A Receptive Vocabulary Knowledge Test for French L2 Learners With Academic Reading Goals

Roselene Batista

A Thesis

in

the Department

of

Education

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Arts (Applied Linguistics) at

Concordia University

Montreal, Quebec, Canada

March 2014

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Roselene Batista

Entitled: A Receptive Vocabulary Knowledge Test for French L2 Learners With Academic Reading Goals

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Applied Linguistics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Chair
Elizabeth Gatbonton

_____ Examiner
Sara Kennedy

_____ Examiner
Joanna White

_____ Supervisor
Marlise Horst

Approved by Richard Schimid

Chair of Department

February 14, 2014 Joanne Locke

Dean of Faculty

ABSTRACT

A Receptive Vocabulary Knowledge Test for French L2 Learners

With Academic Reading Goals

Roselene Batista

Recent studies in second language acquisition have confirmed a positive correlation between L2 learners' lexical knowledge and their language abilities. In more specific terms, researchers are increasingly confirming the idea that vocabulary size greatly impacts reading comprehension. In order to estimate English L2 learners' vocabulary size, several kinds of receptive vocabulary tests have been developed. But what about French L2 learners? How is their vocabulary size measured? There appear to be few well designed measures available. This study describes the development and validation of a new measure for French, the Test de la taille du vocabulaire (TTV). The TTV is closely modeled on Nation's (1990) widely used Vocabulary Levels Test (VLT) and follows the guidelines written by Schmitt, Schmitt and Clapham (2001). The TTV draws on recent corpus-based frequency lists for French (Baudot, 1992; Lonsdale & Le Bras, 2009). Initially, a pilot version was trialled with 63 participants, then an improved version was administered to 175 participants at four levels of proficiency. Results attest to the TTV's validity: scores indicate that the higher the group, the larger its vocabulary size. Moreover, the mean scores across the four word sections decrease as the test sections became more difficult. This assessment tool also proved to be reliable as

performance on the test confirmed learners' level as determined by the institutional placement test. Post-test interviews with the participants confirmed their knowledge of the test words. Recommendations for improving the TTV, implications for theory and practice, and limitations will be discussed.

Acknowledgements

This thesis project would not have been possible without the encouragement, guidance and generosity of my supervisor, Professor Marlise Horst. Her remarkable expertise in the field, her relevant and helpful suggestions as well as her immense patience made all the difference during the course of writing this thesis. I also attribute the success of this thesis to the great relationship we developed throughout this process. It all started at the last day of classes, when vocabulary tests were discussed. This discussion inspired me to learn more and to propose the idea of developing a test to Marlise, who gladly accepted it without any hesitation. From this day on, my interest in vocabulary research has increased greatly, all thanks to a passionate professor who gave me the "vocabulary bug."

I would also like to sincerely thank my thesis committee consisting of Professor Sara Kennedy and Professor Joanna White for their invaluable attention and insightful comments when this project was in its inception. I really appreciate the fact that they took interest in my work, imparting their knowledge on me, which greatly contributed to the completion of my thesis.

I also give thanks to Dr. Pavel Trofimovich and Dr. Walcir Cardoso, who played an important role in my academic pursuits. It was truly a privilege to be their student. Special thanks go to the Department of Education staff.

I would also convey my gratitude to Cécile Hernu, the pedagogical advisor in the department of Continuing Education at the CEGEP de Saint-Laurent, who showed a deep interest in this research since the beginning and allowed me to go back to my former

workplace and conduct my research there. In addition, I would like to thank the teachers in the *Francisation* program, who welcomed me into their classrooms so I could invite their students to participate in this study. My most heartfelt thanks go to the students who kindly and eagerly accepted my invitation and came in large numbers to take my test.

Finally, I am specially grateful to my husband, Valdir Jorge, for enduring years of boring and incomprehensible explanations of my countless academic projects in general, and of this research study in particular; my son, Ivan Jorge, for his unrelenting search of native speaker participants, and my daughter, Larissa Jorge, for being my devoted research assistant and for helping me organize the large amount of data I collected. Each of you can share in this accomplishment, and I am forever indebted to all of you.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

How did I become interested in creating a vocabulary test? Vocabulary became the focus of my lessons in a gradual manner, during the course of several years of teaching French as a second language to newly arrived immigrants in Quebec. On many occasions, I realized that I needed to assess the size of my students' vocabulary more precisely, but I did not have a suitable or reliable tool to measure it. Frequently, these second language (L2) learners approached me at the end of the term to ask if I thought they were ready to pursue higher level studies in French (in their case, CEGEP or university level studies), and at those moments I wished I had a vocabulary test to assess, for example, whether they had the lexical resources necessary to cope with certain language tasks, such as reading academic materials. I answered the question hesitantly, knowing that my verdict was solely based on the learners' overall language performance, not on their specific vocabulary knowledge, due to the fact that the tests they took measured their general proficiency rather than their vocabulary knowledge in particular. It is known that measures of general proficiency may tell us whether a learner is able to communicate efficiently in daily interactions, but they seem less helpful as indicators of readiness to read and understand academic textbooks in the new language. In my studies as a graduate student, I became aware of the strong correlation between L2 vocabulary knowledge and reading comprehension (Nation, 2001). Therefore, it became apparent to me that a vocabulary-size test in French had the potential to be very useful in determining student's readiness for doing academic work.

The French test I envision would be able to estimate how many words learners

know, and its results would indicate with reasonable accuracy whether the learners' vocabulary knowledge is adequate for undertaking academic study. This would help them make more informed decisions concerning their studies in the future. Ideally, the test would also be easy to build, simple to administer and fast to score. I have seen that such measures have been developed for estimating the vocabulary sizes of L2 learners of English, and I began to wonder what was available for L2 learners of French. These thoughts led to this study to create and test a new measure of vocabulary size for French.

In the next chapter, research literature relevant to measuring L2 vocabulary size will be reviewed, beginning with definitions of key concepts. Then recent views on the key role of vocabulary in developing L2 proficiency will be presented. In subsequent sections, vocabulary frequency lists and their uses in building measures of vocabulary size will be examined, and special attention will be given to lists and tests of French.

Chapter 2. Literature Review

Words, numbers, definitions

As Wilkins (1972) rightly put it: "Without grammar very little can be conveyed; without vocabulary nothing can be conveyed" (p. 111). Vocabulary is a crucial part of a language, "an essential building block" as Schmitt, Schmitt & Clapham (2001) assert. The lexicon of a language may contain hundreds of thousands of words or even millions (depending on how "word" is defined). According to Michel et al. (2011), the size of the English lexicon is 1,022,000 words. This number was obtained by the Harvard University and Google team, who ran a computational analysis on a 361 billion word corpus, the product of millions of digitized books in English. In this project, the team defined word as "one string of characters uninterrupted by a space" (Michel et al., 2011, p. 176), and included *speed*, *learnt*, and *netiquette* as good examples of strings.

In addition to strings, words can also be counted as tokens, types, lemmas or families. Nation (2001), for example, defines token as every word form in a text. When this unit of counting is used, the same word form will be counted as many times as it occurs in a page. By counting tokens, it is possible to obtain a raw number of words. So the sentence:

How do they do a word count?

contains seven tokens, since each occurrence of *do* is counted separately. However, if another unit of counting is used, one which treats repetitions as a single item, that same sentence would contain six words or "types." By counting types, it is possible to obtain the number of categories of words. When researchers choose "lemmas" as the basis of

counting, they include all the inflected forms of a word as a single headword or lemma (Lonsdale & Le Bras, 2009). In the case of the verb *learn*, for instance, it may include the inflected forms *learns* (-s, third person singular present tense), *learning* (-ing, progressive), and *learned*/*learnt* (-ed/-t, past tense) as a single lemma. Yet another way of counting words is based on the notion of "word family." A family includes all the common inflections and derived forms of a word (Milton, 2009). For instance, the word family of *learn* includes the headword (*learn*), its inflected forms (*learns*, *learning*, *learnt, learned,*), and also its closely related derived forms (*learner*, *learners*, *learnable*, *relearn*, *relearns*, *relearned, relearnt*, *relearning*, *unlearn*, *unlearns*, *unlearned,* etc.).

The same Google team and Harvard professors mentioned earlier have also created a French corpus containing 45 billion words, but no estimation of number of types, lemmas or families was reported. A different source, though, the Office québécois de la langue française (OQLF) (2002), reports that the French language contains 500,000 words. Differently from the Harvard/Google team, however, the OQLF does not specify which unit of counting they used to obtain their number.

No native speaker of these languages can possibly know all of these many words, let alone an L2 learner. Nevertheless, in order to communicate successfully in an L2, a portion of them must be known. Not surprisingly, many second language acquisition (SLA) researchers have focused their attention on L2 learners' vocabulary size, trying to answer the following recurring questions: How much vocabulary does a learner need to know? Which are the most important words to study? Answering these questions requires a definition of what it means to know a word, which will be explained in the next section.

Knowing a word

According to Nation (2001), "there are many things to know about any particular word and there are many degrees of knowing" (p. 23); knowledge can range from realizing one has heard a certain word before to being able to use it in ways that reflect "deep" awareness of the word's register, collocation and associated concepts. Nation (2001) divides these various aspects of vocabulary knowledge into two basic categories: receptive and productive. When learners recognize a word that they read or hear and understand its meaning, they demonstrate their receptive knowledge. On the other hand, when learners are able to produce a word in written or spoken manner, they demonstrate their productive knowledge. In this study, the focus is on receptive vocabulary knowledge, which involves learners recognizing the form of a word while reading and simultaneously connecting it with its meaning, or, in other words, making the form-meaning link (Schmitt, 2010). Given the importance of having a large receptive vocabulary size, it comes as no surprise that a number of tests have been developed to measure this basic connection (Schmitt, 2010). The next sections feature discussions on research that investigated the vocabulary size variable in relation to learners' developing language proficiency and present the history of word frequency lists that have been used in developing tests of vocabulary size for learners of English.

Vocabulary size, frequency, and L2 learning

Simply put, vocabulary size refers to the number of words a learner knows receptively (Milton, 2009). Most researchers measure this in terms of the number of word families known. While an educated native speaker of English, by the age of 20, knows

approximately 20,000 word families (Nation, 2001), L2 learners need to know between 8,000 to 9,000 word families in order to understand unsimplified written texts in English (Nation, 2006) adequately. Based on Nation's (2001) estimate, a native speaker learns on average a thousand words per year of existence up to the age of 20 - simply by exposure. If the same principle is applied to L2 learners, they would only be able to read authentic materials in English in eight years! That is why teachers and learners should make a priority of ensuring students learn large numbers of useful word as quickly as possible. But what are considered useful words?

With recent advances in computing, the answer to this question has become clear. Researchers have been able to analyze large corpora of texts in a given language and identify lists of the words that occur most frequently. This work also identifies the coverage provided by lists of highly frequent words. The concept of coverage is illustrated in Table 1.

Table 1

The Percentage of Text Cover Necessary to Read Different Kinds of Texts

Levels	Fiction	Newspaper	Academic text
1st 1000	82.3%	75.6%	73.5%
2nd 1000	5.1%	4.7%	4.6%
Academic	1.7%	3.9%	8.5%
Total	89.1%	84.2%	86.6%

Note. Adapted from Nation (2001).

The figures in the first row indicate that learners who know all of the 1,000 most frequent English word families will understand over 82% of the words that occur in fiction texts, almost 76% of the words in newspapers, and about 74% of the words in academic textbooks. These 1,000 families are clearly important for learners of English to know. The table also shows that with knowledge of the 2,000 most frequent families and the families on Coxhead's (2000) *Academic Word List*, learners can attain a high level of known word coverage for the three different kinds of written productions (novels, press, and academic materials).

Frequency-informed selection of words that are important for learners of English to know is not a new approach. In the early 1930's, C. K. Ogden and I. A. Richards created the *Basic English* list, containing only 850 words. This list was designed with the purpose of limiting "English vocabulary to the minimum necessary for the clear statement of ideas" (Schmitt, 2000, p. 15). But due to the very limited number of words, there was an over reliance on paraphrase, which sometimes resulted in a very unnatural language use. For instance, the expression *put a question* replaced the word *ask*. A more useful list was compiled by West (1953). His *General Service* List (GSL) consists of 2,000 frequent words drawn from a 5 million word corpus. In addition to being corpus-based, this list was also informed by range (frequency across text genres). This list greatly influenced the work of leading researchers in L2 vocabulary assessment such as Nation and Meara, and it still is a surprisingly valuable list today, even though a few of its words are out of date.

Another important contribution was made by Coxhead (2000), who created a list with the needs of academic learners in mind: the *Academic Word* List (AWL). By

gathering short and long written texts from four different fields of knowledge (arts, commerce, law, and science), the author created a collection of academic texts in English, consisting of 3.5 million tokens or 70 thousands types. She identified 570 word families that occurred frequently across all four subject areas; in other words, this list was also informed by range.

Assuming that academic learners know the 2,000 words in the GSL (which are the most frequent and the "easiest" words of a language, such as *read*, *word, family*) and the technical vocabulary of their fields of expertise (which are considered "difficult" words by laymen, such as *schwa*, *washback*, *lemmatization*), one can note that there is a vocabulary gap between these two distinct parts of their lexicon. The AWL with its 570 word families (e.g. *assess*, *percent, select*) attempts to fill this gap, bridging the sets of easy and difficult words (Cobb & Horst, 2004). These words are clearly very useful for learners who wish to expand their vocabulary in order to read academic texts, and "become 'successful candidates' in the conventional forms of education" (Nation, 2001, p. 26). How knowledge of frequent vocabulary on lists such as the GSL and AWL impacts reading comprehension will be discussed in the next section.

L2 vocabulary size and reading comprehension

It is now widely known that knowledge of thousands of words is necessary to use a language efficiently, and that vocabulary knowledge correlates strongly with language skills. Work by SLA researchers is increasingly demonstrating that vocabulary size has an impact on the ability to use an L2 (Schmitt, 2010). Generally, L2 learners with larger vocabulary sizes perform better on measures of language skills compared to L2 learners

with smaller vocabularies. Looking at reading skills in particular, most researchers agree that receptive vocabulary size plays a major role in reading comprehension activities. The idea behind this is that learners who have a larger vocabulary size will also know a higher percentage of words that occur in a text.

A recent study by Staehr (2008) confirms this relationship. He looked at the effects of pre-university students' vocabulary size and three types of skills, namely, reading, listening, and writing, and found that of the three, reading comprehension is the most dependent on vocabulary size (p. 148). By using the Vocabulary Levels Test (VLT) to measure 88 Danish L1 learners' receptive vocabulary size, the author discovered that most learners who knew the first 2,000 word families performed well in the reading test; in fact, knowledge of these words at this level was a key predictor of a passing score. Further confirmation of the connection between vocabulary size and reading comprehension comes from the study by Qian (2002). The author tested 217 adult learners attending an intensive ESL program (all pre-university, undergraduate and graduate students from numerous L1 backgrounds), and discovered that scores on the VLT correlated highly with scores on a reading comprehension measure. In other words, the larger the learners' vocabulary size was, the better they performed on the test that presented five short written texts about academic content (e.g. biology, geography, art history) and 30 multiple-choice questions. However, in another study investigating the relationship between L2 learners' vocabulary size and their reading comprehension, Laufer and Ravenhorst-Kalovski (2010) did not obtain consistent confirmation of the idea that a larger vocabulary size leads invariably to increased reading comprehension. They tested 735 learners from three L1 backgrounds (Hebrew, Arabic, and Russian) enrolled in

an English for Academic Purposes course. In order to measure the learners' vocabulary size, the VLT was used; in order to measure reading comprehension, they administered the English part of a university entry test. This English test assessed learners' understanding of words, sentences, and textual information. Although most participants followed the pattern previously confirmed by research, the participants with greater vocabulary knowledge (those who knew 7,000-8,000 words) obtained lower scores on the reading test than the ones with smaller vocabulary knowledge (those who knew 6,000 words). In other words, the expected advantage for knowing more words was not found in all groups under investigation. The researchers concluded that knowing a larger amount of words does not always guarantee better reading comprehension. In spite of that finding, most studies do confirm a connection between larger vocabulary sizes and higher comprehension scores (e.g. Anderson & Freebody, 1983; Laufer, 1992; Qian, 1999; Schmitt, Jiang & Grabe, 2011).

As briefly explained earlier, in order to read and comprehend authentic texts in English, L2 learners need to know 8,000 to 9,000 frequent word families (Nation, 2006). This estimate is based on the idea of coverage. By knowing this number of words, learners will be able to recognize 98% of the words in an unsimplified text, which means that, for each 100 words, only two will be unfamiliar (and there will be enough support available - in principle - to allow them to be guessed from context). Schmitt, Jiang and Grabe's (2011) study of over 600 university learners of English confirmed the 98% figure with actual ESL readers. They tested the readers on their knowledge of a large proportion of the words that occurred in two texts that the participants eventually read. Results showed that when known-word density amounted to 98%, the mean reading

comprehension score (based on two measures) amounted to 75%. The authors considered this to be "adequate" comprehension. This work and the study by Nation (2006) mentioned above converge on the figure of 8,000 to 9,000 frequent words as the vocabulary size needed to attain 98% known word coverage. It is worth noting that both readers of general and academic texts need the same percentage of word coverage, namely 98%, to accomplish their reading tasks adequately.

A study by Cobb and Horst (2004) looked more closely at the vocabulary needed for reading academic texts. This work shows that the first 2,000 most frequent word GSL families covers around 80% of written productions in English (see also Table 1), and that the 570 words in the AWL accounts for 10% of the academic corpus (Coxhead, 2000, p. 222); these 2,570 words together translate into a text coverage level of almost 90%. But, according to research by Schmitt, Jiang and Grabe (2011) described above, this is not nearly enough for adequate reading comprehension (since as stated earlier, 98% coverage is needed). So if learners only master these two groups of vocabulary (GSL+AWL), they clearly will not attain the coverage threshold level needed to make successful inferences about unfamiliar words, and they will have to add at least another 5,430 words to their lexical repertoire if they want to reach the 8,000 figure Nation (2006) has determined is the minimum needed to adequately comprehend the materials they are required to read in their pre-university or university courses. It is important to note that these findings pertain to English. Now, if the focus is turned to L2 learners of French, this research may not be relevant, since it is not known whether they need to know a similar number of French word families to comprehend authentic academic texts, which are required readings for CEGEP or university courses. If these numbers do not apply to them, then

how large does an academic learner's vocabulary size need to be in order to comprehend materials written in French? In the next section, some answers to these questions will be provided.

How many words do learners need to know to understand academic texts in French?

The study by Cobb and Horst (2004) considered the numbers of words learners of French and learners of English need to know in order to read academic texts in their respective L2s. As mentioned earlier, ESL learners need to know almost 2,600 high frequency word families (the first 2,000 most frequent words plus the AWL) in order to achieve 90% known word coverage of academic texts. Their investigation showed that learners of French only need to know the 2,000 most frequent word families in order to reach the same 90% level of coverage of the vocabulary that occurred in the academic materials they analyzed. In other words, by knowing a smaller amount of basic words in French, learners will gain 90% of text coverage; they do not need to learn an additional set of academic words. The authors concluded that "French seems to use its frequent words even more heavily than the English does," since part of the "academic" words in French are used in both everyday language and academic discourse (p. 35).

To explore how many words in each frequency level learners need to know to read academic texts in French, the following mini-experiment was undertaken. An academic text was selected: the research study called "Dynamiques familiales et activité sexuelle précoce au Canada" (Assche & Adjiwanou, 2009). This article was one of the required readings for students in the Psychology Program at Sherbrooke University

during the Winter term in 2013. Preliminarily, a sample of the document (a 4,127 word text) was manually stripped of unneeded material, such as images and graphs. Then the resulting text was fed into *Vocabprofil*, a French-based version of *Vocabprofile* (available at Cobb's *Lexical Tutor* website). The results of the analysis are shown in Table 2, where all the words in the sample text are classified into bands of frequency: K1 stands for the first 1,000 most frequent words in French, K2 for the second 1,000 most frequent words, and the Lower frequency words include the third 1,000 most frequent words and the words that did not appear on the frequency lists.

Table 2

Vocabprofil Output for an Academic Text - Psychology Program

	Families	Types	Tokens	Percent	Cumulative
K1 Words	375	477	4127	80.91	80.91
K2 Words	128	152	416	8.16	89.07
Lower frequency words	119	248	558	10.93	100%
Total	622	877	5101	100%	100%

As can be seen in the last column, with knowledge of only the 2,000 most frequent families, French L2 learners can be expected to understand 89% of the words in this academic text. This is broadly consistent with the findings reported by Cobb and Horst (2004), which identified a coverage level of nearly 89% for the 2,000 most frequent words.

Milton (2009) says that "95% or 98% coverage of a text might be required for full comprehension in any language" (p. 69), but assuming this is true in the case of French,

how many word families would learners of French need to know in order to reach the full 98% coverage level needed for comprehension of academic texts? It is simply not possible to answer this question because the software tools do not (yet) exist to analyze French texts for the occurrence of words beyond the 1K and 2K levels. However, a new corpus-based frequency list (Lonsdale & Le Bras, 2009) for French includes five 1,000-word frequency bands and may be able to help answer the question soon.

So far, this chapter has presented some evidence from the research literature showing that successful reading comprehension is closely associated with performance on measures of receptive vocabulary size. Also, it has detailed studies that attempt to show how many word families learners of English need to know to be able to read and understand academic texts adequately. But in the case of French, the answer is far less clear. Table 2 shows that the 1K and 2K offer an important amount of coverage, but researchers do not (yet) know how many additional bands of what is termed "lower frequency vocabulary" in the table need to be known in order to support comprehension. It is safe to conclude that a receptive vocabulary size measure for French L2 learners will be useful in answering the question. With a size test in place, it will be possible to investigate learners' vocabulary size in relation to their comprehension of texts of various known words densities. Thus, for example, a learner's performance on the test might show that she has receptive knowledge to the 5,000 most frequent French words, and that this gives her 93% known word coverage of a particular reading passage. A subsequent reading comprehension test could then determine the extent to which the learner comprehended the text adequately.

If a study were conducted to measure French L2 learners' vocabulary size, which

test should be used? Which words should be included? How should they be selected from the incredibly large lexicon of a language? In the next section, some types of existing L2 vocabulary size tests will be examined with a view to understanding how they can inform the process of constructing a principled test for French L2 learners.

Receptive vocabulary testing

Tests aiming at measuring learners' vocabulary size (e.g. Meara & Buxton, 1987; Meara & Jones, 1988; Nation, 1983; Nation & Beglar, 2007) are often based on samples of words from word frequency lists with each tested word requiring only a single answer (Read, 1993). Performance on a sample is then extrapolated to the entire set of words in a particular frequency band. Thus a learner who demonstrates knowledge of 16 of 20 (80%) words sampled from a frequency band of 1,000 words is assumed to know 800 of the 1,000. These language tests measure receptive vocabulary knowledge, since they only require learners to recognize words being tested. In most of them, learners are just asked to make the form-meaning link (Schmitt, 2010), demonstrating that they know the form of a lexical item and its corresponding meaning. In several receptive tests, the target words are presented in isolation (no context), which means that they are measuring decontextualized vocabulary knowledge.

Different types of tests using several response formats have been developed throughout the years. To name a few, those most extensively used in testing L2 receptive vocabulary knowledge are multiple-choice, checklists, word association, and matching words with definitions. In terms of evaluation criteria, tests may also vary in relation to their methods: some ask learners to choose written synonyms of target words, others ask

learners to self-report whether they know words on a checklist that contains some nonsense words (Wesche & Paribakht, 1996).

Among all the recognition vocabulary tests that were developed from frequency lists, three will be discussed in this study: the Vocabulary Levels Test, the Yes/No checklists, and the Vocabulary Size Test. They were chosen for their proven strengths and for being extensively used by researchers and teachers to assess vocabulary knowledge of L2 learners. They are discussed in chronological order below.

1) The Vocabulary Levels Test

The Vocabulary Levels Test (VLT, or the Levels Test) was first developed by Nation (1983) as a diagnostic test intended to help learners with vocabulary learning (Nation, 1990). Using a word-definition matching format and including only real words in English, the test results can give a good indication of where learners need to develop their vocabulary knowledge, showing the gaps in their lexicon. Then a number of researchers started using it as a measure to estimate L2 learners' receptive vocabulary size in experimental studies (Staehr, 2008).

The VLT has a number of distinct advantages. First, due to the fact that it only contains single decontextualized words and short definitions, the test does not require sophisticated grammatical knowledge or higher reading skills (Schmitt, Schmitt & Clapham, 2001). Second, the definitions are written using simple vocabulary, which helps ensure that the options are comprehensible. Third, the format (see Figure 1) has been designed to minimize the role of guessing. For each definition on the right, there are six possible answers. Fourth, this test can be scored objectively by comparing learners'

responses with an established set of acceptable responses (scoring key); no subjective judgements are required on the part of the marker. Finally, this type of test format allows a large number of words to be tested in a short time.

As Figure 1 shows, learners simply write numbers to match the three definitions on the right with three of the words on the left. Figure 1 illustrates a cluster from the 5,000 word level, in two different versions: (a) paper-and-pencil and (b) electronic:

(a)

1 desolate
2 fragrant
3 gloomy
4 profound
5 radical
6 wholesome

_____ good for your health
_____ sweet-smelling
_____ dark or sad

(b)

1. desolate
2. fragrant
3. gloomy
4. profound
5. radical
6. wholesome

☐ good for your health
☐ sweet-smelling
☐ dark or sad

Figure 1. Two sample clusters from the VLT - paper and electronic versions.

In the original paper-and-pencil version, learners write the number of the target word next to its meaning; in the electronic one, they type the number instead. All the responses are verifiable, meaning that the test words have only one acceptable response.

The VLT contains five sections: four assessing the 2,000, 3,000, 5,000 and 10,000 frequency levels and one assessing academic vocabulary. Containing 30 clusters (six per section), it has a total of 180 words, 90 test words (18 per section, three per cluster), and it takes 20 minutes on average to complete. Besides being fast to score, and simple to produce, it is low cost (free of charge), and easy to procure: both printable and electronic versions are available at Cobb's (2000) Compleat Lextutor website.

Schmitt et al. (2001) created an updated and bigger version of the VLT. In fact, they created two alternate versions: Version 1 and Version 2. Each consists of 50 clusters (10 clusters per section), 150 test words (30 per section, three per cluster), and it takes 30

minutes on average to complete. Version 1 is available in Schmitt (2000); Version 2 in Schmitt et al. (2001).

2) The Yes/No checklists

Meara and Buxton (1987) developed this paper-and-pencil test as an alternative to multiple-choice vocabulary tests, which are considered to be difficult and time-consuming to produce (Meara, Lightbown & Halter, 1994). The authors attained their objective, since the Yes/No checklist is considered to be the simplest type of test format used in receptive vocabulary testing (Read, 1993). Besides being very easy to construct, it requires only a few minutes to be completed (Meara & Buxton, 1987). This test relies entirely on self-report by the learners, who are asked to read a list of words in isolation (real and invented ones), and mark the ones they think they know. The reason for adding invented words to the test was to avoid learners overestimating their vocabulary knowledge (Beeckmans et al., 2001). If learners say they know a nonsense word, their score is reduced. If learners mark all real words as known, they obtain the maximum score. Figure 2 illustrates a sample of the checklist (prompt and test items). The test words were selected from frequency-informed lists, and the invented ones followed phonotactic and morphological rules for word formation in English (Beeckmans et al., 2001).

Tick the words you know the meaning of e.g. milk ✓			
gathering	forecast	descent	revenge
strap	conscious	wodesome	heartless

Figure 2. Sample questions in the Yes/No Test.

It takes approximately a second for learners to decide whether they know the word or not (Beeckmans et al., 2001). These fast self-reported responses allow hundreds of words to be tested within the 10 minute period assigned for the testing.

Meara (1992) developed a computer-based version of the Yes/No test called the EFL Vocabulary Test. The checklist contains 300 target words divided in five levels. The instruction of this test reads 'For each word: if you know what it means, check the box beside the word; if you aren't sure, do not check the box'. Figure 3 illustrates a sample of words from level 5 of the computerized Yes/No test.

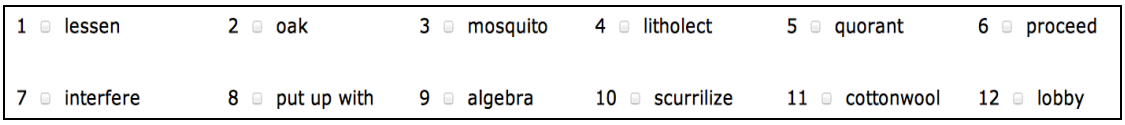


Figure 3. Sample questions in the computerized Yes/No Test or EFL Vocabulary Test.

After examining the Yes/No vocabulary test, Beeckmans et al. (2001) concluded that this measure is not reliable, because learners may adopt different strategies when dealing with invented or real words. In addition, Meara et al. (1994) found out that a small group of learners self-reported knowing more invented words than real ones, resulting in "unscoreable" verdicts. Such non-results present a clear threat to the applicability of this kind of test. Similar problems are reported by Cameron (2002); she compared Meara's (1992) Yes/No test to Nation's (1990) VLT and considered the latter more useful, due to the unreliable results that the inclusion of invented words produced.

Nevertheless, this methodology has been used by Meara and his colleagues to produce vocabulary size tests for a variety of European languages including French. English and French versions are available at Cobb's (2000) Compleat Lextutor website.

3) The Vocabulary Size Test

Created by Nation and Beglar (2007), this proficiency test called the Vocabulary Size Test (VST) determines learners' English receptive vocabulary size from the first 1,000 frequency level to the 14th 1,000, using Nation's (2006) updated frequency lists based on the British National Corpus (BNC). Its ability to assess how many words the learner knows from the most frequent 14,000 word families makes it a more comprehensive measure than the earlier VLT, which tests only up to 10,000 word families. Including 14 sections, each one corresponding to a word level band, the VST contains 140 clusters (10 clusters per section), which amounts to 140 test words. It also has the advantage of drawing on one of the largest and most current corpora of English available. However, 12 of its 14 sections (85% of the test) used the lists of the BNC spoken section, and even the authors recognized that developing a written text based on a spoken corpus is somewhat unusual (Nation et al., 2007). Their choice was based on the assumption that the spoken ordering was closer to the order in which the learners might learn the words.

Similar to the VLT, the test requires the learners to select one response from the options given in each question, and the answers given by the learners are verifiable. In the prompt, the test word is accompanied by a simple non-defining context, which indicates its part of speech, facilitating the establishment of the form-meaning connection. However, the distractors are rather closely related in meaning to the test word; as can be seen in Figure 4, all the meaning options for *miniature* use the word *small*. Since this requires learners to have a more precise idea of a word meaning, the VST is slightly more demanding than the VLT (Nation & Beglar, 2007).

Differently from the VLT, the VST is a standard multiple-choice format that requires the writer to create four definitions for each test word. In addition to making the test more time consuming to construct, this presents a number of problems. Due to the fact that the test contains 560 sentences (140 test words X four definitions), it requires more reading on the part of the test-takers, who also need to know more complex syntax to understand the definitions. Besides, there is a larger scope for guessing on the VST than on the VLT: in the former, there are only four answer options (for each test word), while in the latter, there are six answer options (for three test words). Figure 4 shows a sample question from the 5,000 word section of the VST. This test is available at Cobb's (2000) Lexical Tutor website.

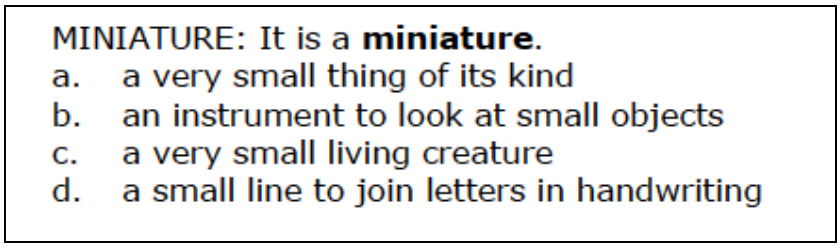


Figure 4. Sample question from the Vocabulary Size Test.

As seen in this overview, existing studies have tended to focus on assessing learners of English with only the Yes/No checklist test, a format that has proved to be somewhat problematic, available in French. In the next section, some criteria will be considered in order to determine which is the best model for creating an improved vocabulary size test in French.

A useful measure for academic learners

When one is designing a vocabulary test, Read (1993) specifies four dimensions that should be taken into account: (1) simple to more complex test formats; (2) verifiable versus self-report responses; (3) breadth and depth of knowledge; and (4) isolation versus contextualization of test items. Based on these dimensions, the three vocabulary tests described above, namely the VLT, the Yes/No test and the VST can be described as follows:

1) In relation to the first dimension, all three tests are on the simple end of the continuum, since they ask test-takers to indicate the correct answer rather than to perform complex tasks such as 'write an essay on a specific subject';
2) In terms of types of responses, the second dimension, two of the three tests use the verifiable response format (VLT and VST) and one uses self-report (Yes/No test);
3) Regarding the third dimension, all three tests are measures of vocabulary breadth rather than vocabulary depth; and
4) As far as the fourth dimension goes, two of them present the test words in isolation (VLT and Yes/No test), one presents them in context (VST).

Brown (2007) specifies three main qualities (or criteria) for testing a test, namely practicability, reliability and validity. According to Bachman and Palmer (1996), test practicability involves three types of resources: material resources (e.g. room for test administration, computers, paper); human resources (e.g. test developers, raters); and time (e.g. time for test development or scoring). First, in terms of material resources, the Yes/No checklist test, the VLT and the VST are very practical, because they are all pencil-and-paper tests, which require only a photocopy machine or a printer to make

paper copies of the test document. Also, they can be administered in traditional classrooms with no need to use any special equipment whatsoever. Second, in terms of human resources and from the developer's point of view, clearly the VLT and the VST are not as simple to construct as the Yes/No checklist test, which, as its name suggests, is simply a list of decontextualized words. Nevertheless, the cluster format of the VLT does not require writing as many definitions as the multiple-choice format of the VST, in which every test item must have a set of four answer options. Nor does it require the creation of short non-defining context sentences as in the VST. From the test-takers' point of view, the Yes/No test is very practical. Read (1993) asserts that there is no simpler test format than this checklist test, where students are only required to read isolated words and mark whether they know them or not. Multiple-choice formats, such as the VST, and word-definition matching formats, such as the VLT, require more extensive reading on the part of the test-taker (Read, 1993). However, the VLT does not require test-takers to read as much as the VST, since its clusters contain individual words and short defining phrases. Third, in terms of time, the Yes/No checklist test is the fastest test to sit: 10 minutes on average. On the other hand, due to the fewer number of clusters (50 cluster), the VLT takes less time to sit (30 minutes on average) than the VST which includes 140 clusters. The larger the number of clusters, the longer the test takes to complete and some test-takers might be tired or bored towards the end of the assessment period, which might affect the test results. In sum, among the three tests described above, the Yes/No test has the highest practicality level but the VLT comes in second with distinct advantages over the more time-consuming VST.

A good test needs to measure consistently, that is, it should be reliable. If test-

takers perform the same test twice, on two different occasions, and obtain similar scores, this is evidence that the test has high reliability. The Yes/No checklist test's use of invented words designed to resemble real words as close as possible in terms of spelling and morphology has raised reliability concerns (Read, 2000). It is easy to imagine that a learner might say yes ('I know this word') to a nonsense word that was a near match to real word in one test setting and no ('I don't know this word') in another. In the case of the VLT and VST, different answers to the same question might be less likely to happen, since test-takers are asked to demonstrate their vocabulary knowledge by choosing real words only. Besides, the responses are verifiable (a scorer or a computer compares learners' answers with a scoring key), which greatly reduces the cases of measurement errors, making both tests have high marker reliability. In short, there is reason to favour the VLT and the VST over the Yes/No test, in terms of potential reliability.

In terms of validity, the Yes/No test is clearly problematic in that it simply asks test-takers to respond "yes" if they recognize a written word form as being an English word. It is difficult to verify what a yes response means. Does the learner have deep knowledge of multiple meanings of the word and its many associations? Or does she simply recognize it as a familiar string of letters? The test-takers are not required to recognize a connection to any kind of actual definition of the real words. In contrast, the VLT and the VST do ask test-takers to explicitly show that they have established the form-meaning links of the test words. In sum, both the VLT and the VST are more likely to be valid reflections of learners' recognition vocabulary knowledge than the Yes/No test, but the VST appears to make finer knowledge distinctions. Table 3 contains a summary of Brown's three criteria and how the three tests described above meet them.

Table 3

Summary of Characteristics of Three Vocabulary Tests - Brown's (2007) qualities

	Practicality	Reliability	Validity
Levels Test	✔✔	✔✔✔	✔✔✔
Yes/No Test	✔✔✔	✔	✔
Size Test	✔	✔✔✔	✔✔✔

Note. ✔✔✔ = very high, ✔✔ = high, ✔ = not so high

In summary, the VLT shows a better compromise among the main criteria commonly used in measuring test efficiency: First, this word definition matching test is practical for both test-writers and test-takers; secondly, it is more likely to be reliable, since it only presents real words, and its response format elicits verifiable responses, reducing the scope for measurement errors; and third, measuring learners' lexical knowledge in terms of their ability to make connections using simple definitions is more likely to be a valid reflection of learners' recognition vocabulary knowledge than unspecifiable yes responses on the checklist test. Although the VST also presents itself as a good candidate for developing a new measure in French (since it requires students to have detailed knowledge of words), it seems so far that the VLT is the ideal model for creating a vocabulary size test for French L2 learners.

In the next section, some reasons why such a measure has not yet been developed in French will be outlined. Moreover, the Yes/No checklist test in French, one of the few size measures that do exist for French, will be examined.

Receptive vocabulary testing in French

Milton (2006) asserts that research on French vocabulary is scanty and little is known about French vocabulary learning in school. The few studies that were conducted to measure learners' vocabulary size in French paid attention to children and adolescents learning this language in European schools, particularly in England and Wales. Due to the fact that these studies have adopted similar methodologies for estimating vocabulary size in French, it was possible to make comparisons among their findings.

According to Milton's (2006) results, 69 British university entrants knew 1,930 words in French, on average, which leads to the conclusion that, with this small amount of vocabulary, these learners were not ready to pursue their studies in French at the university level. In David's (2008) study, findings were similar. The learners of French in the UK were 32 pre-university students attending Year 13 in Newcastle schools. The author found out that they knew 2,108 words in French on average.

In both studies, the French vocabulary test used to measure learners' vocabulary size was the French X-Lex test, an adaptation of the Yes/No test developed by Meara and Milton (2003). Figure 5 illustrates this test format. The French X-Lex test is based on Baudot's (1992) frequency list (see details in the next section), which is a lemmatized vocabulary list of modern French (Milton, 2009). It only tests knowledge of the first five 1,000 word frequency bands, and presents the test-takers with 120 words: 100 real words, drawn on Baudot's (1992) list, as stated earlier, and 20 invented ones, that were built following phonological and morphological rules to resemble real words. The French X-Lex took 5 to 10 minutes to be completed, and was found to be a reliable and valid measure of receptive vocabulary size according to David (2008).

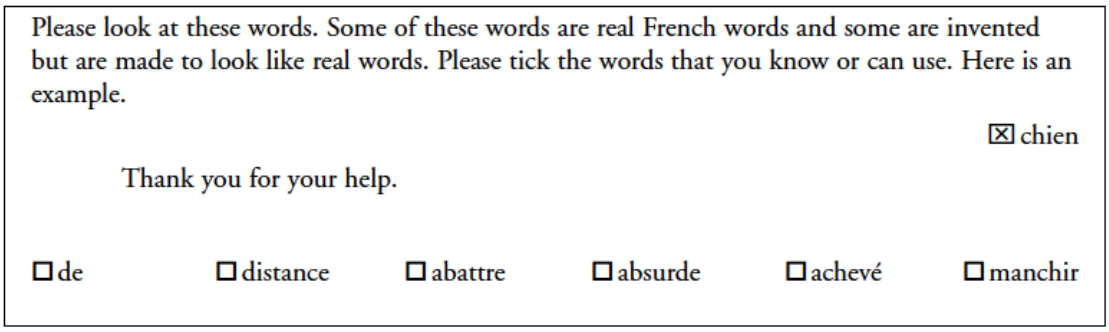


Figure 5. Sample of the French X-Lex Vocabulary Test - paper version.

In order to estimate learners' receptive vocabulary in French, Richards, Malvern, and Graham (2008) also used the X-Lex test. However, they chose to use the computerized version of this test (Meara et al., 2003). Figure 6 shows the opening screen of this receptive vocabulary test. The 23 test-takers were students attending the first year of their non-compulsory education (Year 12), which corresponds roughly to the first year of CEGEP in Quebec. These test-takers were shown 120 real and invented words, one by one, and needed to click on a button to indicate if they knew the word or not. They took between four and seven minutes to complete it. The authors found out that these test-takers knew, on average, 2,437 words.

It is important to point out that this higher average of known words, compared to the 1,930 and 2,108 known words from Milton's (2006) and David's (2008) studies, respectively, can be explained by the fact that Richards et al. (2008) did not deduct points from the yes responses to invented words. Therefore, their results are almost certainly inflated (David, 2008). Nevertheless, Richards et al. (2008) showed that acquisition of vocabulary in French correlates with word frequency; that is, the most frequent words were the most known words among the participants. This was evident by the gradual decline across the five frequency bands shown in the results: the test-takers knew more

words from the 1K level than from the 2K level; they also knew more words from the 2K level than from the 3K level, and so on.

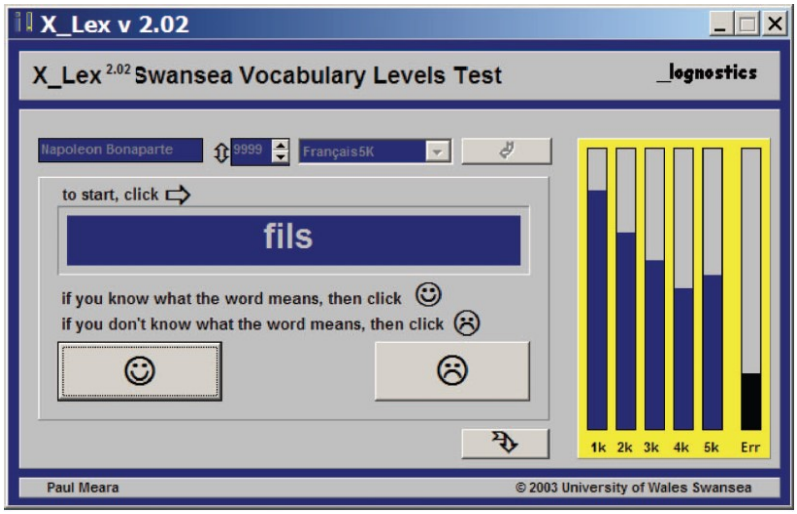


Figure 6. Opening screen of the French X-Lex Vocabulary Test - computerized version.

Milton (2010) sees the development of comparable X-Lex checklist tests in a variety of languages as an important contribution. Test results can be tied to the levels of the Common European Framework of Reference for Languages (CEFR) (p. 228). The CEFR is a document of reference containing proposals for language instruction and common standards to be achieved by L2 learners across Europe. In a number of studies carried out in the UK, Greece and Spain, mainly conceived to estimate how many words English, Greek and Spanish learners of French knew, the X-Lex tests were used to measure their lexical size. Milton (2009) reports that these learners, who were at the CEFR B2 level (which corresponds to 560 - 650 hours of instruction), knew 1,920, 2,450, and 2,630 words, respectively (p. 190). The number of words English learners of French

have, for example, is very similar to the ones the British students on Milton's (2006) and David's (2008) studies know: 1,930 and 2,108, respectively. Milton concludes that taken together, these studies show British students to be rather unimpressive learners of French.

Although the X-Lex test has proven to be suitable for use in comparison studies across different European languages, use of a single measure to estimate learners' receptive vocabulary can be seen as a limitation. David (2008) proposes that these results need to be confirmed by other studies using different measures. This constitutes a further argument for the design and testing of a new measure, such as the Test de la taille du vocabulaire (TTV). This assessment tool is intended to function as the French counterpart of Nation's (1990) VLT, which was identified as a useful model (see discussion in the previous chapter). It is hoped that the TTV will prove helpful to researchers and teachers and will contribute a much needed new instrument for use in L2 vocabulary research.

As mentioned, tests that measure learners' vocabulary size are often based on a sample of words from word frequency lists (Read, 1993). The X-Lex in French, for instance, drew on Baudot's (1992) frequency word list. How about the TTV? Should it use the same list? Are there more recent and up-to-date lists for French, from which the test words in TTV could be selected? In the next session, some of the most important French frequency lists will be examined.

Developments in French frequency lists

Clearly, the success of a test that samples words of differing frequencies depends on a reliable and current frequency list for French. Several frequency lists have been created in this language. Six of those most referred to in the research literature will be

described in order to determine which one can best serve as source from which the words for the vocabulary size test in French, the TTV, will be selected. The treatment here is chronological, and each list will be described based on a number of key criteria that were considered as important for ensuring recent French language usage. Criteria are as follows:

- The list should be generated from an electronic corpus that has been analyzed using frequency software.
- The corpus that the list is based on should be substantial in size (at least one million running words). Corpora smaller than this size cannot be seen as a reliable source of generalizations (O'Keeffe, McCarthy & Carter, 2007).
- The corpus should reflect current French usage. This may seem obvious but many corpora contain large amounts of archaic material.
- The corpus should be representative of the language, reflecting general and academic content from both written and oral sources.
- The lists derived from the corpus should reflect range as well as frequency; that is, a word is considered frequent only if it occurs across a variety of sub-corpora.
- Ideally, the corpus would be relevant to the Quebec setting; that is, it should include materials that reflect French as it is used in Quebec.

1) Le Français fondamental (1er degré)

Containing 1,063 words, this old, precomputational and unlemmatized frequency list, was published in 1954 by the Ministère de l'éducation nationale français (Gougenheim, Rivenc, Sauvageot & Michéa, 1964). This list of supposedly indispensable

words in French emerged from a corpus that included only spoken productions: audio-taped and transcribed conversations about family, health, and vacation, for the most part, with 301 participants, mostly Parisians from all walks of life (engineers, doctors, musicians, nurses, sales persons, farmers, school children, etc.). For a long period of time, almost all the pedagogical materials intended to teach French as a second language used *Français fondamental* as their vocabulary source (Tréville, 2000). This list is clearly a poor candidate for use in building the proposed test. First of all, it is based on corpus of only 312,135 words. Secondly, the corpus, which was built manually, may have errors due to the poor sound quality of the audio-recording (the authors used "light" (6 kg) and inexpensive recording equipment). Thirdly, it is obviously out of date; for example, the words *télévision* ("television") and *ordinateur* ("computer") do not occur. Finally, it also contains a number of chunks, such as *c'est-à-dire* ("that is to say") and *tout d'un coup* ("suddenly"), which are not the type of words that will be tested in the TTV.

2) The Frequency Dictionary of French Words

Published more than four decades ago, the *Frequency Dictionary of French Words* was one of the first lists in French generated by computational analysis. This dictionary was conceived with the sole purpose of becoming a reliable body of evidence for the language, which would present the appropriate structural frameworks, formal procedures, and historical study of French (Juilland, Brodin & Davidovitch, 1970). Parallel to this scholarly purpose, there was a pedagogical one, considered secondary by the authors: to create short lists of most frequent vocabulary for beginner students and longer ones intended for more advanced students. These lists would be a supplement to

advanced reading materials (e.g. academic texts).

To make the list, Juilland et al. (1970) gathered a collection of texts containing 500,000 lexical items from a variety of sources: 90 plays (to compensate for the relative shortness of spoken sources), 50 novels and short stories, 50 essays, 50 newspapers and magazines, and 50 scientific and technical texts. Two word selection criteria were followed: frequency (all words dispersed in fewer than four occurrences were discarded), and range (all words dispersed in fewer than three types of sources were discarded). All these spoken and written productions were part of a larger sample containing 3,078 books, documents, and publications. In order to select the 290 texts that constitute the corpus, which produced a list of 5,082 words, the authors used a computer programmed to generate random numbers. One of the findings reported by the authors is worth mentioning: investigation of the list's coverage indicated that students would be able to understand 90% of the words in a representative text just by knowing their first 1,000 most frequent words. Again, this list is not suitable for the TTV: its corpus of half a million words is not large enough to be considered reliable and it does not contain materials produced in Quebec French.

3) Le Dictionnaire de fréquence des mots du français parlé au Québec

Beauchemin, Martel and Théoret (1992) built an imposing frequency list containing 11,327 words based on a corpus of one million tokens. It was used to create the *Dictionnaire de fréquence des mots du français parlé au Québec* and clearly meets the criterion for Quebec content, since different varieties of French spoken in Montreal, Estrie, Québec and Saguenay-Lac-St-Jean were included; however, it is not a good choice

for building the TTV due to its restricted type of source: it contains only spoken productions, such as transcribed interviews (non-scripted texts), and folktales, theater, television soaps (scripted texts). Moreover, the list includes Quebec forms that differ from standard French (e.g. *char*, "car"), which bring its usefulness for the academic learner into question.

4) Les Fréquences d'utilisation des mots en français écrit contemporain

Aiming at investigating contemporary French vocabulary, Baudot (1992) built a dictionary including 21,684 types, which were the most frequent words found in a larger list of 1,040,150 lemmas. Although this computer-generated dictionary was published in the 1990's, its corpus was created in 1967 by the Bureau des langues du gouvernement canadien. To make this frequency list, Baudot assembled a corpus containing 803 written text samples (from 1,000 to 1,500 words long), dispersed in 15 literary genres (e.g. magazines, books, newspapers, reports, brochures). The corpus consists of written works from the French and Canadian press (62% and 37% respectively), and from other countries (1%).

At first glance, this corpus, including texts written from 1906 up to 1967, could be considered old and out-of-date. However, in fact, it is fairly recent, since 90% of the written productions have their origin in the 60's. For research purposes, it is still considered a list of modern French (Milton, 2009, p. 23).

This list meets almost every criterion specified above: it is generated from an electronic corpus, it is substantial in size, its corpus reflects general and academic content as well as Quebec French usage. However, its texts are restricted to written sources and

contain a portion (10%) of old material. Baudot's (1992) list may be a good candidate for our test. However, a more recent list would be preferable. The search is not over.

5) Word list for a French learners' dictionary

Verlinde and Selva created this substantial frequency list in 2001. It emerged from a 50 million word collection of written texts: 1998 issues of a French and a Belgian newspaper (*Le Monde* and *Le Soir*, respectively). Besides being frequency-informed, this list included words that occurred at least 100 times in both newspapers, which resulted in 12,000 lemmas. These lemmas include single words as well as multiple words (chunks and collocations), and they all represent the most common vocabulary of the current language. Even borrowed words, such as *cool*, *fast-food*, *marketing*, *web*, were incorporated into the list, which was developed with the sole purpose of building a dictionary for a specific kind of L2 French learner: one interested in engaging in everyday conversation and reading the press (Verlinde & Selva, 2001). Even though its sampling is restricted to one type of written source (two national newspapers), this list has the advantage of using recent material, reflecting a more up-to-date vocabulary in French.

After acknowledging some limitations in their list, Verlinde and Selva (2001) concluded their article wishing for a list based on a larger and more varied collection of texts aimed at improving the quality of resources for vocabulary learning. Maybe their wishes came true in the form of the "Frequency Dictionary of French," discussed in the next session.

6) A Frequency Dictionary of French

Conceived by Lonsdale and Le Bras in 2009, this dictionary includes a ranked list of the top 5,000 lemmas in French, derived from a large-scale corpus: the corpus consists of 23 million words contained in a comprehensive range of text types and geographical varieties. Created from transcriptions of spoken and written French sources, it comprises texts used in different French-speaking countries, such as France, Switzerland, and Canada. Examples of spoken texts are governmental debates and hearings, telephone conversations, and film subtitles; of written productions are newspaper and magazine articles, technical manuals, and literature books. It is interesting to note that half of the corpus (around 11.5 million words) is made up of the spoken text portion, the other half, the written one. Although acknowledging the importance of collocations and idioms in L2 learning, the authors decided to build their list with single words only (Lonsdale & Le Bras, 2009).

One of the great advantages of this list is the fact that all texts are recent in origin: only materials that were produced after 1950 were included, reflecting a more modern representation of the French language.

Among the six word lists described above, the Lonsdale and Le Bras (2009) list (the last one) seems to be the most appropriate to be used in the development of the TTV. It was decided to draw the words for the 2,000, 3,000, and 5,000 word sections (henceforth, 2K, 3K, 5K word sections, respectively) from this list. However, a problem remained: how to create the 10,000 word section (henceforth, 10K word section) with the Lonsdale and Le Bras frequency list that only goes as far as 5,000 words? The solution was to use the second best one: Baudot's (1992) list (the fourth discussed above), which

does list words all the way up to the 22,000 most frequent. Only the words included in its 9,000 to 10,000 range were used to create the 10K word section of the TTV.

Research Questions and Hypotheses

The current study has two main purposes: The first is to develop the Test de la taille du vocabulaire (TTV), a French version of Nation's (1983) VLT. This, a receptive vocabulary test, is expected to provide useful estimates of the number of words (lemmas) potential CEGEP and university students know. This test includes four frequency sections, namely, 2K, 3K, 5K and 10K. The second goal is to assess its effectiveness by administering the test to groups of L2 French learners at different proficiency levels, and interpreting the results (quantitatively and qualitatively). The research questions developed for this study are as follows:

1. How do learners perform on the TTV? Is there a proficiency effect such that students in higher groups have higher scores than students in lower groups?
2. Is the TTV implicational such that students' scores are higher on the 2K test, lower on the 3K test, lower still on the 5K and so on?
3. Do learners really know the meanings of words that they have indicated as known on the TTV?

The hypotheses developed for each question are as follows:

1. Research question 1: Learners in higher groups will have higher scores than learners in lower groups.
2. Research question 2: The test will produce higher scores on the 2K section, lower on the 3K section, lower still on the 5K, and even lower on the 10K section.

3. Research question 3: Another measure will show that students know most of the words that they have indicated as known on the test.

The purpose of the first question that investigates whether more proficient learners obtain higher scores than less proficient ones is to determine the extent to which learners' lexical proficiency as determined by their placement in language program relates to their overall vocabulary size as measured by the TTV. This will be explored with groups of L2 learners attending a French course called *Francisation*. This course was designed by the *Ministère de l'Immigration et des Communautés Culturelles* specifically for newly arrived immigrants to aid in their integration into Quebec society.

The second question tackles whether the test provides a scalable profile of vocabulary frequency levels. In particular, it will show how large learners' vocabulary size is at each of the four frequency levels (2K, 3K, 5K, and 10K). It is reasonable to say that the words that recur frequently in the language are most likely to be known (Nation, 2001). Therefore, it is expected that mean score for the entire group on the 2K word section will be higher than the mean on the 3K word section, the mean on the 3K word section will be higher than the mean on the 5K word section and so forth.

As for the third question, which involves gathering qualitative data, it explores the extent to which the test results correspond to the results of an audio-recorded interview. After completing the pilot version of the TTV, learners were interviewed on their knowledge of 48 target words in order to investigate whether they actually knew the words they answered correctly on the test. The goal of this part of the study is to confirm whether they were guessing (or not) while performing the test.

Chapter 3. Methodology

In this chapter, the methodology that was adopted to carry out this research project is discussed. The chapter is divided into four parts: (1) the description of the materials, including a detailed explanation of how the TTV was designed, as well as the kinds of answers elicited by the sociobiographical questionnaire and the interviews with participants; (2) the profiles of two samples of participants: those who took part in the pilot testing phase, and those in the final testing phase; (3) the description of the procedures for administering the pilot test to L2 French learners of different proficiency levels, how their feedback helped to shape the final version of the TTV, and how the results were analyzed; and (4) the key results obtained from the pilot testing. Results based on the improved version that resulted from this process are reported in the next chapter.

Materials

Designing the pilot test

Format

The test followed the model used by Nation (1983, 1990), a matching item test, and drew on guidelines by Schmitt, Schmitt and Clapham, who created an updated and larger version in 2001. Differently from the original VLT, the TTV includes only four sections (2K, 3K, 5K, and 10K word levels), instead of five. The VLT includes these levels plus an academic section, but since no academic word list has been identified for French, this level was not included. Another difference concerns the number of clusters:

although the original 1983 version had six clusters for each of its five sections (a total of 30 clusters), the TTV follows the format by Schmitt et al. (2001), which has 10 clusters per section (a total of 40 clusters). They found that the increase was needed in order for the test to be reliable.

In order to increase the chances of ending up with 10 good clusters per section on the final version of the TTV, the actual pilot version of the test included 12 clusters per section (a total of 48 clusters). Asking test-takers to answer a larger number of clusters made it possible to eliminate the two poorest functioning ones in each section and produce a final version containing the 40 best clusters (10 per section).

Similarly to the original VLT, each cluster consists of six words and three definitions. In sum, the pilot version of the TTV consists of four sections, 48 clusters (12 per section), and 144 test words (36 per section).

Selecting the words and writing the definitions

The lexical content of three sections of the test, namely, the 2K, 3K, 5K levels, was extracted from the relevant sections of the Lonsdale and Le Bras (2009) 5,000 most frequent word list, while the lexical content of the 10K level was extracted from Baudot's (1992) 10th 1,000 most frequent word list. A random number generator was used to help in the selection of words at each frequency level. Four lists of words were produced, one per level. From each randomized word list, six groups of six nouns, three groups of six verbs and three groups of six adjectives were drawn in order to build 12 clusters per level. Interestingly, in the randomized word list used to build Nation's (1983) VLT, the distribution of word classes in English was a ratio of 3 (nouns): 2 (verbs): 1 (adjectives),

according to Schmitt et al. (2001). In the case of the randomized word lists used to build the TTV, the proportion of nouns, verbs and adjectives differs from the one of English with verbs being less frequent. According to the counts obtained from the randomized word lists in French, the proportion of nouns, verbs and adjectives on the TTV should be a 2 : 1 : 1 ratio. In order to respect this distribution as closely as possible among the 48 clusters, the cluster distribution in each section is as follows: 6 noun clusters, 3 verb clusters, and 3 adjective clusters (2K word section), 5 noun clusters, 4 verb clusters, and 3 adjective clusters (3K word section), 6 noun clusters, 2 verb clusters, and 4 adjective clusters (5K word section), and 6 noun clusters, 2 verb clusters, and 4 adjective clusters (10K word section). As a result, the TTV includes 23 noun clusters, 11 verb clusters and 14 adjective clusters (see Figure 7), which taken over the whole test approximates the 2 : 1 : 1 ratio.

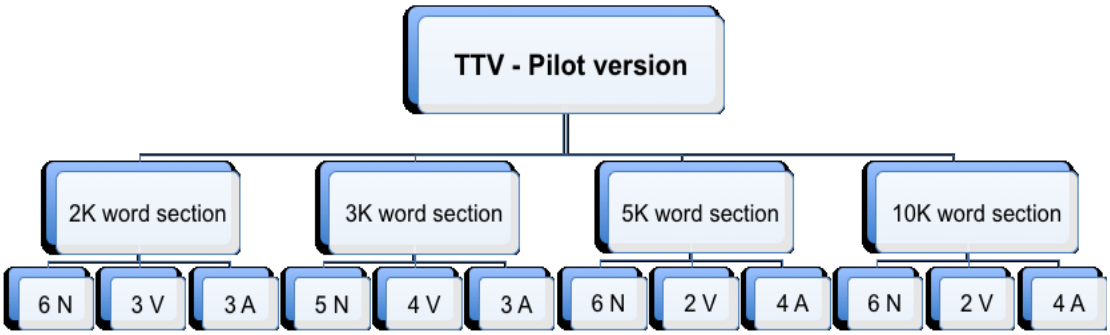


Figure 7. Distribution of clusters in each section on the TTV - pilot version.
Note. N stands for noun (clusters), V for verb, and A for adjectives.

Once the six words in each cluster were chosen, they were then presented in alphabetical order as in the original VLT. Following that, the definitions of all six words in every cluster were written with the help of a paper version dictionary called

Dictionnaire de Synonymes et Contraires (Du Chazaud, 1998), and two on-line dictionaries: *Le Grand dictionnaire terminologique* (Gouvernement du Québec, 2012) and *Portail Lexical* in the Centre national de ressources textuelles et lexicales (CNRTL, 2012). Instead of simply copying the "best" definition of a word chosen among these three sources (which was done in the case of single word definitions), the researcher compared the three and came up with a new definition, which contained as much simple vocabulary as possible (see definition-writing guidelines below).

As stated earlier, the TTV was conceived to assess one single meaning sense of each word. However, the vast majority of the words used in the test were polysemous. The solution to this problem was to choose the sense of a test word according to the meaning used in the French usage context, as stated in the Lonsdale and Le Bras (2009) 5,000 word list, which was said to reflect the most frequent meaning in their corpus (p. 6). This procedure worked very well for three sections of the TTV, namely 2K, 3K and 5K. For the 10K section, which drew words from Baudot's (1992) frequency word list, another solution was needed, since this list did not include definitions or usage contexts: the first meaning sense that was listed in the entry on the on-line dictionary *Portail Lexical* (CNRTL, 2012) was chosen as the sense to be measured in the 10K word section.

Once the definitions of the six words in each cluster were written, then it was time to select three of them randomly (a random number generator was used again), since only three test words per cluster were needed. This exercise showed that a number of the selected definitions did not comply with some definition-writing guidelines, since they contained low frequency words or complex grammar. In the TTV, a definition was considered to be appropriate if it was short (to reduce reading to a minimum),

syntactically simple and used words from a higher frequency level than the test words. Test words from the 2K word section are accompanied by definitions using words from the first 1,000 words in the Lonsdale and Le Bras (2009) list; test words at the 3K, 5K, and 10K word sections were defined using the first and second 1,000 most frequent words from the same list. However, it should be pointed out that a small portion of the test words at the 10K word section was defined using words from the third 1,000 word level (a total of 14). Creating short definitions for the tenth 1,000 most common words in French, having only the 2,000 most frequent words on the Lonsdale and Le Bras (2009) list at one's disposal, turned out to be a very difficult task. Therefore, some definitions were rewritten using some words of the third 1,000 word level, but none that ranked higher than the 3,500 position in the Lonsdale and Le Bras (2009) list. All the initial and rewritten definitions were subsequently checked by a native French speaker with a professional background in Education.

The three definitions in each cluster followed the "length rule" suggested by Schmitt et al. (2001) for ordering answer options: the shortest one precedes a longer one as seen in Figure 8. A sample version of the entire TTV (the 5K word section) that implements these design principles and was piloted with French L2 learners can be seen in Appendix A. The answer key can be seen in Appendix B.

1. brouillard	
2. coïncidence	
3. farce	_____ une histoire qui fait rire
4. instituteur	_____ ce qui empêche de voir loin
5. pneu	_____ un professionnel de l'éducation
6. soumission	

Figure 8. A noun cluster from the Test de la taille du vocabulaire.

The key steps to build the TTV can be summarized as follows:

Building the TTV
1. Identify a frequency word list
2. Select words (nouns, adjectives and verbs) randomly from the list for each section of the test
3. List the words (from the same class of speech) in groups of six (six words = one cluster)
4. Create 10 groups of words on each section (following the 2:1:1 ratio for nouns, verbs and adjectives)
5. Put each group of six option words in alphabetical order
6. Write short definitions for each option word, using high frequency words only
7. Among the six definitions, choose three of them randomly
8. Put the three definitions in length order (from shortest to longest)
9. Display the option words in one column (to the left) and the definitions in another (to the right)

Questionnaire

Before taking the TTV, participants completed a short sociobiographical questionnaire in French (Appendix C), which was used to collect background information (e.g. name, gender, country of origin, L1, language spoken at home, etc.), as well as their attitudes towards learning French as an L2. The first part of the questionnaire (background information) included closed-item and open-item questions. The second part (learning preferences) was mainly developed to elicit the participants' perceptions of the importance of learning grammar and vocabulary. It included two multiple-choice items, each containing a statement and a five-point Likert scale ranging from "very useful" to "not useful."

Interview

In order to investigate whether test-takers actually know the words they answered correctly on a test where it was possible to guess (research question 3), audio-recorded interviews were carried out with a subset of the participants. The individual interviews, which were 30 minutes long, considered three aspects: (1) participants' perceptions of the TTV; (2) how they proceeded to answer the clusters, in particular, an easy cluster and a difficult one (see Appendix D), and (3) meanings of one third of test words, contained in the TTV they had performed the day before. A list of the 48 test words can be seen in Appendix E.

Participants

Pilot testing phase

The pilot test was administered to 63 adult immigrant learners from a variety of first language backgrounds in the spring of 2013. All were enrolled in the *Francisation* program, at a CEGEP in Montreal. There were two proficiency levels: intermediate and advanced. The intermediate level students were attending, at the time of the study, the intensive full-time French course (30 hours per week), and had already attended 700 hours of instruction on average, while the advanced ones, the part-time French writing course (12 hours per week), had already attended 1,200 hours of instruction on average. From the sample of participants, 12 took part in the post-test interviews: five from the intermediate group and seven from the advanced group. It is interesting to note that in this program, vocabulary is taught using theme-based lists (countries, numbers, family professions, etc.), which are not frequency-informed.

Final testing phase

The improved version of the test was administered to 175 learners, 115 female and 60 male in the fall of 2013 at the same institution. They were also enrolled in the same French course as the participants who took the pilot test version. In contrast to the two levels in the pilot testing, they belonged to four different proficiency levels, ranging from 1 to 4 (beginner, low-intermediate, upper-intermediate, and advanced). The students' levels were determined by the institutional placement test: the higher the level, the higher the number of hours the participants attended this French course (see Table 4). This sample population was linguistically and culturally diverse, which is an important requirement to satisfy when investigating the behaviour of this type of vocabulary test (Schmitt et al., 2001).

Table 4

Description of Groups of Participants

Proficiency level	Number of hours of instruction	n	First language	
			Romance	Non-Romance
1	330	31	19%	81%
2	660	55	31%	69%
3	990	64	41%	59%
4	1320	25	72%	28%

Note. N = 175

These learners of French for general and academic purposes come from 39 countries, mainly China, Colombia, Iran, and Moldova. They represented 21 L1 backgrounds: Spanish (38), Persian (26), Romanian (26), Mandarin (23), Russian (20),

Arabic (15), Tagalog (9), Portuguese (3), Ukrainian (2), Vietnamese (2), Amharic (1), Bangla (1), Berber (1), Bulgarian (1), English (1), Hungarian (1), Korean (1), Kyrgyz (1), Nepali (1), Tamil (1) and Teochew (1).

As can be seen in Table 4, the proportions of Romance language speakers vary following a particular pattern: the higher the level, the higher the proportion of participants whose L1 is a Romance language (Portuguese, Romanian, and Spanish).

Although most of the participants (62%) reported residing in Canada for one year or less, the average length of stay was one year and nine months (ranging from three months to 13 years). Also, slightly under half of them (46%) said they had studied French in their countries of origin before coming to Canada, and the same percentage of participants indicated that they do not speak English.

Interestingly, the majority uses French outside the classroom (64%), and thinks that both grammar and vocabulary are important or very important in order to learn French (71%).

Procedures and Analyses

Preliminary version of the TTV

Prior to the administration of the pilot test, five native speakers of French (three graduates from Quebec and two graduates from France) took a preliminary version in order to ensure that all the test items were acceptable to educated native speakers.

They were given as much time as needed to complete the test; average completion time was 15 minutes. Their scores ranged from 141 to 144 with a mean of 142.2 (the maximum score was 144, one point being attributed for each test item). Since these

proficient French speakers reached maximum or near-maximum scores, it is safe to say that the TTV format was not a problem for them. However, some test words posed some comprehension problems for these L1 speakers. In the cases where two of them answered the same test word incorrectly, the definition of that word was rewritten in order to remove the ambiguity. For instance, in the 5K section, the test word *inverser* ("to reverse") should match the definition *changer de place* ("to change the order"). Two test-takers, nevertheless, chose the distractor *desservir* ("to clear the table"), which prompted the test-writer to modify the definition to *changer le sens* ("change the way"). This definition was chosen because none of its senses can be associated with the other distractors, which seemed likely to reduce the confusion. Furthermore, if three or more of these test-takers indicated that they had considered a definition confusing (even though they had answered it correctly), it was also rewritten. For instance, three native speakers disliked the definition *unité de mesure* ("unit of measurement") attributed to the test word *baril* ("barrel") in the 10K section. According to them, this definition is not the first one that comes to mind when they see the word *baril*. In order to come up with an answer, they said they had guessed, choosing their last option after eliminating the others. In view of these comments, a new definition was written: *utilisé pour transporter des liquides* ("used to transport liquids"), which was created in order to avoid test-takers resorting to guessing.

Pilot version of the TTV

This part of the data collection period took 60 minutes. Right before the test administration, the 63 participants spent 10 minutes completing a sociobiographical

questionnaire. While collecting the completed questionnaires, the researcher announced that dictionaries and cellular phones were not permitted and reinforced the idea that the test was individual.

Most of the students completed the entire test of 48 questions in 30 to 35 minutes. This attests to the TTV's practicality.

Post-test Validation Interviews

Interviews with the participants were used in this pilot study to investigate the validity of the TTV. In order to determine whether an item within a cluster is valid or not, it is necessary to determine whether the meaning of this target word on the test is indeed known to a test-taker who has identified a correct definition (Schmitt et al., 2001).

The individual interviews took place on the day after the written test. All interviews were audio-recorded using a digital Edirol Recorder (model: R-09HR) so they could be reviewed and examined for insights into the test taking.

Five participants from the intermediate level group and seven from the advanced group volunteered to take part in it. In this three part interview, which followed a similar procedure used in Schmitt et al.'s (2001) study, the researcher started by posing a question conceived to probe the participants' perception of the vocabulary test they had performed the day before: "Is the TTV a good test of vocabulary?" More often than not, participants answered this question briefly, not spending too much time on it. This initial conversation worked as a warm-up, putting the participants at ease, and prompting them to answer the next questions in a more elaborated way.

In the second part of the interview, which was designed to find out how the

participants proceeded to answer two types of clusters (Appendix D), the researcher always started by showing an easy cluster to the participant, who read the question and spent some seconds trying to answer it silently. Then the researcher asked: "Can you describe what you did while you were answering this question?" and the participant answered it "thinking out-loud." After that, the researcher showed a difficult cluster to the participant and followed the same procedure used with the easy cluster.

In the last part of the interview, the participant's knowledge of the test words was explored. Each was given a sheet of paper containing a list of 48 test words (12 words from each section, see Appendix E), which corresponds to one third of the 144 test words in the TTV. Then the researcher pointed to the first word on the list and asked: "Can you tell me what this word means?" If the participant was not able to answer the question orally (that is, he or she could not come up with an acceptable synonym or definition), the participant was given a card containing the test word and five answer options (the correct definition on the TTV and four distractors). Figure 9 illustrates the card with the word "division" on it, along with its key (option e) and its distractors.

5 division	a. le résultat attendu
	b. une attitude positive
	c. un document officiel
	d. un jour de la semaine
	e. séparation en deux parties

Figure 9. Card used to confirm learners' knowledge of the word "division."

Note that responding to cards such as the one shown in Figure 9 involves answering multiple-choice questions. This format is useful for the validation process in that it presents a different kind of opportunity to demonstrate knowledge than is available on the TTV. That is, test-takers choose from five meaning options for the target word, whereas on the TTV, test-takers must choose one word from a set of six to match to a meaning.

When participants answered this multiple-choice question, the researcher wrote down in her worksheet (Appendix F) whether they had answered it correctly or not. The same procedure was repeated with each of the 47 remaining words on the list. By comparing the participants' responses both in the TTV and in the interviews, it is possible to determine whether they really knew the words. A participant, who answers an item correctly in the TTV and demonstrates knowledge of this same item during the interview, confirms her knowledge of the target word meaning. If the correspondence between the written test and the interview is high, this can be interpreted as the TTV being a good predictor of learners' vocabulary knowledge. This last part of the interview took most of the time allotted. Interestingly, the participants did not seem intimidated by the audio-recorder. For the most part, they were eager to talk and seemed disappointed when the end of the interview was announced.

The final version of the TTV

In order to investigate the validity of the TTV (pilot version), two types of statistics were used: facility and discrimination indices. These indices were calculated for each item in order to explore each cluster's behaviour (recall that there were 144 items or

individual word-to-definition matches within the 48 three-part clusters). But in deciding which clusters to include, the performance of each cluster was considered as a whole.

By calculating the facility index (FI), the proportion of test-takers who answered an item correctly was obtained. Calculating the FI involved adding up the number of correct responses for an item and dividing by the number of test-takers (Fulcher, 2010, p. 182). For the item *sagesse* ("wisdom"), 41 test-takers out of 63 have answered it correctly, and 41/63 = .65. The facility indices for individual items within the clusters ranged from as low as 0.127 (obtained for the item *moisi*, "mouldy") to as high as 1.0 (obtained for the items *hiver*, "winter"; *crier*, "to scream"; and *bras*, "arm"). Facility indices for the 10 best functioning clusters ranged from 0.376 to 0.910. Mean facility indices and standard deviations for each section of the test are shown in Table 5.

By calculating the discrimination index (DI), it was known how well an item discriminated between the top scorers and the bottom ones. Prior to calculating the DI, it is necessary to group the test-takers into three groups (top, middle and bottom scorers) in order to calculate two facility indices: one for the top scorers, which is called FI Top, and another for the bottom scorers, which is called FI Bottom. The DI is obtained by subtracting the FI Top from the FI Bottom, in other words, DI is FI Top − FI Bottom (Fulcher, 2010, p. 182). The discrimination indices for individual items within the clusters ranged from as low as 0.000 (obtained for the items *faim*, "hunger"; *bras*, "arm" and *crier*, "scream" to as high as 0.895 (obtained for the items *ordure*, "trash"; and *fragmentaire*, "fragmentary"). Discrimination indices for the best functioning clusters ranged from 0.175 to 0.684. Mean discrimination indices and standard deviations for each section of the test is shown in Table 5.

Based on the FI and DI values, the poorest functioning two of the 12 clusters in each section were eliminated. Table 5 shows the mean facility and discrimination indices for the 10 remaining clusters in each section.

The item facility means in the third column of Table 5 show that the test is functioning as intended; the means decrease as the test items become more infrequent, with about 80% of the test-takers able to answer the items on the 2K section correctly and the 10K section proving the most difficult. However, the mean FI of the 10K section was higher than expected. Test-takers clearly found these questions harder to answer than the 2K, 3K and 5K items, but the mean of 0.549 is considerably higher than the 0.289 figure Schmitt et al. (2001) found for their 10K section. In fact, it was expected that this section containing low frequency words would be very difficult for most of the participants in this pilot study.

Table 5

Facility Values and Discrimination Indices for Pilot Version

	Number of items	Item facility		Discrimination index	
		M	(sd)	M	(sd)
2K section	30	0.818	0.091	0.318	0.078
3K section	30	0.723	0.147	0.469	0.110
5K section	30	0.654	0.144	0.458	0.099
10K section	30	0.549	0.140	0.549	0.076

Note. N = 63

It was surprising to see that even some participants from the intermediate level group answered a considerable number of items on this section correctly. Closer

inspection of the words targeted in the 10K section and drawn from Baudot's (1992) list, revealed that four of them were not as infrequent as might have been expected. On the more comprehensive and current Lonsdale and Le Bras (2009) list, these same words were identified as 3K and 5K words. For instance, while Baudot (1992) listed *pêcheur* ("fisherman") and *expertise* ("skill") as 10K words, Lonsdale and Le Bras (2009) listed them as a 3K word (rank 2759) and a 5K word (rank 4310), respectively. In view of this discrepancy, the 10K word section was rewritten entirely. A new randomized set of 10K words was drawn from Baudot's (1992) list and checked against both the Lonsdale and Le Bras (2009) list and Cobb's (2000) *Vocabprofil* software to ensure the new items were truly infrequent. The new 10K section consisting of 13 clusters was taken by two native speakers of French, who obtained perfect scores. Their consensus on the questions that they found both clearly written and yet challenging to answer informed the selection of the 10 clusters for the final version of the test. Appendix G shows the new 10K word section of the TTV (under the title *La dixième tranche de mille mots*).

Concerning the mean discrimination indices shown in Table 5, it is important to point out that even though they are above the ideal .3 mark (Fulcher, 2010), the DI for some individual matching items within clusters fell below this criterion in all four sections of the test (recall that each cluster is made up of three definition-matching items). There were 12 such items (40%) in the 2K section, 7 items (23%) in the 3K section, 7 items (23%) in the 5K section, and 6 items (20%) in the 10K section. Thus even though a cluster as a whole discriminates adequately, it may contain items within it that have low DIs. When a test item has a low DI, this means that it is not discriminating well between the top scorers and the bottom ones (e.g. almost all participants answered

an item correctly). The fact that the TTV has retained some items with such low indices could be interpreted as a threat to the validity. However, it is worth noting that many participants in this pilot study were expected to know many (if not all) words included in the "easiest" parts of the test, namely, the words from the 2K and 3K sections, due to their present proficiency level in French. The least proficient participants in this pilot study had already spent 700 hours in class, on average, and could be expected to know as many as 2,800 words. This expectation is based on the learning rate of four words per hour of class time reported by Milton (2009) (700 hours X 4 words per hour = 2,800). In other words, it is not surprising that the TTV, which was conceived to measure learners' vocabulary size over a wide range of proficiency levels, contained some items that were known by most of the test-takers. Schmitt et al. (2001) report low DIs in the development of their test and make a similar argument for the inclusion of such items. In reality, these low DIs attest to the (positive) fact that nearly all of the participants in the pilot testing knew some of the most frequent words in French, which means that they had spent their time wisely, learning the most important words of a language, namely, the high frequency words (Nation, 2001).

A few questions where students were strongly attracted to a particular wrong distractor were rewritten. For instance, in a verb cluster of the 3K section, the distractor *arracher* ("pull out") was flagged due to the high number of participants who had matched this word to the definition *faire naître* ("be born"), instead of *engendrer* ("generate"). The reason why 12 participants (19% of the test-takers) might have been attracted to this distractor could be as follows: Not knowing what *engendrer* meant, they might have tried to come up with another answer, and *arracher* might seem to be the

right answer, since, in the context of a mother giving birth, the baby needs to be pulled out (by an obstetrician). To avoid this kind of "educated guesses" and further confusion, the verb *arracher* was replaced by the next verb in the randomized word list: *commenter* ("comment"). In sum, some of the questions that were removed were omitted on statistical grounds (e.g. low DIs), some because they were confusing (e.g. the distractors were chosen often).

Even though two of the original 12 clusters were eliminated from each section of the pilot test, the distribution of the remaining 40 still follows the 2 : 1 : 1 ratio. On the final version of the TTV, the cluster distribution in each section is as follows: 5 noun clusters, 3 verb clusters, and 2 adjective clusters (2K word section), 5 noun clusters, 3 verb clusters, and 2 adjective clusters (3K word section), 5 noun clusters, 2 verb clusters, and 3 adjective clusters (5K word section), and 5 noun clusters, 3 verb clusters, and 2 adjective clusters (10K word section). As a result, the final version of the TTV includes 20 noun clusters, 11 verb clusters and 9 adjective clusters (see Figure 10), which taken over the whole test approximates the 2 : 1 : 1 ratio.

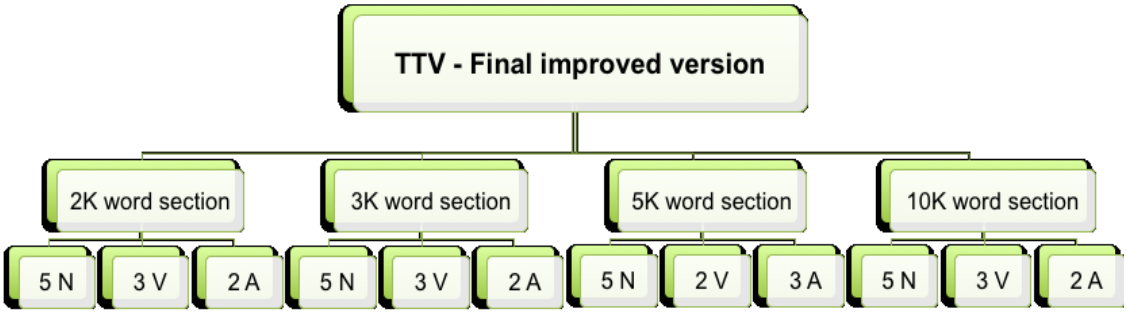
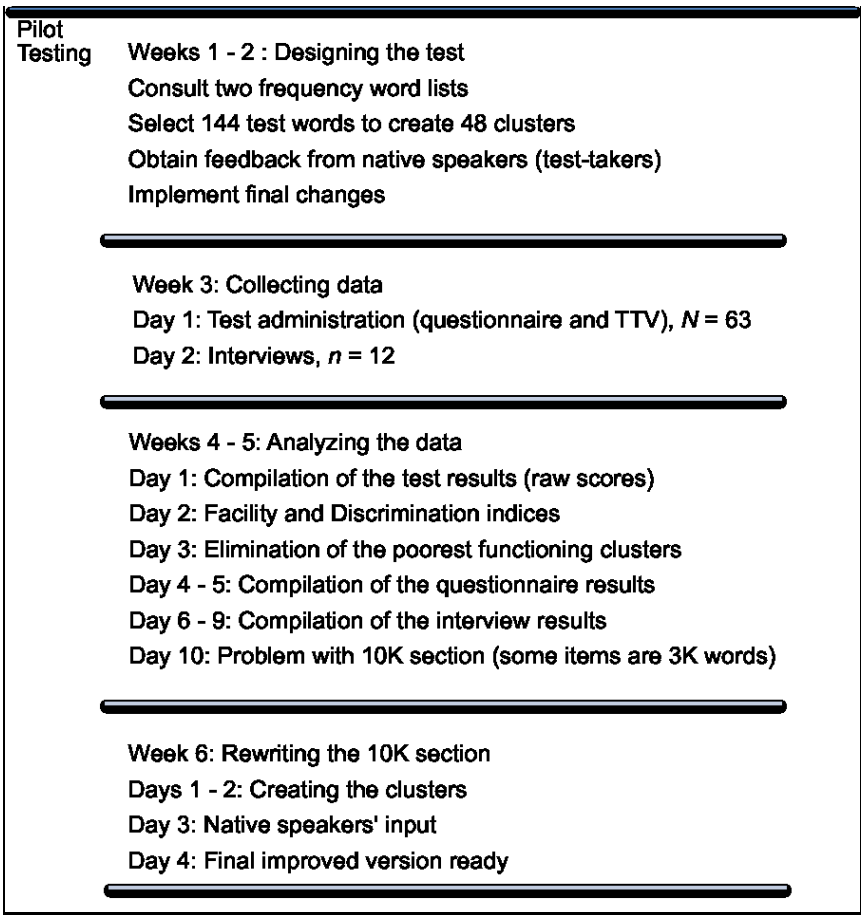


Figure 10. Distribution of clusters in each section on the TTV - final improved version. *Note.* N stands for noun (clusters), V for verb, and A for adjectives.

The participants' feedback provided in the post-test interviews also prompted further revision of the TTV. When they were asked to explain how they had proceeded to answer a cluster, more than half of them mentioned that they resorted to guessing (by using their *doigt magique*, "magic finger," as some would describe their behaviour in a joking way), particularly when completing the 10K section. This behaviour might be due to the assumption that all questions needed to be answered. Based on this kind of input and on the low number of blank responses (995 out of 9,072 responses, 11%), a modification was made to the instructions of the TTV: the sentence *Si vous ne connaissez pas un mot, laissez la réponse en blanc* ("If you do not know the word, leave the answer space blank.") was added in order to avoid blind guessing. Schmitt et al. (2001) also recommend that these instructions be reinforced orally prior to the test administration.

Finally, based on Fulcher's (2010) recommendation of starting the test with the easiest questions in order to motivate the participants, further revisions at this stage included changing the sequence of the clusters in each section in order to display the questions with the highest item facility first.

The sequence of this procedure as well as the steps to analyze the data are summarized below:



The final improved version of the entire TTV that implements all the changes described above can be seen in Appendix G. The answer key can be seen in Appendix H.

This version was given in Fall 2013 to different groups of French L2 learners attending the *Francisation* program. It was held in four sessions on two consecutive days. Similar to the pilot test administration, each session of this final administration of the TTV took 60 minutes: the first 10 minutes was allotted to complete the sociobiographical questionnaire, and the remaining 50 minutes to answer the test questions. Even though the final version of the test was shorter (it included 40 questions instead of the 48 on the pilot version), the participants had the same amount of time to complete it. In each session, all participants started the test at the same time, right after the researcher had

directed their attention to the instructions and reinforced orally the idea that they did not need to resort to guessing.

Participants in three sessions were timed and they averaged 33 minutes (range 14 - 50) to finish the test. Once again, TTV's practicality was confirmed.

No post-test interviews were scheduled at this point of the study.

Results of the pilot testing

The results of the pilot study confirmed that the test was appropriate for the purpose and feasible to administer. Following are some of the most important findings:

- The test was completed in less than 35 minutes by all 63 participants.
- The testing resulted in a range of scores, which was useful for the analysis. Scores across all participants ranged from 31 to 135 (maximum score = 144) with a mean of 99.98 and a standard deviation of 26.23.
- The mean score on all 144 items was 91.36 (SD = 27.45) in the intermediate group and 109.47 (SD = 21.5) in the more advanced group. An independent samples *t*-test showed that this difference was statistically significant ($t = -2.89$, $df = 61$, $p < .003$). Thus these preliminary results confirm the idea that the higher the proficiency level of a group, the greater their vocabulary size.
- When these scores are extrapolated to numbers of words (lemmas), the mean vocabulary size in the intermediate level group amounts to 5,838, and in the advanced group to 7,169.
- Average scores based on all of the piloted clusters in each section (i.e. 36 definition matches per section) show the expected gradual decline across the

frequency levels: from 29.71 (2K), through 26.22 (3K), 24.30 (5K) to 19.75 (10K).

- In the interviews, the TTV was considered a good test of vocabulary by all participants (n = 12).
- All interviewees said that they did not have any problem with the cluster format.

Results based on the final version of the test and the interview data are reported in the next chapter; they are discussed in detail in chapter 5 of this study.

Chapter 4. Results

Research question 1

The first research question asked: "How do learners perform on the TTV? Is there a proficiency effect such that students in higher groups have higher scores than students in lower groups?" Group means for the various proficiency groups are shown in Table 6. As shown in the third column, learners in higher groups obtained higher scores than learners in lower ones (maximum total score = 120): the Level 1 group mean is the lowest at 38.87 (SD = 20.83); as proficiency level increases so do the means, with the highest mean of 92.44 (SD = 13.50) obtained in Level 4. According to the results of a one-way ANOVA, there were significant differences in the data ($df = 3$, $F = 40.97$, $p < .0001$). Post hoc comparisons indicated that all of the between-group differences were significant ($p < .01$).

Table 6

Number of Words (Means) by Proficiency

Proficiency level	Number of items	M	(sd)	%	Extrapolation to 10K
1	120	38.87	20.83	32%	2699
2	120	56.29	22.39	47%	4068
3	120	73.88	29.50	62%	5274
4	120	92.44	13.50	77%	6891

Note. N = 175

The rightmost column in Table 6 shows the mean vocabulary sizes in the various groups; these are obtained by extrapolating the group means to the 10,000 words that the test samples. These data point to a direct relationship between proficiency level and number of "known" words.

Hypothesis 1, which predicted that learners in higher groups would have higher total scores (and larger vocabulary sizes) than learners in lower groups, appears to be confirmed.

Research question 2

The second research question asked: "Is the TTV implicational such that students' scores are higher on the 2K test, lower on the 3K test, lower still on the 5K and so on?" Answering this question involved calculating mean performance for each section of the test in the entire participant group ($N = 175$). The means for each 30-item section are shown in Table 7. The figures show the expected pattern with scores for more frequent words being higher than scores for less frequent words. According to the results of a one-way ANOVA, there were significant differences in the data ($df = 3$, $F = 422.82$, $p < .0001$). Post hoc pairwise comparisons showed that all of the differences in means were significant ($p < .01$). As shown in Table 7, the learners as a group know more than two thirds of the words at the 2K section (69%), but little less than a third of those at the 10K section (32%), and progressively fewer in the sections in between. The declining scores across the word sections clearly indicate that the TTV provides a scalable profile of vocabulary frequency levels.

Table 7

Test Scores (Means) by Section

Section	Number of items	M	(sd)	%
2K	30	20.72	6.59	69%
3K	30	18.25	7.53	61%
5K	30	16.25	7.82	54%
10K	30	9.58	6.00	32%

Note. N = 175

Figure 11 shows the means on the various sections of the test in the four groups. As this figure illustrates, the percentages of correct responses show a consistent pattern of declining scores from the highest frequency level (2K) to the lowest (10K) regardless of learners' proficiency level. It is clear that hypothesis 2, which predicted that the test would produce higher scores on the 2K section, lower on the 3K section, lower still on the 5K, and even lower on the 10K section, is confirmed.

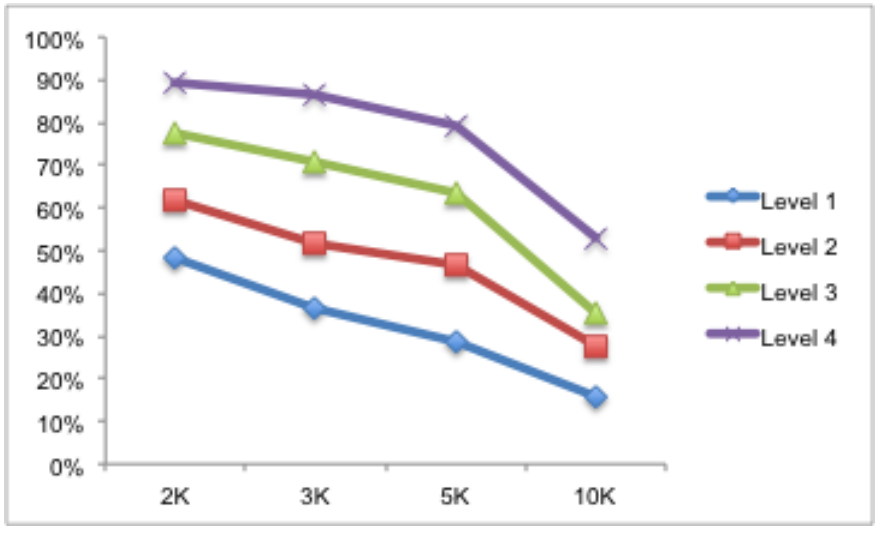


Figure 11. Test score mean per word section and proficiency level.

Research question 3

The third research question asked: "Do learners really know the meanings of words that they have indicated as known on the TTV?" Interviews with a subset of the participants were conducted to answer this question. The post-test interviews elicited 576 answers (12 participants X 48 test words), which fit four different scenarios: (a) correct response on the test and in the interview; (b) incorrect response on the test and correct response in the interview; (c) correct response on the test and incorrect response in the interview; and (d) incorrect response on the test and in the interview. Table 8 shows that scenario a, the perfect match, occurred 68% of the time (390 answers out of the total answers). This indicates that there is a fairly strong correspondence between the test results and the results of the interview.

Table 8

Comparison of Interview Results With TTV's Results

		TTV				
		Correct		Incorrect		
Interview	Knew	a	390	b	60	450
	Did not know	c	46	d	80	126
			436		140	576

Note. N = 12

The hypothesis 3, which predicted that a qualitative measure would show that students knew most of the words that they had indicated as known on the test, is largely

true. The finding that 14% of the words not known on the test were also not known in the interviews (scenario d) also supports the test's validity. Taken together (68 + 14 = 82%), these figures indicate that test performance is a reasonably accurate reflection of what students do and do not know.

The mismatches (scenario b and c), in which participants did not demonstrate knowledge of target words while performing one of the elicitation tasks, occurred 18% of the time (106 answers out of the total answers). However, scenario b occurred slightly more often than scenario c (10% and 8%, respectively), which means that a number of participants performed better in the interview than on the test. In terms of validating the test, there would ideally be very few mismatches. As can be seen by the results, the proportion of mismatches was fairly small, which attests to the TTV's validity.

Chapter 5. Discussion

The goals of this study were to develop and validate a research-informed measure of receptive vocabulary size for French L2 learners. Based on Nation's (1983) VLT and Schmitt et al.'s (2001) guidelines, the TTV was created to estimate learners' vocabulary size up to 10,000 words. This new assessment tool was administered to adult L2 French learners at different proficiency levels and their test scores confirmed its validity. Also, learners' lexical proficiency as determined by their placement in language program relates to their overall vocabulary size as measured by the TTV, which attests to its reliability. It is worthwhile mentioning that the institutional placement/achievement tests are held during a week-long period, and include four subtests, which evaluate students' oral comprehension, written comprehension, oral expression and written expression. The fact that performance on the TTV confirmed learners' level as determined by the results obtained from this thorough testing process can be seen as a strength of this new measure.

In answer to the first research question which explored the relationship between vocabulary size as indicated by the test and L2 proficiency, the results indicate that there was a positive relationship between proficiency level and number of "known" words; in other words, the higher the proficiency level of a group, the greater its mean vocabulary size. At this point, it is pertinent to reiterate that, in the context of this study, the expression "knowing a word" refers to knowing the correspondence between one form and one sense of the word, also known as establishing the word's form-meaning link at the receptive level. The TTV was solely conceived to indicate initial learning of a word, mainly recognition, since only two word knowledge aspects (written form and meaning)

were measured in the current experiment. It is recognized that additional measures, including a measure of production, are needed to give a more complete picture of a learners' vocabulary knowledge.

The second question that explored whether the frequency sections of the test would function as expected was answered in the affirmative. That is, mean performance on the most frequent (2K) words was higher than performance on 3K words and so on with the 10K section proving to be the most difficult. This is an important finding in terms of the test's usefulness. But it is important to note that while this was true of the group, performance of individuals did not always follow this pattern. For example, the individual profiles for six Level 4 learners (24% of this group) did not form a perfect descending scale. One learner, in particular, scored 24 (2K), 29 (3K), 27 (5K), and 18 (10K). Milton (2009) noted similar results in his 2009 overview of vocabulary size testing (p. 33). He identifies word frequency as the strongest predictor of acquisition, with more frequent words being consistently acquired earlier than less frequent ones – across a variety of languages and learning settings – and the findings reported here confirm that. But he points out that individuals cannot be assumed to always follow this trajectory.

Answering the third question involved the investigation of a subset of participants to determine whether responses to test items on the TTV corresponded to knowledge that was assessed in another format (interviews). Results provided strong evidence for the validity of the new test; there was a great deal of consistency between interview and test performance with learners able to demonstrate knowledge of words they had answered accurately on the TTV. A number of other interesting insights emerged from this part of

the research:

First, the test seems to have been received positively. As mentioned, all interviewees answered affirmatively when asked if the TTV was a good test. Some of them, however, considered the test difficult (three interviewees) or long (two interviewees). One student, in particular, added a comment that the TTV was too general, and that she would have preferred to be tested on words from her own field of expertise. Secondly, the interviews revealed students' answering strategies. When asked about their procedures on answering the clusters, two distinct test-taking behaviours were observed: either they use the same strategy to tackle both easy and difficult clusters (eight participants) or they use different strategies for each type of cluster (four participants). Within the former group, for instance, seven participants reported that they read definitions (second column) first and then alternatives (first column), and only one did the opposite (first column then second column). When presented with the easy cluster, the majority of the interviewees answered the three items quickly and correctly. They usually chose the correct alternative without considering the others, treating the items in an independent way. The only exception was two interviewees who answered only two of the three items: both of them did not find among the test words (*évidence*, *hommage*, *maître*, *richesse*, *totalité*, *trafic*) the answer to be associated with the definition/synonym *commerce* (the right response was the last option). They knew that *commerce* meant *magasins* ("stores") or *affaires* ("business"), but when the researcher asked them what *trafic* meant, they answered promptly: 'The circulation of cars,' which is a valid answer. This indicates that they knew one of the meanings of this polysemous word, but not the meaning selected in the TTV: 'circulation of goods,' which prevented them from making

the appropriate form-meaning connection. On the other hand, when presented with the difficult cluster, a small portion of the interviewees answered its three items correctly. However, all interviewees said that, at some point, they had to resort to guessing in order to respond to this type of cluster. For the most part, they read the alternatives more than three times and tried to eliminate the least plausible options, which made them treat the items in a dependent way and spend more time on a cluster. Not surprisingly, this procedure was more time-consuming and oftentimes unsuccessful. In summary, the interviews reveal certain limitations of the TTV's matching format. The single definitions mean that the test may underestimate the knowledge of students who know alternative meanings. It is also clear that the test may overestimate knowledge because the format allows for guessing and elimination.

On the point of guessing, it is interesting to note that instructions added to the TTV's improved version asking participants to avoid blind guessing seem to have been effective. In the pilot testing most students tended to complete the test, which appeared to involve a great deal of guessing. However, at the administration of the improved version, written instructions to leave unknown questions blank were added and reinforced orally by the researcher. This intervention is important to emphasize in future uses of this test for the sake of validity: less guessing means that the measure reflects participants' word knowledge more accurately.

In the next sections, the findings are discussed in relation to two topics: previous estimations of L2 French vocabulary size and the role of L1 influence.

Comparing estimates of vocabulary sizes

Earlier in this thesis, in order to give the reader a preliminary idea of French L2 learners' vocabulary sizes, initial estimates were calculated using Milton's (2009) rule of thumb (learners gain four words/lemmas per classroom hour on average). When comparing this kind of estimates to the ones obtained by the TTV (see Table 9), a clear difference can be observed between the learners Milton investigated and the Quebec participants in the thesis study.

Table 9

Number of Known Words by Proficiency Level: Extrapolations Based on Milton and TTV Results (N = 175)

Proficiency level	Number of hours of instruction	Estimates of vocabulary sizes		
		Milton (2009)	TTV	%
1	330	1320	2699	104%
2	660	2640	4068	54%
3	990	3960	5274	33%
4	1320	5280	6891	31%

In this study, Level 1 students, who attended 330 hours of instruction, know 1,320 words by Milton's (2009) count, but 2,699 words, on average, according to the TTV's results. This amounts to a sizable difference of 1,379 words (104%) or slightly more than the double the claim based on Milton (2009), though it is interesting to note that the higher the proficiency level, the smaller the difference between both estimates. A number of reasons can explain this variation in vocabulary estimates. These include the role of

the learning context and the design of the size test itself. As for context of learning, Milton's investigation of learners of French singled out the Spanish speakers in his sample and this allows for a close comparison to similar test-takers in the current study. The learners in Milton's study had had around 560 - 650 hours of instruction and are therefore roughly comparable to Level 2 group in the TTV study (see Table 9). According to their test scores on the X-Lex (frequency bands 1 through 5), they knew 2,630 lemmas on average (Milton, 2009). For the sake of comparison, the number of words known by the participants in the current study having a similar profile was calculated: the eight learners of French who were at the Level 2 and whose L1 was Spanish knew, on average, 3,875 out of the 5,000 most frequent words (taking together the 2K, 3K and 5K word sections). This large difference in estimates may be explained by the fact that the two groups of participants were fairly different: in Milton's study, the participants were learning French in non-French speaking societies, while the TTV sample were living and working in a French-speaking society. Learners who live in a French speaking society benefit from being exposed to their L2 outside the classroom, and this may explain why their recognition lexicons are much larger compared to those reported by Milton (2009). Future research that tests the TTV in a variety of learning settings may be able to shed light on these issues.

Another reason that might explain the wide variation in vocabulary estimates pertains to the different types of frequency lists used to build the tests the French L2 learners took in the studies. As mentioned, in Europe, vocabulary sizes were measured by the X-Lex test, but in Canada they were measured by the TTV. The former test drew words from the Baudot (1992) list, which was based on a corpus containing written

materials only. The latter test drew 75% of the test words from the Lonsdale and Le Bras (2009) list, which was based on a more comprehensive corpus with a large spoken component (50%). It is possible that by assessing some target words typically used in spoken language, the TTV presents test-takers with higher frequency words, with which they may be more familiar. If indeed the TTV assesses more of these "easier" words than the X-Lex test, then learners have less difficulty in making the form-meaning link, which translates into better scores, and consequently, larger vocabulary sizes.

Previous learned languages and vocabulary sizes

A further point of interest in these results is the role of learners' L1 on their receptive vocabulary knowledge. In order to explore how the L1 impacted test scores, three language groups were investigated: Romance speakers, Asian speakers and non-Romance. The first group consisted of 67 speakers of Portuguese, Romanian, and Spanish, who account for 38% of the population. This particular group is known for having a distinct advantage when learning French (also a Romance language), because they can recognize a great number of cognates, mainly Greco-Latin origin words (Milton, 2009; Schmitt, 2010). The second group consisted of 27 speakers of Korean, Mandarin, Teochew, and Vietnamese (in fact, East Asian languages), who account for 15% of the population. Their L1s are typologically very distant from French and have not been strongly influenced by Latin (as English has been, for instance). The third group consisted of 81 speakers of Persian, Russian, Tagalog and eleven other languages (see Participants subsection), who account for 46% of the population. As seen in Table 10, average scores (maximum score = 30) based on each frequency section showed the

expected gradual decline across the frequency levels, with mean scores on more frequent words consistently higher than those for lower frequency words. However, the analyses show that the mean scores for Romance speakers are substantially higher than the other two groups on all sections of the test. These numbers suggest that these learners used their knowledge of helpful L1 cognates to their advantage. As for the Non-Romance and Asian speakers, mean scores are nearly equivalent in the 2K section in these two groups (17.8, SD = 5.68, and 17.85, SD = 7.49, respectively), and in the 3K section (14.41, SD = 6.06, and 14.30, SD = 7.19, respectively). There is a difference, however, in the most difficult sections, in which the non-Romance group performs slightly better than the Asian group. Perhaps the European (e.g. Russian) language background of some participants in this group offered them a small cognate advantage over the Asian language speakers. A close examination of the role of L1 background on performance on the TTV is beyond the scope of this study, but it is an interesting avenue for further research.

Table 10

Mean Correct Scores for Different Language Groups

Section	Romance speakers	Non-Romance speakers	Asian speakers
2K	25.40 (4.16)	17.80 (5.68)	17.85 (7.49)
3K	24.48 (4.53)	14.41 (6.06)	14.30 (7.19)
5K	21.94 (5.03)	13.04 (6.76)	11.78 (8.23)
10K	13.97 (4.53)	7.02 (5.11)	6.37 (5.30)

Note. Romance speakers (n = 67), non-Romance speakers (n = 81), Asian speakers (n = 27).

It is clear that the learners whose L1 is more closely related to French have an advantage over the other linguistic groups when performing this kind of receptive vocabulary test. This corroborates previous empirical findings (Schmitt et al., 2001) and confirms the importance of L1 in L2 vocabulary development.

Another interesting aspect is the role of previous L2 knowledge (of English and French, in particular). As mentioned, just under half of the participants (46%) said that they did not speak English and the same percentage of them reported that they had studied French in their countries of origin before coming to Canada. Among the Asian speakers, knowing English affected their performance on the test greatly: when looking only at Asian speakers' scores in each proficiency level, all the test-takers who answered the most questions correctly were also the ones who said they knew English. For example, among all the Asian speakers at the Level 3 (nine participants), the highest score (84%) was obtained by a test-taker who knew English. On the other hand, in the Romance group, this pattern was not observed. For example, among all the Romance speakers at the Level 3 (26 participants), the highest score (87%) was obtained by two test-takers: one knew English, the other did not. Studying French before coming to Canada also played a role in test-takers' performance: among the Asian speakers who answered the most questions correctly, 56% of them said they had studied French previously; a similar figure was obtained from Romance speakers (55%). The effects of L1 background and previous study of English and/or French on test performance is an interesting topic for further exploration.

Limitations and tentative solutions

Although the TTV is expected to be useful to SLA researchers and L2 teachers and learners, there are some limitations to this study that should be taken into account.

First, it was quite unexpected to see that six participants outperformed themselves when they were interviewed on their knowledge of the 48 target words. One way of explaining this difference between test and interview performance might be the re-test effect. Although participants were not informed that they would be interviewed on their knowledge of some of the test words, during the interview, four of them said they consulted a dictionary to look up some of the words they had encountered for the first time on the test. The apparent reason for this sudden motivation to learn more words was the realization that, after taking the TTV, their vocabulary size was not as big as they thought it would be. Due to time constraints, the interviews were not scheduled immediately after the test, as Schmitt et al. (2009) scheduled theirs. Since the interviews occurred the following day after the test, some participants had plenty of time to check some words and remember their definitions during the interview. As a result, they were more knowledgeable about these words than they might have been, and this may explain their strong performance on the interview. On the other hand, a few participants performed better on the test than in the interview. One way of explaining this type of mismatch might be that participants resorted to guessing. However, this does not seem to be too much of a problem, since it occurred only 8% of the 576 total cases, as mentioned in the previous chapter.

A second possible measurement limitation is the use of a lemmatized frequency list to build the test. Most L2 researchers measure vocabulary size in terms of word

families. As mentioned in the review of the literature, a vocabulary of 8,000 to 9,000 word families is necessary in order to understand academic written texts in English. Since the frequency word lists used to build the TTV are lemmatized rather than 'familized', the test measured learners' vocabulary size in terms of lemmas known. Thus if a student identifies the correct definition of the lemma *représenter* ("to represent") on the test, strictly speaking, one can assume that she knows the verb forms of this word only. It is reasonable to assume that she also knows words in the same family such as *représentant* ("representative," noun), *représentation* ("performance"), and *représentatif* ("representative," adjective), but, since the frequency list on which the test is based is ordered according to lemmas rather than families, that assumption is questionable. In principle, it is possible to investigate lemma knowledge in relation to reading comprehension: As determined by the TTV, the most proficient learners (Level 4) know an average of 6,891 lemmas. Researchers could explore how much text coverage this knowledge of almost 7,000 French lemmas would yield and whether it would enable these Level 4 learners to understand academic texts adequately. Nonetheless, testing knowledge of lemmas rather than families presents a definite limitation. Can the student who answered the question about *représenter* (a 1K lemma) correctly read and understand a text that contains *représentatif* (a 4K lemma)? The present lemma-based form of the TTV does not allow us to answer that question.

Another important limitation of the study concerns the development of the 10K section of the TTV. The use of two lists, drawn from two very distinct corpora (Lonsdale and Le Bras for the 2K, 3K and 5K sections and Baudot for the 10K) was problematic due to conflicting information they contained: in some instances, a word that was

considered low frequency in one list was high frequency in another. Due to such inconsistencies, the entire 10K section that had been piloted was discarded and another one was created from scratch for the final version of the test. Ideally, this new section would have been piloted and poorly functioning clusters would have been eliminated after calculating the facility and discrimination indices. Fortunately, the participants performed on the new 10K test as expected: their scores on this section containing low frequency test words were the lowest, which indicates that its level of difficulty was the highest. Nonetheless, it is clear that it would have been better to have been able to base the entire test on a single recent and comprehensive word list in French that includes at least 10,000 frequent words. Since such a list does not exist, it seems clear that more corpus work in French is of great necessity.

Although the Lonsdale and Le Bras (2009) frequency list was a very helpful (though incomplete) tool in the development of the TTV, its 23 million word corpus of modern spoken and written French had some limitations. Due to the fact that more than 20% of the corpus consists of European and Canadian parliamentary debates, which require a more formal register, some rare words were ranked as high frequency words, and some basic ones became low frequency. For instance, *clore* ("to close," as in *clore la session*, "close the session"), which is not a basic word in the French language for most users, was ranked at 1827 (or a 2K word), while *cahier* ("notebook"), which is one of the words learned earliest in classroom instruction, was ranked at 4001 (or a 5K word). There is little evidence that such problematic rankings may have affected the test scores, but, again, in order to avoid this problem, it is necessary to create a new frequency count derived from a corpus that reflects a less formal spoken register of the French language.

As mentioned, the research only explores whether learners know the most frequent meaning of the most frequent words. The fact that the test does not permit the testing of multiple meanings of a given words, namely, polysemous words, may be considered a limitation. For example, a learner might know that *bureau* ("desk" or "office") can mean "desk" (particularly, if she is a student and is most likely exposed to expressions such as *bureau du professeur* ("teacher's desk") and *ordinateur de bureau* ("desktop computer") but not know the more frequent meaning "office." One solution that could be envisioned to assess different meanings of a single word is the creation of several versions of the TTV, which would be administered longitudinally. The subsequent versions would contain the words previously tested, with the difference that they would include a new definition, assessing a different meaning sense. For instance, in one version, *bureau* would be matched to its sense of "piece of furniture" as in *bureau du professeur* ("teacher's desk"), and in another version, it would be associated to its sense of "room for work," as in *bureau de l'avocat* ("lawyer's office"). It should be mentioned that this solution was previously described by Beeckmans et al. (2001) as a way to tackle polysemous words in the Yes/No checklist test.

Finally, another limitation of this study concerns the fact that the TTV focuses on measuring receptive vocabulary knowledge only. It does not measure learners' ability to use the test words productively, since it does not assess their pronunciation or spelling skills, for instance. Also, test words are presented in their written forms, so it is not clear whether test-takers who are able to identify correct definitions of words are also able to recognize these words when they hear them in use. In addition, because the TTV only demands test-takers to do a basic and simple task, that is, demonstrate receptive mastery

using test words in isolation, this exercise may give learners a false sense of what knowing a word really is. A way to solve this problem would be to use the TTV as a sub-test included in a larger test, which would consist of several measures, such as a reading comprehension test, presenting the words in different types of contexts of use. It should be pointed out, however, that even if the TTV only asks learners to make basic form-meaning connections, making these fundamental links accurately is crucial in expanding their L2 vocabulary. Essentially, the TTV is a useful starting point in a longer process.

Chapter 6. Implications and conclusions

Implications for research

In the previous chapter, some avenues for future research were suggested or implied. These included various ways of improving the test. First, it is recognized that further validation of the TTV is needed. For instance, administering the test to larger and more varied samples will make it possible to investigate the test clusters more closely. In Schmitt, Schmitt and Clapham's validation of the VLT, they were able to administer the test to 801 participants in five countries, which was clearly beyond the scope of this thesis. But as the TTV becomes available to the research community, it may be possible to conduct a larger-scale investigation and gain further insights on the test's usefulness. If the TTV is tried in other settings where there are different levels of learners, its efficacy in predicting proficiency will also become more clear. A second venue for improving the test pertains to the lists and corpora that were used. As larger and more representative French corpora and frequency lists become available, an improved version of the TTV will become possible. As mentioned earlier, a list of frequent families is important in this regard. An important next step for software developers is to incorporate existing (or improved) French frequency lists into lexical frequency profiling software. This will make it possible to answer questions posed at the beginning of this thesis about matching learners to texts. For instance, it would be useful to know how well a student whose TTV score indicates that she knows the 1K through 3K words in French is able to understand texts written at that level. This kind of research can also answer the question posed at the outset about the level of French needed to read texts at the CEGEP or university level.

Implications for teaching

According to Schmitt (2010), in order to promote vocabulary learning, teachers need to teach words explicitly and expose students to large amounts of language input, especially at the beginning of the learning process. At that point, learners will be inclined to pick up the most basic word knowledge aspects, such as meaning and form, and at the same time it will be appropriate to measure the meaning-form links using a test of receptive vocabulary size as the TTV.

Differently from Szabó (2008), who asserts that "testing hinders rather than facilitates the learning process" (p. 11), the TTV was perceived by participants as a task that encouraged them to learn vocabulary. Most of the interviewees reported that, thanks to the TTV, they became aware that their vocabulary size was smaller than they thought it was and that they were willing to fill this gap by engaging in vocabulary learning. Even the teachers considered the TTV a task that improves vocabulary learning. On an anecdotal note, all teachers, whose students took part in the pilot testing phase of the TTV, allowed the researcher to return to their respective classrooms in the following session and invite their new groups of students to take the final version of the TTV. Surprisingly, some teachers went beyond that and told their colleagues about their learners' positive experience with the TTV, which resulted in a far greater number of participants in this study. On the basis of this experience, it seems clear that the need for a test such as the TTV is recognized by both learners and teachers. Its potential for motivating learning is also evident.

The relatively successful use of the TTV reported in this study indicates that it

attained one of the goals, namely, to devise a test to gain meaningful and reliable information concerning learners' receptive vocabulary size. With this kind of information, teachers can identify gaps in their students' lexical knowledge and focus attention on what appears to be a problematic area. Also, with knowledge of their students' vocabulary size, teachers can choose suitable texts and design level-appropriate instruction, which has the potential to enhance learning. For that reason, efficient testing is of great importance. Only the information obtained by efficient testing overrides the negative side effects caused by poor testing. As Fulcher and Davidson (2007) put it, "[t]he usefulness of assessment, the validity of interpretation of evidence, is meaningful only if it results in improved learning" (p. 35).

Conclusion

Previous research indicates that L2 vocabulary is a problematic area for learners continuing with their study in higher education (Richards et al., 2008). Thus, learners hoping to move beyond basic reading skills (e.g. to read academic texts) must know how large their vocabulary is so they can focus expanding their lexicons in the right direction.

The present study aimed to fill a gap by developing and validating a research-based vocabulary size-test in French. It is hoped that it contributes to the second language acquisition field in providing evidence that receptive vocabulary tests are good tools for L2 learning and furthering understanding of L2 vocabulary development.

References

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. A. Hudson (Ed.), *Advances in Reading/Language Research*. Volume 2 (pp. 231-256). Greenwich, CT: JAI Press.

Assche, S. B.-V., & Adjiwanou V. (2009). Dynamiques familiales et activité sexuelle précoce au Canada. *Cahiers québécois de démographie, 38*(1), 41-69.

Bachman L., & Palmer, A. (1996). *Language testing in practice.* Oxford: Oxford University Press.

Baudout, J. (1992). *Les fréquences d'utilisation des mots en français écrit contemporain.* Montréal: Presses de l'Université de Montréal.

Beauchemin, N., Martel, P., & Théoret, M. (1992). *Dictionnaire de fréquence des mots du français parlé au Québec*. New York: Peter Lang.

Beeckmans, R., Eyckmans, J., Janssens, V., Dufranne, M., & Van de Velde, H. (2001). Examining the Yes/No vocabulary test: Some methodological issues in theory and practice. *Language Testing, 18*(3), 235-274.

Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy* (3rd ed.). White Plains, NY: Pearson Education.

Cameron, L. (2002). Measuring vocabulary size in English as an additional language. *Language Teaching Research, 6*(2), 145-173.

Centre National de Ressources Textuelles et Lexicales (2012). Portail Lexical [Web site]. Retrieved from http://www.cnrtl.fr/definition/

Cobb, T. (2000). The compleat lexical tutor [Web site]. Retrieved in April 5th, 2013, from http://www.lextutor.ca/

Cobb, T., & Horst, M. (2004). Is there room for an academic word list in French? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: selection, acquisition, and testing.* (pp. 15-38). Philadelphia, PA: John Benjamins.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

David, A. (2008). Vocabulary breadth in French L2 learners. *Language Learning Journal, 36*(2), 167-180.

Du Chazaud, H. B. (1998). *Dictionnaire de synonymes et contraires*. Montreal, QC: Dicorobert.

Gougenhein, G., Rivenc, P., Sauvageot, A., & Michéa, R. (1964). *L'élaboration du français fondamental (1er degré).* Paris: Didier.

Gouvernement du Québec (2012). Le Grand dictionnaire terminologique [Web site]. Retrieved from http://gdt.oqlf.gouv.qc.ca/index.aspx

Fulcher, G. (2010). *Practical language testing.* London, UK: Hodder Education.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. New York, NY: Routledge.

Juilland, A., Brodin, D., & Davidovitch, C. (1970). *Frequency dictionary of French words.* Paris: Mouton.

Laufer, B. (1992). Reading in a foreign language: how does L2 lexical knowledge interact with the reader's general academic ability? *Journal of Research in Reading, 15*(2), 95-103.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). 'Lexical threshold revisited: Lexical coverage, learners' vocabulary size and reading comprehension.' *Reading in a Foreign Language*, *22*, 15-30.

Lonsdale, D., & Le Bras, Y. (2009). *A frequency dictionary of French*. New York, NY: Routledge.

Meara, P. (1992). *EFL Vocabulary Tests*. Swansea: Centre for Applied Language Studies, University of Wales.

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing, 4*, 142-154.

Meara, P., & Jones, G. (1988). Vocabulary size as a placement indicator: In P. Grunwell (Ed.), *Applied Linguistics in Society* (pp. 80-87). London: Centre for Information on Language Teaching and Research.

Meara, P., Lightbown, P. M., & Halter, R. H. (1994). The effect of cognates on the applicability of Yes/No vocabulary tests. *The Canadian Modern Language Review, 50*(2), 296-311.

Meara, P., & Milton, J. (2003). *X_Lex, The Swansea Levels Test.* Newbury, UK: Express.

Michel, J.-B., et al. (2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*, 176-182.

Milton, J. (2006). Language lite? Learning French vocabulary in school. *French Language Studies, 16*, 187-205.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. In I. Bartning, M. Martin & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 211–232). Eurosla Monographs Series, 1. Retrieved from http://eurosla.org/monographs/EM01/211-232Milton.pdf.

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines, 5*(1), 12-25.

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 listening? *Canadian Modern Language Review, 63*(1), 59-82.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Office québécois de la langue française. (2002). Combien de mots ? [Text]. Retrieved from http://www.oqlf.gouv.qc.ca/actualites/capsules_hebdo/phistoire_nombre_20030123.html

O'Keeffe, A., McCarthy M., & Carter R. (2007). *From corpus to classroom: Language use and language teaching.* Cambridge: Cambridge University Press.

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review, 56*(2), 282-397.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513-536.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10*, 355-371.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Richards, B., Malvern, D., & Graham, S. (2008). Word frequency and trends in the development of French vocabulary in lower-intermediate students during year 12 in English schools. *Language Learning, 36*(2), 199-213.

Szabó, G. (2008). *Applying item response theory in language test item bank building*. Frankfurt: Peter Lang.

Schmitt, N. (Ed.). (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schmitt, N. (2010). *Researching vocabulary: a vocabulary research manual*. Hampshire: Palgrave and Macmillan.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal, 95*(1), 26-43.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55-88.

Staehr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning, 36*(2), 139-152.

Tréville, M.-C. (2000). *Vocabulaire et apprentissage d'une langue seconde: recherches et théories*. Montreal: Logiques.

Verlinde, S., & Selva, T. (2001). Corpus-based versus intuition-based lexicography: defining a word list for a French learners' dictionary. In *Proceedings of the Corpus Linguistics 2001 conference*, Lancaster University, 594-598.

Wesche, M. B., Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: depth versus breadth. *Canadian Modern Language Review, 53*, 13-28.

West, M. (1953). *A general service list of English words*. London: Longman, Green.

Wilkins, D. A. (1972). *Linguistics in language teaching*. Cambridge, MA: MIT Press.

Appendix A

Test de la taille du vocabulaire (TTV) - 5,000 word section

La cinquième tranche de mille mots

1. brouillard		1. aviser	
2. coïncidence		2. capter	
3. farce	_____ une histoire qui fait rire	3. desservir	_____ informer
4. instituteur	_____ ce qui empêche de voir loin	4. empresser	_____ changer le sens
5. pneu	_____ un professionnel de l'éducation	5. inverser	_____ formuler des conditions précises
6. soumission		6. stipuler	

1. coffre		1. balancer	
2. convoi		2. copier	
3. duc	_____ abri	3. épanouir	_____ mettre en équilibre
4. inondation	_____ caisse où l'on met de l'argent	4. lasser	_____ faire comme l'original
5. nuance	_____ différence entre deux choses semblables	5. naviguer	_____ prendre une certaine partie
6. tente		6. prélever	

1. affection		1. analogue	
2. chancelier		2. barbare	
3. fouet	_____ premier ministre	3. échéant	_____ pareil
4. golfe	_____ un type de sentiment	4. intact	_____ qui n'a pas été touché
5. lance	_____ compétence et expérience	5. philosophique	_____ qui cause une grande surprise
6. qualification		6. stupéfiant	

1. configuration		1. amusant	
2. émeute		2. boursier	
3. harmonisation	_____ vêtement	3. irrégulier	_____ mauvais
4. pantalon	_____ une masse de pierre	4. malin	_____ qui n'est pas uniforme
5. rocher	_____ objet qui ressemble à une boîte	5. naïf	_____ qui concerne le marché financier
6. valise		6. solidaire	

1. cage		1. confiner	
2. panneau		2. énumérer	
3. repli	_____ défaut	3. imputer	_____ isoler
4. sou	_____ prison	4. octroyer	_____ être suspendu
5. traite	_____ monnaie	5. pendre	_____ dresser une liste
6. vice		6. venger	

1. cote		1. anormal	
2. fourniture		2. convaincant	
3. goutte	_____ le milieu entre deux extrêmes	3. éminent	_____ froid
4. hépatite	_____ très petite quantité d'un liquide	4. insensible	_____ bizarre
5. juridiction	_____ des objets utilisés en salle de classe	5. mensuel	_____ qui arrive une fois par mois
6. moyenne		6. séparatiste	

Appendix B

Test de la taille du vocabulaire (TTV) - 5,000 word section - Answer key

La cinquième tranche de mille mots

1. brouillard			1. aviser		
2. coïncidence			2. capter		
3. farce	**__3__**	une histoire qui fait rire	3. desservir	**__1__**	informer
4. instituteur	**__1__**	ce qui empêche de voir loin	4. empresser	**__5__**	changer le sens
5. pneu	**__4__**	un professionnel de l'éducation	5. inverser	**__6__**	formuler des conditions précises
6. soumission			6. stipuler		
1. coffre			1. balancer		
2. convoi			2. copier		
3. duc	**__6__**	abri	3. épanouir	**__1__**	mettre en équilibre
4. inondation	**__1__**	caisse où l'on met de l'argent	4. lasser	**__2__**	faire comme l'original
5. nuance	**__5__**	différence entre deux choses semblables	5. naviguer	**__6__**	prendre une certaine partie
6. tente			6. prélever		
1. affection			1. analogue		
2. chancelier			2. barbare		
3. fouet	**__2__**	premier ministre	3. échéant	**__1__**	pareil
4. golfe	**__1__**	un type de sentiment	4. intact	**__4__**	qui n'a pas été touché
5. lance	**__6__**	compétence et expérience	5. philosophique	**__6__**	qui cause une grande surprise
6. qualification			6. stupéfiant		
1. configuration			1. amusant		
2. émeute			2. boursier		
3. harmonisation	**__4__**	vêtement	3. irrégulier	**__4__**	mauvais
4. pantalon	**__5__**	une masse de pierre	4. malin	**__3__**	qui n'est pas uniforme
5. rocher	**__6__**	objet qui ressemble à une boîte	5. naïf	**__2__**	qui concerne le marché financier
6. valise			6. solidaire		
1. cage			1. confiner		
2. panneau			2. énumérer		
3. repli	**__6__**	défaut	3. imputer	**__1__**	isoler
4. sou	**__1__**	prison	4. octroyer	**__5__**	être suspendu
5. traite	**__4__**	monnaie	5. pendre	**__2__**	dresser une liste
6. vice			6. venger		
1. cote			1. anormal		
2. fourniture			2. convaincant		
3. goutte	**__6__**	le milieu entre deux extrêmes	3. éminent	**__4__**	froid
4. hépatite	**__3__**	très petite quantité d'un liquide	4. insensible	**__1__**	bizarre
5. juridiction	**__2__**	des objets utilisés en salle de classe	5. mensuel	**__5__**	qui arrive une fois par mois
6. moyenne			6. séparatiste		

Appendix C

Questionnaire

Cochez la bonne réponse ou écrivez l'information appropriée.

1. Information personnelle

- Nom _____

- Pays d'origine _____

- Langue maternelle _____

- Langue parlée à la maison _____

- Date d'arrivée au Canada _____

- Genre: Masculin □ Féminin □

- Avez-vous appris le français dans votre pays d'origine? Oui □ Non □

- Parlez-vous français à l'extérieur des cours? Oui □ Non □

- En francisation, vous êtes au niveau 1 □ 2 □ 3 □ français écrit □

- Parlez-vous anglais? Oui □ Non □

2. Préférences d'apprentissage de la langue française

- Apprendre la grammaire est

  Très utile 1 2 3 4 5 Inutile

- Apprendre du vocabulaire est

  Très utile 1 2 3 4 5 Inutile

Merci d'avoir rempli ce questionnaire !

Appendix D

Two clusters used in the interviews:

(1) Easy (2K word band)

1. évidence	
2. hommage	
3. maître	_____ commerce
4. richesse	_____ professeur
5. totalité	_____ beaucoup d'argent
6. trafic	

(2) Difficult (10K word band)

1. bordée	
2. désignation	
3. loucheur	_____ ce que l'on met au pied
4. peluche	_____ personne qui regarde de travers
5. sandale	_____ un objet très apprécié des enfants
6. vaurien	

Appendix E

Word list used in the validation interviews

Lisez les mots suivants et définissez-les:

1. désir
2. mécanisme
3. distribuer
4. circuit
5. division
6. global
7. prudent
8. distinguer
9. traverser
10. fondamental
11. animal
12. joie
13. brutal
14. avertir
15. rigoureux
16. médecine
17. engendrer
18. race
19. record
20. obligatoire
21. caractériser
22. remporter
23. barre
24. cible
25. balancer
26. affection
27. insensible
28. rocher
29. coffre
30. qualification
31. valise
32. goutte
33. mensuel
34. analogue
35. pendre
36. malin
37. récrimination
38. palper
39. cadrer
40. ventilateur
41. isotherme
42. consultant
43. ordure
44. expertise
45. douanier
46. détestable
47. fragmentaire
48. inébranlable

Appendix F

INTERVIEW - TTV

Student: _____ Group: _____ Date: _____ 2013

1) Good test? (Y) (N)

2) Easy cluster: Guessed? (Y) (N) Difficult cluster: Guessed (Y) (N)

3) Confirmation of lexical knowledge:

2000			3000			5000			10000		
1. désir	Y	N	**13. brutal**	Y	N	**25. balancer**	Y	N	**37. récrimination**	Y	N
2. mécanisme	Y	N	**14. avertir**	Y	N	**26. affection**	Y	N	**38. palper**	Y	N
3. distribuer	Y	N	**15. rigoureux**	Y	N	**27. insensible**	Y	N	**39. cadrer**	Y	N
4. circuit	Y	N	**16. médecine**	Y	N	**28. rocher**	Y	N	**40. ventilateur**	Y	N
5. division	Y	N	**17. engendrer**	Y	N	**29. coffre**	Y	N	**41. isotherme**	Y	N
6. global	Y	N	**18. race**	Y	N	**30. qualification**	Y	N	**42. consultant**	Y	N
7. prudent	Y	N	**19. record**	Y	N	**31. valise**	Y	N	**43. ordure**	Y	N
8. distinguer	Y	N	**20. obligatoire**	Y	N	**32. goutte**	Y	N	**44. expertise**	Y	N
9. traverser	Y	N	**21. caractériser**	Y	N	**33. mensuel**	Y	N	**45. douanier**	Y	N
10. fondamental	Y	N	**22. remporter**	Y	N	**34. analogue**	Y	N	**46. détestable**	Y	N
11. animal	Y	N	**23. barre**	Y	N	**35. pendre**	Y	N	**47. fragmentaire**	Y	N
12. joie	Y	N	**24. cible**	Y	N	**36. malin**	Y	N	**48. inébranlable**	Y	N
Total			Total			Total			Total		

Total (Y):_____

Total (N):_____

Date of analysis: _____

Appendix G

Test de la taille du vocabulaire

Nom: _____ Groupe: _____ Date: le _____ octobre 2013

Dans ce test de vocabulaire, vous devez trouver le mot qui correspond à la définition. Mettez le bon numéro à côté de chaque définition choisie. Voici un exemple:

1. accomplir
2. bloquer
3. dormir _____ arrêter
4. encourager _____ aider quelqu'un
5. mêler _____ réaliser ou terminer
6. ressentir

Vous devez inscrire votre réponse de la façon suivante:

1. accomplir
2. bloquer
3. dormir **__2__** arrêter
4. encourager **__4__** aider quelqu'un
5. mêler **__1__** réaliser ou terminer
6. ressentir

Complétez toutes les sections du test. Si vous ne connaissez pas un mot, laissez la réponse en blanc.

La deuxième tranche de mille mots

1. concours			1. brûler		
2. division			2. distinguer		
3. joie	_____	grand plaisir	3. examiner	_____	imaginer
4. phase	_____	un moyen de transport	4. mentionner	_____	remarquer
5. stade	_____	séparation en deux parties	5. rêver	_____	détruire par le feu
6. véhicule			6. supprimer		
1. autorisation			1. fondamental		
2. bonjour			2. global		
3. confusion	_____	erreur	3. moderne	_____	complet
4. faim	_____	le besoin de manger	4. prudent	_____	qui est la base
5. rupture	_____	la maison de la justice	5. récent	_____	qui ne prend pas de risques
6. tribunal			6. traditionnel		
1. adapter			1. attaque		
2. crier			2. contribution		
3. distribuer	_____	partager	3. dommage	_____	institution
4. formuler	_____	parler très fort	4. église	_____	action violente
5. procéder	_____	aller d'un côté à l'autre	5. incident	_____	ensemble de pièces
6. traverser			6. mécanisme		
1. bras			1. actif		
2. circuit			2. inutile		
3. détermination	_____	tour	3. fier	_____	occupé
4. match	_____	principe et règle	4. majeur	_____	qui ne sert à rien
5. réception	_____	une partie du corps	5. puissant	_____	qui a un grand pouvoir
6. théorie			6. scolaire		
1. bâtiment			1. dégager		
2. consultation			2. élire		
3. essence	_____	enquête	3. mériter	_____	choisir
4. habitude	_____	construction	4. persuader	_____	convaincre
5. leçon	_____	la façon normale de faire	5. résumer	_____	raconter de façon brève
6. prestation			6. soulever		

La troisième tranche de mille mots

1. ambassadeur			1. chéri		
2. enfance			2. décisif		
3. portrait	_____	conflit	3. magnifique	_____	strict
4. rayon	_____	première partie de la vie	4. rigoureux	_____	terrible
5. trouble	_____	représentant du gouvernement	5. tragique	_____	celui qui est très aimé
6. vœu			6. vital		
1. aube			1. douleur		
2. docteur			2. minorité		
3. issu	_____	lever du soleil	3. permanence	_____	passage
4. législation	_____	science des lois	4. rédaction	_____	texte écrit
5. préparation	_____	article d'un journaliste	5. sagesse	_____	bon sens et connaissance
6. reportage			6. transition		
1. amateur			1. annuler		
2. cellule			2. cultiver		
3. expansion	_____	petite chambre	3. défaire	_____	travailler la terre
4. profil	_____	un visage vu de côté	4. mentir	_____	ne pas dire la vérité
5. sondage	_____	des questions et des réponses	5. plaider	_____	prendre la défense d'une cause
6. terrorisme			6. siéger		
1. brutal			1. commenter		
2. formidable			2. engendrer		
3. impressionant	_____	dur	3. promouvoir	_____	faire naître
4. mobile	_____	juste	4. remporter	_____	gagner un jeu
5. obligatoire	_____	nécessaire	5. songer	_____	élever à un rang supérieur
6. raisonnable			6. téléphoner		
1. automobile			1. avertir		
2. barre			2. blesser		
3. dominant	_____	espèce	3. caractériser	_____	définir
4. paquet	_____	objet long et étroit	4. déclencher	_____	signaler
5. race	_____	activité d'une personne qui voyage	5. serrer	_____	provoquer un phénomène
6. tourisme			6. trahir		

La cinquième tranche de mille mots

1. cote		1. analogue	
2. fourniture		2. barbare	
3. goutte	_____ le milieu entre deux extrêmes	3. échéant	_____ pareil
4. hépatite	_____ très petite quantité d'un liquide	4. intact	_____ qui n'a pas été touché
5. juridiction	_____ des objets utilisés en salle de classe	5. philosophique	_____ qui cause une grande surprise
6. moyenne		6. stupéfiant	

1. anormal		1. brouillard	
2. convaincant		2. coïncidence	
3. éminent	_____ froid	3. farce	_____ une histoire qui fait rire
4. insensible	_____ bizarre	4. instituteur	_____ ce qui empêche de voir loin
5. mensuel	_____ qui arrive une fois par mois	5. pneu	_____ un professionnel de l'éducation
6. séparatiste		6. soumission	

1. affection		1. coffre	
2. chancelier		2. convoi	
3. fouet	_____ premier ministre	3. duc	_____ abri
4. golfe	_____ un type de sentiment	4. inondation	_____ caisse où l'on met de l'argent
5. lance	_____ compétence et expérience	5. nuance	_____ différence entre deux choses semblables
6. qualification		6. tente	

1. amusant		1. cage	
2. boursier		2. panneau	
3. irrégulier	_____ mauvais	3. repli	_____ défaut
4. malin	_____ qui n'est pas uniforme	4. sou	_____ prison
5. naïf	_____ qui concerne le marché financier	5. traite	_____ monnaie
6. solidaire		6. vice	

1. aviser		1. confiner	
2. capter		2. énumérer	
3. desservir	_____ informer	3. imputer	_____ isoler
4. empresser	_____ changer le sens	4. octroyer	_____ être suspendu
5. inverser	_____ formuler des conditions précises	5. pendre	_____ dresser une liste
6. stipuler		6. venger	

La dixième tranche de mille mots

1. amas		1. astiquer	
2. dortoir		2. foisonner	
3. loque	_____ mots, phrases, règles	3. incarcérer	_____ enfermer quelqu'un
4. huissier	_____ chambre à plusieurs lits	4. marteler	_____ avoir en grande quantité
5. nain	_____ personne de très petite taille	5. railler	_____ faire hésiter entre deux choses
6. syntaxe		6. tirailler	

1. armure		1. bourrer	
2. charpentier		2. crouler	
3. épinette	_____ espèce d'arbre	3. émanciper	_____ remplir
4. granit	_____ vêtement en métal	4. mijoter	_____ détruire ou tomber
5. hypothèque	_____ celui qui construit des immeubles	5. patauger	_____ revêtir d'une couleur
6. torsade		6. teinter	

1. cafard		1. culbuter	
2. étang		2. décupler	
3. fenouil	_____ lac	3. grelotter	_____ souffrir du froid
4. ouïe	_____ plante	4. poncer	_____ couper un arbre
5. répit	_____ pause	5. pulvériser	_____ briser ou réduire en morceaux
6. sangle		6. scier	

1. buée		1. âpre	
2. dicton		2. fulgurant	
3. fût	_____ ce que l'on utilise pour frapper	3. hébété	_____ violent ou rapide
4. garniture	_____ destiné à contenir des liquides	4. rusé	_____ fort, gros et solide
5. maillet	_____ homme qui marche dans les rues	5. trapu	_____ qui est désagréable au toucher
6. piéton		6. utopique	

1. avortement		1. ahuri	
2. matrice		2. dru	
3. persévérance	_____ jeune soldat	3. nutritif	_____ étonné
4. ravitaillement	_____ morceau coupé d'un objet	4. opaque	_____ qui n'est pas transparent
5. recrue	_____ qualité de celui qui a de la patience	5. pondéré	_____ très animé ou mouvementé
6. tronçon		6. trépidant	

Appendix H

La deuxième tranche de mille mots

1. concours			1. brûler		
2. division			2. distinguer		
3. joie	__3__	grand plaisir	3. examiner	__5__	imaginer
4. phase	__6__	un moyen de transport	4. mentionner	__2__	remarquer
5. stade	__2__	séparation en deux parties	5. rêver	__1__	détruire par le feu
6. véhicule			6. supprimer		
1. autorisation			1. fondamental		
2. bonjour			2. global		
3. confusion	__3__	erreur	3. moderne	__2__	complet
4. faim	__4__	le besoin de manger	4. prudent	__1__	qui est la base
5. rupture	__6__	la maison de la justice	5. récent	__4__	qui ne prend pas de risques
6. tribunal			6. traditionnel		
1. adapter			1. attaque		
2. crier			2. contribution		
3. distribuer	__3__	partager	3. dommage	__4__	institution
4. formuler	__2__	parler très fort	4. église	__1__	action violente
5. procéder	__6__	aller d'un côté à l'autre	5. incident	__6__	ensemble de pièces
6. traverser			6. mécanisme		
1. bras			1. actif		
2. circuit			2. inutile		
3. détermination	__2__	tour	3. fier	__1__	occupé
4. match	__6__	principe et règle	4. majeur	__2__	qui ne sert à rien
5. réception	__1__	une partie du corps	5. puissant	__5__	qui a un grand pouvoir
6. théorie			6. scolaire		
1. bâtiment			1. dégager		
2. consultation			2. élire		
3. essence	__2__	enquête	3. mériter	__2__	choisir
4. habitude	__1__	construction	4. persuader	__4__	convaincre
5. leçon	__4__	la façon normale de faire	5. résumer	__5__	raconter de façon brève
6. prestation			6. soulever		

La troisième tranche de mille mots

1. ambassadeur			1. chéri		
2. enfance			2. décisif		
3. portrait	__5__	conflit	3. magnifique	__4__	strict
4. rayon	__2__	première partie de la vie	4. rigoureux	__5__	terrible
5. trouble	__1__	représentant du gouvernement	5. tragique	__1__	celui qui est très aimé
6. vœu			6. vital		
1. aube			1. douleur		
2. docteur			2. minorité		
3. issue	__1__	lever du soleil	3. permanence	__6__	passage
4. législation	__4__	science des lois	4. rédaction	__4__	texte écrit
5. préparation	__6__	article d'un journaliste	5. sagesse	__5__	bon sens et connaissance
6. reportage			6. transition		
1. amateur			1. annuler		
2. cellule			2. cultiver		
3. expansion	__2__	petite chambre	3. défaire	__2__	travailler la terre
4. profil	__4__	un visage vu de côté	4. mentir	__4__	ne pas dire la vérité
5. sondage	__5__	des questions et des réponses	5. plaider	__5__	prendre la défense d'une cause
6. terrorisme			6. siéger		
1. brutal			1. commenter		
2. formidable			2. engendrer		
3. impressionant	__1__	dur	3. promouvoir	__2__	faire naître
4. mobile	__6__	juste	4. remporter	__4__	gagner un jeu
5. obligatoire	__5__	nécessaire	5. songer	__3__	élever à un rang supérieur
6. raisonnable			6. téléphoner		
1. automobile			1. avertir		
2. barre			2. blesser		
3. dominant	__5__	espèce	3. caractériser	__3__	définir
4. paquet	__2__	objet long et étroit	4. déclencher	__1__	signaler
5. race	__6__	activité d'une personne qui voyage	5. serrer	__4__	provoquer un phénomène
6. tourisme			6. trahir		

La cinquième tranche de mille mots

1. cote			1. analogue	
2. fourniture			2. barbare	
3. goutte	**__6__**	le milieu entre deux extrêmes	3. échéant	**__1__** pareil
4. hépatite	**__3__**	très petite quantité d'un liquide	4. intact	**__4__** qui n'a pas été touché
5. juridiction	**__2__**	des objets utilisés en salle de classe	5. philosophique	**__6__** qui cause une grande surprise
6. moyenne			6. stupéfiant	
1. anormal			1. brouillard	
2. convaincant			2. coïncidence	
3. éminent	**__4__**	froid	3. farce	**__3__** une histoire qui fait rire
4. insensible	**__1__**	bizarre	4. instituteur	**__1__** ce qui empêche de voir loin
5. mensuel	**__5__**	qui arrive une fois par mois	5. pneu	**__4__** un professionnel de l'éducation
6. séparatiste			6. soumission	
1. affection			1. coffre	
2. chancelier			2. convoi	
3. fouet	**__2__**	premier ministre	3. duc	**__6__** abri
4. golfe	**__1__**	un type de sentiment	4. inondation	**__1__** caisse où l'on met de l'argent
5. lance	**__6__**	compétence et expérience	5. nuance	**__5__** différence entre deux choses semblables
6. qualification			6. tente	
1. amusant			1. cage	
2. boursier			2. panneau	
3. irrégulier	**__4__**	mauvais	3. repli	**__6__** défaut
4. malin	**__3__**	qui n'est pas uniforme	4. sou	**__1__** prison
5. naïf	**__2__**	qui concerne le marché financier	5. traite	**__4__** monnaie
6. solidaire			6. vice	
1. aviser			1. confiner	
2. capter			2. énumérer	
3. desservir	**__1__**	informer	3. imputer	**__1__** isoler
4. empresser	**__5__**	changer le sens	4. octroyer	**__5__** être suspendu
5. inverser	**__6__**	formuler des conditions précises	5. pendre	**__2__** dresser une liste
6. stipuler			6. venger	

La dixième tranche de mille mots

1. amas			1. astiquer	
2. dortoir			2. foisonner	
3. loque	**__6__**	mots, phrases, règles	3. incarcérer	**__3__** enfermer quelqu'un
4. huissier	**__2__**	chambre à plusieurs lits	4. marteler	**__2__** avoir en grande quantité
5. nain	**__5__**	personne de très petite taille	5. railler	**__6__** faire hésiter entre deux choses
6. syntaxe			6. tirailler	
1. armure			1. bourrer	
2. charpentier			2. crouler	
3. épinette	**__3__**	espèce d'arbre	3. émanciper	**__1__** remplir
4. granit	**__1__**	vêtement en métal	4. mijoter	**__2__** détruire ou tomber
5. hypothèque	**__2__**	celui qui construit des immeubles	5. patauger	**__6__** revêtir d'une couleur
6. torsade			6. teinter	
1. cafard			1. culbuter	
2. étang			2. décupler	
3. fenouil	**__2__**	lac	3. grelotter	**__3__** souffrir du froid
4. ouïe	**__3__**	plante	4. poncer	**__6__** couper un arbre
5. répit	**__5__**	pause	5. pulvériser	**__5__** briser ou réduire en morceaux
6. sangle			6. scier	
1. buée			1. âpre	
2. dicton			2. fulgurant	
3. fût	**__5__**	ce que l'on utilise pour frapper	3. hébété	**__2__** violent ou rapide
4. garniture	**__3__**	destiné à contenir des liquides	4. rusé	**__5__** fort, gros et solide
5. maillet	**__6__**	homme qui marche dans les rues	5. trapu	**__1__** qui est désagréable au toucher
6. piéton			6. utopique	
1. avortement			1. ahuri	
2. matrice			2. dru	
3. persévérance	**__5__**	jeune soldat	3. nutritif	**__1__** étonné
4. ravitaillement	**__6__**	morceau coupé d'un objet	4. opaque	**__4__** qui n'est pas transparent
5. recrue	**__3__**	qualité de celui qui a de la patience	5. pondéré	**__6__** très animé ou mouvementé
6. tronçon			6. trépidant	