are two parts to it.

Student: Chain?

Teacher: Like a ...? No, a chain is the, the long thing. Okay, but the clasp is the part that keeps things together.

Given the active engagement of the students in this exchange, the multiple examples and illustrations provided, the clear feedback given (both positive and negative), and the number of repetitions (10 total including a student use), this K10 word seems very likely to have been learned at least to recognition level by learners who did not already know it. Also noteworthy is the question at the outset of the exchange: The word *clasp* appears to have been met previously. Here and elsewhere in the corpus, there is evidence that the teacher reviewed previously taught vocabulary. These observations are worthy of further study: While a strength of the list-based profiling methodology is its ability to evaluate all of the vocabulary that was spoken regardless of any intent on the part of the teacher to draw attention to particular words, it is clear that a more complete picture of the word learning opportunities available in the speech could be gained by complementing this approach with an investigation of the words the teacher attended to and the techniques she used to explain and review them. A possible framework for investigating teaching techniques is Laufer and Hulstijn's *involvement load hypothesis* (2001). Their scheme evaluates vocabulary tasks in terms of the extent to which they engage learners in effortful cognitive processing; thus a task such as looking an unfamiliar word up in a dictionary carries a greater cognitive load (and is more likely to lead to retention) than reading a text that provides a gloss of the item. The hypothesis and the experimental studies that test it have focused on reading and writing tasks (e.g., Hulstijn & Laufer, 2001; Kim, 2008) but the relevance to vocabulary episodes in teacher talk is clear. In the *clasp* example above, the teacher begins by eliciting the definition from the students rather than providing it, and she goes on to ask questions that engage the students in elaborating their knowledge of the word. If vocabulary-focused teaching episodes are shown to regularly impose these kinds of cognitive processing demands, then there is reason to think that opportunities for learning may be greater than those

available through simply hearing words used in context (as in the 2008 study by Brown et al., where word learning gains achieved through listening were found to be negligible).

While the explicit vocabulary teaching episodes may prove to be good news in terms of opportunities to learn new words, the finding that some families on the K1 list were never heard in the entire corpus of teacher talk highlights a sobering reality: It might be expected that with enough exposure to spoken classroom input, learners might eventually encounter all of the frequent words they need to know, but given the way words are distributed in natural language this is virtually impossible. In fact, the problem of words not heard becomes more pronounced with families on lists beyond the K1 list. By way of illustration, we might consider the 79 families on the K6 list that were found in the teacher speech (Table 1) and ask this question: What would it take for the learners to be exposed to all of the 1,000 families on this list? As Nation (2006) has shown, this is an important set of families for learners to know; unassisted comprehension of authentic input—both written and spoken—depends on knowledge of at least 6,000 frequent families. However, a minute fraction (0.26%) of the 121,000-word corpus investigated here consisted of K6 words, so even with a great deal of additional exposure, opportunities to hear K6 words in use would be very few. With the introduction of new topics, the learners would probably hear some new K6 families in another 32 hours of classroom speech, but the number would probably not reach 79. Instead, there would likely be recycling of K6 families already heard in the first 32 hours, words that typically occur in communicative language teaching contexts like *delicious*, *pizza*, and *sweater*, along with previously known technical vocabulary like *noun* and *pronunciation*. As in the case of the missing K1 words, the many K6 words that are typical of written text like *deduce*, *entity*, and *implicit* would be very unlikely to come up at all—even in many more hours of spoken input. The implications for pedagogy are clear: One not lost on readers of this journal is that L2 learners need to read in their new language; spoken classroom input alone cannot do the job of providing exposure to the vocabulary of written English. But written text is like speech in that meeting all of the high frequency families requires exposure to a great deal of input. Reading enough text to meet all of the families repeatedly presents an even greater challenge. Cobb (2007) has argued that it is virtually impossible for L2 learners to learn 3,000 high-frequency words through extensive reading alone; the prospects for repeated encounters with all of the less frequent but still important families on the K4, K5, and K6 lists are slimmer still. A way out of the dilemma is the solution long advocated by Nation (1990, 2001): In addition to providing opportunities to learn through exposure to meaning-focused input, teachers should devote class time to the direct instruction of high frequency vocabulary, informed by useful lists such as the BNC 20.

Conclusion

A strength of the research was the completeness of the corpus, which contained all of the teacher speech directed to a class of ESL learners during a 9-week session. In contrast to previous research that explored short listening passages or discontinuous excerpts of teacher talk, this study examined the lexical characteristics of a large body of spoken input, a corpus of 121,000 words gathered over 18 successive classes. It is recognized that the teaching and learning did not take place in a sealed environment such that the teacher talk was the only source of L2 input for the learners, but it is safe to say that the speech represented a substantial proportion of the spoken

L2 input the learners attended to during the 9-week period. Results of the analysis were conclusive. In terms of its potential for supporting incidental vocabulary acquisition, the teacher speech proved to be long on richness—hundreds of words likely to be unfamiliar were used—but short on redundancy, with few words recycled often enough to be remembered. Given the known rates of uptake in listening contexts, incidental L2 word gains achieved in the course are likely to have been minimal.

This is hardly the fault of the teacher, who appeared to be skilled in explaining vocabulary and was aware of the importance of review. Rather, it appears that exposure to the natural spoken language of communicative teaching alone simply cannot provide enough repeated exposures to enough unfamiliar words to be considered an efficient method for acquiring new vocabulary. As in the case investigated here, teacher talk in such settings seems likely to follow the patterns of natural speech whereby the limited number of words that are used frequently are often just that: frequently used and therefore less likely to be new. Even in communicative teaching contexts that include explicit attention to vocabulary as was the case here, there are reasons to question the efficacy of the learning environment. First, a large proportion of teacher talk (the majority in this study) is likely to be taken up by the relatively impoverished discourse of setting up communicative activities and managing them. There is also the problem of the attention given to infrequent words that naturally arise in communicative contexts where topics range freely. If questions arise about the meanings of *slinky* (K11), *lingerie* (K14), and *chemise* (off-list), then these are likely to be given just as much attention as common words that would be more useful (though perhaps not as interesting) to know.

Finally, although the study indicates that communicatively-oriented ESL teacher talk offers little support for incidental vocabulary acquisition, I recognize that language learners experience many well attested advantages in communicative classrooms and the study was not intended to cast doubt on those benefits. However, it is clear that in order for L2 vocabulary acquisition to be efficient, exposure to meaning-focused spoken input needs to be supplemented by other kinds of learning activities. The idea that these activities should include direct teaching and study of high frequency words is now familiar to teachers and researchers everywhere thanks to the work of Paul Nation.

Acknowledgments

The teacher speech data for this project were collected by Laura Collins. Funding for the current study was provided by a research grant from the *Fonds québécois de recherche sur la société et la culture* awarded to Laura Collins, Pavel Trofimovich, Joanna White, Marlise Horst, and Walcir Cardoso. I would also like to recognize the valuable assistance of student research assistants Nancy Dytynshyn and Christi Milsom and project team members (listed above)—all colleagues at Concordia University. Finally, I am indebted to Tom Cobb for adapting software tools used in this study.

Note

1. In fact, the analyses indicate that the most frequently repeated “unfamiliar” item was *Canada* (72 times). This word is classified as a K3 family on the BNC-20 lists. However, I have not reported it as a possible learning target due to its proper name status. See Cobb (this issue) for an argument for not including proper names on the frequency lists used by corpus analysis tools. In the case of *Canada*, the decision seems warranted given the strong probability that it was known to the students.

References

- Adolphs, S., & Schmitt, N. (2004). Vocabulary coverage according to spoken context. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 39–49). Amsterdam: John Benjamins Publishing.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, Essex: Pearson.
- British National Corpus*. Described online at <http://www.natcorp.ox.ac.uk>
- Brown, C. (1993). Factors affecting the acquisition of vocabulary: Frequency and saliency of words. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 263–286). Norwood, NJ: Ablex.
- Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, 20, 136–163.
- Burgess, A. (1972). *A clockwork orange*. Hammondsworth: Penguin.
- Chung, T. M., & Nation, I. S. P. (2004). Identifying technical vocabulary. *System*, 32, 251–263.
- Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, 25, 301–315.
- Cobb, T. (1999). Breadth and depth of vocabulary acquisition with hands-on concordancing. *Computer Assisted Language Learning*, 12, 345–360.
- Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, 11, 38–64.
- Cobb, T. (2009). The compleat lexical tutor (version 6.2) [Computer software]. Retrieved from <http://www.lex tutor.ca/>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.
- Duquette, L., & Painchaud, G. (1996). A comparison of vocabulary acquisition in audio and video contexts. *The Canadian Modern Language Review*, 53, 143–172.
- d’Ydewalle, G., & Van de Poel, M. (1999). Incidental foreign language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research*, 28, 227–244.
- Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, 24, 174–187.
- Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, 21, 285–301.
- Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). Range [Computer software]. Retrieved from <http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx>

- Hirsh, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8, 689–696.
- Horst, M. (2009). Revisiting classrooms as lexical environments. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in second language learners: Papers and perspectives in honour of Paul Meara* (pp. 53–66). Clevedon, England: Multilingual Matters.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond *A Clockwork Orange*: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11, 207–223.
- Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13, 403–430.
- Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Malden, MA: Blackwell.
- Hulstijn, J., & Laufer, B. (2001). Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. *Language Learning*, 51, 539–558.
- Hwang, K., & Nation, I. S. P. (1989). Reducing the vocabulary load and encouraging vocabulary learning through reading newspapers. *Reading in a Foreign Language*, 6, 323–335.
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23, 83–103.
- Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, 22, 285–325.
- Laufer, B. (1989). What percentage of lexis is necessary for comprehension? In C. Lauren & M. Norman (Eds.), *From humans to thinking machines* (pp. 316–323). Clevedon, England: Multilingual Matters.
- Laufer, B. (2000). Task effect on instructed vocabulary learning: The hypothesis of “involvement.” *Selected papers from AILA '99 Tokyo* (pp. 47–62). Tokyo: Waseda University Press.
- Laufer, B. (2006). Comparing Focus on Form and Focus on FormS in second language vocabulary learning. *The Canadian Modern Language Review*, 63, 149–166.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1–26.
- Markham, P. (1999). Captioned videotapes and second language word recognition. *Foreign Language Annals*, 32, 321–328.
- Meara, P. (1993). Tintin and the World Service. *IATEFL annual conference report, 1993*, 32–37.
- Meara, P., Lightbown, P. M., & Halter, R. H. (1997). Classrooms as lexical environments. *Language Teaching Research*, 1, 28–47.
- Mondria, J.-A., & Mondria-DeVries, S. (1994). Efficiently memorizing words with the help of word cards and “hand computer”: Theory and applications. *System*, 22, 47–57.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston: Heinle & Heinle.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and speaking? *The Canadian Modern Language Review*, 63, 59–82.
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Rott, S. (1999). The effect of exposure frequency on intermediate language learners’ incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, 21, 589–619.

- Saragi, T., Nation, I. S. P., & Meister, G. (1978). Vocabulary learning and reading. *System*, 6, 72–80.
- Springer, S., & Collins, L. (2008). Interacting inside and outside of the language classroom. *Language Teaching Research*, 12, 39–60.
- Sutarsyah, C., Nation, I. S. P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based study. *RELC Journal*, 25, 34–50.
- West, M. (1953). *A general service list of English words*. London: Longman, Green and Co.
- Webb, S., & Rodgers, M. P. H. (2009). Vocabulary demands of television programs. *Language Learning*, 59, 335–266.
- Wodinsky, M., & Nation, I. S. P. (1988). Learning from graded readers. *Reading in a Foreign Language*, 5, 155–161.
- Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, 3, 215–229.
- Zahar, R., Cobb, T., & Spada, N. (2001). Conditions of vocabulary acquisition. *The Canadian Modern Language Review*, 57, 541–572.

About the Author

Marlise Horst is an associate professor in applied linguistics and a member of the Centre for the Study of Learning and Performance at Concordia University in Montreal. Her recent research explores second language vocabulary acquisition from a corpus-informed perspective. She is also interested in the effects of instruction designed to raise learners' cross-linguistic language awareness. E-mail: marlise@education.concordia.ca