It's all about words

Three empirical studies into the role of lexical knowledge and use in French listening and speaking tasks

Proefschrift ingediend tot het behalen van de graad van
Doctor in de Taalkunde
door

Ann-Sophie Noreillie

Promotor: Prof. Dr. Elke Peters
Copromotor: Prof. Dr. Piet Desmet

2019

Contents

Summary of Tables

Summary of Figures

Appendices

List of abbreviations

AWL	Academic Word List
BNC	British National Corpus
CEFR	Common European Framework of Reference
CET4	College English Test in China
CHAT	Codes for the Human Analysis of Transcripts
CHILDES	Child Language Data Exchange System
CIEP	*Centre international d'études pédagogiques*
CLAN	Computerized Language Analysis
COCA	Corpus of Contemporary English
CPE	Cambridge certificate of proficiency in English
DELF	Diplôme d'études en langue française
EFL	English as a foreign language
ESL	English as a second language
HD-D	Hypergeometric distribution of D
ICC	Intraclass Correlation Coefficient
IELTS	International English Language Testing System (measures learners' language proficiency in 9 levels)
L1	First language
L2	Foreign Language
LD	Lexical diversity
LFP	Lexical Frequency Profile
MSTTR	Mean Segmental Type-Token Ratio
MTLD	Measure of Textual Lexical Diversity
PET	Preliminary English Test
PVLT	Productive Levels Test
RLD	Reference Levels Description
SLA	Second Language Acquisition
TAALES	Tool for the automatic analysis of lexical sophistication
TOEFL	Test of English as a foreign language
TTR	Type-Token ratio
TTV	*Test de la taille du vocabulaire*

UWL	University Word List
VLT	Vocabulary Levels Test
VST	Vocabulary Size Test
WAT	Word Associates Test

Voorwoord

Soyons reconnaissants aux personnes
qui nous donnent du bonheur;
elles sont les charmants jardiniers
par qui nos âmes sont fleuries.
(Marcel Proust)

Dit proefschrift is mede tot stand gekomen dankzij het werk en de steun van velen. Alle mensen die, elk op hun eigen manier, deze intense periode voor mij zo onvergetelijk hebben gemaakt, wil ik hierbij dan ook graag bedanken.

Mijn promotores, Elke Peters en Piet Desmet, ben ik zeer erkentelijk voor de gelegenheid die zij mij boden om te promoveren en veelzijdige ervaring op te doen binnen het grote en inspirerende onderzoeksdomein van de lexicale rijkdom. *It's all about words*, maar voor hen toch *a few special words*. Elke, uw expertise, grondige kennis van het woordenschatonderzoek en accurate feedback hebben mijn teksten altijd op een hoger niveau getild. Je hebt me doorheen onze jaren van boeiende samenwerking theoretisch en methodologisch laten groeien. Aan u dank ik ook de verrijkende opportuniteiten om mijn onderzoek op congressen voor te stellen en om op onderzoeksverblijf te gaan in Lund. Hartelijk dank hiervoor!

Piet, ik kan u niet anders bedanken dan in het Frans, *c'est la force de l'habitude*. Un énorme merci de m'avoir parlé de ce projet et de m'avoir ainsi donné l'envie de réaliser cette thèse de doctorat. Merci de m'avoir poussée à donner le meilleur de moi-même. Votre enthousiasme contagieux, votre compétence tant à me motiver dans les moments les plus difficiles qu'à partager sincèrement les joies ou les petites et plus grandes victoires, m'ont aidée à mener à bien cette recherche. Enfin, j'ai été extrêmement sensible à vos qualités humaines d'écoute et de confiance tout au long de ce travail doctoral.

Ik hou eraan om mijn heel oprechte dank uit te spreken aan de leden van mijn adviescomité: Kris Heylen en Fanny Forsberg-Lundell, en ook de overige leden van de jury: Fanny Meunier en Paul Pauwels. Hartelijk dank ook aan voorzitter Jeroen Van Craenenbroeck. Verder ben ik dankbaar dat de Leerstoel Louis Meerts in communicatie een deel van

mijn onderzoek gefinancierd heeft en dank ik Flora Carrijn om mijn verdediging in de ruimtes van Campus Carolus te kunnen organiseren.

Un grand merci au Professeur Jonas Granfeldt pour sa générosité et l'intérêt manifesté à l'égard de ma recherche ainsi que les conseils précieux qu'il a formulés pendant mon séjour de recherche à l'université de Lund.

Zonder onderzoek, geen proefschrift. Mijn oprechte dank gaat dan ook uit naar alle directeurs en leerkrachten van de secundaire scholen waar ik data mocht verzamelen. Ook mijn collega's Carmen, Emmanuelle en Luc wil ik bedanken om hun studenten "uit te lenen" voor mijn onderzoek.

Monique, Serge et Isabeau, un grand merci d'avoir accepté sans problèmes d'évaluer les expressions orales de mes participants, malgré votre charge de travail non négligeable. Serge, merci également d'avoir accepté d'être « médecin pour un jour » pendant ma première étude pilote. Ton intervention a montré qu'il était nécessaire d'avoir un locuteur natif avec lequel les élèves pouvaient interagir.

Ook naar Solène en Ruben gaat mijn bijzondere dank uit voor het verzamelen van een grote hoeveelheid gegevens. Bedankt omdat jullie je master- en bachelorproef hebben gemaakt over woordenschatgebruik en mondelinge interacties. Jullie dataverzameling was een perfecte pilootstudie voor mijn eigenlijke doctoraatsonderzoek.

De verzamelde mondelinge data moesten ook onvermijdelijk getranscribeerd worden, een titanenwerk. Ma chère Marie, un grand merci de m'avoir aidée à transcrire les multiples dialogues. Tu aurais pu faire 1.000 autres choses bien plus intéressantes pendant ta grossesse, mais tu as choisi de sacrifier ton temps pour m'aider. Merci pour ton soutien continuel tout au long de ces années (merci à toi aussi, Gérard !). Je t'en suis bien plus reconnaissante que ce que j'ai déjà pu te le dire déjà.

Britta, mijn *partner in crime* tijdens de eerste twee jaren van mijn doctoraat, jou wil ik in het bijzonder danken voor de mooie tijden die we samen beleefd hebben in Antwerpen, Leuven, Brussel, maar ook in

Utrecht en Aix-en-Provence. Toen je ervoor koos om je doctoraat niet verder te zetten, ben ik je aanstekelijke enthousiasme, vrolijkheid en drive heel hard gaan missen.

Een proefschrift schrijven kan niet zonder de steun en de vriendschap van betrokken collega's. Aan alle collega's van de tweede verdieping, Bert, Isabeau, Maribel, Ine, Annelies, Hans, Serge, Dirk, Anaïs en Frederik, bedankt voor de fijne koffiepauzes en lunchmomenten, voor de onvergetelijke babbels met (meestal) een lach en (soms) een traan. Sofie en Carmen, ook jullie wil ik bedanken voor de vele aanmoedigende babbels en pauzes. Monique, bedankt om rekening gehouden te hebben met mijn doctoraatswerk en je soms over mijn studenten ontfermd te hebben. Eva, zotte momenten genoeg, bedankt om de laatste jaren mijn Antwerpse tijd opgefleurd te hebben. Thao en Marion, bedankt voor de inspirerende onderzoeksgesprekken en feedback.

Bedankt aan Alexis, Eva, Maribel, Matteo, papa en Tom om als welkome vrijwilligers mijn onderzoekswerk grondig na te lezen en van inspirerende suggesties te voorzien.

Maribel, jij bent de ideale overgang tussen collega's en vrienden. Bedankt om er altijd voor me te zijn, voor de vele koffies, lunches, momenten om m'n hart te luchten over van alles en nog wat! Samen met Elisa zijn jullie sinds vele jaren m'n onmisbare lunchpartners. Het plannen van een 'noodzakelijk' shopuitstapje naar Parijs, een koffiemoment of gelijk wat. Gudrun, ook jij verdient een speciale vermelding. Dank je voor je jarenlange vriendschap, voor de vele mailtjes en aanmoedigingen en voor het delen van je (doctoraats)wijsheid. Liefste Kulak-vriendinnen, dat banden gesmeed worden op de Kulak, valt niet te ontkennen. Al meer dan 15 jaar maken jullie deel uit van mijn leven: bedankt om nog steeds samen nieuwe, onvergetelijke herinneringen te creëren! Thomas en Elise, bedankt voor jullie steun en rotsvast vertrouwen in een geslaagd eindproduct.

Lieve END-vrienden, jullie konden vooral het laatste jaar de vooruitgang van m'n doctoraat volgen. Samen mochten we "genieten" van de ups en downs. Bedankt om altijd te luisteren, mee te leven en aan te moedigen.

Jan en Severine, bij jullie kon ik altijd terecht om het doctoraat even te vergeten en meteen meegenomen te worden in de dagelijkse gang van zaken van een warm gezin. Severine, de "club voor echtgenoten van doctorandi" die je wilde oprichten om Alexis te steunen, is daar een mooi voorbeeld van. Clara, mijn liefste metekind, je enthousiaste en lachende gezichtje maakten en maken me nog altijd gelukkig!

Micheline en Jacques, dank omdat we altijd welkom waren om dit doctoraat even te vergeten. Een dosis gezonde zeelucht opsnuiven gaf de magische energie om er weer tegenaan te kunnen. Bedankt ook schoonbroers en -zussen en natuurlijk Stien en Lio omdat jullie me altijd gesteund hebben, meer dan jullie wellicht beseffen.

Mama en papa, bedankt voor jullie niet aflatende aanmoedigingen en jullie onvoorwaardelijke steun, jullie liefde en vertrouwen in me. Voor jullie interesse in mijn onderzoek en voor het soms vooral *niet* informeren naar de voortgang van mijn proefschrift waardoor jullie mij het kostbare gevoel gaven dat mijn leven uit meer bestond dan werk alleen. Marie-Anne, onze "samenwerkdagen" tijdens het weekend motiveerden mij ook om weer aan de slag te gaan en aan het werk te blijven. Pieterjan, ook jij hebt me met je eigenste rustige standvastigheid, je kleine attenties en jouw eigen professionele verhalen geïnspireerd om dit werk af te ronden.

Tot slot wil ik jou, liefste, Alexis, bedanken. Je hebt me tot nu toe enkel gekend in "doctoraatstijden" waar goede en kwade dagen elkaar afwisselen in een even onvermijdelijke beweging als eb en vloed. Ik prijs me gelukkig met jouw onvoorwaardelijke en niet aflatende liefde en steun. Je humor, je kracht om me te helpen alles te relativeren en te geloven in een goede afloop hebben er mee voor gezorgd dat ik dit dankwoord nu kan schrijven. *Tu es le charmant jardinier par qui mon âme est fleurie !* Uit de grond van m'n hart wil ik je gewoon zeggen: duizend maal dankjewel! En wacht maar, een nieuwe start breekt aan!

Chapter 1

Scope and research aims

It's all about words.

It's all about words. The first part of the title of this PhD thesis refers unquestionably to the importance of words in our everyday life. We use words to think, to read, to write, and to express ourselves. It is impossible to imagine a life without words. When you want to learn a new language, one of the first things you will need to do is learning words since they are essential bricks for building a language (e.g., Ellis, 1997; Long & Richards, 2007; Milton & Donzelli, 2013). Indeed, there is nothing more frustrating than wanting to read or say something in a foreign language[1] (L2), but not being able to do so because you do not know the words.

This also explains why lexicon is crucial in language teaching methodology. Whatever method is used, vocabulary is always a component that should be taken into account, not only throughout the learning and teaching activities but also in assessment. One of the main difficulties is to identify which words are crucial at what level of acquisition. Therefore, reference materials like the Reference Levels Descriptors (RLDs) linked to the *Common European Framework of Reference* (CEFR; Council of Europe, 2001) were developed to identify the vocabulary needed at each of the competence levels. However, these reference materials are not always empirically validated and are not linked to the different skills, viz., reading, writing, listening and speaking. Even though research has indicated that there is a strong relationship between vocabulary knowledge and reading (e.g., Laufer, 1992; Stæhr, 2008) and between vocabulary knowledge and writing

[1] Even though L2 theoretically refers to Second Language, it will be used in this PhD thesis to refer to Foreign Language. Although French is one of the official languages in Belgium, Flemish learners' exposure to French in out-of-school contexts is fairly limited (Peters, Noreillie, Heylen, Bulté, & Desmet, 2019; Vlaamse Overheid, 2013), which is why French is considered a foreign language.

(e.g., Laufer & Nation, 1995; Miralpeix & Muñoz, 2018), less is known about the role of vocabulary in listening and speaking.

Therefore, further research into the role of vocabulary in listening and speaking is warranted and was the starting point of the three empirical studies that have been conducted the last four years. More specifically, the aim of this PhD thesis is to investigate the role of vocabulary in French B1 listening comprehension and in B1 dialogic speaking tasks.

1.1 Study object

The main motivation for this PhD project is the challenge of assessing L2 learners' language proficiency, as described in the CEFR, and more specifically the assessment of the four skills, viz., reading, writing, listening and speaking. The CEFR provides "a common basis for the explicit description of objectives, content and methods" (Council of Europe, 2001, p. 1). It was considered a solution to the difficulty that professionals working in the field of modern languages encountered when developing textbooks, examinations and curriculum guidelines across Europe.

Previous research has pointed to a number of issues with the CEFR (e.g., Alderson, 2007; Hulstijn, 2007; Little, 2007; Wisniewski, 2017), such as the lack of an empirical validation of the different scales in the CEFR and the fact that the CEFR does not seem to be based on a theoretical framework. Moreover, the CEFR is language neutral and rather broadly defined, and has, as a consequence, been interpreted in many ways. This is for instance illustrated in the RLDs, which are language-specific interpretations of the CEFR that were developed for particular languages in order to make the guidelines and objectives in the CEFR more concrete and specific. However, there has not been much empirical research into the vocabulary that is needed for specific CEFR levels (Decoo, 2012), which has led to this PhD project. The focus

of this project is threefold: it concentrates on one specific level, viz., the B1 level, one language, viz., French, and vocabulary.

The B1 level has been chosen because it is the attainment target for French at the end of secondary education, at least in general secondary education[2]. Since this is the expected level in secondary schools and for students entering university, learners in their last year of secondary education or in their first year at university were the target participants of our studies.

Next, French is the first foreign language taught in Flanders. It is a compulsory subject from the fifth year in primary school onwards (from 10 years) (Lamote, Desmet, & Janssen, 2014). Given its importance in foreign language teaching in Flanders, we decided to focus on this language.

Finally, vocabulary was chosen because it is explicitly mentioned in the CEFR, albeit in a vague way. As a consequence, the RLDs greatly differ with regard to the vocabulary that is specified and also the number of words that should be known in a specific language at a specific level (Decoo, 2012; Kusseling & Decoo, 2010; Milton & Alexiou, 2009). Moreover, since vocabulary has been shown to be important for reading, listening, writing and speaking (e.g., Schmitt, 2008; Stæhr, 2008), further research into this relationship seemed warranted.

This PhD thesis will focus on spoken communication, i.e., listening and speaking. Recently, there has been an increasing interest in listening and speaking[3] and more specifically in the relationship

[2] The following website provides information on the CEFR and the attainment targets for French and English in secondary schools in Flanders: http://www.erk.nl/docent/erkineuropa/00002/.

[3] In 2012, there has been a large-scale study in Flanders investigating learners' level and motivation in French listening and speaking in the last two years of secondary education (16-18 years) (Vlaamse Overheid, 2013). The report is available on the following website:

between these skills and vocabulary knowledge. However, because most of the studies on listening and speaking focus on English (e.g., Miralpeix & Muñoz, 2018; Stæhr, 2009; Uchihara & Clenton, 2018; van Zeeland & Schmitt, 2013), the question then is whether and to what extent those research findings can be generalized to other languages such as French. Indeed, as was pointed out by Treffers-Daller et al. (2008, p. 271), it cannot be assumed that the results for English can be transferred by definition to other languages. Studies focusing on other languages such as French, Spanish or Dutch have highlighted differences between languages. A second gap, specifically for speaking, is that studies have mostly investigated one type of speaking tasks, i.e., monologic tasks (e.g., Koizumi & In'nami, 2013; Uchihara & Clenton, 2018). Therefore, more research with other task types seems warranted. Third, previous studies researching the relationship between vocabulary knowledge and listening and between vocabulary knowledge and use in speaking has mostly focused on university students (e.g., Milton, Wade, & Hopkins, 2010; Miralpeix & Muñoz, 2018) and not on secondary school students. More research seems needed with other types of participants. Fourth, more research is also warranted with regard to the predictive value of lexical measures in speaking tasks. Even though some studies have investigated the predicting value of lexical measures for experts' ratings of spoken output (e.g., Crossley, Salsbury, McNamara, & Jarvis, 2011a; Iwashita, Brown, McNamara, & O'Hagan, 2008; Kyle & Crossley, 2015), this has not yet been done in French with dialogic speaking tasks.

To summarize, this PhD project concerns the relationship between vocabulary knowledge and listening at the B1 level on the one hand and vocabulary use in dialogic speaking tasks at the B1 level on

---

https://www.vlaanderen.be/publicaties/peiling-frans-in-de-derde-graad-aso-kso-en-tso.

the other. We hope that this project leads to a better understanding of this CEFR level.

1.2 Rationale

This section provides a rationale for carrying out the research presented in this PhD thesis. It is situated in the field of second language acquisition (SLA) and more specifically in the field of vocabulary. Some gaps in the research field will be pointed out and it will be shown how we have tried to answer these gaps in this PhD thesis.

There is a rich tradition in vocabulary research and especially in relation to reading and writing (e.g., Laufer, 1992; Laufer & Nation, 1995; Milton, 2010; Stæhr, 2008). However, some aspects regarding the relationship between vocabulary knowledge and the oral skills of listening and speaking have remained under-researched, as summarized in Figure 1.1 below. The middle column (light blue) of Figure 1.1 displays what has mainly been studied before, whereas the right hand column (dark blue) indicates what we have added or addressed in this PhD project. The figure thus indicates to what extent our studies differ from what has already been done. From a methodological point of view, the studies in this PhD project combine different approaches, i.e., receptive and productive vocabulary knowledge tests, spoken data of two dialogic tasks from L1 and L2 speakers (= vocabulary use), experts' ratings of that vocabulary use and objective lexical measures such as the number of words, their frequency and lexical diversity.

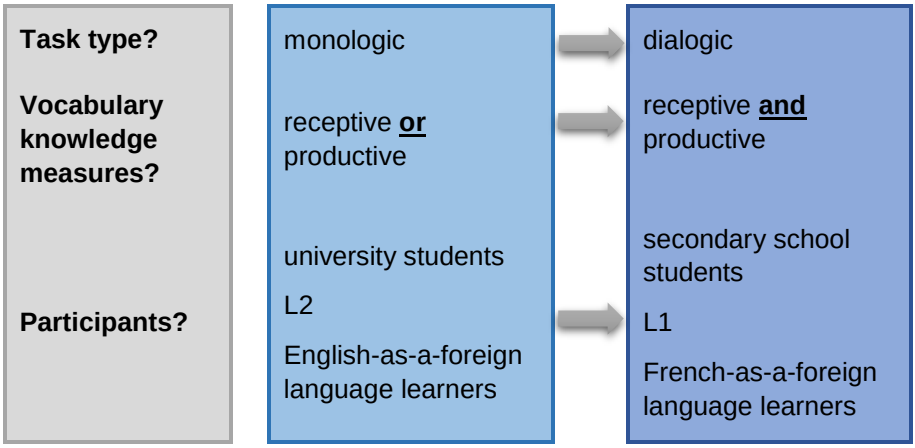


Figure 1.1 Overview of what has been done before (middle column) and what was the focus in the present PhD project (right column)

1.2.1 Task type

For most learners, the aim of learning a language is to be able to use the language in real life situations, which consist, at least for speaking, of interactive communicative settings (Council of Europe, 2001, p. 14). This finding was for instance illustrated in a large-scale research project in Flanders, which showed that 82% of the learners in general secondary education seem to have a larger motivation for learning French for practical reasons (Vlaamse Overheid, 2013, p. 25), such as being able to use the language in daily life. Taking into account the importance of interactive communication in language learning, there seems to be a clear need for more research on dialogic tasks, rather than on monologic tasks such as picture describing tasks, as used in earlier research (e.g., Koizumi & In'nami, 2013; Uchihara & Clenton, 2018).

1.2.2 Vocabulary knowledge measures

Even though there is a growing interest in the relationship between vocabulary and speaking (e.g., de Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012; Uchihara & Clenton, 2018), most studies

have only administered one receptive vocabulary test (e.g., Koizumi, 2005; Miralpeix & Muñoz, 2018; Uchihara & Saito, 2016). Since speaking is a productive skill, it is expected that the relationship between a productive vocabulary knowledge test and this skill should be higher than between a receptive vocabulary knowledge test and speaking. However, no study has combined the use of a receptive and a productive vocabulary test in order to measure vocabulary knowledge in relation to vocabulary use in speaking.

1.2.3 Participants

Most of the previous studies that have looked into the relationship between vocabulary and L2 skills have focused on English-as-a-foreign language and on high proficiency learners, viz., mainly university students. Therefore, studies focusing on another language, e.g., French, are warranted. In addition, research targeting other proficiency levels, for instance low-intermediate learners is needed. Finally, adding data of L1 speakers could provide insights on the cross-linguistic comparison of the relationship between vocabulary and speaking.

Drawing on the findings presented in Figure 1.1 above, two general research questions can be expressed:

- To what extent are vocabulary knowledge and use related with the two oral skills, viz., listening and speaking?
- Are lexical measures a good predictor of L2 learners' vocabulary use in spoken output?

These two questions will be subdivided into more detailed research questions for the three empirical studies that have been conducted. They will follow in the next section.

1.3 Global architecture and research objectives

Figure 1.2 visualizes the global architecture of the three studies that are part of this PhD thesis. As can be seen, the structure of this research project can be organized around two major axes: lexical knowledge and use on the one hand and two oral skills, i.e., listening and speaking on the other. Lexical knowledge refers to learners' knowledge of vocabulary, whereas lexical use refers to how well learners can use the words they know in speaking. Listening comprehension is defined as the capacity to understand the meaning of spoken words in context and the ability to understand them in a specific context. In this PhD thesis, we focus on speaking as the capacity to retrieve words from memory and to use them in interactive communication. The two axes are represented in the blue columns.

As illustrated in Figure 1.2, study 1 investigates the relationship between vocabulary knowledge and listening comprehension and studies 2 and 3 the relationship between vocabulary knowledge and vocabulary use in speaking. Additionally, study 3 also focuses on the prediction of learners' rated output.

In this PhD project, different data collection instruments were used for operationalizing the constructs. First, receptive and productive vocabulary knowledge tests were used to measure learners' written receptive and productive vocabulary knowledge. Results on these tests were correlated with learners' results on a listening comprehension test (study 1) and with objective lexical measures that evaluated learners' spoken output (study 2). The same lexical measures were further used to predict human ratings of learners' spoken output (study 3).

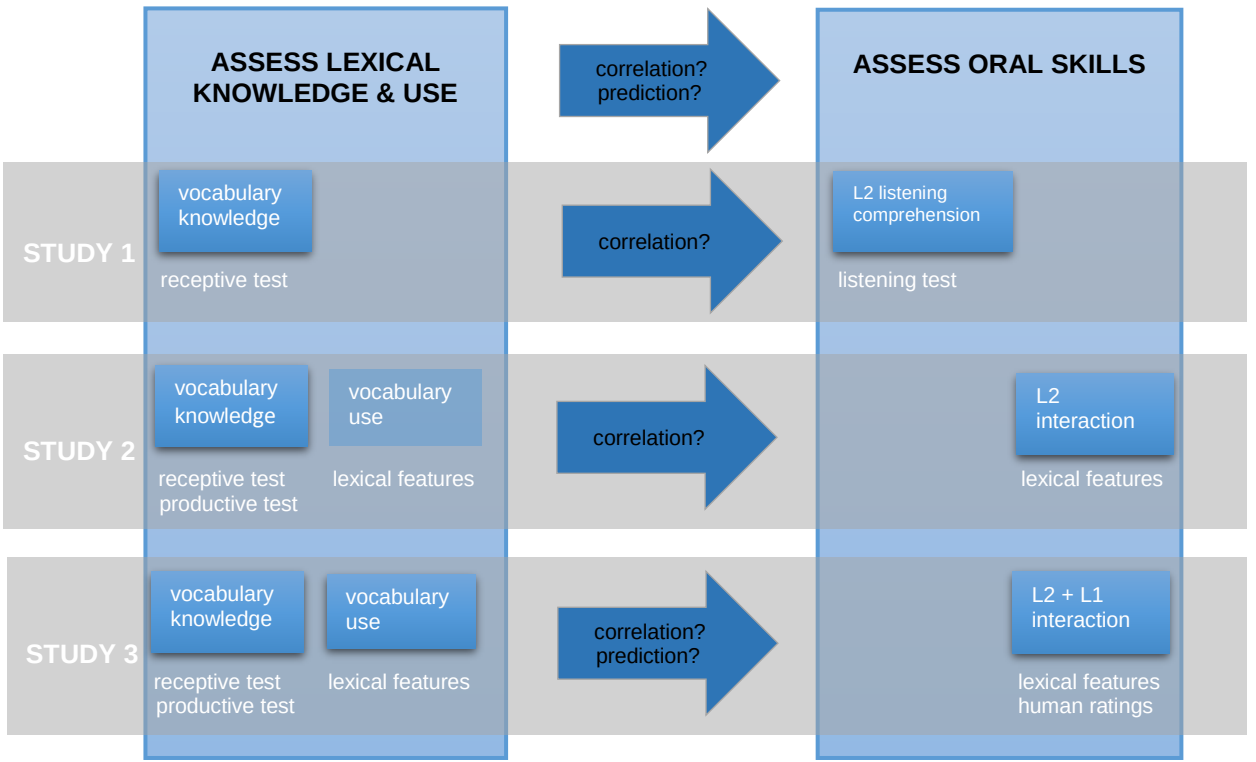


Figure 1.2 Visualization of the global architecture of and interaction between the three studies.

Drawing on the global architecture of this PhD project, several detailed research questions have been studied within three empirical studies. They can be found in Table 1.1 below.

Table 1.1

Research questions that guided the three empirical studies

Study 1	1.	To what extent is **vocabulary knowledge** associated with **listening comprehension** at the B1 level?
	2.	**How much vocabulary** is needed for adequate **listening comprehension** at the B1 level?
Study 2	3.	What is the relationship between learners' **receptive and productive vocabulary knowledge** and the number of **types and lemmas** produced in two semi-structured dialogic speaking tasks?
	4.	What is the relationship between learners' **receptive and productive vocabulary knowledge** and the **lexical frequency profile** of their spoken output in two semi-structured dialogic tasks?
	5.	What is the relationship between learners' **receptive and productive vocabulary knowledge** and the **lexical diversity** of their spoken output in two semi-structured dialogic tasks?
Study 3	6.	Which **lexical factors predict experts' holistic rating** scores of L1 and L2 speakers' oral output in two semi-structured dialogic speaking tasks?
	7.	How does **L1 and low-intermediate L2 speakers' output** in two semi-structured dialogic tasks differ as far as the number of (1) tokens, types and lemmas, (2) the lexical frequency profile and (3) lexical diversity are concerned?
	8.	What is the relationship between low-intermediate L2 learners' **receptive and productive vocabulary knowledge** and **raters' judgement** of holistic ratings of L2 learners' spoken output in two semi-structured dialogic tasks?

1.4 Structure of this PhD thesis

This PhD thesis is organized in seven chapters. After this first chapter specifying our scope and research aims, the next two chapters focus on the theoretical and methodological framework. The following three chapters are built around the three empirical studies. Each study is described in one chapter, which can be read on its own. There might, therefore, be some overlap between them. Finally, the last chapter provides a discussion of the most important findings and some global conclusions.

Chapter 2 focuses on the theoretical grounds on which our experimental studies are based. We will elaborate on vocabulary acquisition and more specifically on what it means to know a word and to use it in the receptive and productive skills of listening and speaking.

Chapter 3 features the methodology of the three studies. It gives information on the methodological choices that were made and the different data collection materials that were used.

In Chapter 4, we present the first study in which we replicated Stæhr's (2009) study on the relationship between vocabulary knowledge and listening comprehension. Two experiments were conducted, one for English and one for French. The results of study 1 are published in *ITL - International Journal of Applied Linguistics* (Noreillie, Kestemont, Heylen, Desmet, & Peters, 2018).

Study 2 (Chapter 5) focuses on speaking. In this chapter, we investigated the relationship between L2 learners' receptive and productive vocabulary knowledge and their vocabulary use in two semi-structured dialogic tasks by looking at the lexical frequency profile of learners' output and the lexical diversity of their output. This chapter is being prepared for submission to a peer-reviewed journal.

In study 3 (Chapter 6), we build on the previous study by adding spoken data of French L1 speakers and by asking raters to assess both

the L1 and the L2 speech data. The study focused on the differences in lexical output between L1 and L2 speakers and investigated whether lexical factors (i.e., number of tokens, types, lemmas, the lexical frequency profile of the oral output and its lexical diversity) could predict raters' holistic lexical scores. The study has been submitted to a peer-reviewed journal.

The conclusion (Chapter 7) provides a summary of our main findings. In addition, the chapter consists of a general discussion and of the conclusions that can be drawn from the three studies. Finally, we highlight the implications of our studies and formulate some recommendations for future research.

Chapter 2

Theoretical grounds

2.1 Introduction

This chapter is concerned with the theoretical grounds on which the three empirical studies of this PhD project are based[4]. The second section (2.2) focuses on one of the main constructs under investigation, viz., vocabulary The third section (2.3) deals with the relationship between vocabulary and the four skills, i.e. reading, writing, listening and speaking. The last section (2.4) zooms in on the *Common European Framework of Reference*.

2.2 Vocabulary acquisition

In what follows, the meaning of a "word" is defined (2.2.1) and it is discussed what it means to know a word (2.2.2). Further, this section also deals with the different ways in which vocabulary knowledge (2.2.3) and use (2.2.4) can be measured.

2.2.1 What is a word?

Before defining what it means to know a word, it should be made clear what a word is. The *Longman Dictionary of language teaching and applied linguistics* (Richards & Schmidt, 2002) defines a word as follows: "the smallest of the linguistic units which can occur on its own in speech or writing" (p. 588). It is elucidated, however, that it is hard to apply this criterion consistently, since it is questionable whether function words (e.g., *a, the, of*) can occur on their own or whether contractions (e.g., *don't = do not*) should be considered one word or two distinct words. It is further specified that "in writing, word boundaries are usually recognized by spaces between the words. In speech, word boundaries may be recognized by slight pauses" (ibid.).

[4] It should be noted that there might be a certain overlap between the theoretical grounds that will follow in this chapter and the theoretical background given in each of the three empirical studies, since the three chapters concerning the studies can be read as separate chapters.

When the words in a sentence have to be counted, the numbers will depend on the counting unit that is used. How many words does the following sentence count ?

"A good book is a book that you like."

This depends on the way words are counted. In this sentence, there are 9 tokens, viz., 9 words in the sentence, 7 types, viz., 7 different words (the words 'a' and 'book' are only counted once) and 7 lemmas. The counting unit that is chosen has important consequences for test development (see section 2.2.3).

2.2.2 What is knowing a word?

The question of what it means to know a word has been extensively studied (e.g., Nation, 2013; Read, 2000; Schmitt, 2000, 2010). The answer is, however, complex and multifaceted. Nation's framework of word knowledge (2013) is probably the most referred to (Figure 2.1). It identifies three main aspects in the knowledge of a word, i.e., form, meaning and use, which are each divided into three components. For each of the nine components of Nation's model, he distinguishes between a receptive (R) and a productive aspect (P), in which the receptive aspect is assumed to be easier than the productive one. If a word is part of learners' receptive vocabulary, many aspects of both knowledge and use are involved, as well as the skills of reading and listening (p. 50).

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognizable in this word?
		P	What word parts are needed to express the meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	constraints on use (register, frequency,…)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

Figure 2.1 Nation's (2013, p. 49) word knowledge framework.

As mentioned before, Nation (2013) identifies two types of vocabulary, i.e., receptive and productive vocabulary. He points out that receptive vocabulary "involves perceiving the form of a word while listening or reading and retrieving its meaning" (p. 47), whereas productive vocabulary "involves wanting to express a meaning through speaking or writing and retrieving and producing the appropriate spoken or written form" (p. 47). Learners' receptive vocabulary is usually thought to be twice as large as their productive vocabulary (Laufer, 1998; Laufer & Goldstein, 2004; Laufer & Paribakht, 1998; Waring, 1997; Webb, 2008).

However, instead of using the distinction between receptive and productive vocabulary, some researchers use the labels passive and active vocabulary (e.g., Laufer, 1998; Meara, 1990). Passive vocabulary refers to receptive vocabulary knowledge. It involves knowing and understanding the most frequent meanings of a word, and is associated with reading and listening. In that case, learners receive information from others through listening or reading, whereas active vocabulary on the other hand refers to productive vocabulary knowledge. Learners can use the words without any given prompts, they produce language output by writing or speaking in order to convey a message to other people.

Read (2000), Schmitt (2010) and Nation (2013) also distinguish between recognition and recall. In recognition, learners "are presented with the target word and are asked to show that they understand its meaning" (Read, 2000, p. 155), whereas recall refers to retrieval from memory. Learners "are provided with some stimulus designed to elicit the target word from their memory" (ibid.). Further, meaning recognition and meaning recall usually refer to receptive vocabulary knowledge and are associated with reading and listening. Form recognition and form recall refer to productive vocabulary knowledge and are associated with speaking and writing. In this PhD thesis, the

recognition-recall distinction will be further used in relation with the tests measuring vocabulary knowledge.

Another important term for the current PhD project that should be defined is lexical competence. Even though it is a daunting project to try to define lexical competence given its multifaceted nature, some researchers have attempted to. According to Meara (1996), lexical competence consists of two parts, viz., vocabulary size and organization. The first important component of lexical competence, vocabulary size, is also called vocabulary breadth and concerns the number of words that are known. The second component, i.e., organization, is related to the depth of vocabulary and investigates how well words are known. However, researchers have proposed to add a third component (Laufer & Nation, 2001): fluency, which is the ease with which learners can access and use the words they know (Daller, Milton, & Treffers-Daller, 2007; Laufer & Goldstein, 2004; Schmitt, 2010). The faster and the more accurate a word can be used in writing or speaking, the better it is 'known'. Speed and accuracy could thus be tested in order to measure the degree to which a learner masters a word. These three components form a 'lexical space' consisting of three axes (vocabulary breadth, depth and fluency) in which learners' vocabulary is situated. Attempts to convert this metaphor into a model have, though, failed (e.g., Read, 2000, 2004).

Further, the distinction between vocabulary size or breadth on the one hand and vocabulary depth on the other can be made. This distinction was first made by Anderson and Freebody (1981, p.92-93) and has since than been widely adopted (e.g., Read, 2000). Learners' vocabulary size refers to the total number of words they know, whereas vocabulary depth refers to how well the words are known. While some researchers have shown that the distinction is valid as depth measures add explanatory power compared to size measures (e.g., Qian, 1999, 2000), other evidence has shown that there exists a strong correlation between size and depth (e.g., Vermeer, 2001) and that they could be

considered one construct. However, most researchers agree that the two constructs are different and complementary and the distinction therefore still seems valid up till now.

2.2.3 Measuring vocabulary knowledge

In this section, we will first look at the distinction between lemma and word family, the importance of frequency in vocabulary learning and the most well-known vocabulary tests. Finally, attention will be paid to how vocabulary use can be measured by studying its lexical diversity.

Lemma versus word family

As has been mentioned in section 1.2, vocabulary knowledge is often defined as the number of words learners know, i.e., vocabulary size or vocabulary breadth (Schmitt, 2014). It is important to note that, in order to measure vocabulary size, several measuring units can be used, i.e., lemmas or word families.

A lemma consists of words that are closely related, i.e., "a headword and its inflected and reduced forms (*n't*)" (Nation, 2013, p. 10), which usually belong to the same part of speech group (e.g., *arbre, arbres; lire, lisons, lu*). An argument in favor of the use of lemmas instead of word families is the learning burden, which can be defined as the effort that is necessary for learning a word. Thus, if learners know the word *livre* (book), the learning burden for *livres* (books) will be low, whereas the learning burden for learning an irregular plural like *yeux* (eyes) (*œil* = eye) will be much higher. Nation (2013, p.10) points out that it should be decided on whether these words are considered as separate lemmas (*œil*, *yeux*) or as one same lemma. Further, the headword should be chosen, which can be the base word or the most frequent form of the base word.

Like the lemma, a word family includes a headword and its inflected forms, but also "closely related derived forms" (Nation, 2013,

p. 11). Since derivations often change the word class of a base word, the words that are part of one word family can belong to different word classes (e.g., *développer, développement, redeveloper*) (Read, 2000). Lemmas would thus be *lire, lisons* and *lu*, but *livre, lecteur* and *lecture* would additionally be part of the same word family.

Different estimates of vocabulary size thus depend on the counting unit that is being used, i.e., the lemma or the word family. In case the lemma is being used as the counting unit, it should be noted that the numbers are much higher than when word families are the counting unit. For English, most vocabulary tests for English use word families as counting unit (Nation, 2001, p. 585). For French, on the contrary, the counting unit that has been used most often is the lemma, because the few frequency lists that exist are based on lemmas.

Recently, Kremmel (2016) has questioned the word family as a counting unit. He argues that it cannot be taken for granted which words should be part of one word family or not (Nation, 2013, p. 11). There are different levels of affixation, going from some obvious affixes such as *-able* or *-ly* to more difficult suffixes such as *-wise, -ment* or *-like,* that are more difficult to acquire (Bauer & Nation, 1993) and that might not be known by beginner learners. Moreover, words like *deceive* and *deception* are part of the same word family, but it cannot be assumed that learners will store these two words under the same single mental entry (Schmitt, 2010, p. 190). Even though the most frequent prefixes and suffixes are thought to be known by beginning learners, the development of less frequent prefixes and suffixes evolves as learners' knowledge of the language develops (McLean, 2017). Albeit the assumption that learners know all the words of a family when they know one word family member seems to make sense for receptive vocabulary, it might not be valid in all contexts, viz., with productive vocabulary and especially with beginner to intermediate English language learners (Kremmel, 2016). Schmitt & Zimmerman (2002) demonstrated that EFL learners could only provide word family

members for around 19% of the tested words. Recent research by Kremmel and Schmitt (2016) showed that Austrian English-as-a-foreign language (EFL) learners could only combine the meaning of a base word and its derivative forms in 73% of the cases. In spite of learners having some knowledge of the connection between word family members, these studies show that knowledge of a headword does not guarantee knowledge of all word family members.

In his 2010 book, Schmitt already argued that the researchers' choice of counting unit should depend on the research questions, even though he also shows a preference for choosing the lemma as the main counting unit (p. 193). Four arguments are given: (1) the definition of a lemma is in most cases unambiguous; (2) it is easier to compare and replicate studies; (3) studies on receptive and productive vocabulary knowledge can be compared easily; and (4) when using word families as counting unit, the vocabulary size needed to use a language seems to be underestimated, given the fact that word families are often interpreted as 'words'. The number of lemmas will be higher and might seem more realistic. To sum up, the choice for the best counting unit should be justified by the purposes, participants, and resources of the study (Schmitt, 2010; Nation, 2016).

Recently, a new term or counting unit has been added, the flemma. McLean (2017) argued more recently in favor of the *flemma*, which would include all the words that are counted as a lemma, i.e., a headword and its inflected forms, but words that do not belong to the same part of speech are also included. Consequently, the verb *developed* and the adjective *developed* would belong to the same flemma, even though they are two distinct lemmas. In his study with Japanese EFL students, McLean found that the flemma was an appropriate word counting unit for the target population. However, additional evidence still has to show whether this is also true in other contexts with other learners and for other L2s.

Frequency in vocabulary learning

Frequency is crucial in language learning because the more frequent a word is, i.e., the more you encounter it, the more likely it is you will know that word, or as Ellis (2002, p. 152) phrased it: "the recognition and production of words is a function of their frequency of occurrence in the language". Cobb and Horst (2004) also pointed out that "learners tend to acquire L2 vocabulary in rough order of frequency" (p.17). Based on a word's frequency, Nation (2001) distinguishes different kinds of vocabulary. Although he initially differentiated between four categories, i.e., high-frequency words, low-frequency words, academic words and technical words (2001, p. 15-17), this distinction was revised in the second edition of his book *Learning vocabulary in another language* (2013). Based on the results of a study he conducted in 2006, he distinguishes only three kinds of vocabulary, solely based on frequency, i.e., high-frequency, mid-frequency and low-frequency words. High-frequency words are considered the 2,000 most frequent words of a language and contain for instance function words, such as *the*, *for*, *of*, *in*. The mid-frequency words concern the words between 2,001 and 9,000, and the low-frequency words concern all the words beyond the 9,000 boundary.

Schmitt and Schmitt (2014) suggested to reassess the frequency boundaries for pedagogical reasons and they argue that care should be taken when defining the boundaries of high- and low-frequency words. Indeed, they argue that the 2,000 word family boundary, which is seen as the traditional cut-off point for high-frequency words, should be reconsidered. The first 3,000 word families might be considered high-frequency words, as they "represent an important milestone in language development" (p. 492). However, as the arguments provided by Schmitt and Schmitt (2014) are founded on a very limited number of studies and only focus on English, the commonly adopted definition considering high-frequency vocabulary as the 2,000 most frequent words will be used in this PhD thesis.

As frequency is crucial to vocabulary learning (e.g., Nation, 2013; Schmitt, 2000), it is also used as a parameter for composing word lists. In order to select words for a frequency-based word list, large corpora or frequency lists derived from these corpora should be available. According to Nation (2013), "corpus based studies draw on language in use" (p. 531). As a result, the word lists drawn from these corpora should "provide a good representation of the high-frequency words of the language" (ibid.). He points out that the low-frequency words might not be included, which is not a major issue when focusing on low level non-native speakers. While many corpora and frequency lists[5] exist for English such as the British National Corpus (BNC 2000) or the Corpus of Contemporary English (COCA) and the corresponding BNC/COCA word lists (Nation, 2012), or the Essential Word list (Dang & Webb, 2016), there is a major lack of resources for French. Almost no corpora are freely available and the number of frequency lists is also rather limited. Three existing frequency lists for French are the ones by Baudot (1992), Verlinde and Selva (2001) and Lonsdale and Le Bras (2009). While Baudot's and Verlinde and Selva's work is mainly based on written language, the frequency list that was composed by Lonsdale and Le Bras also contains a substantial spoken component (50%). However, as Verlinde and Selva (2001) pointed out, "there is a strong need to design and construct for French […] a carefully selected corpus with a large variety of texts". No other frequency lists were available at the time we started this PhD project. As a consequence, the different studies that are part of our PhD project use Lonsdale and Le Bras' (2009) *Frequency Dictionary of French: core vocabulary for learners*, since it is probably the best that is available for French at the moment. It is also the frequency list that is used in French vocabulary tests, such as the *Test de la taille du vocabulaire* (Batista &

[5] Only some examples of corpora and word lists are indicated, as it is beyond the scope of this PhD project to elaborate on English corpora and frequency lists.

Horst, 2016) and in VocabProfile on Lextutor (Cobb, n.d.) (see section 3.6.2).

From a pedagogical point of view, word lists are important, especially for beginner learners because they are often a starting point for their learning. They provide "common vocabulary items that occur frequently across different texts" (Brezina & Gablasova, 2015, p. 1). However, even though many of the word lists are frequency based, some are thematically organized, such as the *Référentiels* for French. Indeed, researchers have pointed out that, apart from frequency, thematic vocabulary also plays an important role in word lists (Bardel, Gudmundson, & Lindqvist, 2012; Milton, 2006, 2009). Both approaches can be complementary.

Vocabulary tests

Given that vocabulary knowledge is such an important part of language proficiency, it is crucial to develop tests that can adequately measure the words learners know. Several vocabulary tests have been developed the last decades. One of the first and perhaps one of the most used tests is Nation's (1983) **Vocabulary Levels Test** (VLT), which was updated by Schmitt, Schmitt, and Clapham (2001), and by Webb, Sasao and, Ballance (2017). The objective of the VLT is to measure learners' vocabulary at four frequency levels, viz. 2,000, 3,000, 5,000 and 10,000 and at an academic level. The items were sampled from the lists that were available at that moment: Thorndike and Lorge (1944), Kučera and Francis (1967), the General Service List (West, 1953) and the University Word List (Xue and Nation, 1984). The test is a meaning recognition test in which learners have to match six words with three definitions. Three of the words correspond to a definition and three are distractors. The changes made to the original VLT by Schmitt et al. (2001) aimed at improving the validity of the test by adding items for each section. As the original VLT consisted only of 18 items per section, this was now expanded to 30 items. Second, the academic word list

section was now based on Coxhead's (2000) Academic Word List (AWL), which relies on a more balanced academic corpus and gives "better coverage of academic texts whilst listing fewer words than the University Word List (Xue and Nation, 1984)" (Schmitt et al., 2001, p. 63), the one used in Nation's (1983) original test. It is also possible to only administer one level to beginner learners for instance. An example of a noun from the 3K section of the **New VLT** (Schmitt et al., 2001) can be found below.

Example

1.	*blanket*		
2.	*contest*	_____	*holiday*
3.	*generation*	_____	*good quality*
4.	*merit*	_____	*wool covering used on beds*
5.	*plot*		
6.	*vacation*		

More recently, Webb et al. (2017) updated Schmitt et al.'s (2001) VLT. Two major shortcomings of the VLT were addressed by Webb and colleagues in this **Updated VLT**. A first concern was the corpus on which the previous test was based, because it mainly consisted of texts from the 1930s and 1940s and could thus be considered outdated. The corpus that was used for choosing the items was changed to Nation's (2012) British National Corpus/Corpus of Contemporary American English word lists (BNC/COCA). These lists should better reflect contemporary English. Second, the original test did not test the 1,000 most frequent words in a separate section. Therefore, the word frequency levels were adapted and were split into five sections of 1,000 words, i.e., 1-1,000, 1,001-2,000, 2,001-3,000, 3,001-4,000, 4,001-5,000 and the section measuring knowledge of academic words was not included in the test anymore. Finally, the third change that was made regards the presentation of the items. An example of the new test format can be

found below. The instruction given to the learners is to "put a check under the word that goes with each meaning".

Example from the 1K section of the Updated VLT (Webb et al., 2017)

	alone	***bad***	***cold***	***green***	***loud***	***main***
most important						
not good						
not hot						

McLean and Kramer (2015) also adapted the VLT (Schmitt et al., 2001) and called it **The New VLT**. They addressed the same shortcomings as did Webb and colleagues (2017). However, the major difference with Webb et al. is that McLean and Kramer argued that the test format was questionable. According to them, there are four main problems with the test format: (1) there is no item independence, (2) the format is inaccurate compared to a multiple choice format with four options, (3) learners have difficulties in understanding the test format, and (4) it is hard to adapt the tests to other testing mediums. As a consequence, they completely adapted the test format, which now resembles the test format of the Vocabulary Size Test (VST) (see infra, p. 29), as can be seen in the example below (McLean & Kramer, 2015, p. 4). The frequency lists from which the test items are taken are the same as the ones used by Webb et al., viz., the BNC/COCA word lists (Nation, 2012).

Example

*TIME: They have a lot of **time**.*

1. *money*
2. *food*
3. *hours*
4. *friends*

Another type of test is the **The Eurocentres Vocabulary Size Test** (Meara & Jones, 1990), also known as the Checklist or Yes/No test. It is a computerized yes/no-test in which learners have to indicate whether they know a word or not. The test measures the 10,000 most frequent words in English. Twenty words, 10 real words and 10 non-words, appear on the screen one by one for each of the frequency bands. The main advantage of such a test is that a lot of words can be tested in a short time and, as a result, that this is interesting when measuring vocabulary size since testing many words is required to be able to obtain valid estimates. However, there are also several drawbacks concerning this type of test. Just because learners recognize the words, this does not mean that they also really know the meaning of the words (Elgort, 2013). As mentioned before, if learners only recognize the form of a word, we cannot conclude that they really know the word in all its aspects. Hence, this type of test probably overestimates learners' vocabulary size and is less reliable than the VLT.

The aforementioned tests focus on learners' familiarity with different frequency levels, whereas the **Vocabulary Size Test** (Beglar & Nation, 2007) aims at estimating learners total vocabulary size. The VST aims to measure both native and non-native speakers' written receptive vocabulary size. In this test, largely decontextualized words are measured in a non-defining context. The choice of the items is based on a spoken ordering of the BNC since the use of the lists in Nation (2006) showed that the formal written nature of the BNC mainly affected the

high frequency levels. Therefore, the first 12,000 "word lists were revised using word family range and frequency figures from only the 10 million token spoken section of the British National Corpus" (p. 10).

The test formation of the VST is a multiple choice test format in which learners receive four short descriptions or single words among which the right answer and three distractors. There are different versions of the test, monolingual and bilingual versions (e.g., Korean, Arabic, Japanese), measuring up to the 14,000 or 20,000 most frequent words[6]. The former (14K) is mainly used for non-native speakers, the latter (20K) can also be used for native speakers.

Examples from the 2K and 10K section of the VST (Beglar & Nation, 2007)

*PUB: They went to the **pub***

1. *place where people drink and talk*
2. *place that looks after money*
3. *large building with many shops*
4. *building for swimming*

*AWE: They looked at the mountain with **awe**.*

1. *worry*
2. *interest*
3. *wonder*
4. *respect*

In sum, the previous tests all focus on receptive vocabulary and target English vocabulary knowledge. Recently, Batista and Horst (2016) developed a receptive vocabulary size test for French, **the *Test de la taille du vocabulaire*** (TTV), which is modeled on Nation's (1990) VLT for English. The test targets the 10,000 most frequent words of

[6] Different versions of the tests can be found on Paul Nation's web page: https://www.victoria.ac.nz/lals/about/staff/paul-nation#vocab-tests.

French and measures words at four frequency bands (2K, 3K, 5K and 10K). The frequency list that was used to choose the items was Lonsdale and Le Bras' (2009) frequency list containing the 5,000 most frequent words in French. The frequency dictionary was taken from a 23 million word corpus of both written and spoken text (50% each). However, since Lonsdale and Le Bras' list only lists up to the 5,000 most frequent words and the TTV also aimed at testing the 10,000 most frequent words, the older list of Baudot (1992) was used to this objective. It should also be mentioned that the TTV did not include a section testing academic words, since such a list does not exist for French and might not even be necessary (Batista & Horst, 2016, p. 218).

Example from the 5K section of the TTV

1.	brouillard *(fog)*		
2.	coincidence *(coincidence)*	_____	une histoire qui fait rire *(a story that makes one laugh)*
3.	farce *(farce)*	_____	ce qui empêche de voir loin *(hinders you to look far)*
4.	instituteur *(primary school teacher)*	_____	un professionnel de l'éducation *(a professional in education)*
5.	pneu *(tyre)*		
6.	soumission *(submission)*		

The test was validated by Batista and Horst (2016). For the studies in this PhD project, we preferred using (adapted versions of) the receptive and productive VocabLab tests (Peters, Velghe, & Rompaey, 2019) in order to be consistent with the type of vocabulary test throughout this PhD project (see section 3.5.2).

The **French VocabLab test** is a frequency-based meaning recognition test. It was developed using Lonsdale & Le Bras' (2009) frequency list for French. The test consists of 120 French items and

targets the 5,000 most frequent words in French. The 120 items are divided into four parts of 30 items, in which the first 30 items correspond to the 1-2,000 (2K) most frequent items, the next 30 items to the 2,001-3,000 (3K) most frequent items, the 30 following items to the 3,001-4,000 (4K) most frequent items and the last 30 items to the 4,001-5,000 (5K) most frequent items. For every French item, four definitions are provided in the target language, among which one correct answer and three distractors. There is also an 'I don't know'-option in order to avoid guessing. An example can be found below.

Example from the monolingual receptive VocabLab-test

lourd

- *qui a beaucoup de poids*
- *qui est dur pour les autres*
- *qui ressemble très bien à quelque chose*
- *qui regroupe beaucoup d'éléments, grand*
- *je ne connais pas la réponse*

In addition to the previous tests, which all measure receptive vocabulary, Laufer and Nation (1999) developed the **Productive Levels Test** (PVLT), a test measuring productive vocabulary. Since productive vocabulary implies different degrees of knowledge, contrary to receptive vocabulary, the test type should be adapted accordingly. The structure of the test adopts the one of the VLT: 18 items are chosen for the 2K, 3K, 5K, University Word List (UWL) and 10K sections. The test items that were used in the A version of the PVLT were the same items as the original VLT, and the items in three parallel versions reused the items of versions of the VLT developed by Schmitt. This also implies that the same corpora and word lists are used. Since Schmitt et al. have only updated the VLT in 2001, the rather old word lists have thus been consulted. The PVLT provides meaningful sentences in which a gap has to be filled. Learners receive the first letters of the word, the minimal

number of letters necessary to avoid other responses. Two examples from the 2,000 and 3,000 word level of the PVLT are given below.

> Example
>
> *Teenagers often adm_____ and worship pop singers.*
> *Sudden noises at night sca____ me a lot.*

The test is composed in the same way as the original VLT (Nation, 1983; Nation, 1990) and tests words at the 2,000, 3,000, 5,000, 10,000 word levels and academic words. Laufer and Nation (1999) do not decide on the score learners have to obtain in order to master a level, but remark that a score of 15 or 16 out of 18 (which is around 85% or 90%) should be alright, depending on a personal judgement. The percentage scores that can be obtained at a certain level are a "very rough indication of the number of words known at that level" (Laufer & Nation, 1999, p. 41).

2.2.4 Measuring vocabulary use: lexical diversity

Lexical diversity (LD) is a way to measure learners' productive vocabulary **use**. This PhD project focuses on different measures of LD as operationalizations of lexical complexity.

Lexical complexity, also called lexical richness (e.g., John Read, 2000) or sometimes vocabulary richness (e.g., Malvern, Richards, Chipere, & Durán, 2004) covers several aspects of vocabulary use. Even though the term lexical richness initially referred to "the number of words in a person's mental lexicon" (Yule, 1944, as cited in Jarvis, 2013), and has been used later as a synonym of lexical diversity (Daller, Van Hout, & Treffers-Daller, 2003, as cited in Jarvis, 2013), it is now mainly used as a term covering several lexical constructs. According to Read (2000), lexical richness consists of four aspects: lexical density, number of errors, lexical sophistication and lexical variation or diversity (p.

200). Lexical density is defined as the proportion of content and function words, whereas lexical sophistication or rareness focuses on the number of low-frequency words that is used. Lexical diversity, then, is defined as "the variety of active vocabulary deployed by a speaker or writer" (Malvern & Richards, 2002, p. 87). In this PhD project, lexical sophistication (frequency) and lexical diversity will be the two focal points in this regard.

Calculating lexical diversity in an 'appropriate' way has not been easy and methodological problems have been encountered throughout the years. The Type-Token Ratio (TTR) is one of the earliest measures for calculating lexical diversity. It consists of the division of the number of different words (types) by the total number of words (tokens). However, the measure is dependent on text length (for a discussion, see Vermeer, 2000; Malvern et al., 2004). Longer texts, which are often produced by learners with a higher proficiency level, usually produce lower TTR-values because learners use fewer new words (types) as text length increases. As a consequence, shorter texts, often attributable to learners with a lower proficiency level, appear to have a higher TTR-value. Thus, this measure is to be avoided, especially with spontaneous spoken or written texts, which are more likely to have different text lengths. In order to establish a measure of lexical diversity that was not affected by text length, different mathematical transformations of TTR have been tested, such as Guiraud (1960), Corrected TTR (Carroll, 1964) and Herdan and Uber indexes (Herdan, 1966), but the problem with text length still seemed unresolved. Richards and Malvern (2000) used the Mean Segmental Type-Token Ratio (MSTTR), a measure originally suggested by Johnson (1944). This measure has been used widely in different types of linguistic research, in order to overcome the problem of text length. Richards and Malvern describe it as follows: "the average TTR for successive segmentation of text containing a standard number of word tokens" (Malvern and Richards, 1997, p. 35). The problem of the variable sample size was

solved, but other problems remained. Therefore, Malvern and Richards (2002) developed the *D* measure more recently. This measure compares empirically derived curves from language data with theoretical curves in which *D* is adjusted until the best curve is obtained for the transcript. The best *D* value will represent the lexical diversity of the oral or written production. Texts receiving a higher *D* will display a higher lexical diversity. This measure overcomes the problem occurring with different sample sizes. Further, since it takes multiple random samples from the entire text, all data is used and all lexical repetitions are taken into account.

McCarthy and Jarvis (2010), however, argue that the D value as measured by the vocd-command is "a complex way of approximating the hypergeometric distribution" (p. 383) and in order to prove this, they described an index called HD-D, which is the hypergeometric distribution of D. Still according to McCarthy and Jarvis (2010), "the hypergeometric distribution represents the probability of drawing (without replacement) a certain number of tokens of a particular type from a sample of a particular size" (ibid.). In order to obtain a value for HD-D they calculated "for each lexical type in a text, the probability of encountering any of its tokens in a random sample of 42 words drawn from the text" (ibid.). As a consequence, the number obtained by doing this is the sum of the probabilities for all lexical types in a text. A text containing more lexical types would thus correspond to a more diverse text. Very high correlations between both measures are observed. Even though other indices such as the measure of textual lexical diversity (MTLD; McCarthy, 2005; McCarthy & Jarvis, 2010) and Maas (Maas, 1972) could also be used and seem quite robust when it comes to text length, they will not be used in the empirical studies that are part of this PhD project. The reasons for this are that there are no tools available to calculate them for French (MTLD) or that other measures have been shown to be more proficient for French. Indeed, Treffers-Daller (2013) compared the usefulness of different measures of LD

(MTLD, HD-D, D and Maas) in predicting different aspects of language ability. The results indicated that HD-D and D are better measures for French than MTLD and Maas.

Finally, and more in general, Jarvis (2013) questioned whether a text containing more types is more lexically diverse in comparison with a text containing fewer types. He gives the following example (2013, p. 100).

Example

(a) *We run every morning.*
(b) *We run up and down the slope of that hill every morning before sunrise.*

These two sentences differ of course in the number of tokens, but are comparable in that they consist both of the same number of tokens and types (4 tokens and types for sentence a and 14 tokens and types for sentence b). As a consequence, these sentences have a TTR rate of 1, which means that they reach the maximum value of lexical diversity. However, on the question which sentence was more lexically diverse, 90% of the raters picked the second sentence. With this example, Jarvis' objective is to show that diversity should, perhaps, be looked at as more than just a statistical comparison of types and tokens, but that additional properties such as richness, density or dispersion should also be taken into account, in imitation of ecologists who consider diversity as a multidimensional phenomenon. This is an interesting consideration that invites the SLA field to think beyond the borders of language acquisition alone. However, it was beyond the scope of this project to further develop this idea.

2.3 How does vocabulary relate to the four skills?

This section will briefly describe the relationship between vocabulary and the four skills of reading, listening, writing and speaking. In this PhD project, we explored the relationship between vocabulary knowledge and listening comprehension and between vocabulary knowledge and speaking with low-intermediate French-as-a-foreign language learners. This section will briefly report the research that has been done in the other skills and on which the studies in this PhD were based to a certain extent.

2.3.1 Receptive skills

It has been shown that there is a strong relationship between vocabulary knowledge and reading comprehension (e.g., Laufer, 1992) and between vocabulary knowledge and listening comprehension (e.g., Stæhr, 2008, 2009; van Zeeland & Schmitt, 2013). Most research has, however, focused on English as a foreign language (EFL) and on highly proficient (university) students.

Reading comprehension

In this section, we will first elaborate on the relationship between vocabulary knowledge and reading comprehension. Next, we will look at the relationship between lexical coverage and reading comprehension to finally inform about the associated vocabulary size that is necessary for reading comprehension.

Relationship vocabulary knowledge and reading comprehension

It has been shown that vocabulary knowledge and reading comprehension highly correlate. Correlations range between .50 and .85 depending on learners' proficiency level (e.g., Laufer, 1992; Milton et al., 2010; Qian, 1999, 2002; Stæhr, 2008).

Laufer (1992) demonstrated that vocabulary knowledge and reading comprehension are strongly related. Positive, significant

correlations (.50 and .75) were found between reading comprehension tested with two standardized reading tests and vocabulary size tested with the VLT and the Eurocentres Vocabulary Test. Participants were university students that were supposed to have reached the B2 level in English[7]. In a study by Qian (1999) with Korean and Chinese learners of English as a second language (ESL) who were familiar with at least the 3,000 most frequent words in English, significant correlations between .64 and .82 were found confirming the strong relationship between vocabulary knowledge and reading comprehension. A TOEFL reading comprehension test was used to assess reading comprehension and three vocabulary tests measured vocabulary, viz., the VLT (Nation, 1983) measuring vocabulary size (.78), and two vocabulary tests measuring depth, Read's (1983) Word Associates Test (.82) and a self-created test looking at morphological knowledge (.64). In line with these results, Qian's (2002) study with intermediate level ESL students, showed correlations ranging from .73 to .77 between three vocabulary tests and a version of the TOEFL reading comprehension test. The vocabulary tests used were the VLT (Nation, 1983) (.74), an adapted version of the Word Associates Test (Read, 1993) (.77) and the TOEFL vocabulary item measure, measuring knowledge of English synonyms, i.e., vocabulary depth (.73). Further, Stæhr (2008) found a similar, significant correlation of .83 between the VLT and reading comprehension measured as part of the Danish national school leaving examination with Danish EFL learners. Finally, Milton et al. (2010) also found a positive, significant correlation (.70) between reading comprehension (IELTS test score) and the X_Lex test (Meara & Milton, 2003), estimating vocabulary size with students of English, whose level of English ranged between intermediate and advanced.

[7] The learners' level was supposed to be at an equivalent of the Cambridge First Certificate of English, which tests the B2 level (see also https://www.cambridgeenglish.org/exams-and-tests/first/).

Relationship lexical coverage and reading comprehension

The previous studies clearly demonstrated that the relationship between reading comprehension and vocabulary size and depth is well established, even though the numbers are not entirely comparable because of different test measures and proficiency levels of the learners. Researchers wanted to go beyond this relationship and aimed at knowing how many words learners would need in order to be able to read adequately. First, in order to establish a threshold, research looked into the relationship between reading comprehension and lexical coverage. Lexical coverage refers to the percentage of words in written or spoken discourse that enables successful comprehension. If there is a lexical coverage of 99%, a learner will come across one unknown word in every 100 words or ten sentences; 95% lexical coverage corresponds to one unknown word in every twenty words or two sentences (Hu & Nation, 2000). Depending on the difficulty of a text, which will contain more low frequency words, a higher number of words should be known in order to have a good understanding of that text. It has been shown that between 95% and 98% lexical coverage is necessary to be able to read a text (e.g., Hirsh & Nation, 1992; Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006). Ninety-five percent lexical coverage should suffice when minimal reading comprehension, viz., needing help and guidance, is required (Laufer & Ravenhorst-Kalovski, 2010), whereas 98% seems necessary for optimal, viz., independent, reading comprehension (Laufer & Ravenhorst-Kalovski, 2010) and for unassisted reading comprehension (Hu & Nation, 2000). A study by Schmitt, Jiang and Grabe (2011) suggested that 98% lexical coverage would be a reasonable coverage figure for understanding academic texts. They found that the relationship between coverage and reading comprehension of two academic texts is a linear relationship, with higher vocabulary knowledge leading to a higher degree of reading comprehension. Nonetheless, the degree of lexical coverage that is necessary for reading comprehension seems to be determined by how adequate

comprehension is defined and the degree of comprehension that is necessary, which will be different when reading for pleasure or having to understand classes at university.

Vocabulary size needed for adequate reading comprehension

After having defined the necessary lexical coverage for adequate reading comprehension, the vocabulary size corresponding to the appropriate coverage levels could be determined. Laufer (1989) found that 95% lexical coverage best distinguished between Israeli EFL learners achieving the necessary 55% for passing the reading comprehension test and learners who did not. Five thousand words were necessary to reach this level.

Early research (Laufer, 1992) suggested that knowledge of 3,000 word families would constitute the minimum lexical threshold for reading academic texts, corresponding to a reading score of 56%. Depending on the reading score that should be obtained, learners would need to have higher vocabulary knowledge, e.g., 5,000 word families corresponding to a score of 70% in this study. Hirsch and Nation (1992) found that a threshold of 5,000 words was associated with 97-98% lexical coverage necessary to adequately read short unsimplified novels "with reasonable ease" (p. 693). These figures were however readjusted by Nation (2006), who established a higher threshold of 8,000-9,000 word families necessary for the adequate reading comprehension of newspapers and novels, and 6,000 to 7,000 word families for spoken text, both corresponding to a lexical coverage of 98%. The lower figures for spoken language are expected since spoken language makes more use of high-frequency words than written language.

A later study by Laufer & Ravenhors-Kalovski (2010) displayed different results. In that paper, it was shown that minimal reading comprehension could be obtained with 95% lexical coverage, corresponding to 4,000 to 5,000 word families (including proper

nouns). Ninety-eight percent lexical coverage corresponded to 8,000 word families (including proper nouns), which would lead to optimal comprehension. These differences could be explained by the use of different definitions for adequate reading comprehension, or because of a different research design using for instance different reading tests. Schmitt et al. (2011) concluded that the 98% coverage figure would be the preferable target for learners when at least 60% is expected on a reading comprehension test.

Listening comprehension

As in the previous section, we will first discuss the relationship between vocabulary knowledge and listening comprehension before addressing the relationship between lexical coverage and listening comprehension. The final subsection focuses on the number of words necessary for adequate listening comprehension.

Relationship vocabulary knowledge and listening comprehension

Different studies have shown that vocabulary knowledge does not only play an important role in reading comprehension but in listening comprehension as well (e.g., Stæhr, 2008, 2009; van Zeeland & Schmitt, 2013; Wang, 2015), even though vocabulary seems to explain less variance in listening than in reading (Mecartty, 2000). Stæhr (2008, 2009) reported significant correlations (.69 and .70) between the VLT (Schmitt et al., 2001) and listening comprehension with EFL learners. In his 2008 study, listening was tested with Danish EFL learners after seven years of English as part of the school leaving exam. Stæhr observed, however, that 65% of the learners who did not reach the 2,000 most frequent word level did at least obtain the average score in listening comprehension. He concluded that learners were able to compensate for their lack of vocabulary knoweldge and that 2,000 words might be an important threshold for learners in that specific context. In the 2009 study, advanced Danish EFL learners took a standardized listening comprehension test. Depth, as measured by a

test developed following Read's Word Associates Test, also strongly correlated (.51) with listening comprehension. Milton et al. (2010) found lower correlations between two receptive vocabulary size tests and an IELTS listening test with intermediate to relatively advanced students of English. Even though the written X_Lex-test correlated significantly with the listening test (.48), the correlation between the aural vocabulary size test (A_Lex) and the listening test was higher (.67). This comparison of both tests shows that it might be better to use an aural vocabulary test in order to test an aural skill. Similarly, aural vocabulary knowledge and IELTS listening comprehension with EFL university students was reported to correlate highly and significantly (between .67 and .73) in two later studies by Matthews and Cheng (2015) and Matthews (2018).

Furthermore, Teng (2014) revealed a highly significant correlation (.86) between the VST (Nation, 1983) and an IELTS listening comprehension test for Chinese EFL learners who had reached the 5,000 most frequent word level. However, the correlation dropped to .41 for learners who had mastered only the 3,000 most frequent words in English. Teng also investigated the relationship between vocabulary depth and the listening test and identified a significant correlation of .91. In this study, depth explained more of the variance in listening comprehension than size alone. In a study by Afshari and Tavakoli (2017) with high-intermediate to advanced Iranian EFL students, however, the correlation between vocabulary size (VLT) and a TOEFL test (.85) was higher than between vocabulary depth (WAT) and the same listening test (.83).

Finally, Wang (2015) and Wang and Treffers-Daller (2017) reported moderate correlations between receptive vocabulary size and listening comprehension with Chinese EFL university students. In Wang (2015), a correlation of .36 was found between the VLT (Nation, 1983) and listening comprehension as measured by the CET4, a formal English proficiency test in China. A correlation of .44 was found in

Wang and Treffers-Daller (2017) between the VST (Beglar & Nation, 2007) and the same CET4 listening comprehension test.

Relationship lexical coverage and listening comprehension

Although the reading comprehension studies mentioned in the previous section might provide useful insights, transferring these results to listening comprehension is questionable as reading and listening are different skills. Recently, there has been an increasing interest in the relationship between vocabulary and listening comprehension. It should be noted that research in listening has adopted the same coverage figures as for reading comprehension. Ninety-eight percent lexical coverage corresponds to very high comprehension, whereas 95% lexical coverage corresponds to a lower, but acceptable level of comprehension, especially with L2 learners (e.g., Bonk, 2000; van Zeeland & Schmitt, 2013).

Bonk (2000) reported that adequate listening comprehension would already be possible at 80 to 89% lexical coverage. According to him, 95% lexical coverage might, thus, not even be necessary. However, Schmitt (2008) further analyzed Bonks results and showed that learners with 95% coverage reached better comprehension than learners with 90% coverage or less. Van Zeeland & Schmitt (2013) found that 98% lexical coverage was needed to achieve very good comprehension of informal narratives, while 95% coverage might already be a satisfying goal for good but not entirely complete comprehension.

Vocabulary size needed for adequate listening comprehension

When comparing both skills of reading and listening by looking at the number of words necessary to reach 95% lexical coverage, it seems that fewer words are necessary for adequate comprehension in listening than in reading. Indeed, whereas at least 4,000 word families would be needed for reading comprehension (Laufer & Ravenhorst-Kalovski, 2010), only around 3,000 word families seem necessary for listening to narrative stories (van Zeeland & Schmitt, 2013) and for

understanding television programs (Webb & Rodgers, 2009a, 2009b). Around 2,000 to 3,000 word families were also needed to reach a coverage of 95% in spoken corpora (Adolphs & Schmitt, 2003; Nation, 2006). According to Nation (2006), 6,000 to 7,000 word families are necessary to reach 98% lexical coverage of unscripted spoken English. These results show that knowledge of the 3,000 most frequent word families in English should be sufficient for understanding colloquial English. Dang & Webb (2014) found that 4,000 word families were necessary to reach 95% lexical coverage in the BASE corpus, a corpus of academic spoken English, whereas 8,000 word families provided 98% coverage in that same corpus. Even though 3,000 word families might be sufficient for understanding colloquial English, it seems insufficient for other purposes such as understanding political debates or academic courses.

In sum, the relationship between vocabulary knowledge and reading and listening seems to be established for English-as-a-foreign language learners at high proficiency levels. Given that such figures do not exist for French, it seemed necessary to fill that gap and to investigate whether these figures could be transposed to French and more specifically to lower proficiency levels. Moreover, as Schmitt, Cobb, Horst, and Schmitt (2017) suggested, it is important to "start looking into what learners can actually do with various vocabulary sizes" (p. 13).

2.3.2 Productive skills

Unlike the large number of studies conducted on the relationship between vocabulary knowledge and the receptive skills of reading and listening, less research has been conducted into the relationship between vocabulary knowledge and the productive skills of writing and speaking. In what follows, we will briefly report on what has been done.

Writing

In the nineties, it had already been demonstrated that vocabulary plays a major role in learners' written output (Astika, 1993) as far as the connection between L2 vocabulary knowledge and writing is concerned. Laufer and Nation (1995) demonstrated that, in EFL learners of low-intermediate level and beyond in New-Zealand and Israel, higher scores on the active version of the VLT (Nation, 1983) were associated with the use of more sophisticated and less frequent words. A larger productive vocabulary thus tends to lead to the use of more low frequency words. More recent research has shown that there is, as in reading and listening, a strong relationship between both receptive and productive vocabulary knowledge and writing (e.g., Laufer & Nation, 1995; Miralpeix & Muñoz, 2018; Stæhr, 2008). Receptive vocabulary size has been shown to correlate well with EFL writing, correlations range between .53 and .73. In two studies, Schoonen et al. (2003) and Schoonen, van Gelderen, Stoel, Hulstijn, and de Glopper (2011) found large correlations between a receptive vocabulary size test that was tailored to the context of Dutch EFL learners in secondary schools (i.e., 13 to 14 year old learners) and rated writing tasks (.53 to .63). Further, larger correlations were also found by Stæhr (2008). He measured a correlation of .73 between the VLT (Schmitt et al., 2001) and writing as tested with Danish 15-16 year old secondary school students who had passed a national school leaving examination. He observed that "more than half of the variance in the ability to perform above average in the writing test was explained by vocabulary size" (p. 148). A similar correlation was found by Milton et al. (2010). They found a .76 correlation between receptive vocabulary size (X_Lex; Meara, 2005) and written compositions of the IELTS with intermediate to advanced EFL learners. Finally, a somewhat lower but still large correlation (.57) between receptive vocabulary size as tested with the X_Lex (Meara, 2005) and the Y_Lex (Meara & Miralpeix, 2006) and writing was found in Miralpeix and Muñoz (2018) with Spanish EFL university students. They were assumed to be at the B2 level. In

this study 30% of the variance could be explained by vocabulary size, which is lower than wat Stæhr found.

To summarize, learners with a larger receptive vocabulary perform better in writing (e.g., Miralpeix & Muñoz, 2018; Stæhr, 2008). Further, a higher productive vocabulary knowledge leads to the use of more low-frequency words in written productions, as measured with the Lexical Frequency Profile (Laufer & Nation, 1995). Moreover, Laufer and Nation (1999) reported that the 1,000 most frequent words in English make up around 75% of the running words in formal written texts and about 84% in informal spoken language. The words in the lower frequency bands decreasingly add less coverage to texts, which indicates the importance of learning and teaching at least the 2,000 most frequent words in English. This is also what Stæhr (2008) concluded for low level EFL learners. He noted that the 2,000 most frequent words in English might be a crucial learning goal since almost all learners knowing the 2,000 most frequent words performed adequately on the writing task, whereas this could not be said for the learners that did not yet reach the 2,000 word level.

Speaking

Previous studies investigating speaking have shown that vocabulary knowledge plays an important role in learners' speaking performance (e.g., Koizumi, 2005; Milton et al., 2010; Miralpeix & Muñoz, 2018). Milton et al. (2010) reported a positive and significant correlation (.71) between an aural receptive vocabulary size test (A_Lex) and IELTS speaking scores with intermediate to advanced EFL learners, not between an written test (X_Lex; Meara, 2005) and speaking scores, however. This contradicts Miralpeix & Muñoz' (2018) findings, who found a moderate correlation (.49) between receptive vocabulary (X_Lex; Meara, 2005 and Y_Lex; Meara & Miralpeix, 2006) and scores on a semi-guided interview with B2 Spanish EFL students. A similar correlation (.55) was observed by Uchihara & Clenton (2018) between a

receptive Yes/No test and rated vocabulary in a monologic speaking task with advanced EFL learners. They noted that the 2,000 most frequent words in English seem to cover a large part in speaking proficiency, i.e., 95% of the words speakers use in monologic tasks.

Compared to receptive vocabulary size, higher correlations were reported by Koizumi (2005) and De Jong et al. (2012) between productive vocabulary size and speaking. Koizumi (2005) found a positive and significant correlation (.78) between a productive vocabulary size test designed for Japanese beginner EFL learners and rated monologic and dialogic speaking tasks. Similarly, a strong correlation (.79) was observed by De Jong et al. (2012) between productive vocabulary size and rated monologic speaking tasks with intermediate to advanced learners of Dutch.

To summarize, a strong relationship exists between both receptive and productive vocabulary knowledge and speaking, for English as a foreign language. The question is, however, whether these results can be transposed to non-university students of French and whether there is a difference in learners' output between monologic and dialogic tasks.

2.4 The Common European Framework of Reference and the B1 level

The *Common European Framework of Reference* (Council of Europe, 2001) is one of the most influential documents in educational language policy in Europe aiming to provide "a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe" (p. 1). It offers a language neutral description of competences (e.g., linguistic, pragmatic competence) that learners should master in order to be able to perform functional language tasks at six proficiency levels. These proficiency levels are divided into three bands, A, B and C, which are in turn subdivided in two levels each: A1 and A2 (Basic User), B1 and B2 (Independent User)

and C1 and C2 (Proficient User) (see Appendix 1). According to Hulstijn (2007), the notion of language proficiency, and as a result also the six proficiency levels, contain a quantitative (*what can learners do?*) and a qualitative (*how well can learners do it?*) dimension. As far as lexical competence is concerned for instance, its progress is defined in terms of range of vocabulary (*what*) and vocabulary control (*how*). These terms are not defined but could be linked to the different labels that are commonly used in the literature. Range of vocabulary could be defined as vocabulary size or breadth, and vocabulary control could be associated with depth, since it refers to the accuracy and appropriate use of the lexicon. The lexical descriptors corresponding to the different proficiency levels are given below in Table 2.1 (Council of Europe, 2001, p. 112).

Table 2.1

Illustrative scale for the range of vocabulary knowledge.

	Vocabulary Range
C2	Has a good command of a very broad lexical repertoire including idiomatic expressions and colloquialisms; shows awareness of connotative levels of meaning.
C1	Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions; little obvious searching for expressions or avoidance strategies. Good command of idiomatic expressions and colloquialisms.
B2	Has a good range of vocabulary for matters connected to his/her field and most general topics. Can vary formulation to avoid frequent repetition, but lexical gaps can still cause hesitation and circumlocution.
B1	Has a sufficient vocabulary to express him/herself with some circumlocutions on most topics pertinent to his/her everyday life such as family, hobbies and interests, work, travel, and current events.

A2	Has sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics. Has a sufficient vocabulary for the expression of basic communicative needs. Has a sufficient vocabulary for coping with simple survival needs.
A1	Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.

Table 2.2

Illustrative scale for the control of vocabulary knowledge.

	Vocabulary Control
C2	Consistently correct and appropriate use of vocabulary.
C1	Occasional minor slips, but no significant vocabulary errors.
B2	Lexical accuracy is generally high, though some confusion and incorrect word choice does occur without hindering communication.
B1	Shows good control of elementary vocabulary but major errors still occur when expressing more complex thoughts or handling unfamiliar topics and situations.
A2	Can control a narrow repertoire dealing with concrete everyday needs.
A1	No descriptor available.

As can be seen from Table 2.1, vocabulary range is described in terms such as "very broad lexical repertoire" at the C2 level, "good range of vocabulary" at the B2 level to "basic vocabulary repertoire" at the A1 level. As learners go through the levels, they are expected to gain much vocabulary knowledge. It is implied that vocabulary depth

would only be important at the C1 and C2 levels: "idiomatic expressions and colloquialisms" and "connotative levels of meaning" only appear at these levels. Milton (2013) noted that an implicit assumption within this framework is that a growth in vocabulary size is also associated with progress in depth, fluency and vocabulary use, as can be seen in the descriptors of vocabulary control (Table 2.2). At the C2 level, for instance, "appropriate use" is expected, whereas at the A2 level, a learner should be able to "control a narrow repertoire".

The descriptors above also employ a mixed presentation: both the *what* ("routine, everyday transactions involving familiar situations and topics") and the *how* ("sufficient", "some circumlocutions") are present in the can-do statements. At the B1 level, for instance, learners dispose of "sufficient vocabulary", "good control of elementary vocabulary" and use "some circumlocutions" (*how*), and can talk about "most topics pertinent to his/her everyday life such as family, hobbies and interests, work, travel and current events" (*what*). These illustrative scales are, however, not empirically validated and there is no reference to any theoretical paradigm (Alderson, 2007; Hulstijn, 2007; Wisniewski, 2017).

In the CEFR, it is also implicitly taken for granted that learners' proficiency grows in a linear way and that learners at the B2 level have passed through all the other levels. Moreover, it is implied that learners at a certain level should have attained the same level on different scales, for listening as well as for interacting and vocabulary range for instance. This is, however, only one type of L2 learners, as pointed out by Hulstijn (2007). He argues that there could be two other types of L2 learners, viz., those learners performing with high linguistic quality, but limited in quantity on the one hand and learners able to do many things in terms of quantity but few in terms of linguistic quality on the other. The fact that a linear growth might not exist should be further investigated empirically.

Moreover, as can be seen, these descriptions are quite general. Little (2007) already pointed out that the CEFR describes for each of the levels "the communicative functions that learners should be able to perform at different proficiency levels, but does not specify how those functions might be realized in, say, French or German" (p. 646). As a consequence, given the abstract and language neutral character of these descriptions, there is room for interpretation when people want to operationalize the CEFR for pedagogical purposes. It is therefore not surprising that professionals have attempted to make these descriptions concrete and language specific by creating *Reference Levels Descriptions* (RLDs) (e.g., Decoo, 2012; Hulstijn, 2014), such as the *Niveau B1 pour le français. Un référentiel* (Beacco, Blin, Houles, Lepage, & Riba, 2011) for French. These RLDs are being developed for different languages. However, it is not clear how they have been composed and how RLD makers decided on the lexical items to include given the fact that the illustrative scales of the CEFR are open to interpretation and that there is, in the CEFR, no definition of what lexical competence entails. In addition, these RLDs are based on linguists' and teachers' intuition but have not been empirically validated (Alderson, 2007; Decoo, 2012; Hulstijn, 2007, 2014; Kusseling & Lonsdale, 2013).

A second empirical problem of the CEFR is that it is not clear how many and which words should be known at a certain level, which is what professionals have attempted to find out (Decoo, 2012). Regardless of skill (reading, writing, listening or speaking), there is a discrepancy between vocabulary sizes proposed at a specific level by several studies. For instance, according to Milton (2006), the B1 CEFR-level would correspond to 800 to 1,000 words for French, whereas it would correspond to 952 to 2,422 words according to research done by Milton and Alexiou (2009). They investigated the number of words needed for EFL, French as a foreign language and Greek as a foreign language at different CEFR levels by administering the X_Lex vocabulary test (Meara & Milton, 2003), which existed for all three

languages. The results indicate that learners' vocabulary size at the same levels differs according to the language that is studied, but that there are also differences according to the country in which one same language was tested. For EFL, vocabulary sizes at the B1 level range from 2,750 to 3,250 words compared to a range that starts at 952 to 2,422 for French as a foreign language. First, fewer words seem necessary at the B1 level for French than for English. Second, the results suggest that there might be systematic differences between the number of words required in different foreign languages at a certain level. For instance, learners seem to need fewer words in order to achieve adequate listening comprehension in French than in English. Finally, the data in all languages reveal that learners tend to progress in vocabulary knowledge as they go through the CEFR proficiency levels. As argued by Hulstijn (2014), the discrepancy between languages could be attributed to the lack of consensus on what an intermediate level specifically entails and a lack of a common, sound methodology. Further, Little (2007) pointed out that the CEFR does not make any reference to distinct languages as its aim was to be language independent. However, this assumes that a similar language proficiency level would be necessary for a given communicative activity in different languages (Alderson, 2007), which does not seem to be in line with the results of Milton and Alexiou (2009).

Further, there seems to be a major theoretical problem with the CEFR. Hulstijn (2014) argues that the CEFR confounds the six proficiency levels with intellectual abilities since it does not take into account to what extent proficiency levels interact with intellectual abilities. One could wonder whether advanced argumentation and presentation skills, as being part of the C-levels, could be reached by native speakers with lower educational backgrounds. A lack of underlying theoretical framework for language proficiency might be part of the explanation (Alderson, 2007; Fulcher, 2004; Hulstijn, 2007, 2014).

In conclusion, the main problems of the CEFR are that no empirical validation has been done, that there is no consistency in the terminology used and that it lacks a theory of language proficiency (Alderson, 2007; Fulcher, 2004; Hulstijn, 2007; Wisniewski, 2017). In order to answer the empirical validation, Kusseling and Lonsdale (2013) aimed at comparing the French RLDs to French corpus language resources. This process contained two steps: (1) a quantitative comparison of the lexicon in the different resources and (2) a qualitative judgment of the vocabulary. Even though there was considerable overlap between the lexical content in the RLDs and the corpus resources, there was also quite some variation. The results show that, at the B1 level for instance, 7,490 types (= different words) are recommended to be included in the *Référentiel*, whereas 4,423 types should be excluded. This study points out the importance of this type of empirical studies. There are also some studies that investigated the relationship between functional language activities and linguistic competences even though they have mainly focused on grammatical competence and foreign language writing (Gyllstad, Granfeldt, Bernardini, & Källkvist, 2014; Kuiken, Vedder, & Gilabert, 2010). In this PhD project, the relationship between foreign language learners' vocabulary knowledge and listening comprehension and vocabulary use in speaking was explored. Learners were situated at the pre-intermediate level, as operationalized by the B1 CEFR level.

2.5 Conclusion

It has been clearly demonstrated to what extent vocabulary knowledge is important in the four skills of reading, writing, listening and speaking. It has been shown how different types of vocabulary (receptive vs. productive) can be measured and how vocabulary tests can be used for that purpose. This has, for the major part of the studies, been done for English with university students. In our PhD thesis, we focus on both receptive and productive vocabulary knowledge in relation to listening on the one hand and speaking on the other. Our

studies mainly differ from other studies in that our main target is L2 French with secondary school students who have a low-intermediate level of the language.

Chapter 3

Methodology

3.1 Introduction

This chapter deals with the methodology we employed in the three empirical studies in this PhD project. First, we briefly describe in a nutshell the three empirical studies (3.2). Next, we focus on the research context (3.3), the methodological choices common to the three studies, that is, the participants (3.4), the target language (3.5) and the data collection instruments (3.6). Finally, we elaborate on the scoring and analyses used in each of our studies (3.7).

3.2 Overview of the three empirical studies

In this section, we provide a short overview of the three studies included in this dissertation. They are summarized in Table 3.1. As can be seen from Table 3.1, the common ground of the three studies is L2 French vocabulary, which is operationalized as vocabulary knowledge in studies 1 and 2 and as vocabulary use in studies 2 and 3. In the first study, we investigate the receptive skill of listening whereas we focus on the productive skill of speaking in studies 2 and 3. The number of materials used is expanded as the studies go. In the first study, we focus on a receptive vocabulary knowledge test and a listening comprehension test. A productive vocabulary knowledge test was added in studies 2 and 3 and the listening comprehension test was replaced by two semi-structured dialogic speaking tasks. Finally, whereas studies 1 and 2 only involved L2 learners, study 3 also focused on L1 speakers. In this last study, our measurement instruments also included ratings of experts who scored the spoken output of both L1 and L2 speakers.

Table 3.1

Overview of the three empirical studies

Study	Focus	Skill	French	Participants	Participants' level	Measuring instruments
1	Vocabulary knowledge	Listening comprehension	L2	*N* = 351	A2 B1	- Receptive vocabulary test - Listening comprehension test
2	Vocabulary knowledge and use	Speaking	L2	*N* = 51	B1	- Receptive vocabulary test - Productive vocabulary test - Speaking tests
3	Vocabulary use	Speaking	L1 L2	*N* = 27 *N* = 51	Native B1	- Receptive vocabulary test - Productive vocabulary test - Speaking tasks - Ratings

Study 1 – Vocabulary knowledge and listening comprehension at an intermediate level in English and French as foreign languages. An approximate replication study of Stæhr (2009)

In the first study, that focused on vocabulary knowledge and listening comprehension with L2 learners of French and English, we replicated Stæhr's (2009) study to see whether his results could be expanded to another context. More particularly, we investigated (1) whether there is a correlation between vocabulary knowledge and listening comprehension at the (pre-)intermediate level and (2) how much vocabulary is necessary for listening comprehension as measured by a standardized B1 listening test. In order to address these questions, we recruited 351 French-as-a-foreign-language learners and 199 English-as-a-foreign-language learners from the second and fourth year of secondary education and first-year university in Flanders (Belgium). Learners' French and English vocabulary knowledge was measured by means of the VocabLab test, a frequency-based receptive vocabulary test. Listening comprehension was tested by the *Diplôme d'Étude en Langue française* (DELF) for French and the Preliminary English Test for schools (PET), which both correspond to the B1 CEFR-level. The details of this study can be found in Chapter 4 (p. 99).

Study 2 – The relationship between low-intermediate French-as-a-foreign language learners' receptive and productive vocabulary knowledge and their vocabulary use in semi-structured dialogic speaking tasks

The second study focused on (1) the correlation between low-intermediate French-as-a-foreign language learners' receptive and productive vocabulary knowledge and their speaking output at the B1 CEFR-level, (2) the correlation between their vocabulary knowledge and the lexical frequency profile of their output and (3) its lexical diversity in two semi-structured dialogic speaking tasks. Data were collected from 51 Dutch-speaking learners of French in their last year

of secondary education. Learners' vocabulary knowledge was measured by means of a frequency-based receptive and productive vocabulary knowledge test. Learners' speaking proficiency was measured in two B1 speaking tasks. This study is presented in Chapter 5 (p. 127).

Study 3 – Low-intermediate French-as-a-foreign language learners' rated vocabulary use in semi-structured dialogic speaking tasks

In our third study, with both L1 and L2 learners, we had three main interests: we determined the differences between L1 and L2 speakers' vocabulary use in two semi-structured dialogic speaking tasks by looking at the vocabulary use of their spoken output. Vocabulary use was measured by means of the lexical frequency profile and lexical diversity of the output. Further, we focused on the relationship between L2 learners' receptive and productive vocabulary knowledge and their speaking output. Third, we aimed at determining which lexical factors could predict the holistic scores of participants' lexical proficiency as given by experienced raters (3). The L2 data from the second study were also used in this study but additional data were collected with 27 L1 French-speaking learners of the same age group as the L2 learners. In order to allow for comparison, L1 learners' speaking proficiency was measured in the same B1 speaking tasks as in the second study. The report on the last study can be found in Chapter 6 (p. 159).

3.3 Research Context

French is the first L2 Flemish learners have in a formal learning context. It is considered a foreign language and not a second language, even though French is one of the three official languages in Belgium. However, learners have a fairly limited contact with French outside the classroom, as has been demonstrated by Peters et al. (2019) and by the *Peilingsonderzoek Frans* (Survey for French listening and speaking; Vlaamse Overheid, 2013).

French classes usually start in the fifth year of primary school, at the age of 10, even though a minority of schools already offer French classes from the third year onwards (age = 8 years)[8]. During two years, learners have one hour of French a week. When entering into secondary education in Flanders, they have three or four classes of fifty minutes of French (Vlaamse Overheid, 2013, p. 28), depending on their curriculum.

3.4 Participants

The data for our three empirical studies were collected in different schools and universities in Flanders and targeted L2 speakers of French of different levels. In this regard, our studies are also innovative with respect to the proficiency level of the participants. Indeed, it has been argued that most studies on L2 learning and the four skills have focused on university students (e.g., Stæhr, 2009; Uchihara & Clenton, 2018; van Zeeland & Schmitt, 2013). In order to expand on previous research, we focus on both university and non-university students.

For the first study, a convenience sample of 351 participants was recruited from different proficiency levels. We recruited participants from different university campuses studying languages or enrolled in other programs such as Law or Business administration. Data were also collected with L2 learners of French in second and fourth year of secondary education in different sections of general secondary education (ASO), e.g., Greek-Latin, Sciences, Modern Languages. For the second study, 51 participants who were in their last year of secondary education and enrolled in different programs (e.g., Greek-Science, Latin-Modern languages, Mathematics-Modern languages, Science-Modern languages) were recruited. They were supposed to have an intermediate level of French. Finally, in study 3, the data of the

[8] More information can be found on the following website: http://www.erk.nl/docent/erkineuropa/00002.

same 51 participants of study 2 were used. Additionally, 27 L1 speakers were engaged from a secondary school in Wallonia, the French speaking part of Belgium. The details are outlined below (see Table 3.2).

Table 3.2

Number of participants per study and per level

Studies	Total *N*	SEC2	SEC4	SEC6	UNIV	French
1	*N* = 351	*N* = 65	*N* = 53		*N* = 233	L2
2	*N* = 51			*N* = 51		L2
3	*N* = 78			*N* = 51		L2
				N = 27		L1

3.5 Target language

The target language of our studies is French (and also L2 English in study 1), which is the L2 (studies 1, 2 and 3) or the L1 (study 3) of the participants we recruited. Several reasons inspired us to focus on French. By focusing on French as L2 in our studies, we want to take into account the following aspects: first, most research into listening and speaking has focused on English-as-a-foreign language (e.g., Iwashita et al., 2008; Stæhr, 2009; Uchihara & Clenton, 2018; van Zeeland & Schmitt, 2013), but fewer studies have focused on French (e.g., Lindqvist, 2010; Treffers-Daller, 2013). We therefore wanted to extend previous research to a less studied language, i.e., French. Second, French is a compulsory course in secondary schools as well as in several programs in the first year at university. The data for our studies could thus be collected during a regular French class.

3.6 Data collection instruments

This section focuses on the different data collection instruments. We follow the order of our studies and first discuss the listening

comprehension test (3.6.1) before presenting the different vocabulary tests (3.6.2). We conclude with the speaking tasks (3.6.3).

3.6.1 Listening comprehension test

The listening comprehension test that we used in study 1 is part of the DELF, developed by the *Centre International d'Études Pédagogiques* (CIEP). It is an internationally recognized test and provides certifications for different CEFR-levels. The diplomas are awarded by the French Ministry of Education. The DELF examinations test the four skills, i.e., listening, reading, writing and speaking. We only administered the listening part to our participants. In order to obtain the level that the examination tests, viz., B1, learners need to obtain 50% in total and at least 20% on each of the parts according to the instructions of the CIEP. Learners had to listen to a conversation between a mother and her daughter and two testimonies, from an actor telling about his job and from a worker. They could always listen twice and received time to answer after the first and after the second listening of each part. In total, the test consisted of 23 questions of which 19 multiple choice questions and 4 open questions. An example for each of the categories can be found below.

Example of a multiple-choice question (1 point)

Il se souvient des numéros de téléphone de:
(He remembers the telephone numbers of:)

- *tous ses amis.* (all his friends)
- *tous ses enfants.* (all his children)
- *toute sa famille*. (his whole family)

Example of an open-ended question (2 points)

A quoi compare-t-il la mémoire ? Donnez au moins un exemple.
(What does he compare memory to? Give at least one example.)

The listening test was piloted with 4 students studying languages at a Flemish university and did not reveal any problems. The mean score they obtained was 91%. The listening test can be found in Appendix 2.

3.6.2 Vocabulary tests

VocabLab-test

In our first study, we administered the recently developed and validated VocabLab-test (Peters et al., 2019), a monolingual receptive vocabulary test specifically designed for Dutch-speaking learners. The test was developed for two target languages (English and French) and addresses a number of concerns that were raised by different researchers concerning the VLT and the VST (Gyllstad, Vilkaite, & Schmitt, 2015; Webb & Sasao, 2013). They mainly regard the overestimation of the learners' vocabulary knowledge, the rather outdated frequency lists on which the tests are based, the lack of a 1K frequency band and the ratio of adjectives, nouns and verbs. Apart from these concerns, the English tests that existed did not seem to be tailored to the Flemish context with (low-)intermediate learners. As this test was adapted to the Flemish context, cognates and loanwords such as 'sport' were not included. Furthermore, there was a differentiation in frequency bands. The VLT addresses words in 4 sections, viz., 1-2,000, 2,001-3,000, 3,001-5,000 and 5,001-10,000, whereas the VocabLab-test focuses on more accurate frequency bands for the target public, viz., 1-2,000, 2,001-3,000, 3,001-4,000 and 4,001-5,000 since the 3,001-

5,000 frequency band might not be sensitive enough for the Dutch-speaking learners (Peters et al., 2019).

For French, however, the TTV (Batista & Horst, 2016) did not yet exist when the VocabLab-tests were being developed and no other tests for French existed. As the English test, the French test was thus developed from scratch. Both the English and the French test are frequency-based meaning recognition tests, in which learners have to recognize the meaning of French words (see Section 1.3.3). The entire test can be found in Appendix 3.

This version of the VocabLab test had already been validated before we administered the tests (Peters et al., 2019). However, in order to verify the procedure of the data collection, we piloted the vocabulary test as well as the listening comprehension test with 4 first-year university students in language studies. The students' mean score on the vocabulary test was 90.42%. We therefore decided to administer the test as such, without any changes.

Bilingual receptive vocabulary test

As the results on the VocabLab-test in the first study revealed relatively low scores, even with first year university students, we decided to administer a bilingual vocabulary test in studies 2 and 3. Bilingual tests are supposed to be easier because they avoid problems with reading comprehension and only focus on vocabulary instead of also focusing on grammar (Nguyen & Nation, 2011). They might therefore also give a more realistic representation of learners' vocabulary knowledge. In addition, according to Elgort (2013), seeing L1 translations instead of entire definitions in the target language is expected to diminish test anxiety.

Several changes were made compared to the original monolingual test, i.e., the 5K section was removed from the test and additional high-frequency items were added. The section with the low-

frequency words was removed because the results of the previous study had shown that learners obtained very low scores (*M* = 40.13%) on that section and the maximum score obtained was 83.33%. For the 4K section, however, there were learners who obtained more than 90% as a maximum score, which is the threshold that is set for mastering a level, even though the total mean for that frequency band is also low. Therefore, we decided not to test the 5,000 most frequent words but to keep the 4K section, also because previous studies (De Clercq, 2016; Segers, 2015; Vrancx, 2016) had shown that the results significantly increased with bilingual tests and were probably less conservative. Next, a 1K section was added because it turned out that we needed to be able to distinguish more between learners mastering or not the 2K section, as was also pointed out by Webb and Sasao (2013). High-frequency words will have a higher impact on comprehension, especially for low level learners. According to Nation (2013), 84.3% of the words in conversations belong to the 1,000 most frequent word families in English. Webb and Rodgers (2009a) showed that 85.11% of the words in television programs pertain to the 1,000 most frequent word families. It seemed therefore necessary to also test these items.

Therefore, 17 items were added to the 1K level and 13 items to the 2K level in order to obtain 30 items for both levels and to extend the test. The different versions of the monolingual test had been translated and validated beforehand (e.g., De Clercq, 2016; Segers, 2015; Vrancx, 2016). In order to choose the items, we searched for the best items in these different tests by looking at the item facility, which is the mean score of an item. Items with a value below .20 are considered difficult items, items with a value above .80 are considered easy items[9]. We especially looked at the easy and difficult items. Next, we evaluated how much an item correlated with the overall test score. This

[9] The interpretation of the values was based on the information found on the following website of the R Project for statistical computing: https://cran.r-project.org/web/packages/sjPlot/vignettes/sjtitemanalysis.html.

discrimination index "tells us how well a test item is capable of separating high scoring test takers from low scoring test takers" (Fulcher, 2013, p. 320). Items with a correlation of r =.30 or above are considered very good items because they measure the same construct as the other items. Items below r = .20 were removed from the test. Finally, the items that showed a higher correlation than the Cronbach's Alpha of the overall test were removed. Deleting such items would make the vocabulary test more reliable. Nine items were removed from the original bilingual version because they behaved badly according to the criteria mentioned before. The inverse procedure was done for choosing new items to include in the test. The items belonging to the 1,000 and 2,000 most frequent words can be found below in Table 3.3. The test is available in Appendix 4.

Table 3.3

Items that were removed from and added to the bilingual receptive vocabulary test.

Removed items		Added 1K items		Added 2K items	
French	**English**	**French**	**English**	**French**	**English**
sang	blood	*espoir*	hope	*honneur*	honor
promesse	promise	*nécessité*	necessity	*envie*	desire
lumière	light	*coup*	punch	*foyer*	home
lourd	heavy	*tomber*	fall	*distance*	distance
ciel	sky	*taux*	percentage, proportion	*fil*	thread
épreuve	test, examination	*satisfaire*	satisfy	*puissant*	strong
écrivain	writer	*faible*	weak	*retraite*	retreat
boîte	box	*échange*	exchange	*retard*	delay
oreille	ear	*cher*	expensive	*chute*	fall
		prêt	loan	*ferme*	farm
		movement	movement	*trace*	track

Removed items		Added 1K items		Added 2K items	
French	**English**	**French**	**English**	**French**	**English**
		paix	peace	*trait*	line
		ignorer	ignore	*échec*	failure
		pauvre	poor	*foule*	mass, crowd
		absence	absence	*port*	port
		avis	opinion	*poids*	weight
		proche	nearby	*rêve*	dream
		échapper	escape	*sévère*	severe
				bonheur	happiness
				plonger	dive
				éclater	explode

As a result, this bilingual version consisted of 120 items divided into four sections of 30 items, each testing 1,000 words, viz., 1-1,000, 1,001-2,000, 2,001-3,000 and 3,001-4,000. As explained earlier, we did not administer the highest frequency band from the original VocabLab test, i.e., 4,001-5,000 words, because scores turned out to be very low in the first study. In this bilingual test, learners received a French word and four Dutch words. They had to pick the right L1 translation instead of having four definitions to choose of. There were four options in Dutch, among which the right answer and three distractors, and an 'I don't know'-option.

The adapted version of the test was piloted in pilot study 2 with 12 French-as-a-foreign language learners in their last year of secondary education. A bilingual version of the receptive vocabulary test was administered instead of the original, monolingual version. Segers (2015), De Clercq (2016) and Vrancx (2016) compared the French monolingual and Dutch-French bilingual tests and showed that the learners' results on these bilingual tests were significantly higher. They were also found to better represent learners' actual vocabulary knowledge. During this second pilot session, the test was administered online (see Figure 3.1), whereas a paper-and-pencil version was administered in study 1. Learners needed about 10 to 15 minutes to complete the test. Afterwards, we asked the learners orally whether they would have preferred a paper-and-pencil test, or an online version and they all preferred the computer version. Given the fact that participants needed less time to fill out the online test, we could also avoid test fatigue effects.

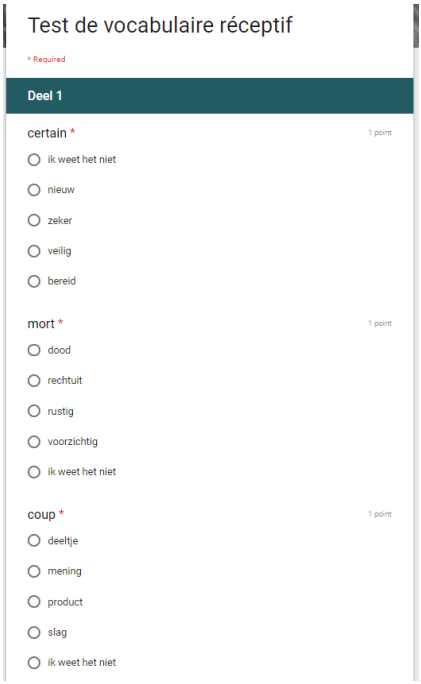


Figure 3.1 Example of the online vocabulary test environment

The results of the receptive vocabulary test can be found in Table 3.4. As expected, the mean percentages on the different K levels decrease as the words become less frequent. The results on the bilingual version of the test are higher than on the monolingual test used in study 1. Even though there were only twelve participants in the pilot study, the test was found to have good internal consistency (Cronbach's Alpha = .89, N = 12). The pilot study showed that the materials were appropriate for our research purposes.

Table 3.4

Descriptives per K level for the receptive vocabulary test (N = 12)

K level	***M* % (*SD*)**	**95% *CI***	***M* raw scores (Max=30) (*SD*)**	**95% *CI***
0-1,000	87.50 (7.93)	[82.46-92.54]	26.25 (2.38)	[24.74-27.76]
1,001-2,000	84.44 (9.25)	[78.57-90.32]	25.33 (2.77)	[23.57-27.10]
2,001-3,000	76.19 (10.91)	[69.26-83.12]	21.33[10] (3.06)	[19.39-23.27]
3,001-4,000	67.50 (13.34)	[59.02-75.98]	20.25 (4.00)	[17.71-22.79]
Total	78.95 (8.67)	[73.45-84.46]	93.17 (10.23)	[86.67-99.67]

Finally, the item facility of the bilingual version can be found in Table 3.5. The item facility is the percentage of participants answering an item correctly and corresponds to the mean percentage of Table 3.4. The results indicate that the item facility for the different frequency bands is quite high, especially for the first 2,000 words, but this might be due to the small sample size on which these values are based.

[10] It should be noted that the total score for this section was only 28, since two items disappeared from the test during the pilot study.

Table 3.5

Item facility per K level for the bilingual receptive vocabulary test (N = 12)

K level	Number of items	Item facility
0-1,000	30	.88
1,001-2,000	30	.84
2,001-3,000	28	.76
3,001-4,000	30	.68
Total	118	.79

Bilingual productive vocabulary test

In studies 2 and 3, participants took a French bilingual productive vocabulary test, measuring learners' productive vocabulary. A bilingual version was used, as was done for the receptive test. As outlined above, it was demonstrated by De Clercq (2016), Segers (2015) and Vrancx (2016) that learners knew significantly more words on a bilingual productive test than on a monolingual productive test. A possible explanation is the way in which learners learn French, i.e., they learn wordlists with translations from the L2 to the L1 and sometimes also the other way around.

The original test is a frequency-based test that used Lonsdale and Le Bras' (2009) frequency list. It measures the 3,000 most frequent words in French productively. In order to do this, 90 items are administered to the learners, among which the first 45 belong to the 2,000 most frequent words in French and the next 45 belong to the 2,001-3,000 most frequent words in French. The test format is a gap-fill exercise. The L1 translation of the word they have to fill out is given in brackets as well as the first letter of the word in French and the exact

number of letters in order to avoid guessing. An example can be found below.

> Example
>
> ***Je n'ai plus confiance en lui parce qu'il ne tient jamais ses p_ _ _ _ _ _. Il ne fait jamais ce qu'il a dit. (beloftes)***
>
> (I don't believe him anymore because he never keeps his **promises**. He never does what he said he would do.)

During the second pilot study with 12 learners, we administered a paper-and-pencil version of the bilingual test that measured up to the 3,000 most frequent words in French. The original test contained 90 items, among which 45 items tested the 2,000 most frequent words and 45 items the 2,001-3,000 most frequent words. However, previous research has shown that testing 30 items provides good reliability for testing a 1,000 word frequency band (Schmitt et al., 2001). Therefore, 15 of the 45 items testing the 2K and the 3K most frequent words were removed. The choice for the items that had to be removed was made based upon the item facility: items with a value below r = .20 or above r = . 80 are not considered being good items. If, additionally, the discrimination index of the item was below .30 and the Cronbach's Alpha value given in the "Cronbach's Alpha if item Deleted" was higher than the Cronbach's Alpha of the entire test, the item was not included in the test. As a consequence, 30 items testing the 2,000 most frequent words and 30 items testing the 2,001-3,000 most frequent words were administered during this pilot study.

Learners needed 30 to 40 minutes to complete the test. The test is reliable, which can be seen from the Cronbach's Alpha value (.90). The results of that test are presented in Table 3.6 and show that, according to our expectations, the results decrease as the word frequency decreases. The item facility can be found in Table 3.7 and

indicates the difficulty of the items within one frequency band. The values range between .42 and .64, which is good since values below .20 and above .80 should be avoided.

Table 3.6

Descriptives per K level for the productive vocabulary test (N = 12)

K level	***M* % (*SD*)**	**95% *CI***	***M* raw scores (Max=30) (*SD*)**	**95% *CI***
0-2,000	64.17 (15.96)	[54.02-74.31]	19.25 (4.79)	[16.21-22.29]
2,001-3,000	41.66 (13.30)	[33.22-50.11]	12.50 (3.99)	[9.97-15.03]
Total	52.92 (14.09)	[43.96-61.87]	31.75 (8.45)	[26.38-37.12]

Table 3.7

Descriptives per K level for the productive vocabulary test (N = 12)

K level	***N* of items**	**Item facility**
0-2,000	30	.64
2,001-3,000	30	.42
Total	60	.53

In order to avoid test fatigue with learners not knowing a lot of words in the 3K word section and for the pragmatic reason that both vocabulary tests had to be administered in one class hour, viz., 50 minutes, we decided to test only the 2,000 most frequent words in the actual data collection. Therefore, the 3K word section was deleted but

a 1K section was added. In order to do so, we added items both in the 1K as well as in the 2K section to obtain 30 items per frequency band. The procedure that was followed to remove items was the same as explained earlier (see section 3.6.2). The items that were removed can be found in Table 3.8. Similarly, the items that had to be added to the test were chosen from other versions of the bilingual test according to the same criteria. Nineteen items were added for the 1,000 most frequent words and twenty items were added to the 2K section. They can be found below (Table 3.8).

In sum, two different versions of the bilingual productive vocabulary test were used, one for the pilot session and one for the actual data collection (study 2). The version measuring up to 3K used during the pilot session is provided in Appendix 5 and the test measuring up to 2K administered during the actual data collection is provided in Appendix 6.

Table 3.8

Items that were removed from and added to the bilingual productive vocabulary test.

Removed items		Added 1K items		Added 2K items	
French	**English**	**French**	**English**	**French**	**English**
beau	beautiful	*coup*	kick	*ferme*	farm
participer	participate	*début*	start	*émission*	program
naissance	birth	*juger*	judge	*accusé*	accused
sourire	smile	*peine*	punishment	*puissant*	powerful
écrivain	writer	*sauf*	safe	*envie*	desire
boîte	box	*objectif*	goal	*île*	island
goût	taste	*tomber*	fall	*retard*	delay
cerveau	brain	*large*	wide	*victoire*	victory
		pied	foot	*trait*	feature
		satisfaire	satisfy	*bois*	wood
		échange	exchange	*meurtre*	murder
		feu	fire	*reprocher*	reproach
		message	message	*vigueur*	strength, vigor

Removed items		Added 1K items		Added 2K items	
French	**English**	**French**	**English**	**French**	**English**
		accès	access	*pêche*	catch
		ressources	resources	*étroit*	narrow
		confier	confide	*trace*	track
		progrès	progress	*caisse*	cash register
		détruire	destroy	*inventer*	invent
		environnement	environment	*joie*	joy
				faim	hunger

3.6.3 Speaking tasks

Selecting the speaking tasks

In our second and third study, two speaking tasks were administered to L2 learners of French (studies 2 and 3) and to L1 speakers (study 3), viz., a doctor's visit and a job interview. As the starting point of our PhD project was the B1 CEFR level, we decided to look for speaking tasks that were explicitly mentioned in the CEFR at that level.

In the CEFR, a task is defined as "any purposeful action considered by an individual as necessary in order to achieve a given result in the context of a problem to be solved, an obligation to fulfil or an objective to be achieved" (Council of Europe, 2001, p. 10). We assumed that the CEFR would contain clear descriptions of functional language activities and tasks. Moreover, we expected that these functional language activities would be clearly linked to one language proficiency level. A close examination of the CEFR, however, showed that this is not the case and finding specific speaking tasks in the CEFR was challenging due to its lack of specificity. As a result, the descriptions given in the CEFR are susceptible to interpretation, as was also pointed out by Alderson (2007). The main problems related to the CEFR have been addressed in Chapter 2, but we will point out some difficulties in finding the appropriate speaking tasks.

First of all, many of the descriptions are dependent on the person or the situation, for example: "Can communicate with some confidence on familiar routine and non-routine matters related to his/her interests and professional field" (Council of Europe, 2001, p. 74) and "can narrate a story" (p. 59). The specific content of these language activities will be heavily determined by the learner – what are his/her interests and professional field, what is a familiar routine – and the situation – which story will the learner narrate? Because of this variety in interpretations, it is difficult to define what a language activity

specifically entails. Secondly, although the CEFR sometimes describes a specific example of a language activity, for instance "asking a passenger where to get off for an unfamiliar destination" (Council of Europe, 2001, p. 80), most functional activities are not described specifically enough. An example of a broadly defined can-do statement is "I can enter unprepared into conversation on topics that are […] pertinent to everyday life (e.g., family)" (Council of Europe, 2001, p. 26). As a result, the specific contents of the activities can be interpreted differently. Moreover, it is not always clear which tasks are included in the language activities and which are excluded as the language activities are rather broadly defined. For example: "communicate on topics that are pertinent to everyday life: family, hobbies, travel, current events, work" (Council of Europe, 2001, p. 26) and "can deal with most situations likely to arise whilst travelling in an area where the language is spoken" (Council of Europe, 2001, p. 24). In these examples, it remains unclear which specific communicative language activities are targeted by the CEFR.

Moreover, based on the aforementioned can-do statements, it is not clear which language activities should be excluded because they are too specific to be used at the B1 CEFR-level: should a learner at the B1 level be able to complain about a parking ticket he/she received because he/she was parked in a parking-disc zone? This is a situation which can be ascribed to the can-do statement of "dealing with most situations likely to arise whilst travelling". It could be argued that this situation is too specific for a B1 level, but the CEFR does not provide a decisive answer to whether or not the language activity is part of the B1 CEFR-level. Related to this issue is the fact that language activities are sometimes mentioned at the A2 as well as the B1 level, for example, at A2 "I can use a series of phrases and sentences to describe in simple terms my family", at B1 "I can enter unprepared into conversation on topics that are […] pertinent to everyday life (e.g., family)" (Council of Europe, 2001, p. 26). Although there are differences between the

descriptors at the A2 and B1-level such as 'describe in simple terms' and an 'unprepared conversation', the CEFR does not explicitly state what simple terms are or which tasks learners can already perform at the B1-level, but not yet at the A2-level.

Finally, it should be mentioned that some examples of concrete speaking tasks are given in the description of the language activities. For instance, according to the CEFR, a learner at the B1 level "can provide concrete information required in an interview/consultation (e.g., describe symptoms to a doctor) but does so with limited precision" (p. 82). Furthermore, when communicating at work, a learner should be able to "attend interviews giving written or spoken information about own personal data, qualifications and experience and answer questions about them" (Council of Europe, 2001, p. 54). Next, a learner should be able to "deal with most situations likely to arise when making travel arrangements through an agent or when actually travelling" (Council of Europe, 2001, p. 34). The three tasks that were finally chosen, i.e., a doctor's visit, a job interview and a hotel reservation correspond to the aforementioned descriptions, which are, however, rather global can-do statements that should be worked out by people using the CEFR for their own purposes. Unlike the tasks used in most speaking studies (e.g., Crossley & McNamara, 2013; Iwashita et al., 2008), the tasks we use are interactive tasks because these better reflect language use in daily life.

The topics of the speaking tasks that were chosen for our studies are part of oral expression examinations of the DELF at the B1 level. We adapted them however according to the learners' suggestions given during qualitative discussions in pilot studies. The tasks are semi-spontaneous dialogic tasks in which a learner interacts with a native speaker. They are semi-spontaneous because they are guided in a certain way. Details can be found in the next section. The tasks as they were used in studies 2 and 3 can be found in Appendix 7.

Piloting the speaking tasks

The aim of the two pilot studies was to finalize and evaluate the procedure, materials, methods and to assess the feasibility of the data collection method (Gass & Mackey, 2007). The data of the pilot studies are thus not further taken into account in our main studies.

Pilot study 1

The three initial speaking tasks, viz., a doctor's visit, a job interview and a hotel reservation, were first piloted in February 2016 with 64 pupils in their last year of secondary education (17-18 years old). Most of the learners had a major in modern languages (N = 41) in combination with Latin (N = 11), science (N = 7) or economics (N = 23). One group was enrolled in a science and mathematics section (N = 23).

As far as the hotel reservation and the job interview are concerned, dialogues were held between two L2 learners. For the hotel reservation, learners received a flash card that contained 9 images, which could guide their conversation (Appendix 8). For the job interview, learners received a flash card with a structure in Dutch that could help them construct the dialogue. There was one flash card for the recruiter and one for the student looking for a job (Appendix 9). For the doctor's visit, there were two options, an interaction between two L2 learners and an interaction with a native speaker. Six dialogues were held between classmates (L2-L2) and six other dialogues between a pupil and a native speaker of French (L2-L1). We told the learners that the native speaker was a medical doctor. Learners also received a flash card with some instructions in Dutch and images that could help them to talk about the topic (Appendix 10). The learners could not choose whether they interacted with a classmate or with the L1 speaker. The selection was made alphabetically.

After having performed the speaking tasks, learners also received a short questionnaire (Appendix 11) to collect information about their mother tongue, their opinions concerning the French

language, their contact with French outside of the classroom, their familiarity with the topics and their anxiety to speak French. These questions were administered in order to investigate whether there would be large differences between learners or whether learners' anxiety to speak French affected the number of words produced. Participants answers to the questions were, however, purely informative and only used as background information for the researcher. They will thus not be analyzed in detail. In previous pilot study 1, similar questions were asked on paper. In order to be more efficient, the questionnaire was adapted and had to be filled out online.

The first pilot study showed that the speaking tasks as such worked well. The semi-structured dialogues were found to be appropriate in interactions between L1 and L2 speakers. However, they did not lead to the desired outcome when used between two L2 speakers. This was due to the fact that it was very difficult for the learners to imagine themselves being a receptionist, a recruiter or a doctor. This is not surprising since these types of roles are not part of our target audience's everyday life. Compared to the L1-L2 conversations, the dialogues between L2 speakers were also found to be much shorter. In the L2-L2 interactions, the mean number of words for the doctor's visit was 184,33 words, 241,64 words for the hotel reservation and 299,83 words for the job interview, whereas the L1-L2 doctor's conversations consisted on average of 614,5 words per conversation, with a range between 127 and 278 words for the L2 speaker. The L2-L2 dialogues for the doctor's visit however contained only between 10 and 131 words for the learner taking the patient's role and between 24 and 178 words for the L2 learner taking the doctor's role. As a result, we decided to only focus on L1-L2 conversations for the actual data collection.

In addition, we found that pupils were not used to making a hotel reservation. Hence, they did not know which questions to ask. This might be due to the absence of a detailed instruction for the hotel

reservation (i.e., only images) in comparison to the job interview. Two other possible explanations for the lack of elicited words in this task might be that learners' parents make hotel reservations or that they are more used to surfing on the internet for making a (hotel) reservation on websites such as booking.com or Airbnb.com, where you only have to tick some boxes. That is why the decision was made not to include this topic in the actual data collection.

Pilot study 2

The second pilot study took place in November 2016 with a group of 17-year old pupils. The goal of this pilot study was to validate the speaking tasks that were adapted according to the results of the first pilot study. We thus wanted to verify whether the changes that were made to the materials were appropriate.

The following changes were made: first, concerning the speaking tasks, as mentioned earlier (see section 3.5.3), only two topics were used in this second pilot study, viz., the doctor's visit and the job interview. Second, another change that was made compared to pilot study 1 was that we administered both speaking tasks in Dutch as well as in French in order to control for cognitive problems. Since the first pilot study had shown that pupils did not necessarily speak much during these conversations in French, we wanted to make sure that the lack of spoken output was due to a lack of French vocabulary, not to their impossibility of speaking about that topic. Finally, the flash card for the doctor's visit did not contain images and the flash card for the job interview contained less information compared to the previous version. The materials can be found in Appendix 12.

The pilot study showed that learners' performance on the speaking tasks in both Dutch and French did not seem to reveal any cognitive problems. They did not say much more in the Dutch conversations compared to the French ones. Consequently, it was decided to administer the speaking tasks only in French in the actual

data collection. This pilot study also allowed us to adapt the timing of the speaking tasks. It was clear that more time was needed for the dialogic tasks than was initially foreseen. Two sessions had to be planned, i.e., separate sessions for the vocabulary tests and for the oral tasks. Furthermore, images were added on the flash card of the doctor's visit, since learners suggested that it would help them to know what to talk about.

The figure below indicates the different steps and adjustments. It specifies the topics that were tested, the languages in which the dialogues were performed, whether the interactions were held between two L2 speakers of between L1 and L2 speakers and the type of flash cards that were used.

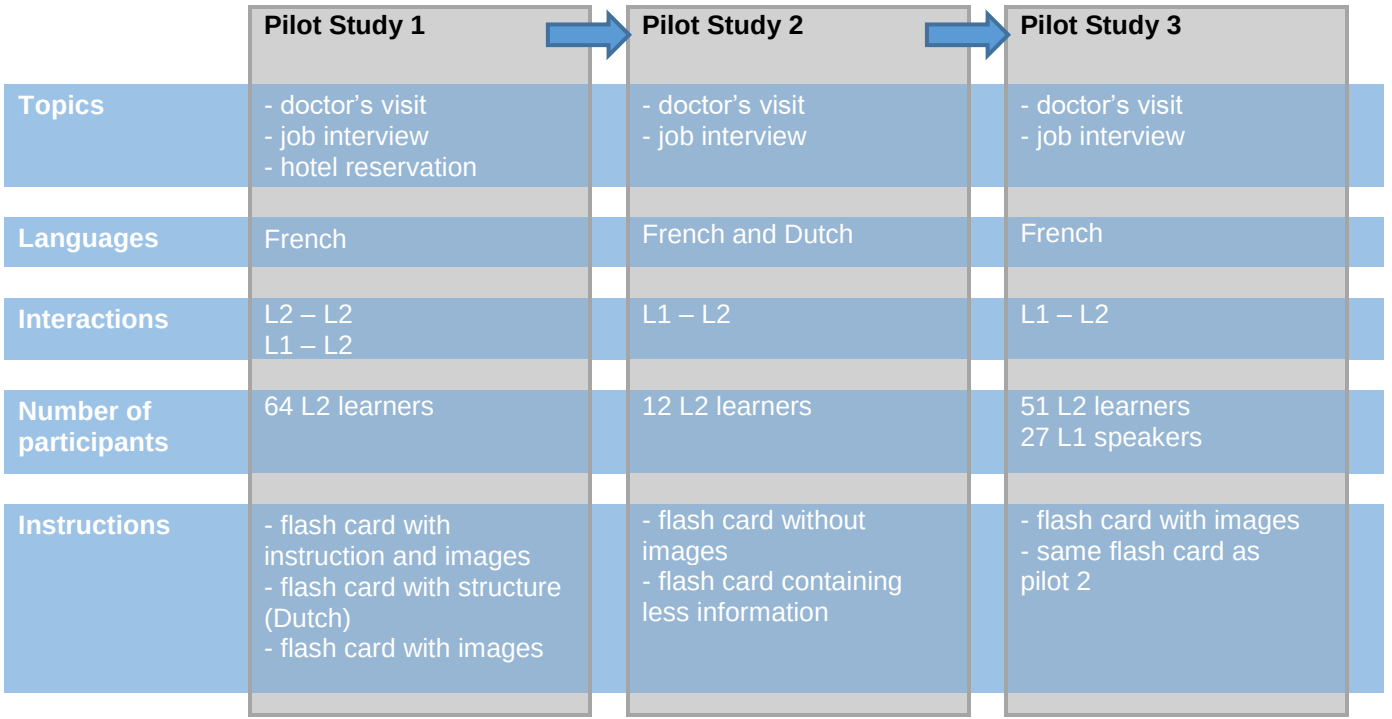


Figure 3.1. Schematic overview of the changes made during the pilot studies.

3.7 Scoring and analyses

In this section dedicated to the scoring of the different tests and the statistical analyses, we first focus on the scoring of the different tests (3.7.1). Next, we present the materials we have used to analyze our data (3.7.2). In the last part, we introduce the statistical methods used to analyze our data and to answer our central research questions (3.7.3).

3.7.1 Scoring of the tests

Listening comprehension test

The listening comprehension test was scored using the instructions given by the CIEP. The test was mainly composed of multiple-choice questions (19 out of 23 questions), which were scored dichotomously: 1 point was awarded for a correct answer and 0 for an incorrect answer. There were also four open-ended questions. One point could be attributed to two of the questions and two points to the two other questions. Learners received one point or half of a point when their answer was partially correct. The maximum score on the listening test was 25.

Receptive vocabulary tests

Both on the monolingual as well as on the bilingual receptive vocabulary test, 120 points could be obtained. The test was scored dichotomously and learners received a score of 1 when they ticked the right answer, while they received a score of 0 when they ticked one of the three distractors or the "I don't know"-option. Thirty points were attributed to each of the frequency levels, viz., 1-2,000, 2,001-3,000, 3,001-4,000 and 4,001-5,000 for the monolingual vocabulary test and 1-1,000, 1,002-2,000, 2,001-3,000 and 3,001-4,000 for the bilingual version. In order to determine whether participants mastered a given frequency band, we used the scoring proposed by Nation (n.d., as cited in Stæhr, 2009) and used 27 (max = 30) as threshold.

Productive vocabulary test

The maximum score that a learner could obtain on the productive vocabulary test was 60 points. Spelling mistakes were not taken into account because we wanted to test whether learners knew the meaning of the word, not its spelling. For the example below, many leaners gave the answer *resources*, with the English spelling instead of *ressources*, as it should be spelled in French. In spite of the exact number of letters that was given, participants still gave the answer they thought was right. They were not penalized for given this answer.

> Example
>
> ***Ce pays n'a plus de r_ _ _ _ _ _ _ _ _ (bronnen) naturelles et doit donc importer son énergie.***
>
> (This country does not have any natural **resources** anymore and has to import its energy.)

Speaking tasks

The spoken output on the semi-structured dialogic speaking tasks that were carried out with last year secondary school students were evaluated by expert raters (Study 3). The latter received a holistic lexical rating scale, that was translated to French and adapted from the rating scale conceived by Crossley, Salsbury, and McNamara (2015) (see Appendix 13). A holistic scale is a type of scale in which a single score is awarded and which "reflects the overall quality of the performance" (Fulcher, 2013, p. 208). The scale consisted of a 5-point Likert scale that prompted experts' evaluations. A descriptor was given for each of the 5 categories and it was specified that the distance between all of the scores should be considered equal.

Every performance was rated by two experts from a panel of three native speakers of French who had a background in language studies and/or teaching. They rated both L2 and L1 performances and

did not know beforehand what type of performance they would listen to. After having participated in a training session, experts carried out the ratings of the audio files in their own time, on a voluntary basis.

3.7.2 Analyzing materials

VocabProfile

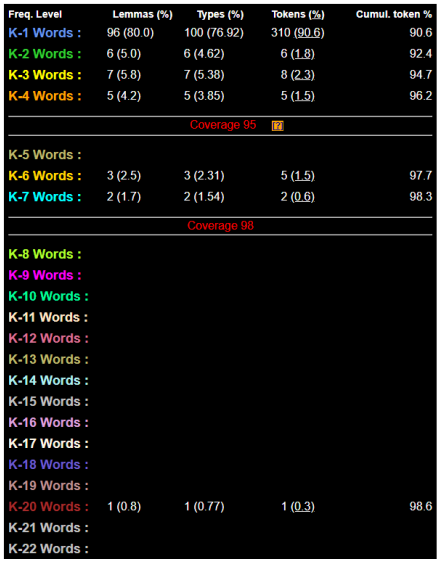
The Lexical Frequency Profile (LFP) was created by Laufer and Nation (1995) for English and provides an estimate of the proportion of high- and low-frequency words in a text by categorizing the different words of the output into frequency bands, which have been established based on written language corpora. The assumption behind this tool is that a higher proportion of low-frequency words would reflect a more advanced vocabulary.

Frequency is calculated for word families, i.e., a headword and its inflections and most common derivations, based on the BNC. The VocabProfile-program (Laufer & Nation, 1995), made available for free on Lextutor (Cobb, n.d.) is an online tool allowing to analyze the word frequency in a text. It provides information on the first 2,000 most frequent words, divided into two frequency band of 1,000 words and the Academic Word List (AWL; Coxhead, 2000), which consists of 570 words from academic texts. There is a "off-list" category for the words that do not belong into those three categories. According to Laufer and Nation (1995), the tool allows to differentiate between language proficiency levels. The tool was used in different studies (e.g., Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006; Uchihara & Clenton, 2018). More recently, several other word frequency lists are available for English, among which the BNC/COCA word frequency lists containing 25.000 word families. More information about the frequency lists for English can be found on the website[11].

[11] The website can be accessed on the following address: www.lextutor.ca/vp/comp/.

In 2004, Cobb and Horst decided to develop a French version of the LFP (Cobb & Horst, 2004). It provides the LFP of a text based on 1,000 word frequency bands (see Figure 3.2). The frequency bands obtained are determined by an analysis of research corpora, which is Lonsdale and Le Bras' 25K-list for French. According to the information on the Lextutor-website[12], the French lists are based on

> "a dedicated 23-million-word corpus of French which includes a balanced sample of both written and spoken material, both literary and non-literary material, from both France and other places (mainly Canada) where French is spoken, and employs criteria of both frequency and range (distribution throughout the corpus rather than just in one part of it)"



Freq. Level	Lemmas (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	96 (80.0)	100 (76.92)	310 (90.6)	90.6
K-2 Words :	6 (5.0)	6 (4.62)	6 (1.8)	92.4
K-3 Words :	7 (5.8)	7 (5.38)	8 (2.3)	94.7
K-4 Words :	5 (4.2)	5 (3.85)	5 (1.5)	96.2
Coverage 95				
K-5 Words :				
K-6 Words :	3 (2.5)	3 (2.31)	5 (1.5)	97.7
K-7 Words :	2 (1.7)	2 (1.54)	2 (0.6)	98.3
Coverage 98				
K-8 Words :				
K-9 Words :				
K-10 Words :				
K-11 Words :				
K-12 Words :				
K-13 Words :				
K-14 Words :				
K-15 Words :				
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :				
K-20 Words :	1 (0.8)	1 (0.77)	1 (0.3)	98.6
K-21 Words :				
K-22 Words :				

Figure 3.2 Example of the Vocabprofile output received after introducing a text

However, it is important to note that, compared to the English counterpart of the tool, the highest word level that is used is the lemma rather than the word family. Lemmas are base words and their

[12] Retrieved from: www.lextutor.ca/vp/comp/lonsd.html

inflections within the same part of speech (e.g., *enfant*, *enfants*; *ris*, *rit*), whereas word families also include derivations which means that a change in the part of speech is possible (e.g., *lire*, *lu*, *lecteur*, *lecture*).

Finally, it should be mentioned that the Lextutor tool is most widely used and validated for English (Laufer & Nation, 1995) although it has also been validated for French (e.g., Lindqvist, 2010; Lindqvist, Bardel, & Gudmundson, 2011; Ovtcharov, Cobb, & Halter, 2006). The aforementioned studies were based upon the old version of Lextutor, which only measured up to the 3,000 most frequent words. Researchers obtained a percentage for the 1,000, 2,000 and 3,000 most frequent words and for the words that were 'off-list'. In 2013, however, major changes were made to the French corpus and, as far as we know, no study has validated since then the new Lextutor for French. The corpus has now been largely extended and measures up to the 25,000 most frequent words in French. This is, of course, an additional reason for using the program.

CHAT and CLAN

In our second and third study, we made use of the Codes for the Human Analysis of Transcripts (CHAT) transcription and coding format and the Computerized Language Analysis (CLAN) program for several reasons. First, it is a free program that is well-supported since there is a thorough follow-up by the developers. Questions can be asked and are answered rapidly. Second, as several thousands of studies using these tools have been published[13], the tools have international standing. Finally, they are not language-specific and contain over 40 basic commands that can be adapted to the researchers' needs.

[13] Over 3,000 publications making use of the tools are available on http://talkbank.org/info/usage/childesbib.pdf.

The CHAT and CLAN tools were developed, together with a database, within the CHILDES Project (Child Language Data Exchange System), that was initiated in 1984 by Brian MacWhinney and Catherine Snow (University Carnegie Mellon, Pittsburgh, USA). It initially aimed at transcribing naturalistic child language data and it has now become a huge online database of language data[14].

The CLAN program, "designed to make full use of the CHAT format to facilitate a wide variety of searches and analyses" (CHAT manual, p. 13; MacWinney, 2000a), is a tool that allows the transcription of spontaneous spoken language in the CHAT format, starting from audio or video input. The aim is to transcribe the language as much and as completely as possible (see Figure 3.3). The program also allows to perform a certain number of analyses on the transcribed spoken output, e.g., frequency analyses and calculation of lexical diversity.

[14] The database can be accessed on the following website: https://childes.talkbank.org/.

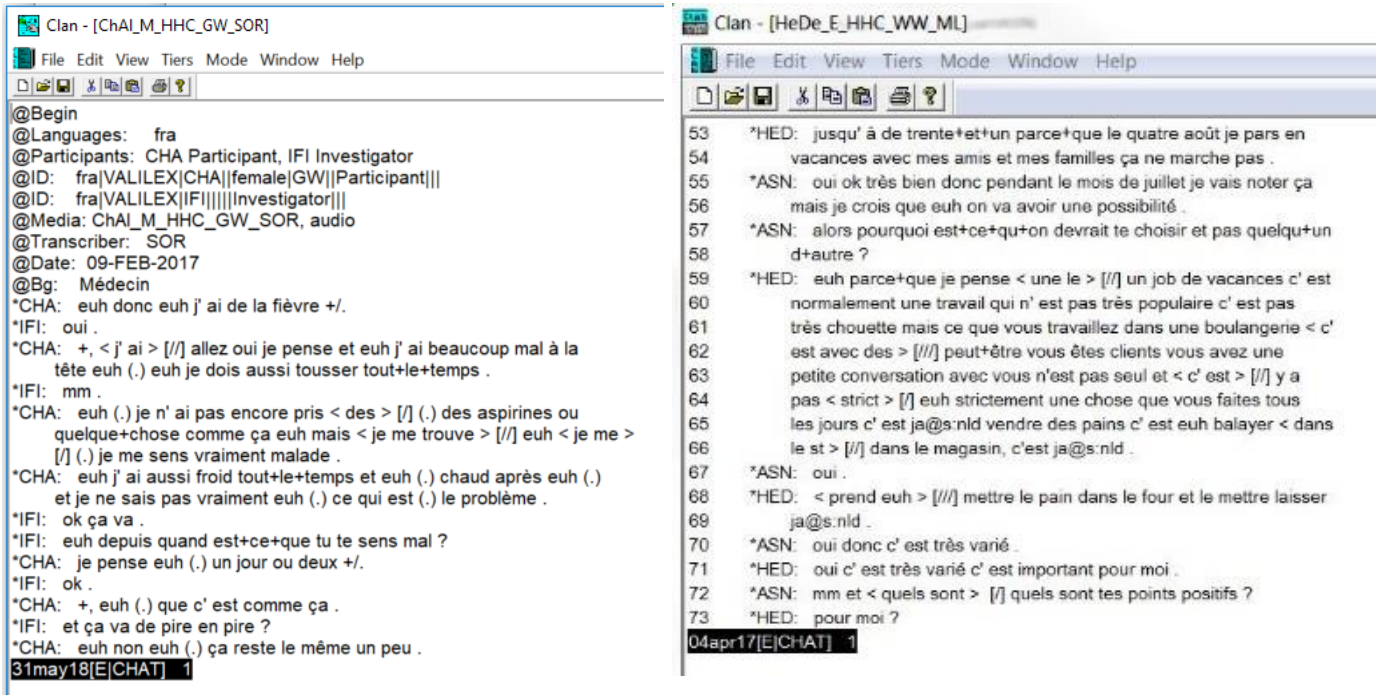


Figure 3.3 Part of a doctor's visit transcription (left) and of a job interview transcription (right) in CLAN

There are a number of conventions proper to the CHAT format, regardless of the language under research and additionally some conventions that are specific to French:

- Negation: when a part of the negation is not said by the speaker, it will not be transcribed, e.g., (ne)… pas
- Subject: the subject is pronounced in its abbreviated form but will be transcribed entirely, e.g., t'as → tu as
- Elision at the beginning or the end of a word, or inside a word is noted between brackets, e.g., (en)fin, p(e)tit
- Hesitations are always transcribed as 'euh', independently of the speakers' way of pronouncing them, e.g., euhm, heu, heum
- Confirmation markers are always transcribed as 'mm', independently of the speakers' way of pronouncing them, e.g., uhu, hmm
- Multiword units are transcribed containing a '+' between both words so that they are counted as one word, e.g., quand+même, job+d+été, qu+est+ce+que
- The same holds for compounds having a '-' in the word, e.g., peut-être → peut+être
- Words presenting an apostrophy inside the word (e.g., aujourd'hui) will be considered as one word. In the case of *aujourd'hui*, this will be fine, but for all other cases, there should be a space between both words, e.g., l' image (= la image, two separate words).

3.7.3 Statistical analyses

Before starting our analyses, we verified the normal distribution of the data by looking at the skewness and the kurtosis of the different variables. Positive values of skewness are indicative of a large number of low scores on the left of the distribution, whereas a negative skewness is indicative of a large number of high scores on the right. A pointy and heavy-tailed distribution involves positive values of kurtosis, whereas a flat and light-tailed distribution involves negatives values of kurtosis. Values that are far away from zero indicate that data

might not be normally distributed. For small samples ($N < 50$), *z*-score values above 1.96 indicate that data is likely not to be normally distributed. For larger samples ($50 < N < 300$), however, the null hypothesis should only be rejected when values of 3.29 are obtained (Kim, 2013, p. 53).

Table 3.9 summarizes the different types of analyses per study. We will start with the parametric tests we conducted. In all three studies, correlations were calculated in order to determine the relationship between different variables. A Pearson correlation coefficient measures the strength of the linear relationship between two variables. In order to be able to conduct a Pearson correlation, several assumptions should be met. First, data should be continuous interval data, which means that a score that can take any value on the measurement scale is attributed to each person. The intervals between the numbers on the scale should "represent equal differences in the property being measured" (Field, 2009, p. 10). Moreover, in order to conduct a Pearson correlation, there should be a linear relationship between both variables, no significant outliers should be present in the dataset and normal distribution should be guaranteed. If one of the variables is not normally distributed, however, Pearson correlations can still be computed if the sample size is big enough. The outcome values usually lie between -1 and +1, in which a coefficient of +1 indicates a perfect positive relationship and -1 a perfect negative relationship. A coefficient of 0 shows that there is no linear relationship at all between the two variables. Cohen's (1988) interpretation is used for these *r* values: .1-.3: small effect, .3-.5: medium or intermediate effect, .5 and higher: large or strong effect. In studies 2 and 3, Pearson correlations were computed because data were normally distributed, whereas in the first study the assumption of normality was violated. Therefore, Spearman's Rank Order correlations were computed between the results on the vocabulary knowledge test and the listening comprehension test.

Table 3.9

Statistical methods used in the three studies

Study 1	***Study 2***	***Study 3***
Spearman's Rank Order correlation Mann-Whitney-U tests	Pearson correlations	Pearson correlations Independent *t*-tests Mann-Whitney-U tests Mixed linear effects modelling

Whereas the correlation (*r*) explains the strength of the relationship between two variables, r^2 explains the amount of variability in one variable that is shared by the other. It thus represents the proportion of the variance for one variable that is explained by another variable. Effect sizes, which are "standardized, quantitative indication[s] of a relationship or an effect" (Plonsky, 2015, p. 31), are useful because "they provide an objective measure of the importance of an effect (Field, 2009, p. 57). It has become common practice in SLA to report effect sizes in addition to inferential statistics, as outlined by De Keyser and Schoonen (2007) and Plonsky (2015). According to Plonsky (2015, p. 36), there are three main reasons why we should include effect sizes. First, unlike *p* values, they are not influenced by sample size. Second, they inform about the strength of a relationship and finally, they are "continuous, standardized […] and scale-free", as opposed to the arbitrary nature of the *p* value.

Further, in order to compare two groups with each other, both the parametric as well as the non-parametric variants of the tests were used. As data were normally distributed, parametric independent *t*-tests were used to compare different groups, i.e., the L1 and the L2 speakers (study 3). In this type of analysis, the means of two independent groups are compared. If the significance value is less than .05, the means of the two groups are significantly different. In study 3, L1 and L2 speakers' number of words and their frequency was

compared using mainly independent *t*-tests. In the first study, however, the Mann Whitney-U test, the non-parametric version, was used because the assumption of normality was not met. This analysis aimed at verifying whether the scores on the listening test for the L2 learners belonging to different frequency levels, i.e., mastering a different number of words, were statistically significant or not. In this study, the different independent groups constituted by the frequency levels to which the learners belonged constituted the categorical variable.

In study 3, a multivariate linear mixed effects model was run to determine which lexical variables could predict experts' ratings of L1 and L2 speakers' oral output. As linear mixed models allow both fixed and random effects, they are an extension of simple linear models and are one way of dealing with hierarchy in the data (Linck & Cunnings, 2015). The holistic rating score, a continuous variable, was our dependent variable. The covariates that could predict the rating scores were the lexical measures that were under investigation. Next, the fixed effect in this analysis was the topic of the output, as there were two topics (i.e., a doctor's visit and a job interview) per participant. Finally, the participants were added as random effects in the analysis. In mixed models, the random variable is added because of the assumption that observations within the participants-level (i.e., the random variable in this study) are correlated. These types of models are designed in a way that they address this correlation and do not violate the assumption of independence of observations.

All analyses were conducted with IBM SPSS Statistics (versions 24 and 25). Data were reported according to the guidelines provided by the Publication Manual of the American Psychological Association (6th Edition).

Chapter 4

Study 1: Vocabulary knowledge and listening comprehension at an intermediate level in English and French as foreign languages. An approximate replication of Stæhr (2009)

4.1 Introduction

It has been argued that 2,000-3,000 word families are needed for comprehension of spoken English input, if 95% coverage is sufficient, and 6,000-7,000 word families, if 98% coverage is required (Nation, 2006). One study that has empirically investigated the relationship between vocabulary knowledge, lexical coverage, and listening comprehension is Stæhr (2009). His study showed that knowledge of the 5,000 most frequent word families is needed for 98% coverage and for adequate listening comprehension at an advanced level. The current replication study aims to expand Stæhr's findings to a different population (Flemish students), listening comprehension at a lower proficiency level (intermediate) and two foreign languages (English and French).

4.2 Background

4.2.1 The original study

Stæhr (2009) investigated the relationship between vocabulary and advanced listening comprehension with Danish English-as-a-foreign language (EFL) learners (N = 115). Participants took the advanced listening comprehension test from the *Cambridge Certificate of Proficiency* in English. Two aspects of vocabulary knowledge were measured: breadth (by means of the revised version of the written *Vocabulary Levels Test*, Schmitt et al., 2001) and depth (by means of the written *Word Associates Test*, Read, 1993). The results revealed that there was a strong, positive correlation between vocabulary breadth and advanced listening comprehension (r = .70). A regression analysis showed that breadth explained 49% of the variance and depth 2% of the variance. Stæhr also found that a larger vocabulary knowledge and more lexical coverage resulted in better listening comprehension at an advanced level. Knowledge of the 3,000 most frequent word families provided learners with almost 94% lexical coverage, whereas

knowledge of the 5,000 most frequent word families corresponded to 98% lexical coverage. Additionally, his analyses revealed that there was a significant difference in comprehension score between learners mastering the 3,000 most frequent and learners mastering the 5,000 most frequent word families. Adequate comprehension was defined as a score of 70% or higher. The results showed that learners needed a vocabulary of 5,000 word families for adequate listening comprehension at that advanced level.

4.2.2 Importance of the problem

Given that research has consistently shown that vocabulary is a key factor in reading comprehension, it is essential to investigate its role in other skills as well. Exploring the relationship between vocabulary knowledge, lexical coverage and listening comprehension is relevant if we want to establish realistic vocabulary learning targets and determine how many words a learner needs for listening comprehension in a foreign language (Schmitt et al., 2017). Although no other study has investigated vocabulary knowledge, lexical coverage, and listening comprehension in one study, except for Bonk (2000) and Stæhr (2009), three studies have focused on vocabulary and listening: two corpus-driven studies (Adolphs & Schmitt, 2003; Dang & Webb, 2014) and one intervention study (van Zeeland & Schmitt, 2013).

4.2.3 Relevant scholarship

Bonk (2000) was one of the first to look at the relationship between lexical coverage and L2 listening comprehension, although in an indirect way since he looked at the amount of familiar lexis. Learners had to listen to four passages with increasing lexical complexity. The listening was followed by an L1 written recall test for assessing learners' listening comprehension and a dictation test was administered in order to assess learners' lexical comprehension. The dictation scores were correlated with the L1 recall scores. Bonk found that higher scores on the dictation test corresponded to better listening

comprehension (r = .50) but he did not find a particular lexical threshold for achieving good comprehension, as there was variation in comprehension scores. However, high comprehension scores were rarely associated with low lexical knowledge scores.

Another study researching the lexical coverage of spoken discourse was Adolphs and Schmitt's (2003) corpus study of every day spoken language. Drawing on Bonk (2000) and the figures proposed in reading studies, they concluded that knowledge of the 2,000 most frequent word families is necessary to start to adequately engage in every day communication, if 95% lexical coverage is desired. The 2,000 most frequent words (not word families) provided 92.26% to 93.30% coverage in the CANCODE and BNC.

In another corpus-driven study, Dang and Webb (2014) investigated the lexical coverage of spoken academic English. Their analyses revealed that knowledge of the 4,000 most frequent word families provided 96.05% coverage, although there were differences between sub-disciplines. However, learners familiar with the 3,000 most frequent word families and the *Academic Word List* (AWL) might reach 95% coverage. This study showed that a larger vocabulary knowledge is needed for spoken academic English than for more informal varieties of English. Yet, it remains unclear whether 95% is required for adequate comprehension.

In a highly-controlled study, van Zeeland and Schmitt (2013) investigated lexical coverage in L2 listening comprehension with high-intermediate and advanced EFL learners. All participants listened to four stories, which differed in lexical coverage (100%, 98%, 95%, and 90%). The findings showed that adequate comprehension was achieved at the levels of 90% and 95% coverage. However, there was much more variation in the case of 90% coverage, showing that L2 listeners can differ greatly in how their comprehension is affected by the number of unknown words. Given the lower variation at the 95% coverage level,

van Zeeland and Schmitt proposed that 95% coverage is necessary for good comprehension, which would correspond to vocabulary knowledge of 2,000-3,000 word families. However, if 90% coverage is sufficient, learners would need to know approximately 750 to 2,000 word families.

The above studies add to our understanding of the relationship between vocabulary and listening by showing that the vocabulary knowledge and lexical coverage needed may vary according to the type of spoken input and the degree of comprehension. However, they did not measure vocabulary knowledge directly (Bonk, 2000), they focused on advanced learners of English only (Stæhr, 2009; van Zeeland & Schmitt, 2013), or they did not (really) test listening comprehension (Adolphs & Schmitt, 2003; Bonk, 2000; Dang & Webb, 2014).

4.2.4 Statement of purpose

The aim of the present approximate replication study was to further evaluate Stæhr's findings. An approximate replication study is a study which "involves repeating the original study exactly in most respects, but changing non-major variables" (Porte, 2012, p. 8) with the aim to generalize research findings to other contexts. This replication study aims to investigate whether Stæhr's findings can be generalized to another population (Flemish learners), another proficiency level (intermediate) and to another language (French) because it cannot be taken for granted that the vocabulary knowledge and lexical coverage figures for advanced listening comprehension can be applied to intermediate listening comprehension in English as well as French. Schmitt et al. (2017) have also pointed to the need for replication research into listening and vocabulary in order to set realistic vocabulary goals for listening comprehension. Given the lack of replication research, a study aiming to evaluate previous findings is essential because we need sound figures based on empirical research to determine the vocabulary demands for listening at different proficiency

levels. To that end, two experiments were set up in which non-major variables of Stæhr's study were changed, viz., a different population (Flemish learners), comprehension of listening texts of a different proficiency level (intermediate), and two foreign languages (English and French) (Porte, 2012).

4.2.5 Research questions

1. To what extent is vocabulary knowledge associated with listening comprehension?
2. How much vocabulary is needed for adequate listening comprehension?

We did not include Stæhr's research question "to what extent is depth of vocabulary knowledge associated with listening comprehension". This change was motivated by the lack of a validated *Word Associates Test* for French. Second, depth only played a minor role compared to vocabulary knowledge in Stæhr's study. We will first present the experiment on English, before moving on to the French leg of the study.

4.3 Experiment 1: English

4.3.1 Method

Participants

The original study collected data among 115 Danish EFL learners who were first-year students at the Business school in Copenhagen. These students were supposed to be at least at the B2 level of the CEFR. Unlike Stæhr's study, the focus in the first experiment was on the intermediate proficiency level or the B1 level according to CEFR. In the English experiment, 199 participants in Flanders (= the Dutch-speaking part in Belgium) were recruited from the second and fourth year of secondary education and the first year at university (see Table 4.1). At the end of secondary education, Flemish learners are expected

to have reached a B1 level for listening. However, because previous research has shown that a majority of learners in Flanders reach this level earlier in their school career (European Commission, 2012), especially for listening, we also recruited learners from the second and fourth year of secondary education to avoid a ceiling effect.

Table 4.1

Number of participants per group and years of formal instruction

Group/level	***N***	**Years of formal instruction**
2nd year of secondary education	64	0.5
4th year of secondary education	52	2.5
1st year university	81	6.5
All participants	197	

Most participants had Dutch as their L1. Nineteen participants had another L1 than Dutch. Data of learners whose L1 was English (*N* = 2) were not included in the analysis, bringing the total number of participants for the English experiment to 197.

Materials

The participants took two paper-and-pencil tests: a listening comprehension test and a vocabulary knowledge test.

Listening comprehension test for English: Preliminary English Test for schools (PET)

The original study tested advanced listening comprehension through the *Cambridge certificate of proficiency in English* (CPE), which is a standardized C2-listening test. Following Stæhr's research (2009), we selected an English listening test developed by the same institution (Cambridge English Language Assessment) but targeted at the intermediate level, i.e.,, the *Preliminary English Test* (PET) for schools, corresponding to the B1 CEFR-level. The listening test consists of four parts that each contain a different type of question: seven questions based on pictures, six multiple choice questions, six short-answer questions and six true-false questions. The questions in this test can be linked to topics described in the "overall listening comprehension" scale of the CEFR for B1, i.e., "can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure etc. including short narratives" (Council of Europe, 2001, p. 66). The test deals with the following situations: a discussion about a poster, an arrangement to meet each other, an interview with a student about her passion for the weather, a call for students to participate in a contest and a conversation between two people about their ski holiday. The listening test can be found in Appendix 13.

Receptive vocabulary test

Stæhr (2009) used the second version of the Vocabulary Levels Test (Schmitt et al., 2001) to measure participants' vocabulary knowledge. However, we used a self-designed vocabulary test, the VocabLab test (Peters et al., 2019)[15]. This change was motivated by two factors. We wanted to be able to compare the results of the first experiment (English) with those of the second experiment (French).

[15] The design, development and validation of the English and French vocabulary size test used in this study are reported elsewhere (Peters et al., 2019). The vocabulary test data of the present study were also used in the validation.

Therefore, a test had to be chosen that was developed according to the same principles as the French test. Second, at the time of our data collection, no French vocabulary test was available, which is why we used a self-designed vocabulary test for English and one for French. Third, the use of our own self-designed tests guaranteed that the same counting unit, lemmas, was used in the English as well as the French test, although it should be acknowledged that this makes a direct comparison between our findings and Stæhr's findings more difficult.

Like the VLT, the English vocabulary test provided an estimate of learners' vocabulary knowledge at different frequency levels, viz., 0-2,000 (2K), 2,001-3,000 (3K), 3,001-4,000 (4K), and 4,001-5,000 (5K). Given our focus on a lower proficiency level, it was not deemed necessary to test learners' vocabulary knowledge at the 10K level, as Stæhr did. Lemma frequencies for the English test were sampled from the *Corpus of Contemporary American English* (COCA) (Davies, 2008). In the item selection, the number of cognates and loanwords was minimized as they might inflate test scores (Elgort, 2013; Laufer & McLean, 2016). We acknowledge that, ideally, the number of cognates in a test section corresponds to the number of cognates in the frequency band tested (Laufer & Levitzky-Aviad, 2018), but given the time-consuming nature of such an analysis, this was considered not feasible in the development of the test. We are, however, aware that the test might thus slightly underestimate learners' vocabulary knowledge, although it should be pointed out that the cognate facilitation effect is more pronounced for low-frequency items than for high-frequency items (Elgort, 2013; Laufer & McLean, 2016).

The test, which consists of 120 items, has a written multiple choice recognition format in which test items are presented in isolation. Each item is accompanied by four definitions (one correct answer and three distractors) and one "I don't know"-option to minimize guessing. Given that the test aims to minimize guessing and thus the risk of overestimation, the test probably provides a more conservative

estimate of a learner's knowledge compared to other existing tests, although this still needs to be examined empirically. The vocabulary test can be found in Appendix 15.

Example of test item of the English vocabulary test

Guide

- *A table that you use when you work*
- *Something that you hope will happen*
- *The money that you have to give back to someone*
- *A person or book that gives information about a place*
- *I don't know the answer*

Procedure

The participants from secondary education completed the tests in two sessions as one English class lasts fifty minutes, which was too short to complete both tests. They took the vocabulary test in one session and the listening comprehension test in the second session. The first-year university students took both tests in one session (1.5 hours). They started with the listening comprehension test.

Scoring and analyses

The vocabulary test was scored dichotomously: one point for every correct answer, zero points for every incorrect answer (including the "I don't know"-option). The maximum score on the entire vocabulary test was 120 with a maximum score of 30 on each section (2K, 3K, 4K, 5K). In line with Stæhr's (2009) study, mastery of a vocabulary level was set at 90% or 27 out of 30. The listening comprehension test was corrected according to the key of the *Cambridge English Language Assessment*. One point was given for every correct answer. The maximum score on the English listening comprehension test was 25 points (i.e., one point per question).

Unlike the original study, the data in experiment 1 were not normally distributed. Consequently, we used non-parametric tests. A Spearman's Rank Order Correlation was used to answer the first research question, while a Kruskal-Wallis test was selected to answer the second research question.

4.3.2 Results

Descriptive statistics

The results in Table 4.2 show that the mean score on the listening test is relatively high (almost 86%). The reliability of the listening test (Table 4.2) is higher than Stæhr's Cronbach's alpha of .60, probably because there was less homogeneity in our sample. Plonsky and Derrick (2016) argue that reliability figures for listening are generally lower than for other skills. Scores were lower on the vocabulary test. There was considerable variation in learners' vocabulary knowledge, as can be seen in the standard deviations (in Table 4.3). The reliability coefficient of the vocabulary knowledge test was high (.98) and in line with Stæhr's reliability coefficient for the VLT (.91).

Table 4.2

Descriptive statistics (in percentages) for the English listening test (N = 197)

Test	***M*** **(*SD*)**	**95% CI**	**Min – Max**	**Cronbach's alpha**
Listening test	85.70% (11.58)	[84.08, 87.32]	32-100%	.75

Table 4.3

Descriptive statistics (in percentages) for the English vocabulary test (N = 197)

	Total	2K	3K	4K	5K
Mean (SD)	66.95% (22.34)	83.05% (19.09)	71.78% (23.98)	61.40% (23.04)	51.56% (27.20)
95% CI	[63.80, 70.09]	[80.36, 85.72]	[68.41, 75.15]	[58.17, 64.64]	[47.73, 55.38]

To what extent is vocabulary knowledge associated with listening comprehension?

As our data were not normally distributed, we computed a Spearman's Rank Order Correlation to determine the relationship between the total scores on the vocabulary test and the listening comprehension test. We found a significant, positive correlation of r_s = .63, explaining 40% of the variance (confidence interval: 0.54; 0.71; p < .0001). Although our correlation is slightly lower than Stæhr's (r = .70), our findings show that there is also a large and positive correlation between learners' vocabulary knowledge and listening comprehension at an intermediate level as measured by a B1-test.

How much vocabulary is needed for adequate listening comprehension?

Stæhr found that 98% lexical coverage and knowledge of 5,000 word families is needed for adequate listening comprehension at an advanced level. To explore the relationship between vocabulary knowledge, lexical coverage, and listening comprehension, we duplicated Stæhr's method. We first classified participants on the basis of their vocabulary knowledge. We used the same cut-off point for mastering a vocabulary level, viz., 90% or a score of 27 out of 30. However, our data showed that it was necessary to also distinguish between the 1K and 2K frequency band. To this end, the 2K section,

which included 1K and 2K words, was divided into a 1K section and a 2K section. The former comprised 13 items, while the latter contained 17 items. Participants who did not obtain a score of 11 (or more) out of 13 on the 1K words were labelled -1K. Students who obtained 11 or more out of 13 on 1K items but less than 15 out of 17 on the 2K items were also labelled 1K.

Secondly, we also carried out a lexical coverage analysis of the listening comprehension test using the same frequency list as Stæhr, viz., the BNC-20, which is available on the website *Lextutor* (Cobb, n.d.). It should be pointed out here that the BNC-20 is divided into 20 lists, each containing 1,000 word families (not lemmas), whereas the counting unit for the vocabulary test was the lemma. As can be seen in Table 4, the analysis revealed that the first 1,000 word families will provide almost 92% coverage. The second 1,000 word families will result in 97% coverage of the listening text. This contrasts with Stæhr's advanced listening text, in which knowledge of the 4,000 most frequent word families was needed to reach 97% coverage. The analysis showed that there were only a few words in the PET-test that were not from the 2,000 most frequent word families in English.

Like in Stæhr (2009), the listening comprehension test scores increased, when learners' vocabulary knowledge increased. Participants in the -1K group had an average listening comprehension score of 69.17%. However, it should be pointed out that there was considerable variance at this level, as the lowest score in this group was 32%. Learners who were familiar with the 1K level had a mean score of 85% on the listening test, which is higher than the average score of the learners in the -1K group. As can be observed in Table 4.4, knowledge of the 1,000 most frequent word families will lead to 91% lexical coverage of the listening text. However, given that the vocabulary test (lemmas) and the Lextutor BNC-20 (word families) used different counting units, it is not easy to determine the exact amount of lexical coverage of the learners. Learners who knew the 2,000 most frequent

words (as measured in the vocabulary test) scored on average 91% on the listening test. However, the listening comprehension scores seemed to level off from 2K onwards, which is different from Stæhr. A Kruskal-Wallis analysis showed that there was a significant difference in the listening comprehension score between the six groups ($H(5) = 85.79$, $p < .0001$). Pairwise comparisons with adjusted p-values showed that there was a significant difference in the average listening comprehension scores between participants not mastering the 1K level and participants mastering the 1K level ($p < .0001$; $d = 1.52$) with a large effect and between participants mastering the 1K and participants mastering the 2K level ($p = .008$, $d = .85$) with a medium effect. The other groups did not differ significantly from each other. Stæhr only found a significant difference between the 3K and 5K vocabulary level. To sum up, compared to the CPE-test used by Stæhr, the PET-test contained few words (2.5%) that did not belong to the 2000 most frequent word families in English. Unsurprisingly, knowledge of fewer word families, as indicated by Lextutor BNC-20, was needed to reach 95% and 98% coverage in the PET-test than in the CPE-test. Secondly, the results suggest that knowing or not knowing the most frequent 1,000 words on the one hand and knowing the 2K words on the other has an impact on listening comprehension scores at an intermediate level.

Table 4.4

Descriptive statistics and lexical coverage for the English listening test based on K levels

K level	*N*	*M* listening (*SD*)	95% CI	Cumulative lexical coverage
-1K	36	69.17% (12.96)	[64.78, 73.55]	/
1K	52	85.07% (8.35)	[82.75, 87.40]	91.45%
2K	43	91.23% (5.55)	[89.59, 93.01]	97.42%
3K	37	91.20% (5.64)	[89.36, 93.13]	98.40%
4K	9	91.60% (4.81)	[87.42, 94.81]	98.69%
5K	20	92% (6.07)	[88.56, 96.24]	98.86%
6K	/	/		98.97%
7K	/	/		99.08%
8K	/	/		99.14%
9K	/	/		99.20%
Off-list	/	/		100%

4.4 Experiment 2: French

4.4.1 Method

Participants

In the French experiment, 351 Flemish participants were recruited from the second and fourth year of secondary education and the first year at university (see Table 4.5). At the end of secondary education, Flemish learners are expected to have reached a B1 level for listening in French. Learners of the second and fourth year also participated in the French experiment for reasons of comparability with the English experiment.

Table 4.5

Number of participants per group and years of formal instruction

Group/level	***N***	**Years of formal instruction**
2nd year of secondary education	59	3.5
4th year of secondary education	47	5.5
1st year university	224	8.5
All participants	330	

Most participants had Dutch as their L1. Thirty-two participants had L1's other than Dutch. Data of participants whose L1 was French (N = 15) were eliminated from the analyses. As a result, data of 330 participants were used in the analyses.

Materials

Similar to the first experiment, the participants took two paper-and-pencil tests: a listening comprehension test and a vocabulary test.

Listening comprehension test for French: Diplôme d'études en langue française (DELF)

Drawing on Stæhr and the English experiment, we also selected an internationally recognized and standardized listening test, viz., the B1 DELF-test developed by the *Centre International d'Études Pédagogiques* (CIEP) (Appendix 2). The B1 DELF listening test consists of three parts. The first and second part contain six multiple choice questions each and the third part seven multiple choice questions and four open-ended questions. The themes that are tackled are a telephone conversation between mother and daughter, testimony from a worker, and an actor talking about his job.

Receptive vocabulary test

Like the English vocabulary test, the French vocabulary test provides an estimate of learners' vocabulary knowledge at four frequency levels, viz., 0-2,000 (2K), 2,001-3,000 (3K), 3,001-4,000 (4K), and 4,001-5,000 (5K). Lemma frequencies for the French test were sampled from *A Frequency Dictionary for French: core vocabulary for learners* (Lonsdale & Le Bras, 2009). The French vocabulary test, which was also a meaning recognition test consisting of 120 test items, was developed according to the same principles as the English test and was based upon lemmas since there was no test using word families available for French. The vocabulary test can be found in Appendix 3.

Example of test item of the French vocabulary test

Goût (m)

- *morceau long et fin*
- *période de cent ans*
- *terre entourée d'eau*
- *impression laissée dans la bouche*
- *je ne connais pas la réponse*

Procedure

The procedure was identical to the one used in the English experiment. Details can be found in Section 3.1.3.

Scoring and analyses

The vocabulary test was scored in the same way as the English vocabulary test. As for the French listening comprehension test, scores were given according to the instructions of the CIEP. One point was given for multiple choice questions and for two of the four open-ended questions, whereas two points were given for the two remaining open-ended questions. The participants received partial credit when they provided only a part of the expected answer. A total score of 25 points on 23 questions could be obtained.

The data in experiment 2 were not normally distributed (Shapiro $p < .05$). Consequently, we used the same non-parametric tests as in the English experiment.

4.4.2 Results

Descriptive statistics

Compared to the English test results, Tables 4.6 and 4.7 illustrate that the average score on the vocabulary test and the listening test is lower. Moreover, the standard deviation of the scores on the listening test is also higher for French, indicating that there might be more

variation in the French results. The reliability coefficient of the vocabulary test (.96) and the listening test is in line with the results for English (both vocabulary test and listening test) and Stæhr's research.

Table 4.6

Descriptive statistics (in percentages) for the French vocabulary test (N = 330)

	Total	**2K**	**3K**	**4K**	**5K**
Mean (SD)	51.62% (18.25)	71.85% (19.08)	51.06% (19.33)	43.45% (21.28)	40.13% (18.51)
95% CI	[49.65, 53.60]	[69.78, 73.91]	[48.97, 53.15]	[41.15, 45.76]	[38.13, 42.14]

Table 4.7

Descriptive statistics (in percentages) for the French listening test (N = 330)

Test	***M* (*SD*)**	**95% CI**	**Min – Max**	**Cronbach's alpha**
Listening test	69.51% (14.65)	[67.97, 71.11]	20-96%	.70

To what extent is vocabulary knowledge associated with listening comprehension?

The data of the French experiment were not normally distributed. Therefore, a Spearman's Rank Order Correlation was computed to determine the relationship between the total scores on the vocabulary test and the listening comprehension test. We found a significant, positive correlation of $r_s = .70$, explaining 49% variance (95% confidence interval: 0.64; 0.75; $p < .001$). This is somewhat higher than the correlation found in the English experiment and equal to Stæhr's finding.

How much vocabulary is needed for adequate listening comprehension?

The method we used to explore the relationship between vocabulary knowledge, lexical coverage and listening comprehension in our experiment is the same as for the English experiment. We first classified participants on the basis of their vocabulary knowledge and used the same cut-off point for mastering a vocabulary level, viz.,, 90% or a score of 27 out of 30. However, as in the English experiment, the data for French showed that it was necessary to also distinguish between the 1K and 2K frequency band. To this end, the 2K section, which included 1K and 2K words, was divided into a 1K section and a 2K section. The former comprised 12 items, while the latter contained 18 items. Participants who did not obtain a score of 10 out of 12 on the 1K words were labelled -1K, participants who obtained 10 or more out of 12 on 1K items but less than 16 out of 18 on the 2K items were labelled 1K. Participants who obtained 16 or more out of 18 on the 2K items were labelled 2K.

Secondly, we also carried out a lexical frequency analysis of the listening comprehension test using *Lextutor* (Cobb, n.d.), which uses the same frequency list as the one that was used for the vocabulary test, viz., *A frequency dictionary of French*. This means that the unit of counting (lemmas) was the same in the vocabulary test and in frequency analysis of the listening test. As can be seen in Table 4.8, the analysis revealed that the first 1,000 lemmas provided 86% coverage. Only a few words in the listening test were not from the 3,000 most frequent lemmas in French according to *A frequency dictionary of French*.

Table 4.8 provides the results for the analysis of the relationship between learners' vocabulary knowledge, lexical coverage of the listening test, and listening comprehension.

As in the English experiment, the listening test scores increased, when learners' vocabulary knowledge increased. Participants in the

-1K group had an average listening comprehension score of 61.57%. Knowledge of the 1,000 most frequent lemmas provided the learners with 86% lexical coverage. In addition, learners in the 1K group had a higher score on the listening test (74%). Learners who mastered the 2,000 most frequent lemmas obtained 82.10% on the listening test and had 93.26% lexical coverage. Participants mastering the 3,000 most frequent lemmas scored 86.40% on the listening test, which corresponded to 96.42% lexical coverage. However, it has to be pointed out that there were too few participants in the 3K and 4K groups in order to draw any firm conclusions. A Kruskal-Wallis analysis showed that there was a significant difference in the listening comprehension score between the five groups ($H(4) = 117.06$, $p < 0.001$, $N = 330$). Pairwise comparisons with adjusted p-values showed that there was a significant difference in the average listening comprehension scores between participants in the -1K group and participants in the 1K group ($p < 0.0001$; $d = .67$) with a small effect, and between participants in the 1K and the 2K groups ($p = 0.002$; $d = .84$) with a medium effect. The other groups did not differ significantly from each other. The French results seem to be in line with the results of the English experiment, although it should be acknowledged that we cannot make any claims beyond the 2K level given that few students mastered the 3K, 4K or 5K word level.

Table 4.8

Descriptive statistics and lexical coverage for the French listening test based on K levels

K level	*N*	*M* listening (*SD*)	95% *CI*	Cumulative lexical coverage
-1K	167	61.57% (13.79)	[59.43, 63.83]	/
1K	93	74.00% (9.97)	[71.93, 75.96]	86.24%
2K	62	82.10% (9.08)	[79.55, 84.42]	93.26%
3K	5	86.40% (11.52)	[77.60, 95.20]	96.42%
4K	3	84% (4.00)	[80.00, 88.00]	97.25%
5K	0	/	/	97.94%
6K	/	/	/	98.49%
9K	/	/	/	98.77%
10K	/	/	/	99.05%
11K	/	/	/	99.19%
12K	/	/	/	99.33%
24K	/	/	/	99.47%
Off-list	/	/	/	100%

4.4.3 The relationship between vocabulary and listening comprehension

The first research question in this study sought to determine the relationship between vocabulary and listening comprehension. The results demonstrate that there is a large and positive correlation between intermediate EFL and French-as-a-foreign language learners' vocabulary knowledge and their scores on listening comprehension tests. Learners knowing more words tended to score better on the listening tests. Both the results of the English and the French experiment thus corroborate Stæhr's (2009) study showing a strong correlation between vocabulary knowledge and listening comprehension. Our study indicates that Stæhr's findings can be generalized to another population (Flemish learners), another proficiency level, viz., intermediate proficiency level operationalized by a B1 listening comprehension test, and to another foreign language, French.

It should be pointed out that, as in Stæhr's study, a written vocabulary test was used in the present study, even though listening requires comprehension of spoken words. It could be argued that recognition of a written word form does not entail recognition of a spoken word form. This could be considered a limitation of the present study. However, as this is a replication study, we did not want to make major changes to the test formats used. Additionally, at the time of the data collection, no aural vocabulary tests for English and for French were available. Finally, as pointed out by Stæhr, a written vocabulary test might be an appropriate measure for listening as well as reading comprehension, whereas an aural vocabulary test might only be suitable for listening (p. 597).

4.4.4 How much vocabulary is needed for adequate listening comprehension?

The findings of the current study are consistent with Stæhr's research in that they indicate that learners who knew more words obtained higher listening comprehension scores, even though Stæhr's figures were higher than ours. He found a significant increase from the 3,000 to the 5,000 vocabulary level, which corresponded to 94% or 98% lexical coverage respectively. Nonetheless, our results also differ from Stæhr's, who found a significant difference in listening comprehension scores between learners mastering the 3,000 word level and learners mastering the 5,000 vocabulary level. We only observed significant increases in listening comprehension for the highest frequency bands. Put differently, for English, we found a significant and large increase in comprehension from knowledge of less than 1,000 words to knowledge of 1,000 words and an increase with a medium-effect size from knowledge of the 1K to the 2K word level. No other differences were found. For French, we found a significant increase in comprehension from knowledge of less than 1K words to knowledge of 1K words and a medium-sized increase from the 1K word level to the 2K word level (86% to 93% lexical coverage). The different findings between our study and Stæhr's study can be explained by differences in the lexical profile of the listening tests. The B1-tests used in our study contained more high-frequency words than the C2-listening test used by Stæhr. This was illustrated in the number of word families (English) and words (French) needed to reach 95% and 98% coverage in the PET listening test and the CPE listening test. The vocabulary size that is needed to reach this coverage level seems to depend on the input text and the required degree of comprehension.

The question arising is what constitutes adequate listening comprehension at an intermediate level. If a comprehension score of 69% is seen as adequate for the PET-test, then learners do not even need to know the most frequent 1,000 lemmas in English. Yet, there was

considerable variation at this level. Looking more closely at learners' listening performance in this group, we found that 52% of the learners who did not master the 1K level (N = 19) had a score lower than 70% on the listening comprehension test, which is considered the passing level of the PET-test. Moreover, one learner even had a score of only 32%. It, thus, seems fair to conclude that knowledge of less than the 1,000 most frequent lemmas or 625 word families[16] will not suffice for adequate comprehension of listening text of an intermediate proficiency level. If acceptable comprehension is approximately 85%, then knowledge of the most frequent 1,000 lemmas or 625 word families will suffice for the English PET-test. Further, if adequate comprehension is viewed as a listening score of 91%, then learners will need to be familiar with the 2,000 most frequent lemmas or 1,250 word families in English. Finally, knowing more than 2,000 lemmas does seem to have an effect on listening comprehension at an intermediate level as measured by the PET-test. This is clearly a less daunting challenge than Stæhr's figure of 5,000 word families for advanced listening comprehension. Our figures seem to be closer to van Zeeland and Schmitt's (2013) findings who found adequate comprehension with 90% coverage in narrative texts. They argue that this would require knowledge of between 750 and 2000 word families.

The results for French show a similar picture. If adequate comprehension for French at the DELF-test is defined as 61%, then learners do not need to be familiar with the 1,000 most frequent lemmas in French. However, it should be pointed out that there was quite a lot of variance in this group as was the case for English. Of the learners who were not familiar with the 1K level, 26% had a listening score between 50% and 61% (N = 43). Twenty-three percent of the learners

[16] In order to make a tentative comparison between word families and lemmas, we used Milton's (2009) rule of thumb. He found that multiplying a word family by 1.6 gave an approximate estimation of the number of lemmas in that word family.

who did not master the 1,000 vocabulary level scored lower than 50% (*N* = 38), which is considered the passing level of the DELF-test. If approximately 74% is viewed as adequate comprehension for French, then knowledge of the 1,000 most frequent lemmas or 86% coverage might be enough, which is slightly lower than for English (91% coverage). If a listening score of more than 80% is required, knowledge of the 2,000 most frequent lemmas or 93% lexical coverage will be needed. It seems that 3,000 lemmas would be needed to provide learners with 96% of lexical coverage, which is higher than for English, although care should be taken given the few participants in the 3K group. It should be emphasized that these figures are based upon the results on the DELF-test. Results might be somewhat different with a more difficult listening test or even with another B1-test.

Given the variance in the group not familiar with the 1,000 most frequent lemmas in English and French, it seems that at least 1,000 lemmas should be known for listening comprehension at an intermediate level. Our findings suggest that knowledge of the 1,000 most frequent word families will provide learners with 91% coverage for English and knowledge of the most frequent 1,000 lemmas will result in 86% lexical coverage for French. Knowledge of the 2,000 most frequent English word families and the 2,000 most frequent lemmas in French will result in 97% and 93% coverage respectively. Our findings thus seem to suggest that learners might need less than 98% lexical coverage for listening comprehension of texts at an intermediate level as measured by the PET-test and the DELF-test, although it is not easy to directly compare the coverage figures because of the different unit of counting. Our findings seem to be closer to van Zeeland and Schmitt's (2013) coverage figures of 90% and 95%.

4.5 Limitations

There are a number of limitations in this approximate replication study. The first limitation regards the number of participants mastering the different K levels. As was most clear in the French experiment, the number of participants beyond the 2K level is rather limited, which clearly differs from the original study. As a result, it was not possible to draw any conclusions about those vocabulary levels and potential differences between vocabulary levels beyond the 2K level. Future studies should recruit participants with more vocabulary knowledge to investigate whether knowledge of the 3K and 4K (and beyond) vocabulary level has an effect on learners' listening comprehension.

Secondly, our study is limited in that only one listening test was used for the English and the French experiment. In order to draw any firm conclusions about the lexical demands for listening at an intermediate level it would be better to administer more than one listening test to the participants but this was beyond the scope of the present study, which aimed to replicate Stæhr's research.

Finally, it should be acknowledged that different tests were used in the present study compared to Stæhr's study. As mentioned in the description of the tests, the VocabLab tests might provide a more conservative estimate of learners' vocabulary knowledge compared to other tests. Further, our tests and the VLT used in Stæhr employ a different unit of counting, viz., lemmas versus word families making a direct comparison difficult. According to Milton (2009, p. 12), the rule of thumb that could be used to compare word families to lemmas is to multiply the word family by 1.6. Admittedly, this is a crude estimate but it allows a comparison of both figures to a certain extent. Given that most research so far has used word families as a counting unit, more research into the relationship between vocabulary knowledge and

listening is warranted if we are to generalize the findings to other languages and proficiency levels.

4.6 Conclusion

The aim of the study was to investigate whether Stæhr's (2009) findings could be generalized to another context. In line with his findings, we found a positive and strong correlation between learners' vocabulary knowledge and listening comprehension at an intermediate level both for English and for French. Our results seem to indicate that learners need to know the 1,000 or 2,000 most frequent lemmas in English and French for adequate listening comprehension at an intermediate level as measured by the PET-test and the DELF-test. These figures are clearly below those observed by Stæhr, who proposed knowledge of 5,000 word families as a learning goal for advanced learners. Finally, compared to Stæhr, our approximate replication study also seems to suggest that lower coverage figures (86% for French; 91% for English) might provide adequate listening comprehension at an intermediate level.

Chapter 5

Study 2: The relationship between French-as-a-foreign language learners' receptive and productive vocabulary knowledge and their vocabulary use in dialogic speaking tasks

5.1 Introduction

While researchers have already shown the importance of vocabulary knowledge for second language (L2) reading and listening (e.g., Laufer, 1992; Noreillie et al., 2018; Stæhr, 2009), little is known about the relationship between vocabulary knowledge and speaking. Although some studies on L2 speech have focused on different components of L2 oral production including fluency, grammar, pronunciation (e.g., de Jong et al., 2012; Koizumi & In'nami, 2013), less research has focused on the relationship between learners' vocabulary knowledge and L2 oral production. The aim of this study is to extend previous research into the relationship between vocabulary knowledge and speaking by using a receptive as well as a productive vocabulary test, dialogic speaking tasks, and by focusing on French, a less studied language than English. More specifically, the study targets the relationship between learners' receptive as well as productive vocabulary knowledge and their vocabulary use in speaking, i.e., the number of tokens, types and lemmas used in L2 speech, the relationship of those vocabulary tests with the learners' lexical frequency profile (LFP) and with two measures of lexical diversity (LD).

5.2 Background

In spite of the considerable amount of research into the relationship between vocabulary knowledge and reading and listening, it is only recently that more studies have started to look into the relationship between vocabulary knowledge and speaking (e.g., Miralpeix & Muñoz, 2018; Treffers-Daller, 2013; Uchihara & Clenton, 2018). First, we will discuss studies focusing on the relationship between vocabulary knowledge and reading and listening before moving on to studies on the relationship between vocabulary knowledge and speaking.

5.2.1 Relationship vocabulary knowledge and reading and listening

Studies have shown that learners with a larger vocabulary tend to comprehend written texts better (e.g., Laufer, 1992; Nation, 2006; Schmitt et al., 2011). A study by Laufer (1992) showed high correlations between learners' receptive vocabulary knowledge and their reading comprehension (.50 and .75 depending on the vocabulary test used). In spite of the strong relationship between vocabulary knowledge and reading, it seems difficult to establish a threshold for the exact amount of vocabulary that is needed. Estimates range from 5,000 word families for pleasure reading (Hirsh & Nation, 1992) to 8,000 or 9,000 word families for reading newspapers and novels (Nation, 2006).

Similar results were found in studies focusing on vocabulary and listening comprehension. Stæhr (2009) found a moderate correlation (.70) between vocabulary knowledge and a standardized listening comprehension test. These findings were corroborated in a replication study of Stæhr (Noreillie et al., 2018) that looked into the relationship between vocabulary knowledge and listening comprehension in English (.63) as well as in French (.70) at an intermediate level. Overall, these studies show that there is undeniably a strong link both between vocabulary knowledge and reading as well as between vocabulary knowledge and listening.

5.2.2 Relationship vocabulary knowledge and speaking

Researchers have only recently started to explore the relationship between vocabulary knowledge and speaking (e.g., de Jong et al., 2012; Koizumi & In'nami, 2013; Miralpeix & Muñoz, 2018; Uchihara & Clenton, 2018). The study by de Jong et al. (2012) with intermediate and advanced Dutch-as-a-foreign language learners showed that vocabulary knowledge (size and depth combined) and intonation rating predicted 75% of speaking proficiency and that there was a strong correlation between learners' vocabulary size and their

overall speaking proficiency (r = .79) as rated by a panel of judges. Koizumi and In'nami (2013), who investigated the degree to which size, depth and speed of L2 vocabulary could predict L2 speaking proficiency, found similar results as de Jong et al. (2012). Although the tests, the nationality and the level of the participants (novice and low-intermediate Japanese learners of English) were different compared to de Jong et al.'s study, the results showed in both cases that L2 vocabulary knowledge substantially explains L2 speaking proficiency. Learners at novice and low-intermediate levels with a higher speaking proficiency seem to have a larger vocabulary knowledge in terms of size, depth and speed, which enables them to produce more rapid, accurate and syntactically complex oral performance (p. 910). Koizumi and In'nami (2013) even conclude that "size could be a powerful single predictor of L2 proficiency" (p. 910). The study by Miralpeix and Muñoz (2018) focused on receptive vocabulary knowledge and listening, reading, writing and speaking in English. The results indicated that receptive vocabulary size correlates significantly with general English-as-a-foreign language proficiency on the one hand and with the skills of writing, reading, speaking and listening on the other. Finally, the study by Uchihara and Clenton (2018) investigated the role of receptive vocabulary knowledge in a monologic L2 speaking task. The results confirmed the relationship between receptive vocabulary knowledge and speaking even though the correlations were lower than in the studies mentioned before. Their study confirmed, however, the significant role of high-frequency words in spoken texts: the 2,000 most frequent words in English accounted for around 95% of the learners' texts. Finally, the findings showed that having a larger vocabulary knowledge does not necessarily result in the use of more low-frequency words.

5.2.3 Measuring vocabulary use: Lexical Frequency Profile

One of the aims of learning vocabulary is to be able to use words productively in speaking and writing. However, even though learners might know a word productively, this does not entail that they can also use it properly (Laufer & Nation, 1995). Vocabulary use can be measured in different ways. One way to measure learners' vocabulary use is the Lexical Frequency Profile (LFP; Laufer & Nation, 1995). This measure, which was originally developed for English written production, allows to distinguish between high- and low-frequency words and indicates the percentage of words at different frequency levels in a learners' production. It can be considered a measure that reveals how vocabulary knowledge is reflected in use, since it shows to what extent learners make use of their available vocabulary knowledge. In their validation study of the LFP, Laufer and Nation (1995) examined whether there were significant differences between LFPs of learners with different proficiency levels and whether there was a high correlation between learners' LFP and their score on the active version of the VLT (Nation, 1990). The study showed that LFP is a reliable and valid measure of lexical use in writing as it provides similar results for two writings by the same learner and it discriminates between learners of different proficiency levels. It also correlates well with the active version of the VLT, an independent measure of vocabulary knowledge. Moreover, the study concluded that we can expect learners' vocabulary size as measured by a vocabulary knowledge test to be reflected in the learners' productive use of the language.

Goodfellow, Lamy and Jones (2002) were the first to apply the LFP to the assessment of French learners' writings. They found that the teachers' marks correlated with learners' use of medium-frequency words, which, however, was not completely in line with Laufer and Nation's (1995) findings, where they stated that the use of more low-frequency words in a learners' production could be seen as a sign of a richer vocabulary.

The LFP has not only been shown to be a valid measure of lexical richness in English written production (Laufer & Nation, 1995), but also in French oral production (Ovtcharov et al., 2006). Ovtcharov et al.'s study showed that the LFP could differentiate between different proficiency levels and the frequency levels as measured by the LFP. Further, the study also showed a strong link between the number of rare words, i.e., low-frequency words, used by learners in oral production and their level of linguistic competence. One limitation mentioned, however, is the use of a written corpus for analyzing oral language. Filling that gap, Lindqvist (2010) examined whether the LFP is a valid measure for analyzing oral French. Fourteen advanced and very advanced learners of French, and ten native speakers of French passed a semi-guided interview of about 15 minutes with a native speaker of French on topics such as studies, hobbies and family. First, Lindqvist concluded that all the students' productions consisted of more than 90% of the 1,000 most frequent words (1K) and that a that larger number of 1K words seemed to indicate a lower proficiency level. Second, and contrary to the findings of Ovcharov et al. (2006), the study showed that the LFP was not the most appropriate method for analyzing oral French since the corpus consisted at that time only of written language. However, this issue has meanwhile been addressed (Lextutor, Cobb, s.d.) and the actual corpus on which the LFP is now based is a corpus consisting of both oral and written texts. Finally, a recent study by Uchihara and Clenton (2018) did not find a relationship between EFL learners' receptive vocabulary knowledge and their LFP in one monologic speaking task. Given that the limited research so far has produced contradictory findings, more research seems to be warranted.

5.2.4 Measuring vocabulary use: Lexical Diversity

Although, intuitively, it would seem that the LFP of learners' productions also provides information about the diversity of their production, research into this relationship might be warranted.

However, lately, more attention has been drawn to measuring the LD of learners' written or spoken texts, which can be done through the different measures that will follow.

The term lexical diversity is often used interchangeably with vocabulary richness (Richards & Malvern, 2000), lexical richness (Tidball & Treffers-Daller, 2008) or even lexical variation (Lu, 2012). In this article, we will adhere to the term lexical diversity and adopt the definition of McCarthy and Jarvis (2007, p. 459) who define it as "the range and variety of vocabulary deployed in a text either by a speaker or a writer", and in which a larger range indicates a higher lexical diversity.

Lexical diversity has been operationalized in several ways, e.g., Type-Token Ratio (TTR), D (Malvern et al., 2004), the hypergeometric distribution (HD-D; McCarthy & Jarvis, 2007; Wu, 1993), the measure of textual and lexical diversity (MTLD; McCarthy, 2005; McCarthy and Jarvis, 2010). However, it has been shown that measures such as TTR, Root TTR (Guiraud, 1960) and Corrected TTR (Carroll, 1964), which are mathematical transformations of TTR, tend to be flawed measures to calculate LD because they are dependent on text length (e.g., Malvern et al., 2004; Vermeer, 2000). More recently, researchers sought for more robust approaches to LD assessment, such as D (Malvern et al., 2004), HD- D (McCarthy & Jarvis, 2007; Wu, 1993) and MTLD (McCarthy, 2005; McCarthy & Jarvis, 2010). The validity of these measures has been tested for English by McCarthy and Jarvis (2010) and for French by Treffers-Daller (2013).

D and HD-D are measures that are calculated similarly. D, as measured by the vocd command in CLAN, calculates the relationship between the TTR and token size regarding a third variable, called *D* (Malvern & Richards, 2002). It thus calculates the average of a series of random samplings in a text and looks for the best fit between the empirical curve and theoretical curves calculated through the model.

HD-D, on the other hand, is based on the hypergeometric distribution function and calculates the probability of encountering a token of the lexical types in a text taken from a random sample of 42 tokens of that text (McCarthy & Jarvis, 2010). The probabilities of all lexical types in that text are enumerated and the result is used as a measure of the text's LD. MTLD is a measure that sequentially evaluates a text for its TTR, for which a threshold is set. Whenever that threshold is reached in the language sample, the TTR is reset, reducing in that way the impact of text length on the measure (McCarthy & Jarvis, 2010). Finally, Maas is a log-corrected TTR function, that seems to be quite robust when it comes to text length (McCarthy & Jarvis, 2010).

A few studies have been using these measures. deBoer (2017) for instance evaluated the relationship between receptive vocabulary knowledge and LD in written productions of Asian learners of English. A regression analysis showed that there were only very low correlations between the measures of LD used in the study and students' performance on the vocabulary test. Thus, test scores of a receptive vocabulary test did not seem to be effective for predicting LD in students' writings.

One recent study on LD focused on French. Treffers-Daller (2013) compared HD-D and MTLD with the better-known measures D and Maas. Two groups of L2 learners of French of different proficiency levels had to tell monologic stories to elicit oral production. The results showed that HD-D and D most strongly correlated with the French C-test used as a measure of general language ability, more than MTLD and Maas. Thus, the largest portion of the variance in the C-test is explained by HD-D (62%) and D (58%). The study concluded that "HD-D and D are good indicators of language ability in French" (p. 100) and that they strongly indicate a learners' language ability as scores on the LD measures correlated highly with the number of words learners produced. Even though this study is very informative, it only gives information about the relationship between learners' general

proficiency and the LD of monologic tasks but it does not give any information about the relationship between learners' receptive or productive vocabulary knowledge and the LD of learners' productions, neither monologic nor dialogic tasks.

5.3 Rationale and research questions

Although some studies have started to explore the relationship between vocabulary knowledge and speaking, no study, as far as we know, has used a receptive as well as a productive vocabulary test to measure learners' vocabulary knowledge in relation to vocabulary use in speaking. Further, no study has investigated the relationship between learners' vocabulary knowledge and the lexical frequency profile as well as the lexical diversity of their output in two dialogic speaking tasks. The present study aims to fill that gap.

The main research questions of this study concern the relationship between vocabulary knowledge and learners' vocabulary use in two speaking tasks. It is subdivided into the following research questions:

1. What is the relationship between learners' receptive and productive vocabulary knowledge and the number of types and lemmas produced?
2. What is the relationship between learners' receptive and productive vocabulary knowledge and the lexical frequency profile of their output?
3. What is the relationship between learners' receptive and productive vocabulary knowledge and the lexical diversity of their output?

5.4 Method

5.4.1 Participants

Data were collected with 51 Flemish French-as-a-foreign language learners in two secondary schools in Flanders (Belgium). All the participants were in their last year of secondary education (*M* age = 17) and were halfway their eighth year of French instruction. Depending on their curriculum, learners had three (*N* = 22) or four (*N* = 29) fifty-minute classes of French a week. At the end of that year, they were supposed to have reached at least the B1 level of the CEFR for speaking, as well as for the other skills. They all had Dutch as their mother tongue, except for one participant who had Arabic as her mother tongue and one bilingual learner (French-Dutch). This last learner's data were removed from the dataset. As a result, data of 50 learners were analyzed.

5.4.2 Instruments

For the data collection, a receptive and a productive vocabulary knowledge test were used as well as two dialogic speaking tasks.

Receptive vocabulary knowledge test

In order to measure learners' receptive vocabulary knowledge, we developed an adapted version of the French VocabLab-test (Peters et al., 2019), which is a frequency-based monolingual multiple choice test. The original French VocabLab-test targets words from the 2,000 most frequent, 3,000 most frequent, 4,000 most frequent and 5,000 most frequent words in French. The frequency list which was used was the one developed by Lonsdale and Le Bras (2009). The original test has a written multiple choice format in which test items are presented in isolation. Each item is accompanied by four definitions in the L2, French (one correct answer and three distractors), and one "I don't know"-option to discourage guessing. The first change we made was

the inclusion of a 1K level, as recommended by Webb and Sasao (2013), because the 1,000 most frequent words account for the majority of the words L2 learners encounter. Moreover, as most of the words in conversations belong to the 1,000 word level (Nation, 2013), mastery of this level will have a more important impact on second language learning than mastering the 3,000 or 4,000 word level, as learners will encounter fewer of those less frequent words. Further, a study with learners with a similar profile (Noreillie et al., 2018) showed that it might be more informative to also test the 1K level given the learners' fairly limited vocabulary knowledge in French. Therefore, unlike the original VocabLab-test, which has 30 test items from the 2,000 most frequent words in French, the test that was used in this study had a 1K-test section containing 30 items and a 2K-test section with 30 items. The 30 items for the 3K and 4K level are the same as those used in the VocabLab-test (Peters et al., 2019). The 30 items for the 5K level were not administered since the study by Noreillie et al. (2018) had shown a floor effect with participants of a similar proficiency level. The second change concerns the language of the multiple choice options, which were offered in the learners' L1, viz., Dutch and not in French, contrary to the original test. A bilingual test, in which the choices are given in the learners' mother tongue, avoids possible difficulties in French reading skills and focuses more on vocabulary only instead of also focusing on grammar when definitions in the target language are given (Nguyen & Nation, 2011). The effect of comprehension difficulties is even bigger for low proficient learners. Furthermore, according to Elgort (2013), bilingual vocabulary knowledge tests would reduce learners' test anxiety. The test had good internal consistency (Cronbach's alpha = .91, $N = 50$).

Example
distance (f)

- *afstand* (distance)
- *begin* (beginning)
- *glimlach* (smile)
- *verplaatsing* (movement)
- *ik weet het niet* (I don't know)

Productive vocabulary knowledge test

Learners' productive vocabulary knowledge was measured using a bilingual frequency-based vocabulary test developed according to the format of the English Productive Levels Test (Laufer & Nation, 1999). This test provides an estimate of learners' productive vocabulary knowledge at two frequency levels, viz., 0-1,000 (1K), and 1,001-2,000 (2K). Pilot studies had shown that learners knew only few words from the 3K or the 4K frequency level productively. Therefore, only the 1K and 2K frequency levels were tested. For each frequency level, 30 items were sampled from a lemmatized frequency list, *A Frequency Dictionary for French: core vocabulary for learners* (Lonsdale & Le Bras, 2009), the same frequency list that was used in the receptive vocabulary test. The test has a written fill-in-the-gaps format in which a French sentence is given. The Dutch translation of the word is available in brackets and the first letter as well as the exact number of letters of the word are also shown in order to minimize guessing, as pilot studies had shown that this format worked better than only providing the first letter(s) or only providing the translation. The test had good internal consistency (Cronbach's Alpha = .90, N = 50).

Example

On pouvait lui c _ _ _ _ _ _ (toevertrouwen) n'importe quel secret : il est très discret.
(We could confide him any secret: he is very discrete.)

Speaking tasks

The topics of the two speaking tasks are 'a doctor's visit' and 'a job interview'. These topics were chosen because they are described as such in the CEFR at the B1 level and they are topics that typically appear in standardized B1 examinations, such as the DELF (CIEP). Learners should be able to interact about those two topics at the B1 CEFR-level, which is the level the participants were supposed to have reached at the time they took the tests, i.e., their attainment targets.

> Can provide concrete information required in an interview/consultation (e.g., describe symptoms to a doctor) but does so with limited precision. Can carry out a prepared interview, checking and confirming information, though he/she may occasionally have to ask for repetition if the other person's response is rapid or extended. (Council of Europe, 2001, p. 82)

Two topics were chosen because the topic might affect the results (Pauwels, 2016).

Speaking task 1: a doctor's visit

The learners received the instructions in French (Task 1). They were told that they were not feeling well and that they had to describe their symptoms to the doctor, i.e., the researcher. Images were added (see Appendix 7) in order to increase the comparability between learners' output and to guide the discussion since a pilot study had shown that learners found it useful and easier to talk about the topic when pictures were provided. Learners did not have to use all of the images.

Task 1:

Vous êtes allé(e) à la montagne et en rentrant vous vous rendez compte que vous êtes malade. Vous vous rendez chez le médecin et vous lui décrivez vos symptômes.

Vous pouvez utiliser les images ci-dessous pour vous aider.

You went on a trip to the mountains and now you are not feeling well. You go to the doctor and describe your symptoms.

You can use the images below to help you.

Speaking task 2: a job interview.

The second dialogic task is also one of the topics presented in the DELF at the B1 CEFR-level (CIEP). This speaking task differs from the first speaking task (a doctor's visit) in that there were no images given to the learners since this task was more abstract and it thus seemed difficult to find appropriate images that would provide an added value for performing the task. Learners were told in French that they had to present themselves at an employment agency to find a student job and that they could choose between two different student jobs: one at the bakery or one in the supermarket. Two topics were given in order to allow students to choose but it was limited to two topics because of the comparability between learners. Learners also received the instructions on paper (Task 2).

Task 2

Vous vous présentez à une agence pour l'emploi pour obtenir un job d'étudiant.

Il y a deux propositions :

- *un job d'étudiant dans une boulangerie*
- *un job d'étudiant dans un supermarché*

L'employé vous demande de vous présenter (études, expérience, motivation) et vous essayez de le convaincre.

You present yourself at an employment agency to find a student job. You can choose between a job at a bakery or in a supermarket.

You have to introduce yourself (studies, experience, motivation) and you try to convince them that you are the person they need.

Questionnaire

Before starting the receptive vocabulary test online, learners were administered a questionnaire tapping into their familiarity with both topics that they would have to talk about. They were asked whether they had already performed a doctor's visit or a job interview during language classes at school, in one of the languages taught at school, i.e., Dutch (their mother tongue), French, English, German, Spanish or Italian in order to investigate whether there would be a task effect. They were also asked questions about their familiarity with those tasks in real life in French. The questionnaire data will only be used to facilitate the interpretation of the study's results.

5.4.3 Procedure

The data collection took place on different days and was conducted by two researchers. The participants first took the productive paper-and-pencil vocabulary knowledge test before

continuing with the online receptive vocabulary knowledge test. The latter was offered online so that the responses could be scored automatically. Learners had fifty minutes to complete both tests. In one school, the vocabulary knowledge tests were followed immediately by the speaking tasks, whereas in the second school, the speaking tasks took place on another day that same week and a week later. The learners all started with the doctor's visit. They first read the instructions and received two minutes of planning time to think about what they wanted to say before starting the speaking task with the researcher. Subsequently, they received the instructions for the second speaking task, i.e., the job interview. As in the first task, they had two minutes to prepare themselves for the speaking task. Both speaking tasks took about fifteen minutes per learner.

5.4.4 Scoring and analyses

The receptive test was scored dichotomously. Ticking the "I don't know"-option as well as one of the three other incorrection options would lead to a score of zero points. A correct answer would lead to one point with a total maximum score of 120 points.

The productive test was also scored dichotomously: one point for every correct answer, zero points for every incorrect answer. Spelling mistakes were not taken into account: several students wrote *resources* instead of *ressources* in the example that follows. There was an agreement of 99.99% in the scores given by the two raters who corrected the productive tests.

Example

Ce pays n'a plus de r_ _ _ _ _ _ _ _ _ (bronnen) naturelles et doit donc importer son énergie.
(This country does not have any natural resources anymore and has to import its energy.)

Given that the data were collected by two researchers, a *t*-test was run to verify whether the interlocutor had an effect on the learners' use of vocabulary (number of tokens, types and lemmas; lexical frequency profile; measures of lexical diversity). The independent *t*-test showed that the interlocutor did not have an effect on learners' vocabulary used. As a result, all the doctor-data on the one hand and all the interview-data on the other hand were analyzed together.

The spoken interactions were transcribed in CHAT (Codes for the Human Analysis of Transcripts) format (MacWhinney, 2000a) and our analyses were conducted using the Computerized Language Analysis (CLAN) package of programs (MacWhinney, 2000b), which contains a large number of automatic analyses. Both were developed for the CHILDES project.

As data were normally distributed, parametric tests were used. Pearson correlations were calculated in order to answer our research questions. Lextutor (Cobb, s.d.) was used for determining the LFP of the learners' productions. Furthermore, the CLAN programs as well as the Gramulator 6.0 (McCarthy, Watanabe, & Lamkin, 2012) were used to calculate the measures of lexical diversity. Previous research (Treffers-Daller, 2013) has shown that D and HD-D are good indicators of language ability in French. Therefore, these measures were selected for our analyses.

5.5 Results

5.5.1 Descriptive results vocabulary knowledge tests

Tables 5.1 and 5.2 display the results of the receptive and the productive vocabulary knowledge tests per K level and show that the scores are consistently higher on the receptive test than on the productive test, as learners are familiar with the words in the 1K level (*M* = 91.83%) and almost (*M* = 87.20) with those in the 2K level on the receptive test. The less frequent the words are, i.e., the higher the K level, the lower the participants' scores. The standard deviation of the productive vocabulary test is higher than for the receptive vocabulary test, which shows that there was more variation in learners' productive vocabulary than in their receptive vocabulary.

Table 5.1

Descriptives per K level for the receptive vocabulary test (N = 50)

K level	***M* (%)**	***SD* (%)**	**95% *CI***
0-1,000	91.87	6.57	[90, 93.73]
1,001-2,000	87.20	8.12	[84.89, 89.51]
2,001-3,000	72.13	11.52	[68.86, 75.41]
3,001-4,000	69.73	12.36	[65.22, 74.25]
Total	80.23	9.30	[77.59, 82.88]

Note. M = Mean; *SD* = Standard Deviation; *CI* = Confidence Interval.

Table 5.2

Descriptives per K level for the productive vocabulary test (N = 50)

K level	*M* (%)	*SD* (%)	95% *CI*
0-1,000	72.07	12.36	[68.55, 5.58]
1,001-2,000	61.73	16.01	[57.18, 6.28]
Total	66.90	13.48	[77.59, 2.88]

Note. M = Mean; *SD* = Standard Deviation; *CI* = Confidence Interval.

5.5.2 Descriptive results words produced

In order to answer the first research question, i.e., the relationship between learners' receptive and productive vocabulary knowledge and the number of tokens, types and lemmas used, descriptive results of the number of tokens, types and lemmas of the learners' output were computed. The output of only 38 learners was used for the doctor's-topic and of 49 learners for the interview-topic since McCarthy and Jarvis (2007) have shown that it is advisable to keep text length constant if measures of LD were to be computed. Therefore, following McCarthy and Jarvis (2007), productions counting less than 100 tokens or more than 400 tokens were removed from the data. As can be seen from Table 5.3, participants produced more tokens, types and lemmas in the job interview task than in the doctor task. There was considerable variation in the number of tokens learners used in both tasks, as can be seen in the standard deviations, which suggests that the task topic could have an effect on learners' output.

Table 5.3

Descriptive statistics (in percentages) for both topics (N = 38 Doctor; N = 49 Interview)

	M **(*SD*)**	***Min - Max***	***N***
Doctor			
Tokens	160.61 (60.00)	102-369	38
Types	67.76 (18.88)	34-117	38
Lemmas	62.29 (16.31)	35-106	38
Interview			
Tokens	220.33 (67.69)	110-366	49
Types	93.20 (21.59)	51-141	49
Lemmas	84.41 (22.52)	44-137	49

Note. M = Mean; *SD* = Standard Deviation.

5.5.3 Relationship vocabulary tests and types and lemmas

To investigate the relationship between learners' receptive and productive vocabulary knowledge and the number of types and lemmas they used, Pearson correlations were computed. As can be seen in Table 5.4, the analyses show a small correlation (r = .26) (Plonsky, 2015) between the receptive vocabulary test and the number of lemmas in the doctor task on the one hand and in the number of types in the interview task on the other. There is a significant, positive but small

correlation (r = .34) between the receptive vocabulary test and the number of lemmas in the interview task.

Second, there was a positive and significant correlation (r = .32) between the number of types and the productive vocabulary test in the interview task, but not in the doctor's task. Furthermore, there also was a positive and larger correlation (r = .35 for the doctor's task and r = .41 for the interview task) between the productive vocabulary test and the number of used lemmas in both tasks, with a small effect size for the doctor task and a medium effect size for the interview task (Plonsky, 2015). Learners who have a higher score on the productive vocabulary test and thus know more words productively, tend to use more types and lemmas. Hence, these results suggest that the productive vocabulary test might be more useful when investigating oral language production. The relationship between the vocabulary tests and the number of tokens was not further looked into, as the number of types and lemmas seemed to be the adequate measurement for a strongly inflected language such as French, which has been demonstrated by Treffers-Daller (2013) and Treffers-Daller, Parslow, and Williams (2016).

Table 5.4

Pearson correlation vocabulary tests and types and lemmas (N = 38 Doctor; N = 49 Interview)

	Types doctor	**Types interview**	**Lemmas doctor**	**Lemmas interview**
Receptive vocabulary test	.08 $r^2 = 0.01$	.26 $r^2 = 0.07$	.26 $r^2 = 0.07$	.34* $r^2 = 0.12$
Productive vocabulary test	.31 $r^2 = 0.10$	.32* $r^2 = 0.10$	.35* $r^2 = 0.12$	.41** $r^2 = 0.17$

** correlation is significant at the 0.01 level

* correlation is significant at the 0.05 level

5.5.4 Lexical frequency profile of the words produced

In response to our second research question, i.e., the relationship between learners' receptive and productive vocabulary knowledge and the LFP of their output, and in order to obtain a more fine-grained picture of learners' vocabulary use, we analyzed the LFP of the learners' output (Lextutor, Cobb, n.d.) (see Table 5.5). The findings suggest that learners draw the large majority of their lexical use from the 1,000 most frequent words in French. They also seem to use slightly more 2K words for the doctor's visit but more low-frequency words for the job interview. The cut-off for low-frequency words was put beyond the 2,000 most frequent words in French because the learners only used a limited number of words that belong to the higher K levels. There was also more variation between learners in the doctor task.

Table 5. 5

Mean number of tokens, types and lemmas in different K levels (%) (N = 38 Doctor; N = 49 Interview)

Topic	***M* 1K (*SD*)**	***M* 2K (*SD*)**	***M* 1+2K (*SD*)**	***M* beyond 2K (*SD*)**
Doctor				
Tokens	90.16 (2.86)	4.02 (1.32)	94.18 (2.47)	5.52 (2.36)
Types	79.68 (5.17)	8.04 (2.24)	87.72 (4.43)	10.60 (4.12)
Lemmas	84.37 (4.03)	8.49 (2.81)	92.86 (3.16)	7.11 (3.12)
Interview				
Tokens	89.48 (2.41)	2.70 (0.74)	92.17 (2.21)	7.69 (2.10)

Topic	***M* 1K (*SD*)**	***M* 2K (*SD*)**	***M* 1+2K (*SD*)**	***M* beyond 2K (*SD*)**
Types	77.55 (4.42)	6.15 (1.59)	83.90 (4.11)	14.96 (3.78)
Lemmas	82.08 (3.45)	6.82 (1.84)	88.83 (2.92)	11.10 (2.91)

Table 5.6 shows the correlations between both vocabulary tests and the number of tokens, types and lemmas used when we look at the frequency of those tokens, types and lemmas. We expect a negative correlation between the vocabulary tests and the number of high frequent (i.e., 1K and 2K) tokens, types and lemmas, and a positive correlation between the vocabulary tests and low frequent (i.e., beyond 2K) tokens, types and lemmas. We only found significant correlations between the receptive vocabulary test and the LFP of the tokens in the doctor's task. No other significant correlations were found, which indicates that neither the receptive nor the productive vocabulary test might be good predictors of the LFP of learners' productions in dialogic speaking tasks.

Table 5.6

Pearson correlation vocabulary tests and frequency per number of tokens, types and lemmas (N = 38 Doctor; N = 49 Interview)

Topic	Tokens 1K+2K	Tokens beyond 2K	Types 1K+2K	Types beyond 2K	Lemmas 1K+2K	Lemmas beyond 2K
Doctor						
Receptive vocabulary test	-.36* $r^2 = 0.13$	.37* $r^2 = 0.14$	-.06 $r^2 = 0.001$	.09 $r^2 = 0.008$	-.15 $r^2 = 0.02$	.14 $r^2 = 0.02$
Productive vocabulary test	-.07 $r^2 = 0.005$	.16 $r^2 = 0.03$	.15 $r^2 = 0.02$	-.07 $r^2 = 0.005$	-.09 $r^2 = 0.008$	.08 $r^2 = 0.006$
Interview						
Receptive vocabulary test	-.20 $r^2 = 0.04$	.17 $r^2 = 0.03$	-.21 $r^2 = 0.04$	.23 $r^2 = 0.05$	-.11 $r^2 = 0.01$	.12 $r^2 = 0.01$
Productive vocabulary test	.02 $r^2 = 0.000$	-.05 $r^2 = 0.003$	.04 $r^2 = 0.002$	-.07 $r^2 = 0.005$	-.07 $r^2 = 0.005$	.08 $r^2 = 0.006$

* correlation is significant at the 0.05 level

5.5.5 Vocabulary knowledge and lexical diversity (LD) in two speaking tasks

In answer to our third research question, i.e., the relationship between learners' receptive and productive vocabulary knowledge and the LD of their output, two measures were used to determine the LD in the learners' output: vocd-D and HD-D. As can be seen in Table 5.7, the results suggest that the interview interactions were more lexically diverse than the doctor's interactions, with a higher number for D indicating a higher lexical diversity. For HD-D, the more the number comes close to zero, the more lexically diverse the output is.

Table 5.7

Means of different measures of lexical diversity (N = 38 Doctor; N = 49 Interview)

Topic	***M* vocd-D (*SD*)**	***M* HD-D (*SD*)**
Doctor	27.21 (9.33)	-5.85 (2.49)
Interview	33.00 (8.45)	-3.98 (1.88)

In order to investigate the relationship between the vocabulary tests and these two measures of LD, Pearson correlations were computed. Since French is a highly inflected language, data had to be lemmatized before calculating LD measures.

The analyses (Table 5.8) showed that there was only a significant and positive correlation between the receptive vocabulary test and HD-D for the doctor's task. However, there were no significant correlations between the receptive test and vocd-D, in either task, nor between the receptive test and HD-D for the interview task. The productive vocabulary test, on the other hand, was positively associated with

vocd-D and HD-D in the doctor's task, with a medium effect size (Cohen, 1992). The results in Table 5.8 show that vocd-D and HD-D explain some variance (between 8 and 17 percent) for the doctor's topic. However, since we only found significant correlations in one speaking task, the results are inconclusive.

Table 5.8

Pearson correlation between different measures of lexical diversity and vocabulary knowledge (N = 38 Doctor; N = 49 Interview)

	vocd-D doctor	**HD-D doctor**	**vocd-D interview**	**HD-D interview**
Receptive vocabulary test	0.28 $r^2 = 0.08$	0.36* $r^2 = 0.13$	0.14 $r^2 = 0.02$	0.08 $r^2 = 0.01$
Productive vocabulary test	0.37* $r^2 = 0.14$	0.41** $r^2 = 0.17$	0.26 $r^2 = 0.07$	0.18 $r^2 = 0.03$

**significant at the 0.01 level

*significant at the 0.05 level

5.6 Discussion

The current study explored the relationship between learners' receptive and productive vocabulary knowledge and their vocabulary use in French dialogic tasks. In a sample of fifty learners of French, receptive and productive vocabulary knowledge was assessed as well as their vocabulary use in two speaking tasks. The lexical frequency profile of learners' output was measured using Lextutor (Cobb, n.d.) and the measures of lexical diversity used were D (Malvern et al., 2004) and HD-D (McCarthy & Jarvis, 2007, 2010).

5.6.1 Learners' vocabulary knowledge and spoken output

The first focus of the study was to investigate the relationship between learners' receptive and productive vocabulary knowledge and the number of types and lemmas produced in two dialogic speaking tasks. The results indicate that learners use more tokens, types and lemmas for the interview than for the doctor's visit and that the more words learners know productively, i.e., the higher their score on the productive vocabulary test, the more types and lemmas they seem to use. The findings of the productive vocabulary test are in line with the results of the receptive vocabulary test even though the correlations are stronger. This is not surprising as a productive vocabulary test should be a better predictor of productive vocabulary use because it requires that learners access their L2 productive lexicon, which they also need for speaking (Milton, 2009, p. 121). This also confirms Laufer and Nation's (1995) conclusion that we can "expect learners' vocabulary size as measured by a vocabulary test to be reflected in the learners' productive use of the language" (p. 319). Further, these results are in line with previous research, where knowledge of productive vocabulary was found to be one of the best predictors of speaking proficiency (de Jong et al., 2012) and where productive vocabulary knowledge seemed to have a substantial influence on vocabulary performance in speaking (Koizumi, 2005; Koizumi & In'nami, 2013). In sum, learners with higher scores on the productive vocabulary test seem to use more types and lemmas.

It should, however, also be pointed out that the findings showed differences between the two tasks in terms of the number of lexical items (tokens, types and lemmas) learners produced, which might indicate a possible topic effect. We checked in the questionnaire whether learners' familiarity with either topic could explain the differences between the two tasks. However, topic familiarity does not seem to be an explanation for the higher output in the interview task, because learners' answers in a questionnaire in which they were asked

whether they had already encountered the topics in language classes at school showed that 32 out of 49 (65%) learners had already performed a doctor's visit in French at school and 14 out of 49 (almost 29%) learners had already done a job interview in French during language classes. Thus, even though more learners indicated having already performed a doctor's visit in French language class, they still used fewer lexical items. We might therefore conclude that the lexical output, i.e., the number of types and lemmas depends on the topic of the task and that some topics require the use of more words, as concluded by Pauwels (2016).

5.6.2 Relationship between receptive and productive vocabulary knowledge and the lexical frequency profile in learners' output

The second focus of this study was on the relationship between learners' receptive and productive vocabulary knowledge and learners' LFP in two dialogic speaking tasks. First, the analysis of the LFP of the tokens, types and lemmas in learners' output showed that the majority of the tokens, types and lemmas belonged to the 1,000 most frequent words in French for both tasks. Nation (2013, p. 22) already showed the importance of a small group of high-frequency words because they "cover a very large proportion of the running words in spoken and written texts and occur in all kinds of uses of the language". These findings confirm previous research (Lindqvist, 2010; Milton, 2009; Uchihara & Clenton, 2018), which revealed that learners' productions consist of at least 90% of high-frequency words. In English, the 2,000 word level has usually been set as the most suitable limit for high-frequency words (Nation, 2013, p. 23). Although the interview task was characterized by the use of a higher number of less frequent words compared to the doctor's task, i.e., beyond the 2,000 most frequent words in French, the proportion of low frequent words was not high enough to claim that learners have a high proficiency level, as shown also by the results on the vocabulary knowledge tests. A high

proportion of low frequent or rare words in a text is assumed to indicate a high level of lexical proficiency in the language, which has been shown in previous research by Ovtcharov et al. (2006) and Lindqvist (2010) for French and Laufer and Nation (1995) for English. We should, however, take into account Nation's assertion that the number of high-frequency words is usually higher in spoken language than in written language (2006, p. 79). Consequently, the number of less frequent words should be lower in oral language.

As far as the correlations between the vocabulary tests and learners' LPF is concerned, almost no correlations were found except for the receptive vocabulary test and the LFP of one task. This corroborates previous research (Uchihara & Clenton, 2018) in which no significant association was found between a receptive vocabulary test and the LFP of learners' output in one monologic task. Thus, learners with a larger vocabulary knowledge do not necessarily produce more low-frequency words in an L2 speaking task. This might be explained by the small role of low-frequency words, i.e., beyond the 2,000 most frequent words, in spoken discourse (Milton, 2009). Consequently, no claims can be made about the predictability of learners' results on a receptive or productive vocabulary test for learners' LFP.

5.6.3 Relationship receptive and productive vocabulary knowledge and lexical diversity in learners' output

A third aim was to examine the LD of the learners' output and to determine whether there is a correlation between the receptive and productive vocabulary test scores and LD in two speaking tasks. The results showed that learners' LD, as measured by D and HD-D, was higher for the interview task than for the doctor's task. This result might be explained by the fact that learners used more domain specific vocabulary (e.g., *postuler* (apply for), *épices* (spices), *balayer* (sweep), *agence d'assurance* (insurance agency), *sociable* (social)) regarding their professional experience or personal characteristics used to present

themselves for a future student job. In spite of the higher LD in the interview task, we did not find significant correlations between the vocabulary tests and the LD in the interview task. However, we only found a small significant correlation between the receptive vocabulary test and HD-D in the doctor's task. This finding corroborates deBoer's (2017) conclusions, who found only very low correlations (ranging from -.030 to .220) between the performance of Chinese, Japanese and Korean learners of English on the Vocabulary Size Test (VST; Beglar & Nation, 2007) with LD measures for writing. He states that receptive vocabulary tests might not be the best way to decide on the LD in learners' writing and by extension upon learners' overall communicative competence. Moreover, even though we used both a receptive and a productive vocabulary test, our results are inconclusive because we only found significant correlations in one task.

To conclude, we found some evidence for the relationship between receptive and productive vocabulary knowledge and LD in spoken output, albeit in one speaking task only. It seems that learners' vocabulary knowledge might not be such a good predictor for LD in speaking and other factors might play a more important role. Future research could investigate which factors are better predictors of lexically diverse speech.

5.7 Limitations and suggestions for future research

We should take account of the limitations of our study. One limitation is that we only tested participants from one grade, varying little in their proficiency level. It would be interesting to investigate whether the same results could be replicated with learners having different proficiency levels, especially with respect to the relationship between vocabulary knowledge and spoken output. This brings us to another limitation, namely that we only used two dialogic tasks. It needs to be determined whether the results can be generalized to other task types, i.e., monologic tasks or cognitively more demanding tasks,

as it has been shown that L2 learners' oral performance can differ according to various task conditions (e.g., Derwing, Rossiter, Munro, & Thomson, 2004). Another suggestion for future research might be to expand research in speaking with dialogic tasks to other languages in order to verify whether these results can be replicated.

5.8 Conclusion

This study aimed to investigate the relationship between receptive and productive vocabulary knowledge and vocabulary use in two dialogic tasks by intermediate learners of French. It is, as far as we know, the only study that focused on both the LFP as well as the LD of learners' output in dialogic tasks and that used two vocabulary knowledge tests. Our study shows that it might be better to use a productive vocabulary test than a receptive vocabulary test when investigating the relationship between vocabulary knowledge and vocabulary use in speaking. Further, the findings suggest that learners who know more words productively tend to use more words when they speak. However, the relationship between learners' productive vocabulary knowledge and the LFP and LD of their spoken output is less straightforward. It thus seems that (controlled) productive vocabulary knowledge might not be such a good predictor of learners' vocabulary use in speech. Finally, the study also indicates that it is important to use at least two tasks because of a possible topic effect, as learners systematically produced more tokens, types and lemmas for the interview topic than for the doctor's topic and they produced more lexically diverse speech in the interview task.

Chapter 6

Low-intermediate French-as-a-foreign language learners' rated vocabulary use in semi-structured dialogic speaking tasks

6.1 Introduction

While previous research has investigated the relationship between second language (L2) vocabulary knowledge and reading, listening and writing (e.g., Crossley, Salsbury, Mcnamara, & Jarvis, 2011b; Laufer & Ravenhorst-Kalovski, 2010; Stæhr, 2009), less is known about the relationship between vocabulary knowledge and speaking (see de Jong et al., 2012 for an exception). Although the number of studies on L2 speech is growing, most studies have focused on English L2 learners performing monologic tasks (e.g., de Jong et al., 2012; Uchihara & Clenton, 2018) less research has investigated vocabulary use in dialogic tasks and in French at a low-intermediate level. The aim of this study is thus to extend previous research into speaking by focusing on dialogic tasks with both L1 and low-intermediate L2 speakers of French. We explored the lexical factors that could predict experts' holistic ratings of lexical proficiency for a less studied language and the correlations between L2 learners' scores on a receptive and a productive vocabulary test and experts' ratings of learners' oral output on two semi-structured dialogic speaking tasks.

6.2 Background

First, we will focus on previous research that has investigated experts' ratings of speaking tasks with regard to the number of words used by speakers, the frequency of these words and their lexical diversity. Next, we will briefly report on studies that have explicitly compared L1 and L2 speakers' oral output. Finally, studies that have looked at L2 learners' vocabulary knowledge in relation with experts' ratings of speaking output will be mentioned.

6.2.1 Relationship between number of words and experts' ratings of speaking tasks

Studies investigating the relationship between written texts and vocabulary use by EFL learners (e.g., Crossley & Mcnamara, 2012; Crossley et al., 2011b; Jarvis, 2002) have shown that there is a positive relationship between the number of words learners use in a text and the ratings they received. However, as far as the relationship between the number of words in learners' output and human judgements of that output in speaking is concerned, different studies have shown similar results (Crossley & McNamara, 2013; Iwashita et al., 2008; Lorenzo-Dus & Meara, 2005). A study by Iwashita et al. (2008) investigated the relationship between EFL learners' spoken English in two monologic tasks (= an independent and an integrated task), and holistic scores awarded by raters. In the independent task, learners had to give their opinion on a certain topic and in the integrative task, learners first had to listen to or to read information which they had to explain, describe or recount afterwards. The results indicated that learners of distinct proficiency levels differ in the number of words (tokens) as well as in the range of words (types) they use, i.e., the more tokens or types they use, the higher their level is and the better experts' ratings. Similar results were reported in a study by Lorenzo-Dus and Meara (2005) that looked at the relationship between learners' vocabulary use and vocabulary grades awarded by raters. Twenty-nine British secondary school students took Spanish oral proficiency tests that consisted of two parts: a general conversation with the examiner/rater about their personal interests, studies and plans for the future and a conversation about a prepared topic. The number of types showed to be a distinctive factor for the rating, i.e., the higher the number of types used, the higher the grade awarded. However, it should be noted that the rating scale did not only assess vocabulary, but also pronunciation, intonation, accuracy, and knowledge of society and culture. Consequently, the aforementioned relationship between grade and vocabulary use might have been influenced by one of the other factors too.

Crossley and McNamara (2013) investigated the potential of the use of automated indices for predicting human judgments of speaking proficiency with a corpus of TOEFL speaking data. The dataset consisted of spoken data in a monologic speaking task in which speakers had to talk about a familiar topic and their opinions and experiences. The number of types, i.e., the number of unique words used by speakers as well as their frequency showed to be predictive of the ratings given by experts and explained 61% of the variance in human ratings of proficiency. Thus, speakers producing a larger number of unique words and using more low-frequency words are judged to have greater speaking proficiency. The importance of the words' frequency in speaking will be discussed in more detail in the next section.

6.2.2 Relationship between lexical frequency profile and experts' ratings of speaking tasks

Apart from counting the number of (unique) words in a text, learners' vocabulary use can also be measured by investigating the frequency of the words that are used. This can for instance be done with the lexical frequency profile (LFP; Laufer & Nation, 1995), which measures the percentage of high- and low-frequency words in the output. This tool, originally developed for English written production, reflects to what extent learners' vocabulary knowledge is transferred to vocabulary use. However, research so far has produced conflicting evidence for the relationship between word frequency and holistic ratings.

A study by Kyle and Crossley (2015) explored whether and to what extent automatic indices of frequency could account for the variance in holistic judgments of speaking proficiency. Apart from the frequency indices, other measures were also used. They are all available in the Tool for the automatic analysis of lexical sophistication (TAALES), a tool that automatically measures over 400 indices of

lexical sophistication. Two monologic speaking tasks (e.g., give your opinion on a topic) from the TOEFL public use data set were used in this study. They were taken together because of the short length of the output. The results showed that there was a positive relation between well rated samples and the use of more academic words and more low-frequency words. Uchihara and Clenton (2018) on the contrary, did not find a significant correlation between the number of high or low-frequency words used by EFL learners and experts' ratings in their study with EFL learners performing a picture description task. Both the LFP and TAALES were used to analyze the frequency of the words learners produced. They concluded that learners who were lexically proficient according to experts' ratings "did not necessarily produce lexically sophisticated L2 words in spontaneous speech" (p. 10). The different findings in these two studies might be due to the differences in methodology, viz., the rating scales that were used or the different lexical measures or the different topics of the speaking tasks.

6.2.3 Relationship between lexical diversity and experts' ratings of speaking tasks

In addition to frequency, learners' production has also been analyzed in terms of lexical diversity. The term lexical diversity (LD) is often used interchangeably with lexical richness (Tidball & Treffers-Daller, 2008), vocabulary richness (Richards & Malvern, 2000) or lexical variation (Lu, 2012). In the present study, lexical diversity will be defined according to McCarthy and Jarvis' (2007, p. 459) definition: "lexical diversity can be described as the range and variety of vocabulary deployed in a text by either a speaker or a writer". It is in contrast with the potential vocabulary a learner might have available, but is not using. There are different ways to operationalize lexical diversity, e.g., Type-Token Ratio (TTR) or its transformations (e.g., Root TTR (Guiraud, 1960), Corrected TTR (Carroll, 1964), Advanced TTR (Daller, Van Hout, & Treffers-Daller, 2003), D (Malvern et al., 2004), the hypergeometric distribution HD-D (McCarthy & Jarvis, 2007; Wu,

1993) and the measure of textual and lexical diversity MTLD (McCarthy, 2005; McCarthy & Jarvis, 2010). Research has shown, however, that the TTR measure and its transformations are not appropriate for calculating LD because they are dependent on text length. Indeed, as the number of tokens in a text steadily increases, the relative number of types decreases, since words, i.e., tokens, tend to be repeated after a relative short amount of text. Therefore, researchers have looked for more robust approaches to testing LD, such as HD-D (Wu, 1993; McCarthy & Jarvis, 2007), D (Malvern et al., 2004) and MTLD (McCarthy, 2005; McCarthy & Jarvis, 2010). They have also been validated for English by McCarthy and Jarvis (2010) and for French by Treffers-Daller (2013). The assumption behind LD indices is that more diverse texts are indicative of larger and more proficient lexicons and thus of a higher proficiency level.

The studies that have investigated the relationship between LD and experts' ratings of learners' oral output (e.g., Appel et al., 2019; Crossley & McNamara, 2013; Crossley et al., 2015; Crossley et al., 2011a; Iwashita, 2010; Lorenzo-Dus & Meara, 2005; Lu, 2012; Malvern & Richards, 2002; Yu, 2010) display conflicting evidence, which might be due to the methodological differences. In a number of studies with EFL university students, Crossley and his colleagues (2011a, 2013, 2015) found that LD, measured as D or MTLD, explains a significant percentage of the variance in human ratings of speaking, which means that speakers who produce speech that is more lexically diverse are better rated. Similarly, a study by Lu (2012), in which English L2 oral narratives with university students were computationally analyzed for 25 lexical measures, showed that LD as measured by D correlated most strongly with raters' judgment of L2 speaking proficiency. A study by Appel et al. (2019), in which non-expert L1 English raters judged the comprehensibility and nativeness of L2 English speech samples, showed similar results. In their study, LD (MTLD) explained 5% of the variance in the raters' assessment. Furthermore, a study by Yu (2010)

aiming to understand the differences in LD between performances on EFL writing and speaking showed that LD as measured by D explained around 23% of the variance in the overall quality ratings of speaking but care should be taken given the limited number of participants (N = 25). Similarly, the results of a study by Iwashita (2010), which looked at the distinguishing features of oral proficiency, among which LD as measured by D, in performances of EFL and JFL learners, are in line with this finding. She found that LD could distinguish between low and high proficiency groups, both in EFL as in the JFL groups. The more diverse the output, the more proficient the learner was.

In spite of the positive evidence for LD as a predictor of speaking, some studies did not find statistically significant correlations. A study by Malvern and Richards (2002) that analyzed oral proficiency with British learners of French did not show statistically significant correlations between lexical diversity (D) and rated range of vocabulary. However, according to the authors, the absence of a correlation might be due to difficulty for teachers to assess LD independently of other factors such as fluency, structure complexity and pronunciation. Finally, in their study with British secondary school students learning Spanish, Lorenzo-Dus and Meara (2005) tested whether learners' vocabulary output and grade ratings were related. They found that the LD (D) of learners' output did not discriminate between the different groups that were created after the experts' ratings were attributed. These last two studies indicate that there might be a discrepancy between EFL and French and Spanish as foreign languages since the results do not indicate similar findings.

In sum, there is some evidence that more lexically diverse output is also better rated. However, it is clear that most studies have focused on English university students and monologic tasks. Less is known about other languages and learners' LD in dialogic tasks. Therefore, more research seems warranted.

6.2.4 Differences in oral output between L1 and L2 speakers

In a study with university students studying French and a control group of native speakers, Treffers-Daller (2013) showed, by focusing on a monologic task, that L1 speakers produced a significantly higher number of tokens and types compared to the lower level students. The groups also differed significantly from each other. Similar results were obtained by Vermeer (2000) in a study with four to seven year old children speaking Dutch as a second language and a control group of native Dutch speaking children, who had to tell a story from a picture book and were interviewed about topics such as friends, television and holidays. Vermeer concludes, however, that a more valid measure of lexical richness might be the LFP and that frequency might thus add important information that is not given by only looking at the number of types and tokens.

Lindqvist (2010) explored whether the LFP is a valid measure for analyzing oral French with advanced learners of French and a control group of native speakers. The results of a semi-guided interview indicated that the higher the speakers' proficiency level, the more words they used and the higher the number of low-frequency words in their speech. In addition, even though a larger number of 1K words was related to a lower proficiency level, more than 90% of the students' speech consisted of 1K words. Lindqvist indicated that the low percentage of low-frequency words might be due to the topic speakers interacted about, viz., family, hobbies and studies and that these topics ask for less low-frequency words. However, it might also be due to the fact that the number of high-frequency words is usually higher in spoken language than in written language (Nation, 2006, p. 79). It should also be noted that the number of participants in the study is quite small and that care should be taken when generalizing these results. Finally, it was suggested that the LFP is not entirely trustworthy as a tool for oral French. However, the corpus the LFP is based upon has been updated since.

6.2.5 Relationship between vocabulary knowledge and experts' ratings of speaking tasks

Previous research has shown that productive vocabulary knowledge (sentence completion, L1-L2 translation) is one of the best indicators of learners' speaking proficiency (e.g., de Jong et al., 2012; Koizumi & In'nami, 2013) and that, according to expert raters, vocabulary also constitutes one of the salient features for assessing speaking (e.g., Brown, Iwashita, & McNamara, 2005).

Uchihara and Clenton (2018) investigated the relationship between receptive vocabulary knowledge as measured by a receptive Yes/No test and experts' ratings of EFL speakers' output (= a monologic picture description task). Learners' output was rated on its vocabulary component by means of the IELTS speaking descriptors. The results showed that the receptive vocabulary test correlated (r = .55) positively with the vocabulary ratings. A study by de Jong et al. (2012), in which 181 L2 learners of Dutch and 54 native speakers of Dutch took 8 monologic speaking tasks and 9 tests of linguistic skills, yielded similar results. The study looked at the degree to which L2 knowledge skills and L2 processing skills explain L2 speaking proficiency. The findings indicated that there was a strong correlation between vocabulary knowledge (as measured by productive fill-in-the-gaps tests) and overall speaking proficiency (r = .79).

6.3 Research questions

Although a growing number of studies has explored the relationship between vocabulary use and speaking, more research is needed in other foreign languages, with non-university students and with tasks other than monologic tasks, which have been used most often. The present study aims to fill these gaps by focusing on low-intermediate learners of French, on two semi-structured dialogic tasks, and by comparing L1 speakers' output to that of L2 speakers.

We addressed the following research questions:

1. Which lexical factors predict experts' holistic rating scores of L1 and L2 speakers' oral output in two semi-structured dialogic speaking tasks?
2. How does L1 and low-intermediate L2 speakers' output in two semi-structured dialogic tasks differ as far as the number of (1) tokens, types and lemmas[17], (2) the lexical frequency profile and (3) lexical diversity are concerned?
3. What is the relationship between low-intermediate L2 learners' receptive and productive vocabulary knowledge and raters' judgement of holistic ratings of L2 learners' spoken output?

6.4 Method

6.4.1 Participants

Data in this study were collected from French L1 and L2 speakers. The latter were 51 Flemish low-intermediate learners of French-as-a-foreign language in two secondary schools in Flanders (*N* = 13 for school 1; *N* = 38 for school 2). All the participants were supposed to have reached the B1 level of the CEFR for speaking as well as for the other skills at the end of the year, i.e., after eight years of French instruction. They were in their last year of secondary education (*M* age = 17) and had three (*N* = 22) or four (*N* = 29) fifty-minute classes of French a week depending on their curriculum. Except for two learners, all learners had Dutch as their L1. One learner had Arabic as her mother tongue and there was one Dutch-French bilingual learner. The data of the bilingual learner were removed from the dataset. As a result, data of 50 learners were analyzed. Second, data of 27 native speakers of French were collected in a secondary school in Wallonia, the French-

[17] The number of tokens represent all the words in a text, whereas the number of types indicates the different words in a text. Lemmas are considered as a base words and their inflections.

speaking part of Belgium. Like the non-native learners, these participants were also in their last year of secondary education. Since the L2 as well as the L1 speakers produced two speaking tasks, our spoken corpus contained in total 100 non-native speech samples and 54 native speech samples (N = 154).

6.4.2 Instruments

The data collection instruments consisted of a receptive and productive vocabulary test and two semi-structured dialogic speaking tasks. The vocabulary tests were only administered to the L2 speakers.

Receptive vocabulary knowledge test

In order to measure L2 learners' receptive vocabulary knowledge, a French frequency-based bilingual multiple choice test was created and administered. Since the results on the monolingual receptive VocabLab test (Peters et al., 2019) used in a previous study (Noreillie et al., 2018) had shown that the scores were very low on the 5K band, a vocabulary test only assessing the 4,000 most frequent words in French was created. We used Lonsdale and Le Bras' (2009) frequency list to develop the test. The test targets words from the 1,000 (1K), 2,000 (2K), 3,000 (3K) and 4,000 (4K) most frequent words in French. The 120 items of the test, consisting of 30 items per 1,000 word frequency band, are presented in isolation and are offered in a written multiple choice format. Learners receive four translations in Dutch (i.e., the learners' L1), among which the correct answer, three distractors and an 'I don't know'-option. The test had good internal consistency (Cronbach's alpha = .91, N = 50).

Example
lutte (f)

- *gevecht (battle)*
- *verdriet (grief)*
- *moord (murder)*
- *uitdaging (challenge)*
- *ik weet het niet (I don't know)*

Productive vocabulary knowledge test

Learners' productive vocabulary knowledge was measured by means of a bilingual frequency-based vocabulary test developed according to the format of the English Productive Levels Test (Laufer & Nation, 1999). This test provides an estimate of learners' productive vocabulary knowledge at two frequency levels, viz., 0-1,000 (1K) and 1,001-2,000 (2K) for which 30 items were sampled from the frequency list of Lonsdale and Le Bras (2009), as in the receptive vocabulary test. We only focused on the 2,000 most frequent words because a pilot study had shown that learners obtained only very low scores on the 3K and 4K frequency word bands. Given that learners tend to know fewer words productively than receptively (Laufer, 1998) and to avoid test fatigue because of the lack of knowledge of too many words, we only tested the 2,000 most frequent words productively. The test has a written fill-in-the-gaps format in which a French sentence is given. The Dutch translation of the word is available in brackets and the first or the first two letters are given as well as the number of letters the word contains in order to minimize guessing. The test had good internal consistency (Cronbach's Alpha = .90, N = 50).

Example

Le champion olympique, hospitalisé depuis trois mois et demi après un grave accident de ski, fait de petits p_ _ _ _ _ _ (vooruitgang) encourageants.

(The olympic champion, hospitalized since three and a half months after a terrible skiing accident, is making encouraging progress.)

Speaking tasks

The speaking tasks that were administered were semi-structured dialogic speaking tasks, in which a learner had to interact with the researcher.

The topics of the two tasks, i.e., 'a doctor's visit' and 'a job interview', were chosen because they are described as such in the CEFR at the B1 level, which is the proficiency level the L2 learners in the present study were supposed to have reached, and they are topics that typically appear in standardized B1 examinations, such as the DELF (CIEP).

> Can provide concrete information required in an interview/consultation (e.g., describe symptoms to a doctor) but does so with limited precision. Can carry out a prepared interview, checking and confirming information, though he/she may occasionally have to ask for repetition if the other person's response is rapid or extended. (Council of Europe, 2001, p. 82)

The L2 French learners in the present study were supposed to be able to interact about those two topics at the time they did the tasks. L1 speakers performed the same semi-structured dialogic speaking tasks which would allow us to compare L1 and L2 speakers' output. Finally, two topics were chosen because genre and topic might have an effect on the output (e.g., Bygate, 1999; Pauwels, 2016).

Speaking task 1: a doctor's visit. Both the L1 and the L2 speakers received the instructions in French (Task 1) and were told that they did not feel well. They had to describe their symptoms to the doctor, i.e., the researcher. Images were added in order to increase the comparability between the speakers' output and to guide the discussion (see Appendix 7), since a pilot study had shown that L2 learners found it useful and easier to talk about the topic when pictures were provided. The images were there to help, but the speakers did not have to use all of them.

> Task 1:
>
> *Vous êtes allé(e) à la montagne et en rentrant vous vous rendez compte que vous êtes malade. Vous vous rendez chez le médecin et vous lui décrivez vos symptômes.*
>
> *Vous pouvez utiliser les images ci-dessous pour vous aider.*
>
> You went on a trip to the mountains and now you are not feeling well. You go to the doctor and describe your symptoms.
>
> You can use the images below to help you.

Speaking task 2: a job interview. The second semi-structured dialogic task is also one of the topics presented in the DELF at the B1 CEFR-level (CIEP). The difference with the previous speaking task (a doctor's visit) is that there were no images added since this task was more abstract. Both L1 and L2 speakers were told in French that they had to present themselves at an employment agency to find a student job and that they had two options: a student job at the bakery or in the supermarket. The different options were given in order to provide participants with a choice, which, however, was limited to two topics that were supposed to be close to the participants' daily life. The instructions were also given on paper (Task 2).

Task 2

Vous vous présentez à une agence pour l'emploi pour obtenir un job d'étudiant.

Il y a deux propositions :

- *un job d'étudiant dans une boulangerie*
- *un job d'étudiant dans un supermarché*

L'employé vous demande de vous présenter (études, expérience, motivation) et vous essayez de le convaincre.

You present yourself at an employment agency to find a student job. You can choose between a job at a bakery or in a supermarket.

You have to introduce yourself (studies, experience, motivation) and you try to convince them that you are the person they need.

6.4.3 Procedure

The data collection with the non-native speakers took place on different days and was conducted by two researchers. The data collection with the native speakers also took place on different days but was conducted by only one researcher. The participants all started with the doctor's visit. They first read the instructions and received two minutes of planning time before doing the task. Subsequently, they received the instructions for the second speaking task, i.e., the job interview. As in the first task, they had two minutes to prepare themselves. Both speaking tasks took about fifteen minutes each per participant for the L2 speakers and ten minutes for the L1 speakers. Even though the same questions were asked to both L2 and L1 speakers, L1 speakers answered more quickly and therefore needed less time in total. The procedure was the same for all the participants, regardless of their school or the researcher who administered the speaking task.

6.4.4 Rating instrument and human ratings

Participants' spoken output was measured in terms of holistic features of their performance and was evaluated by two raters from a panel of three. The three raters assessed performances of both topics and of L1 and L2 speakers. The raters, who were native speakers of French and had a background in language studies and/or teaching, did not know in advance whether they would be rating an L1 or an L2 performance. As raters performed the ratings of the audio files on a voluntary basis, they completed them at their own pace, after having participated in a training session, in which they rated audio files from the pilot study in presence of the researcher.

The rating instrument that was used consisted of a holistic rating scale, in which the evaluations were prompted through a 5-point Likert scale. The holistic lexical rating scale was translated to French and adapted from the rating scale used by Crossley et al. (2015). The rating instrument is available in Appendix 13.

6.4.5 Scoring and analyses

Vocabulary tests

The receptive test was scored dichotomously. A score of zero points was attributed when a learner indicated one of the three incorrect options or the "I don't know"-option. A correct answer would lead to one point with a maximal score of 120 points. Learners who obtained a score of 90% on a 1,000 word section, i.e., 27 out of 30 on a frequency band, were supposed to master that level (Nation, n.d. as cited in Stæhr, 2009).

The productive test was also scored dichotomously: one point was given for every correct answer, zero points for every incorrect answer. Spelling mistakes were not taken into account. Several students wrote for instance *resources* instead of *ressources* in the example below. This example shows that even though the number of letters of

each words was given, participants did not hesitate to give an answer that did not contain the exact number of letters. Finally, there was a Pearson correlation of .997 between the two raters who corrected the productive tests.

> Example
> *Ce pays n'a plus de r_ _ _ _ _ _ _ _ _ (bronnen) naturelles et doit donc importer son énergie.*
> (This country does not have any natural resources anymore and has to import its energy.)

Elicited spoken data

Following previous research (e.g., Brown et al., 2005; Malvern & Richards, 2002), the spoken interactions were transcribed in CHAT format (MacWhinney, 2000a) and the CLAN package of programs (MacWhinney, 2000b), which contains a large number of automatic analyses, was also used for our analyses. Both were developed for the CHILDES project. Lextutor (Cobb, n.d.) was used for determining the LFP of the participants' productions. Furthermore, the CLAN programs and the Gramulator 6.0 (McCarthy et al., 2012) were used to calculate the measures of LD. D and HD-D were selected for our analyses since Treffers-Daller (2013) showed that both measures are good indicators of language ability in French.

Given that the non-native data were collected by two researchers and that data were normally distributed, an independent *t*-test was run to verify whether the researcher who served as an interlocutor had an effect on the learners' use of vocabulary (number of tokens, types and lemmas; LFP; measures of LD). The results showed that there was no interviewer effect[18] and that the data collected by the

[18] An independent *t*-test showed that interacting with one or with the other researcher did not have an effect on learners' vocabulary. The differences for

two interviewers could be analyzed together. We also investigated whether following a three or a four hour French course had an influence on learners' vocabulary using an independent *t*-test, which did not seem to be an issue[19]. Therefore, no distinction was made between the two groups of learners. Finally, it should also be mentioned that we controlled for the words that speakers clearly did not know and repeated after the researcher produced them. These were not taken into account for the analyses.

Intraclass Correlation Coefficient (ICC) estimates were calculated using SPSS version 25 based on a random consistency model in order to check the consistency of raters' judgements. The results showed that there was a high degree of reliability between the different raters. The average ICC was .90 with a 95% confidence interval from .88 to .91 $F(1401) = 9.57$, $p < .001$.

First, in order to determine the factors that predict holistic rating scores of L1 and L2 speakers' vocabulary use in oral output in two semi-structured dialogic speaking tasks, we ran a multivariate linear mixed effects model using SPSS (version 25) with (1) the holistic rating score as dependent variable, (2) topic, as fixed main factor, (3) number of types, number of lemmas 1K, number of tokens 2K and HD-D as covariates and (4) participants as random effect. This analysis was used because two tasks and two expert ratings per participant were used. We started with a model including all the aforementioned main factors before carrying out a backward stepwise analysis to arrive at the simplest model possible. Each time a non-significant factor was

---

both groups on the different measures were not significant. Therefore, we analyzed data of those learners as part of the same group.

[19] An independent *t*-test showed that having a three-hour course or a four-hour course did not affect learners' vocabulary. The differences for both groups and both tests were not significant $t(48) = 1.24$, $p > .05$ for the receptive test and $t(48) = .96$, $p > .05$ for the productive test. Therefore, we analyzed data of those learners as part of the same group.

removed, the models were compared using *Chi*-square tests and the AIC values.

In order to answer the second research question, independent *t*-tests were used for the normally distributed variables and Mann-Whitney U-tests for the variables that were not normally distributed. It should be mentioned that the number of tokens was kept constant between 90 and 450 tokens for calculating lexical diversity. Although a cut-off between 100 and 400 tokens is suggested by McCarthy and Jarvis (2007) for calculating D, some studies (e.g., Appel et al., 2019) have suggested slightly more lenient cut-off scores in order to capture "a wider range of linguistic abilities represented in the corpus, as lower level speakers often produce shorter samples than their higher level counterparts" (p. 31). This was also the case in the present study.

Finally, Pearson correlations were computed to evaluate the relationship between the receptive and productive vocabulary knowledge tests and holistic ratings of the L2 speakers' output since data were normally distributed. This should allow us to answer our third research question.

6.5 Results

6.5.1 Predicting factors for experts' ratings of speaking tasks

In order to better understand the importance of holistic rating scores of participants' spoken output in the two semi-structured dialogic tasks and the factors that could predict these ratings, a linear mixed effects model was conducted.

First, we checked for multicollinearity. As for all measures the variables displayed Pearson correlations above .70, which is the threshold commonly used in research (e.g., Appel et al., 2019; Crossley et al., 2011), they could not all be included in the regression model. Therefore, the variables that correlated best with the holistic rating

score were retained as predictor variables for the analyses (Table 6.1). As a result, only four variables, i.e., the number of types used, lemma 1K, tokens 2K and HD-D were used in the analyses.

Table 6.1

Significant correlations with holistic rating scores for both topics

Measures	Lexical variable	Pearson correlation
Word count	Tokens	.25**
	Types	.42**
	Lemmas	.37**
Lexical Frequency Profile	Tokens 1K	-.35**
	Tokens 2K	.44**
	Tokens 2K+	.21*
	Types 1K	-.05
	Types 2K	.34**
	Types 2K+	-.01
	Lemmas 1K	-.43**
	Lemmas 2K	.39**
	Lemmas 2K+	.23**
Lexical diversity	HD-D	.70**
	vocd-D	.68**

** significant at the 0.05 level, ** significant at the 0.01 level*

The results show that three parameters, i.e., HD-D, the number of types, the number of 1K lemmas used are positively and significantly related to raters' holistic scores (see Table 6.2). All the predictors together explained 64.57% of the variance of raters' holistic scores. HD-D predicted most of the variance in human ratings of spoken output based on the null model, viz., 55.54%, followed by the number of types, which explained 30.91% of the variance in comparison with the null

model, and finally the number of 1K lemmas explained 17.57% of the variance in comparison with the null model in ratings of spoken output.

Table 6.2

Parameter estimates linear mixed effects model for holistic rating score

Fixed effects	**Estimate**	***SE***	***t***	***p***	**95% confidence interval**	
					Lower bound	Upper bound
Intercept	7.097	1.084	6.542	.000	4.96	9.23
Topic= Interview	-.702	.109	-6.429	.000	-.92	-0.49
Topic= Doctor	0	0				
Types	.008	.003	3.118	.002	0.00	.01
Lemmas 1K	-.042	.013	-3.232	.001	-.067	-.016
HD-D	.174	.031	5.681	.000	0.114	.235
Residual	.30	.03		.000	.244	.368
Intercept (participants)	.50	.11		.000	.332	.760
-2LL	598.88					

Note. SE = standard error

6.5.2 Number of words produced

Table 6.3 provides the mean number of tokens, types and lemmas used by L2 and L1 speakers for the doctor's topic and the interview topic as well as the results of the Mann-Whitney U-test. The results show that L1 speakers use slightly more tokens, types and lemmas for both topics than L2 speakers even though they needed less time to perform the speaking tasks.

Since most of the variables were not normally distributed, Mann-Whitney U-tests were computed between L1 and L2 speakers to evaluate whether both groups statistically differed in the number of words they produced. The difference in the number of lemmas between L1 and L2 speakers for the doctor's topic was normally distributed and an independent *t*-test was used (Table 6.4). As can be seen in Tables 6.3 and 6.4, there were significant differences between both groups for both topics regarding the number of tokens, types and lemmas used. Thus, native speakers seem to use more tokens, types and lemmas while speaking.

Table 6.3

*Descriptives and summary results of the Mann-Whitney U-test for the number of tokens for the doctor's topic (N L1 = 27, N L2 = 50) and for the interview topic (*N *L1 = 27;* N *L2 = 50)*

Variable	***M* L1 (*SD* L1)**	**95% *CI* L1**	***M* L2 (*SD* L2)**	**95% *CI* L2**	**Mann-Whitney U**	***Z*-score**	***p*-value**	**Effect size *d*[20]**
Doctor								
Tokens	173.78 (67.23)	[147.18, 200.37]	143.30 (62.69)	[125.48, 161.12]	462.00	-2.27	.023	.54
Types	85.37 (25.80)	[75.17, 95.57]	61.80 (19.49)	[56.26, 67.34]	293.50	-4.07	.000	1.05

[20] Effect sizes were calculated using the tools retrieved from https://www.psychometrica.de/effect_size.html, Lenhard, W. & Lenhard, A. (2016). Calculation of Effect Sizes. Dettelbach, Germany: Psychometrica.

Variable	*M* L1 (*SD* L1)	95% *CI* L1	*M* L2 (*SD* L2)	95% *CI* L2	Mann-Whitney U	*Z*-score	*p*-value	Effect size *d*[21]
Interview								
Tokens	362.81 (119.46)	[315.56, 410.07]	238.24 (80.93)	[215.24, 261.24]	242.50	-4.62	.000	1.24
Types	151.93 (33.48)	[138.68, 165.17]	95.70 (24.41)	[88.76, 102.64]	93	-6.21	.000	2.00
Lemmas	128.00 (27.01)	[117.32, 138.68]	86.12 (25.37)	[78.91, 93.33]	165	-5.45	.000	1.58

Note. M = *Mean;* SD = *Standard Deviation*; CI = *Confidence Interval*.

[21] Effect sizes were calculated using the tools retrieved from https://www.psychometrica.de/effect_size.html, Lenhard, W. & Lenhard, A. (2016). Calculation of Effect Sizes. Dettelbach, Germany: Psychometrica.

Table 6.4

Descriptives and summary results of the independent samples t-test between L1 and L2 speakers for the doctor's topic (N L1 = 27; N L2 = 50)

Variable	*M* L1 (*SD* L1)	95% *CI* L1	*M* L2 (*SD* L2)	95% *CI* L2	*t*-value	*p*-value	Effect size *d*
Doctor							
Lemmas	73.07 (21.13)	[64.71, 81.43]	57.62 (17.18)	[52.74, 62.50]	-3.47	.001	.83

Note. M = *Mean;* SD = *Standard Deviation;* CI = *Confidence Interval.*

6.5.3 Lexical frequency profile of the words produced

In order to answer the second part of our second research question, we looked at the LFP of L1 and L2 speakers' output, which is available in Tables 6.5 and 6.6. Independent *t*-tests were used for the variables that were normally distributed (Table 6.5) and Mann-Whitney U-tests for the variables that were not normally distributed (Table 6.6). It should be mentioned that these two task types are supposed to elicit mainly high-frequent words, as learners at the B1 level are supposed to interact about them. The results show this is the case for both L1 and L2 speakers. In addition, L1 speakers use slightly more low-frequency tokens, types and lemmas than L2 speakers for both topics. Even though the differences between both groups were statistically significant, the differences in percentages are small and probably only represent a limited number of tokens, types and lemmas. For the doctor's topic, the number of 2K types produced by L1 and L2 speakers is not significantly different. Both groups seem to use a similar number of unique words. As far as the interview topic is concerned, the results show that there is no statistically significant difference between the number of low-frequency words (i.e., tokens, types and lemmas beyond 2K) L1 and L2 learners use.

Table 6.5

*Descriptives and summary results of the independent samples t-test between L1 and L2 speakers for both topics (*N *doctor L1 = 27,* N *doctor L2 = 50;* N *interview L1 = 27;* N *interview L2 = 50)*

Variable	***M* L1% (*SD* L1)**	**95% *CI* L1**	***M* L2% (*SD* L2)**	**95% *CI* L2**	***t*-value**	***p*-value**	**Effect size *d***
Doctor							
Tokens 1K	85.14 (3.13)	[83.91, 86.38]	90.08 (2.79)	[89.29, 90.87]	7.09	.000	1.69
Tokens beyond 2K	9.93 (3.21)	[8.66, 11.20]	5.41 (2.39)	[4.73, 6.09]	-7.00	.000	1.67
Types 1K	74.37 (4.29)	[72.67, 76.06]	79.96 (4.80)	[78.60, 81.33]	5.65	.000	1.35
Types 2K	9.11 (2.27)	[8.21, 10.01]	8.18 (2.66)	[7.42, 8.93]	-1.54	.128	.37
Types beyond 2K	15.69 (4.10)	[14.07, 17.31]	9.96 (4.12)	[8.79, 11.13]	-5.84	.000	1.40

Variable	*M* L1% (*SD* L1)	95% *CI* L1	*M* L2% (*SD* L2)	95% *CI* L2	*t*-value	*p*-value	Effect size *d*
Lemmas 2K	10.54 (2.81)	[9.42, 11.65]	8.64 (3.06)	[7.77, 9.51]	-2.66	.010	.64
Lemmas beyond 2K	12.85 (3.80)	[11.35, 14.36]	6.78 (3.08)	[5.90, 7.65]	-7.60	.000	1.82
Interview							
Tokens 1K	87.61 (1.84)	[86.88, 88.34]	89.44 (2.40)	[88.76, 90.12]	3.45	.001	.82
Tokens 2K	4.24 (1.26)	[3.74, 4.74]	2.70 (0.73)	[2.49, 2.91]	-5.84	.000	1.40
Tokens beyond 2K	8.03 (1.16)	[7.57, 8.49]	7.72 (2.09)	[7.12, 8.31]	-0.84	.403	.20
Types 1K	77.29 (2.47)	[76.32, 78.27]	77.43 (4.44)	[76.17, 78.70]	.18	.857	.04
Types 2K	8.02 (1.86)	[7.29, 8.76]	6.20 (1.61)	[5.74, 6.65]	-4.49	.000	1.07

Variable	*M* L1% (*SD* L1)	95% *CI* L1	*M* L2% (*SD* L2)	95% *CI* L2	*t*-value	*p*-value	Effect size *d*
Types beyond 2K	14.31 (2.35)	[13.38, 15.24]	15.04 (3.78)	[13.97, 16.12]	.92	.361	.22
Lemmas 1K	77.75 (2.60)	[76.72, 78.78]	82.05 (3.42)	[81.07, -83.02]	5.69	.000	1.36
Lemmas 2K	9.30 (2.22)	[8.42, 10.17]	6.82 (1.82)	[6.30, 7.34]	-5.25	.000	1.25
Lemmas beyond 2K	12.96 (2.38)	[12.02, 13.90]	11.12 (2.88)	[10.30, 11.94]	-2.83	.006	0.68

Note. M = Mean; *SD* = Standard Deviation; *CI* = Confidence Interval.

Table 6.6

*Descriptives and summary results of the Mann-Whitney U-test for L1 and L2 speakers for tokens and lemmas for the doctor's topic (*N *L1 = 27,* N *L2 = 50)*

Variable	***M* L1% (*SD* L1)**	**95% *CI* L1**	***M* L2% (*SD* L2)**	**95% *CI* L2**	**Mann-Whitney U**	***Z*-score**	***p*-value**	**Effect size *d***
Doctor								
Tokens 2K	4.93 (1.11)	[4.49, 5.36]	4.02 (1.62)	[3.74, 4.66]	413.50	-2.79	.005	.67
Lemmas 1K	76.61 (4.35)	[74.89, 78.33]	84.55 (3.84)	[83.46, 85.64]	115.00	-5.98	.000	1.86

Note. M = Mean; *SD* = Standard Deviation; *CI* = Confidence Interval.

6.5.4 Lexical diversity of the words produced

Table 6.7 displays the descriptive results of the LD and the differences between L1 and L2 speakers. It shows that both L1 and L2 speakers produced output that is more lexically diverse for the interview topic than for the doctor's topic, with higher values for D indicating a higher LD and with values closer to zero for HD-D indicating a more lexically diverse output. Moreover, the results indicate that the performance of L1 speakers is more lexically diverse than that of L2 speakers and that the variation between speakers is also higher for the L2 speakers, even though the number of tokens has been kept constant, i.e., between 90 and 450 tokens. As the data were normally distributed[22], we computed independent *t*-tests (Table 6.7), which indicate that there were statistically significant differences for both topics between L1 and L2 speakers concerning LD, as measured by D and HD-D. L1 speakers' output is more lexically diverse than L2 speakers' output. The effect sizes, which are very large (Cohen, 1988) for both topics, are higher for HD-D than for D.

[22] As the data of the non-native speakers for the doctor's topic were lightly skewed for both measures of LD, we also performed a Mann-Whitney U-test in order to have a more robust analysis. The results were very similar and showed that both groups were statistically significant for vocd-D (U = 53.000, z =-6.054, p < .000, r = .74) and for HD-D (U = 40.000, z =-6.224, p < .000, r = .76).

Table 6.7

*Descriptives and summary results of the independent samples t-test of the lexical diversity of L1 and L2 speakers for both topics (*N *doctor L1 = 24,* N *doctor L2 = 43;* N *interview L1 = 21;* N *interview L2 = 50)*

Variable	*M* L1 (*SD* L1)	95% *CI* L1	*M* L2 (*SD* L2)	95% *CI* L2	*t*-value	*p*-value	Effect size *d*
Doctor							
Vocd-D	46.20 (7.66)	[42.96, 49.43]	26.47 (9.02)	[23.70, 29.25]	9.04	.000	2.30
HD-D	-1.75 (1.02)	[-2.18, -1.32]	-6.02 (2.43)	[-6.77, -5.27]	10.05	.000	2.56
Interview							
Vocd-D	58.81 (8.00)	[55.17, 62.45]	33.30 (8.62)	[30.85, 35.75]	11.61	.000	3.02
HD-D	-0.18 (0.71)	[-0.50, 0.15]	-3.93 (1.89)	[-4.47, -3.40]	12.15	.000	3.16

Note. M = Mean; *SD* = Standard Deviation; *CI* = Confidence Interval.

6.5.5 Descriptive results vocabulary knowledge tests and experts' ratings of speaking tasks

Table 6.8 shows the results of the receptive and productive vocabulary knowledge tests of the L2 learners per K level. It displays results that are higher on the receptive vocabulary test than on the productive vocabulary test. As far as the receptive vocabulary test is concerned, learners seem to be familiar with the 1,000 most frequent words in French (*M* = 91.87%) and almost with the 2,000 most frequent words (*M* = 87.20%), as obtaining 90% on a 1,000 word level is considered mastery of that K level (e.g., Read, 1988; Schmitt et al., 2001). Learners did not, however, obtain the 90% benchmark for mastering the 1K level in French (*M* = 72.07%) on the productive test. Moreover, the standard deviation of the productive vocabulary test is higher than of the receptive vocabulary test, which indicates that there might be more variation in learners' productive vocabulary than in their receptive vocabulary. Finally, the mean holistic rating scores for the L2 speakers (*N* = 50) are similar for both topics, i.e., 2.70 for the doctor's topic (*SD* = .78) and 2.72 for the interview topic (*SD* = .79) on a scale from 1 to 5.

6.5.6 Relationship between vocabulary knowledge tests and experts' ratings of speaking tasks

In order to investigate the relationship between L2 learners' receptive and productive vocabulary knowledge and the holistic rating scores given by raters, Pearson correlations were computed. We only found one positive, significant correlation between the receptive vocabulary test and the holistic ratings for the doctor's topic, as can be seen in Table 6.9. No significant correlations were found with the productive vocabulary test.

Table 6.8

Descriptives per K level for the receptive and productive vocabulary tests of the L2 learners (N = 50)

K level	*M* % (*SD*)	95% CI	*M* raw scores (Max=30) (*SD*)	95% CI
Receptive vocabulary test				
0-1,000	91.87 (6.57)	[90, 93.73]	27.56 (1.98)	[27, 28.12]
1,001-2,000	87.20 (8.12)	[84.89, 89.51]	26.16 (2.44)	[25.47, 26.85]
2,001-3,000	72.13 (11.52)	[68.86, 75.41]	21.64 (3.46)	[20.66, 22.62]
3,001-4,000	69.73 (12.36)	[65.22, 74.25]	20.92 (4.76)	[19.57, 22.27]
Total	80.23 (9.30)	[77.59, 82.88]	96.28 (11.16)	[93.11, 99.45]
Productive vocabulary test				
0-1,000	72.07 (12.36)	[68.55, 75.58]	21.62 (3.71)	[20.57, 22.67]
1,001-2,000	61.73 (16.01)	[57.18, 66.28]	18.52 (4.80)	[17.15, 19.89]
Total	66.90 (13.48)	[63.07, 70.73]	40.14 (8.09)	[37.84, 42.44]

Note. M = Mean; *SD* = Standard Deviation; CI = Confidence Interval.

Note. The mean raw scores are out of 30 per frequency band, out of 120 for the total receptive test and out of 60 for the total productive test.

Table 6.9

Pearson correlation vocabulary tests and mean holistic rating score (N = 50 Doctor; N = 50 Interview)

	***M* rating doctor**	***p*-value**	***M* rating interview**	***p*-value**
Receptive vocabulary test	.32* $r^2 = 0.10$	.025	.19 $r^2 = 0.04$	.185
Productive vocabulary test	.27 $r^2 = 0.07$	.054	.05 $r^2 = 0.003$	.742

* significant at the 0.05 level

6.6 Discussion

The current study aimed at determining which factors could predict holistic lexical rating scores of L1 and L2 speakers' spoken output in two semi-structured dialogic speaking tasks. Next, it looked at how the output of L1 and L2 speakers differs and finally aimed at exploring the relationship between low-intermediate L2 learners' vocabulary knowledge and holistic lexical rating scores given by experts on learners' output on two semi-structured dialogic speaking tasks.

6.6.1 Predictors of raters' holistic scores of L1 and L2 speakers' oral output

With respect to the first research question, we investigated the factors that could predict raters' holistic ratings of L1 and L2 speakers' vocabulary use in oral output. The measure that could best predict raters' holistic scores was the LD of participants' output, followed by the number of types and the number of high-frequency lemmas. Speakers' output that contained more low-frequency words was scored higher than output that contained more high-frequency words. This is in line with previous research in speaking and writing (Kyle &

Crossley, 2015), even though the methodologies are not entirely comparable. Nevertheless, despite differences in tasks and languages, these studies show that speaking samples of high proficiency display more low-frequency words.

Moreover, the results showed that the number of words (tokens) did not predict ratings, whereas the number of different words (types) did. These results confirm previous research (Iwashita et al., 2008; Lorenzo-Dus & Meara, 2005) that demonstrated that the number of unique words a speaker uses, i.e., the number of types, was a strong predictor for raters' scores of speaking proficiency, even though no percentages were given. Such trends indicate that raters will better evaluate speakers who produce a higher number of unique words. More proficient speakers can access and use a higher number of (unique) words than low proficient speakers, which is indicative of a larger vocabulary.

Finally, the results of LD are also in line with previous research (Crossley, et al., 2011; Yu, 2010). Whereas HD-D explained up to 55.54% of the variance based on the null model in spoken output in the present study, the results are lower in previous research, which might be due to the LD measure that was used. Even though the numbers are diverse due to the use of different statistical models and different predictors, they all show that speaking samples that are more lexically diverse receive better ratings, which might indicate that a more diverse vocabulary is associated with a larger vocabulary knowledge.

6.6.2 Differences in lexical output between L1 and L2 speakers

The second focus of the study was to explore how L1 and L2 speakers' output in two semi-structured dialogic tasks differs regarding (1) the number of tokens, types and lemmas, (2) the LFP and (3) the LD. The results show that L1 speakers, even though they needed less time to perform the speaking tasks, use a higher number of tokens,

types and lemmas than L2 speakers for both topics. These results are in line with the findings of Treffers-Daller (2013), who found that French native speakers produced more tokens and types than L2 learners of French.

In our study, the data were elicited through two B1 tasks, which should elicit a high percentage of high-frequency words. This is confirmed as both groups of speakers draw the majority of the tokens, types and lemmas used from the 1,000 most frequent words in French. This is confirmed as both groups of speakers draw the majority of the tokens, types and lemmas used from the 1,000 most frequent words in French. These findings are also in line with Nation's (2013, p. 22) claim that a small group of high-frequency words covers a very large proportion of the words in texts. These results indicate that speakers performing tasks at the B1 level mainly employ high-frequency words. However, although L1 speakers tend to use slightly more low-frequency tokens, types and lemmas for the doctor's topic, the results are less straightforward for the interview topic. This indicates that some topics might prompt speakers to use more low-frequency or more high-frequency words (Lindqvist, 2010) and that, therefore, care should be taken when generalizing to other topics. Nevertheless, the results confirm previous research looking into the frequency of the words low and highly proficient speakers use (Laufer & Nation, 1995; Lindqvist, 2010; Ovtcharov et al., 2006) and which concluded that the more low-frequency or rare words speakers use, the more proficient they are. L1 speakers are often assumed to use more low-frequency or rare words. Furthermore, the results of our study indicated that speakers using more low-frequency words were also better rated, which confirms previous research (Crossley & McNamara, 2013; Kyle & Crossley, 2015).

Finally, as far as the LD of both L1 and L2 speakers is concerned, the results show that L1 speakers produce output that is more lexically diverse than L2 speakers' output, as measured by D and HD-D, for both

topics. This finding is in line with previous research (Treffers-Daller, 2013) in which the LD of L1 speakers showed to be higher than that of L2 speakers. In sum, more lexically proficient French speakers do not only produce more words but also use less frequent words and more different words.

6.6.3 Relationship between vocabulary knowledge and experts' ratings of speaking tasks

Finally, we investigated the relationship between L2 learners' vocabulary knowledge and holistic lexical ratings. The results are not conclusive as there was only a significant correlation between the receptive vocabulary test and the holistic rating of the doctor's topic but not in the interview topic. This means that learners who had a higher score on the receptive vocabulary test received a better rating score of their output in the doctor's task. Different correlations were found in Uchihara and Clenton (2018) but this could be due to differences in the task type (monologic versus interactive speaking tasks). It was expected that a productive vocabulary test would be a better predictor of vocabulary use than a receptive test as speakers need to access their L2 or even L1 productive lexicon (Milton, 2009, p. 121). This would also confirm results in a study by de Jong et al. (2012), in which a high correlation (r = .79) was found between productive vocabulary knowledge, testing the knowledge of single words up to the 10,000 most frequent words of a Dutch spoken corpus and multiword units, testing knowledge of 26 verb-noun collocations and prepositional phrases and holistic ratings of speaking with intermediate and advanced learners of Dutch. Surprisingly, we did not find a significant correlation between the productive vocabulary test and holistic lexical ratings, which might indicate that for low-intermediate learners, a productive vocabulary test is not a good predictor of holistic lexical ratings. The absence of a correlation might also be due to the fact that we only measured up to the 2,000 most frequent words in French,

whereas other studies measured more low-frequency words, or because of learners' low scores on the test.

6.7 Limitations and future research

There are a number of limitations to this study that need to be acknowledged and addressed in future research. First, as the aim of this study was to focus on low-intermediate L2 learners, future research should also include other proficiency levels. Second, although two task topics were given to the speakers to control for a topic effect, more research into other task types, such as cognitively more demanding task types, is warranted, since such task types could reveal differences in learners' vocabulary knowledge and lead for instance to the use of fewer words or the use of more high-frequency words. Moreover, the tasks were not counterbalanced since all learners first did the doctor's task. This might have led to the production of a lower number of words due to test anxiety. Third, only D and HD-D were used to investigate the lexical diversity of speakers' output. Unlike for English, there are not many automated measures and lexical tools allowing to lexically analyze spoken language for French. Therefore, studies focusing on the development of such tools would be needed. Fourth, the productive vocabulary test in this study only tested up to the 2,000 most frequent words in French. Future research could include a test measuring up to 4,000 words. Finally, in order to better predict experts' holistic ratings, we should not only take into account complexity as measured by lexical parameters but also focus on accuracy and fluency.

6.8 Conclusion

This study sought to explore which factors could predict holistic lexical ratings of L1 and L2 speakers' oral output in two semi-structured dialogic speaking tasks. In addition, this study investigated the differences between L1 and L2 speakers' output in the same speaking tasks and explored the relationship between L2 learners'

receptive and productive French vocabulary knowledge and experts' holistic ratings of learners' lexical output. As far as the predictors of holistic human ratings of vocabulary use in two semi-structured dialogic speaking tasks are concerned, the results show that the more diverse the spoken output is, the higher the ratings. The number of types as well as the number of high-frequency lemmas also have an influence on how experts perceive speakers' output. Second, a comparison between the lexical output of L1 and L2 speakers does not only show that L1 speakers use more tokens, types and lemmas, but that they also use more low-frequency words and that their output is more lexically diverse. Finally, the results on the relationship between L2 learners' receptive and productive vocabulary knowledge and experts' holistic lexical ratings show that a receptive vocabulary test might predict experts' holistic lexical ratings. This study adds to our understanding of the lexical evaluation of French spoken output and emphasizes the importance of having appropriate indicators for automatically scoring speaking performances.

Chapter 7

General conclusion

7.1 Introduction

In this PhD thesis, the first three chapters provided the theoretical and methodological framework in which the three empirical studies of this PhD project were embedded. The three studies aimed at exploring the relationship between vocabulary and listening on the one hand and vocabulary and speaking on the other. They were developed in the three following chapters. In this final chapter, a summary and discussion of the main results of the three studies is presented (7.2), followed by some limitations (7.3). Next, the implications of the three studies are listed and suggestions for future research are formulated (7.4). This chapter is concluded with some final remarks (7.5).

7.2 Main results

7.2.1 Global summary

This section deals with the most important findings resulting from the three empirical studies in this PhD thesis. Figure 7.1 (see also Chapter 1, Figure 1.2) visualizes the overall architecture of the studies and the interrelation between them. The two main research questions that were outlined in Chapter 1 will be answered first.

- To what extent are vocabulary knowledge and use related with the two oral skills, viz., listening and speaking?
- Are lexical measures a good predictor of L2 learners' vocabulary use in spoken output?

With regard to the first research question, it has been demonstrated in study 1 that intermediate French-as-a-foreign language learners' receptive vocabulary knowledge correlates positively and significantly with L2 listening comprehension, as measured by a B1 listening test, both for English (r = .64) and for French (r = .70). Learners seem to need at least 1,000 lemmas for adequate listening comprehension in English as tested with the PET-test, if a score of 85% on the listening test is considered adequate. For French, at

least 2,000 lemmas should be known in order to obtain adequate listening comprehension as tested with the DELF-test, if a score of 80% is considered adequate. The results also indicate that less than 98% lexical coverage might be needed for listening at the intermediate level. Learners knowing the 1,000 most frequent words in English might already reach a lexical coverage of 91%, whereas knowledge of the 2,000 most frequent lemmas in French might lead to a lexical coverage of 93%.

Further, the answer to the question of how low-intermediate L2 learners of French' vocabulary knowledge and speaking are related in study 2 is threefold, as the relation regards (1) the number of words in the output, (2) the lexical frequency profile of learners' output and (3) the lexical diversity of learners' output. The analyses of the results ask for some nuance. First, the findings showed that productive vocabulary significantly correlated with learners' vocabulary use in two semi-structured dialogic speaking tasks, as measured by the number of types and lemmas. However, the results were only confirmed in one topic. No significant correlations were found between receptive vocabulary knowledge and learners' vocabulary use as measured by the number of words they used. Second, as far as the relationship between receptive and productive vocabulary knowledge and the frequency of the words learners use is concerned, the results were less straightforward. There was only a significant correlation between the receptive vocabulary test and the total number of high- and low-frequency words (tokens), in one topic. No other significant correlations were found. The vocabulary tests might thus not be good predictors of the learners' LFP for the topics of the tasks in our study.

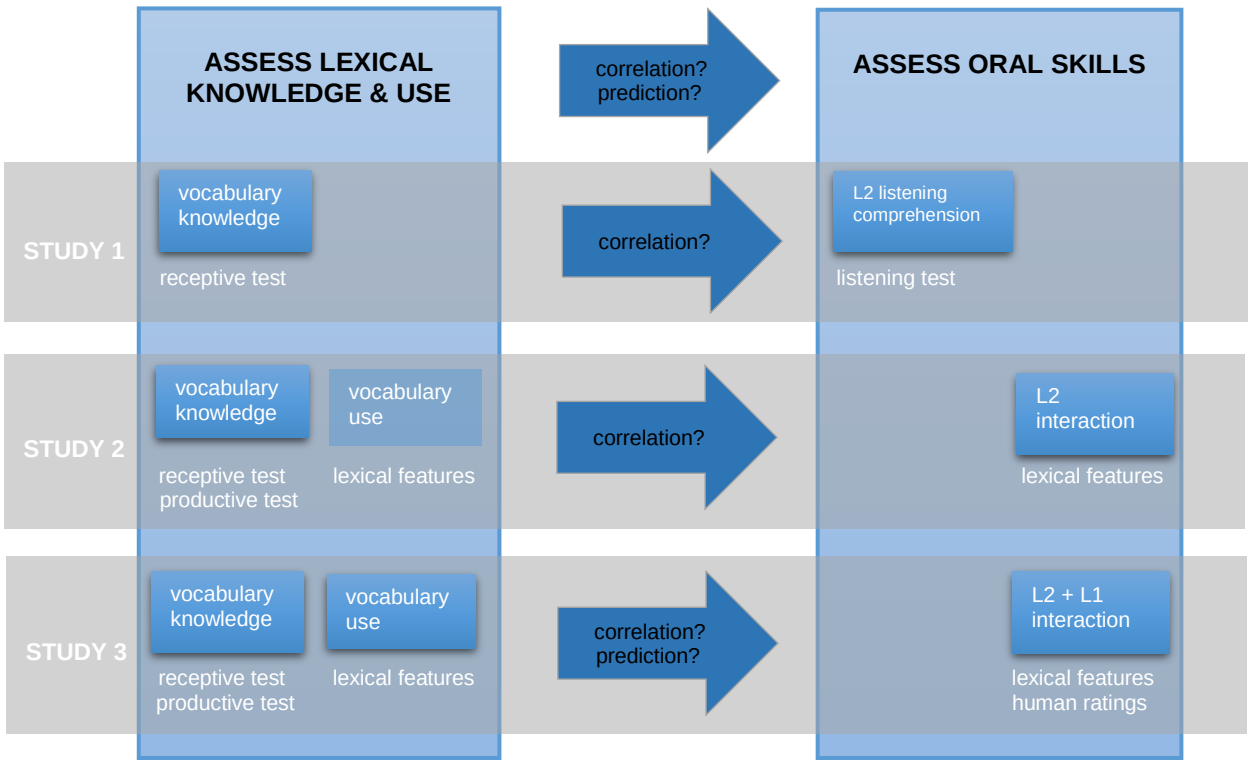


Figure 7.1 *Visualization of the global architecture of and interaction between the three studies.*

Third, a significant correlation was found between the receptive test and one of the LD measures (HD-D), but only in one task, viz., the doctor's task. The productive vocabulary test significantly correlated with the two measures, but also only in the doctor's task. The results are thus inconclusive, given that there were only significant correlations with one of the tasks. Finally, the last research question investigated the relationship between L2 learners' receptive and productive vocabulary use and the scores of that spoken output given by experts. The results displayed a significant correlation between the receptive vocabulary test and raters' scores given in the doctor's topic, but not in the interview topic. No correlation was found, however, between the productive vocabulary test and the ratings, with none of the two topics.

The second main research question concerned the predicting value of lexical measures for experts' ratings of spoken output. The results of study 3 showed that, of the selected measures in our study, three factors predicted experts' holistic ratings of L1 and L2 speakers' output, i.e., lexical diversity as measured by HD-D, the number of different words (types) and the number of high frequency lemmas (i.e., 0-2,000 most frequent) in the output. These three factors explained 64.57% of the variance in the ratings given by experts.

Table 7.1 summarizes whether significant correlations were found between the different vocabulary measures and listening comprehension, vocabulary use in speaking and the ratings. An 'X' indicates that the relationship was confirmed, a hyphen (-) is indicative of the absence of a relationship.

Table 7.1

Summary of relations between receptive and productive knowledge and listening and speaking

Skill	Receptive vocabulary knowledge	Productive vocabulary knowledge	Ratings
Listening comprehension	X	Not tested	Not tested
Number of words			
- tokens	-	-	X
- types	-	X	X
- lemmas	X	X	X
LFP			
- tokens	X	-	X
- types	-	-	X
- lemmas	-	-	X
Lexical diversity			
- HD-D	X	X	X
- vocd-D	-	X	X
Ratings	X	-	/

7.2.2 Discussion

After having presented the global results, this discussion section offers a more fine-grained interpretation and contextualization of the results. This section deals with lexical knowledge and use, listening and speaking and with the predicting factors for vocabulary use in speaking.

Lexical knowledge and use

In studies 1 and 2, vocabulary knowledge tests were administered to L2 learners of French. In the first study, 330 participants completed a monolingual receptive vocabulary test, measuring up to the 5,000 most frequent lemmas in French. The results show that most learners knew only few words from the 2K level onwards. The results are better, though, on the bilingual receptive test administered in study 2 with 51 L2 learners of French. These results probably give a better representation of learners' actual vocabulary knowledge. These findings are in line with previous research, which has already shown that scores on bilingual tests tend to be higher than on monolingual tests (De Clercq, 2016; Elgort, 2013; Segers, 2015; Vrancx, 2016). Further, the results on the bilingual productive vocabulary test are lower than on the receptive vocabulary test. This is an expected result because productive vocabulary knowledge usually represents only a subset of receptive vocabulary knowledge. These results seem to be consistent with the findings of previous research (Laufer, 1998; Laufer & Paribakht, 1998; Webb, 2008).

The second and third study of this PhD project explored French vocabulary use in two semi-structured dialogic speaking tasks. Study 2 aimed at examining the correlation between L2 learners' receptive and productive vocabulary knowledge and vocabulary use in two semi-structured dialogic tasks, with (low-)intermediate learners of French. Learners' vocabulary use in these speaking tasks was determined by the number and frequency of the words that learners used, i.e., the lexical frequency profile of the output, and also by its lexical diversity, as measured by D and HD-D. The research questions at stake investigated the threefold relationship between the vocabulary knowledge and vocabulary use as measured by (1) the number of words used, (2) their frequency and (3) their lexical diversity.

First, the results indicate that learners with a larger vocabulary knowledge, both receptively and productively, use more types and

lemmas when they speak. In addition, the study shows that a productive vocabulary test might be a better predictor of vocabulary use in speaking than a receptive vocabulary test, since the correlations between the productive vocabulary test and the number of types and lemmas are higher than between the receptive vocabulary test and the number of types and lemmas. These results match those observed in earlier studies (de Jong et al., 2012; Koizumi, 2005; Koizumi & In'nami, 2013), in which productive vocabulary knowledge seemed to be a good predictor of speaking. Unlike in previous studies (e.g., Miralpeix & Muñoz, 2018; Uchihara & Clenton, 2018), a receptive as well as a productive vocabulary knowledge test are used in study 2, which allows us to compare learners' receptive and productive vocabulary knowledge.

Second, in line with some previous studies (Uchihara & Clenton, 2018), only small correlations are found between receptive vocabulary knowledge and the lexical frequency profile and lexical diversity of learners' spoken output. The low correlations might be due to the fact that the relationship between vocabulary knowledge and speaking is less direct than between vocabulary knowledge and listening. In speaking, other factors seem to play a role such as time pressure. Compared to results found in writing research (e.g., Laufer & Nation, 1995), the results in this study are lower, but the main difference between writing and speaking is that learners have time to think in writing, which they do not have in speaking.

Listening and speaking

The first study in this PhD project focuses on listening comprehension at the (low-)intermediate level. Its objective is to investigate whether the results of Stæhr's study (2009) examining the relationship between vocabulary knowledge and listening comprehension with advanced Danish EFL learners would provide similar results in another context, with another proficiency level and

another language. First, the participants were Flemish learners compared to Danish learners in the original study. Second, they had an intermediate proficiency level that was operationalized by a B1 test, compared with advanced learners as tested with a C2 listening comprehension test. Finally, the results were also tested with another language, i.e., French.

Apart from looking at the extent to which vocabulary knowledge and listening comprehension are associated, this study sought to determine how much vocabulary is needed for adequate listening comprehension at an intermediate level. In addition, it also investigated the lexical coverage that was needed in order to obtain adequate listening comprehension.

Our results indicate a positive and strong relationship between both English and French L2 learners' vocabulary knowledge and listening comprehension at a (low-)intermediate level, viz., r = .64 for English and r = .70 for French. The results are in agreement with those obtained by Stæhr (2009), who found a correlation of r = .70 between vocabulary knowledge as measured by the VLT and a listening comprehension test with advanced Danish EFL learners. As a consequence, the findings of the first research question of study 1 show that Stæhr's results can be generalized to another context.

Our findings also confirm other previous studies, e.g., Stæhr (2008), who found a correlation of r = .69 between the VLT and a listening comprehension test with advanced Danish EFL learners at university. Nevertheless, lower correlations were found by Milton et al. (2010) in their study with intermediate to advanced EFL learners. Even though a similar correlation of r = .67 was found between a spoken vocabulary test (A_Lex, Milton & Hopkins, 2005) and listening comprehension as measured by IELTS, a lower correlation (r = .48) was found between a written vocabulary test (X_Lex, Meara & Milton, 2003) and the same listening comprehension test. The lower results in Milton

et al.'s study might be due to the type of vocabulary test that was administered. Both the written and the spoken tests were Yes/No tests, in which learners often tend to overestimate their vocabulary knowledge (Mochida & Harrington, 2006) since they do not have to show whether they really know the word (Eyckmans et al., 2007). This type of test might, thus, correlate less well with listening comprehension.

Further, both studies, the original study and the replication study, indicate a similar tendency, viz., higher vocabulary knowledge seems to be related to better listening comprehension scores, although the number of words needed in order to obtain adequate listening comprehension were higher in Stæhr's (2009) study compared to what was found in our study. Our results seem to indicate that the knowledge of 1,000 lemmas in English and 2,000 lemmas in French might suffice for adequate listening comprehension at an intermediate level, as tested with two B1 listening comprehension tests, viz., the PET-test and the DELF-test. Stæhr (2009), however, revealed that 5,000 word families would be needed for advanced learners to achieve adequate listening comprehension on a C2 listening test. Other studies (e.g., Nation, 2006; van Zeeland & Schmitt, 2013) also suggested that a larger number of words would be needed for adequate listening comprehension than what was found in our study. Depending on the listening target that should be attained and the input (e.g., narrative stories, television programs), between 2,000 and 7,000 word families would be needed in order to achieve good comprehension for EFL (Nation, 2006; van Zeeland & Schmitt, 2013; Webb & Rodgers, 2009a, 2009b). Our results seem to indicate that for (low-)intermediate learners, adequate listening as tested with the PET-test and the DELF-test, might be obtained with only 1,000 to 2,000 lemmas, corresponding to 625 and 1,250 word families.

Finally, Stæhr (2009) found that a lexical coverage of 98% was needed for adequate listening comprehension at an advanced level.

Our study seems to suggest that a lower lexical coverage, viz., 91% for English and 86% for French might be required for obtaining adequate listening comprehension at an intermediate level, as tested with the PET-test and the DELF-test. Other studies (e.g., van Zeeland & Schmitt, 2013), however, corroborated Stæhr's (2009) results, viz., that 98% coverage was needed for very good comprehension of informal narratives (van Zeeland & Schmitt, 2013), while 95% would already be sufficient for good comprehension (Schmitt, 2008; van Zeeland & Schmitt, 2013) since it "leads to relatively high comprehension rates" (van Zeeland & Schmitt, 2013, p. 474).

Study 3 investigated the relationship between L2 learners' vocabulary knowledge as measured by a receptive and a productive vocabulary test and their vocabulary use in two dialogic speaking tasks as rated by experts. A moderate, significant correlation (r = .32) was found between the receptive vocabulary test and raters' scores on the doctor's task, not on the interview task, nor between the productive vocabulary test and the two tasks. This result is somewhat lower than the results found by Uchihara and Clenton (2018) between a receptive vocabulary test and rated vocabulary use in speaking (r = .55), but it should be noted that the task administered by Uchihara and Clenton was monologic. It was expected that productive vocabulary would better predict experts' ratings of learners' vocabulary use in speaking, a productive skill, corroborating in that way previous research (de Jong et al., 2012), in which a high correlation (r = .79) was found between productive vocabulary knowledge and speaking with adult intermediate and advanced learners of Dutch. However, a possible explanation for the results in our study might be the learners' intermediate language level and the fact that only the 2,000 most frequent words were tested productively. Future research investigating up to the 4,000 most frequent words could allow to better differentiate between learners. Likewise, these results may be explained by the fact that the vocabulary tests targeted general vocabulary knowledge,

whereas the speaking tasks focused on a specific topic for which a rather specific vocabulary was needed (e.g., a doctor's visit and a job interview).

Towards the predictive evaluation of learners' vocabulary use in speaking

Study 3 sought to explore which lexical factors could predict holistic lexical ratings of L1 and L2 speakers' oral output in two semi-structured dialogic speaking tasks. As far as the prediction of holistic lexical ratings is concerned, the results indicate that lexical diversity, the number of uttered types and the number of high frequency lemmas seem to predict the ratings experts assign to spoken output in two dialogic speaking tasks. Moreover, almost 65% of the variance in experts' holistic ratings was explained by these lexical factors. These results mirror those of previous research that showed that spoken output that is more lexically diverse is better rated et al., 2011a; Yu, 2010). Moreover, the findings are also in line with studies that indicated that the number of unique words used by a speaker is indicative of the rating that speaker will receive (Iwashita et al., 2008; Lorenzo-Dus & Meara, 2005). Finally, the use of more low-frequency words seems to lead to a higher holistic rating score of the spoken output, and the results concerning the predictive value of the use of low-frequency words for experts' ratings also seem to be consistent with previous research (Kyle & Crossley, 2015).

Finally, in study 3, L1 and L2 speakers' oral output in two semi-structured dialogic speaking tasks was compared as far as the number of words was concerned, as well as the frequency of those words and their lexical diversity. The results indicate that L1 speakers produce more tokens, types and lemmas than L2 speakers. Even though L1 speakers tend to use more low-frequency tokens, types and lemmas, both L1 and L2 speakers draw the majority of the tokens, types and lemmas they use in B1 speaking tasks from the 1,000 most frequent

words in French. Moreover, L1 speakers' spoken output is also more lexically diverse than L2 learners' output. These findings corroborate the results found by Treffers-Daller (2013), which showed that L1 speakers produced more words (i.e., tokens, types and lemmas) than L2 speakers. It is also in agreement with the findings of Crossley and McNamara (2013) and Kyle and Crossley (2015), who indicated that the use of more low-frequency words would lead to a higher rating of the output.

In conclusion, the most obvious finding to emerge from the first study is the strong relationship between receptive vocabulary knowledge and listening. The evidence from studies 2 and 3 suggest that the relationship between receptive and productive vocabulary knowledge and vocabulary use in speaking is more complex. In addition, the results of study 3 strengthen the idea that lexical measures such as LD, the number of different words and their frequency may be important predictors of experts' holistic ratings of spoken output.

7.3 Limitations

However important and promising the conclusions of these three studies might be, it goes without saying that empirical studies are always subject to limitations. In what follows, we present the limitations of our three studies. We start with a contextual constraint that is almost unavoidable, and has a certain impact on the results. Some methodological limitations are also provided. They are mainly linked to the lack of validated measurement instruments for French.

The sample of the three studies conducted within this PhD project mainly consisted of secondary school students at the B1 level, which is an underrepresented group in vocabulary research. This sample has shown to be relevant for the study object that was determined. As a consequence, the results emerging from studies 2 and 3 cannot be generalized to other populations. The study object could be expanded to other levels, learners at the A2 and C1 or C2 levels, but

especially the higher level participants cannot be found in the context of secondary schools in Flanders.

In study 1, we investigated the relationship between vocabulary knowledge and listening comprehension. In order to do so, and following the methodology of the original study (Stæhr, 2009), one listening test was used for French and one for English. Further research should be undertaken to determine whether similar results can be found with different listening tests.

Given that the aim of study 1 was to compare the relationship between vocabulary knowledge and listening comprehension in French and English, similar vocabulary knowledge tests were administered. The consequence is that the vocabulary knowledge test for English was not the same as the one used by Stæhr (2009). Instead, the VocabLab tests (Peters et al., 2019) using the lemma as counting unit were given to the participants. Therefore, care should be taken when comparing the results of our study with Stæhr's results, even though conversions were suggested in order to improve the ability to compare the findings.

The focus in studies 2 and 3 was on B1 dialogic speaking tasks. In order to control for a topic effect, two topics were selected. It should be noted, though, that further research is needed to investigate whether other topics might result in a larger output with more high- or low-frequency words. Future work could also indicate whether the results emerging from these studies can be generalized to other task types, since previous research has indicated that the task type might affect learners' output (see Derwing et al., 2004).

The two measures choses here were D as measured by Vocd-D, and HD-D. Even though these showed to be good measures for French (Treffers-Daller, 2013), the development of tools for the automatic analysis of oral French would be warranted, given the fact that less tools are available for French than for English.

7.4 Implications and future research perspectives

7.4.1 Methodological and theoretical implications

The research conducted in this PhD project has an impact on three domains, viz., methodology, SLA and pedagogy. The main implications will be detailed below.

The methodological implications of the three studies concern the population and the investigated level in a real learning context, the vocabulary tests and the task types in speaking and finally the oral corpus that was composed.

First, the methodology of the three studies was not only limited to one vocabulary test, but both receptive and productive tests were administered in studies 2 and 3. Moreover, testing materials were not restricted to a vocabulary test and questionnaires, but also included experts' ratings of learners' oral production (study 3). The use of different testing materials in one study allowed us to deepen our understanding of the link between vocabulary knowledge in listening and between vocabulary knowledge and vocabulary use in speaking. At the start we did not yet know whether learners who knew a lot of words could also use them in a communicative setting or whether they would also produce a more diverse vocabulary. Even though the relationship between vocabulary knowledge and use is not a direct one, the results provide insights for a less studied language, i.e., French and for secondary school students at a low-intermediate level.

Second, given that speaking usually requires interaction with someone, speaking is best assessed in contexts that are as close to reality as possible. Hence, vocabulary use was tested through dialogic tasks, which is a task type that has been less investigated.

Finally, using corpora for analyzing (learner) language allows for more in-depth corpus analyses. As the compilation of a spoken corpus is a fastidious task, this might be a reason why not many oral

corpora exist for French (learner) language (e.g., the *Corpus de français parlé parisien des années 2000* (CFPP2000), the French Learner Language Oral Corpora (FLLOC), *Traitement de corpus oraux en français* (TCOF)). Nevertheless, L2 data were collected on two topics with secondary school students in study 2 and they were completed with L1 data of participants of the same age group on the same tasks in study 3, allowing for a comparison between learner language and L1 language serving as a benchmark.

With regards to SLA, the research conducted within this PhD project has provided new insights in the field of the assessment of spoken skills. One of the main challenges is to come to a more empirically validated assessment of the oral skills and of the role of the lexical component in this. More precisely, it has been demonstrated that a strong relationship exists between vocabulary knowledge and listening on the one hand and between vocabulary knowledge and vocabulary use in speaking for French at the B1 level on the other. Further, lexical diversity seems to be a good predictor of vocabulary use in speaking. The studies in this PhD thesis have shown that this is also true in dialogic speaking tasks, not only in monologic speaking tasks and with low-intermediate learners of French.

7.4.2 Pedagogical implications

Little was known about the relationship between vocabulary knowledge and listening comprehension for French at the B1 level. The results of our first study have pointed out that there exists a strong relationship between vocabulary knowledge and listening comprehension at a (low-)intermediate level in French, as it had already been shown for English at higher language levels. Further, according to Decoo (2012), between 800 and 3,350 words are needed at the B1 level for French, depending on the sources. The first study in this PhD thesis demonstrated that at least 2,000 words are needed for adequate listening comprehension at the intermediate level, viz., B1 for French.

The findings have provided some concrete evidence concerning the number of words that is needed for listening comprehension at the B1 level in French. The slightly divergent numbers for English and French in the first study showed that it is not possible to transfer findings from one language to another, viz., from English to French. Every language seems to require proper evidence.

Bilingual receptive and productive vocabulary tests were developed as part of this PhD project, allowing to better grasp the actual vocabulary knowledge of our target population, i.e., learners at the end of secondary education. The tests focus on assessing the most frequent words in French, which are particularly important for beginner learners of French. These bilingual receptive and productive vocabulary tests can now serve as a valid basis for testing and they can help teachers to determine whether L2 learners are familiar with the 1,000 or 2,000 most frequent words.

Another implication for language assessment and more specifically for the assessment of speaking, is the diversity of learners' vocabulary. The results have shown that the number of different words, as measured by features of lexical diversity and frequency, is an important predictor of how learners' output will be rated. It is therefore recommended that teachers pay attention to the frequency and diversity of the words that are used, especially in learners' spoken output. Teachers could make use of freely accessible tools such as Lextutor[23] in order to give feedback on learners' spoken and written productions. Further, given the importance of frequency in learning vocabulary and the lack of textbooks based on frequency, it might be a good idea for material developers to develop textbooks based on a better understanding of the lexical component including such parameters as frequency or lexical diversity.

[23] Lextutor is freely available online on the following website: https://www.lextutor.ca/vp/comp/.

Finally, the third study in this PhD project showed that lexical diversity seems to be a good predictor of vocabulary use in dialogic speaking tasks as rated by human raters. A last implication concerns the potential of the use of automated measures for scoring learners' oral output. Teachers could make use of freely accessible tools (e.g., Lextutor) for assessing the frequency of learners' spoken or, by extension, written output.

7.4.3 Future research perspectives

Drawing on the findings of the studies presented in this PhD thesis, four main areas are identified, which require further research. The first one concerns the starting point of this PhD project, viz., the CEFR. The second one concerns vocabulary use and more specifically the lexical features that were used to measure lexical complexity. The third one regards the learner corpus of oral French that was composed and finally, the fourth one concerns the prediction of oral production.

As a recent study by Wisniewski (2017) pointed out, more empirical research on the distinctions between the CEFR levels is still required, next to empirical research on the different CEFR scales. Wisniewski emphasizes the importance of conducting studies for other languages than English. One way of providing empirical evidence for the CEFR levels could be by looking at the RLDs. As was pointed out in the theoretical framework (see Chapter 2, section 2.3), the RLDs were developed based on experts' and teachers' intuition, but have not received empirical validation. It will be interesting to verify to what extent the specific vocabulary in learners' spoken output corresponds to what experts consider necessary at that level. The tasks could be different tasks than the ones used in this thesis, but they should be explicitly mentioned in the CEFR at a specific level.

Another study could include an extension of the lexical measures used in studies 2 and 3 for measuring lexical complexity. The focus was only on lexical complexity, but in order to better predict

experts' holistic ratings, we could also focus on fluency (e.g., filled pauses, speed fluency) and accuracy and not only take into account complexity as measured by lexical parameters. These three components could be linked to communicative adequacy (de Jong et al., 2012; Revesz, Ekiert, & Torgersen, 2016). Another study could not only focus on raters' holistic scores, but also on different analytical aspects such as comprehensibility, coherence and cohesion, pronunciation, further investigating the relationship between these different aspects.

Next to the validation of the CEFR and to the complexity, accuracy and fluency of learner data, another study could focus on the extension of the oral learner corpus French that was composed within this PhD project. Data from learners with other proficiency levels or from other tasks could be added. It would also be interesting to compare similar tasks in two languages, for instance in the L1 and the L2 of the learners.

Finally, apart from frequency and lexical diversity, other measures exist to analyze learners' vocabulary use in oral or written output, mainly for English. TAALES (Kyle, Crossley, & Berger, 2018) for instance, is a tool that allows to automatically measure lexical complexity for English with more than 400 indices. For French, more tools and measures are needed for the automatic analysis of (learner) language in order to obtain a more complete and fine-grained overview of vocabulary use in speaking. In this way, SLA and NLP can be combined in order to better assess lexical complexity and diversity in oral or written output.

7.5 Conclusion

Most studies on the relationship between vocabulary knowledge and listening on the one hand, and between vocabulary knowledge and use and speaking on the other have focused on receptive vocabulary knowledge for advanced EFL learners studying at university. This PhD project, however, aimed at contributing to the

development of theory by investigating both receptive and productive vocabulary knowledge, by focusing on dialogic tasks instead of monologic tasks and by targeting language learners at the end of secondary school, at a (low-)intermediate level in a less studied language, viz., French. The findings in this PhD thesis have demonstrated that a strong relationship exists between vocabulary knowledge and both listening and vocabulary use in speaking. Moreover, the results have underlined the value of using a complex methodology with more than one vocabulary knowledge test, different lexical measures and human ratings. Finally, the main drive behind this whole PhD project has become our profound conviction: it's all about words.

References

Adolphs, S., & Schmitt, N. (2003). Lexical Coverage of Spoken Discourse. *Applied Linguistics*, *24*(4), 425–438. doi: 10.1093/applin/24.4.425

Afshari, S., & Tavakoli, M. (2017). The relationship between depth and breadth of vocabulary knowledge and Iranian EFL learners' listening comprehension. *International Journal of Research Studies in Language Learning*, *6*(3), 13–24. doi: 10.5861/ijrsll.2016.1438

Alderson, J. C. (2007). The CEFR and the Need for More Research. *The Modern Language Journal*, *91*(4), 659–663. doi: 10.1111/j.1540-4781.2007.00627_4.x

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Anderson, R.C., & Freebody, P. (1981). Vocabulary knowledge. In J.T. Gutherie (Ed.), *Comprehension and Teaching: Researching Reviews* (pp. 77-117). Newark, DE. International Reading Association.

Appel, R., Trofimovich, P., Saito, K., Isaacs, T., Webb, S., & Appel, R. (2019). Lexical aspects of comprehensibility and nativeness from the perspective of native-speaking English raters. *ITL International Journal for Applied Linguistics*, *170*(1), 24–52. doi: doi: 10.1075/itl.17026.app

Astika, G. G. (1993). Analytical assessment of foreign students' writing. *RELC Journal, 24*(1), 61–72.

Bardel, C., Gudmundson, A., & Lindqvist, C. (2012). Aspects of Lexical Sophistication in Advanced Learners' Oral Production. *Studies in Second Language Acquisition*, *34*(02), 269–290. doi: 10.1017/S0272263112000058

Batista, R., & Horst, M. (2016). A New Receptive Vocabulary Size Test for French. *The Canadian Modern Language Review / La Revue Canadienne Des Langues Vivantes*, *72*(2), 211–233. doi: 10.3138/cmlr.2820

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

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253–279. doi: 10.1093/ijl/6.4.253

Beacco J.-C., Blin B., Houles E., Lepage S. & Riba P. (2011). *Niveau B1 pour le français*. Paris : Didier.

Beglar, D., & Nation, P. (2007). A vocabulary Size Test. *The Japan Association for Language Teaching*, *31*(7), 9–13. doi: 10.1177/0265532209340194

Bonk, W. J. (2000). Second Language Lexical Knowledge and Listening Comprehension. *International Journal of Listening*, *14*(1), 14–31. doi: 10.1080/10904018.2000.10499033

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics*, *36*(1), 1–22. doi: 10.1093/applin/amt018

Brown, A., Iwashita, N., & McNamara, T. (2005). An Examination of Rater Orientations and Test-Taker Performance on English-for-Academic-Purposes Speaking Tasks. *ETS Research Report Series, Monograph 29*, (April). doi: 10.1002/j.2333-8504.2005.tb01982.x

Bygate, M. (1999). Quality of language and purpose of task : patterns of learners ' language on two oral communication tasks. *Language Teaching Research*, *3*, 185–214. doi: 10.1177/136216889900300302

Carroll, J. (1964); *Language and thought.* Englewood Cliffs, NJ: Prentice Hall.

Cobb, T. *Compleat Web VP v.2* [computer program]. Accessed 19 Sept 2018 at https://www.lextutor.ca/vp/comp/

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: John Benjamins. doi: 10.1075/lllt.10.04cob.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Mahwah, NJ: Laurence Erlbaum Associates.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238. doi: 10.2307/3587951

Crossley, S.., & McNamara, D. (2012). Predicting second language writing proficiency : the roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, *35*(2), 115–135. doi: 10.1111/j.1467-9817.2010.01449.x

Crossley, S., & McNamara, D. (2013). Applications of text analysis tools for spoken response grading. *Language Learning & Technology*, *17*(2), 171–192. doi: 10.4039/n06-003

Crossley, S., Salsbury, T., & Mcnamara, D. (2015). Assessing Lexical Proficiency Using Analytic Ratings: A Case for Collocation Accuracy. *Applied Linguistics*, (February 2015), 1–22. doi: 10.1093/applin/amt056

Crossley, S., Salsbury, T., Mcnamara, D., & Jarvis, S. (2011). What is lexical proficiency? Some answers from computational models of speech data. *TESOL Quarterly*, *45*(1), 182–193. doi: 10.5054/tq.201

Crossley, S., Salsbury, T., McNamara, D., & Jarvis, S. (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, *28*(4), 561–580. doi: 10.1177/0265532210378031

Daller, H., Milton, J., & Treffers-Daller, J. (Eds.). (2007). *Modelling and Assessing Vocabulary Knowledge*. Cambridge University Press.

Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical Richness in the Spontaneous Speech of Bilinguals. *Applied Linguistics*, *24*(2), 197–222+267. doi: 10.1093/applin/24.2.197

Dang, T. N. Y., & Webb, S. (2014). English for Specific Purposes The lexical profile of academic spoken English. *English for Specific Purposes*, *33*, 66–76. doi: 10.1016/j.esp.2013.08.001

Dang, T. N. Y., & Webb, S. (2016). Making an essential word list for beginners. In P. Nation (Ed.), *Making and using word lists for language learning and testing.* (pp. 153–167). Amsterdam: John Benjamins.

De Clercq, M. (2016). *Tester l'étendue du vocabulaire français des étudiants néerlandophones : une étude comparative des versions monolingues et bilingues d'un test réceptif et productif* (Unpublished Master's thesis). KU Leuven: Antwerpen.

de Jong, N., Steinel, M., Florijn, A., Schoonen, R., & Hulstijn, J. (2012). Facets of Speaking Proficiency. *Studies in Second Language Acquisition*, *34*(01), 5–34. doi: 10.1017/S0272263111000489

De Keyser, Robert; Schoonen, R. (2007). Editors' announcement. *Language Learning*, *57*(1), ix–x. doi: doi: 10.1111/j.1467-9922.2007.00396_2.x

Deboer, F. (2017). *Evaluating the Relationship Between VST Score and Lexical Diversity in Written Production Among Asian Learners of English*.

Decoo, W. (2012). De Europese pogingen om tot lexicale normen te komen. *Levende Talen Magazine*, 21–23.

Derwing, T., Rossiter, M., Munro, M., & Thomson, R. (2004). Second Language Fluency: Judgements on different tasks. *Language Learning*, *54*(4), 655–679. doi: 10.1111/j.1467-9922.2004.00282.x

Elgort, I. (2013). Effects of L1 definitions and cognate status of test items on the Vocabulary Size Test. *Language Testing*, *30*(2), 253–272. doi: 10.1177/0265532212459028

Ellis, N. (1997). Vocabulary acquisition: word structure, collocation, word-class, and meaning. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: description, acquisition and pedagogy* (pp.122-139). Cambridge: Cambridge University Press.

Ellis, N. C. (2002). Frequency effects in language processing. A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 143–188. doi: 10.1017.S0272263102002024

European Commission (2012). *The European survey on language competences*.

Eyckmans, J., Van de Velde, H., van Hout, R. and Boers, F. (2007). Learners' response behavior in Yes/No vocabulary tests. In H. Daller, J. Milton and J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 59-76). Cambridge: Cambridge University Press.

Field, A. (2009). *Discovering statistics using SPSS (and sex and drugs and rock'n' roll)* (3rd ed.) SAGE.

Fulcher, G. (2004). Deluded by Artifices ? The Common European Framework and Harmonization Deluded by Artifices ? The Common European Framework and Harmonization. *Language Assessment Quarterly*, *1*(4), 253–266. doi: 10.1207/s15434311laq0104

Fulcher, G. (2013). Practical language testing. *Practical Language Testing*, 1–352. doi: 10.4324/9780203767399

Gass, S., & Mackey, A. (2007). *Data Elicitation for Second and Foreign Language Research*. Routledge.

Goodfellow, R., Lamy, M.-N., & Jones, G. (2002). Assessing learners' writing using lexical frequency. *ReCALL*, *14*(01), 133–145. doi: 10.1017/S0958344002001118

Guiraud, P. (1960). *Problèmes et méthodes de la statistique linguistique.* Dordrecht: D. Reidel.

Gyllstad, H., Granfeldt, J., Bernardini, P., & Källkvist, M. (2014). Linguistic correlates to communicative proficiency levels of the CEFR: The case of syntactic complexity in written L2 English, L3 French and L4 Italian. *EUROSLA Yearbook, 14*, 1–30. doi:10.1075/eurosla.14.01gyl

Gyllstad, H., Vilkaite, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *ITL International Journal for Applied Linguistics*, *2*(January), 276–303. doi: 10.1075/itl.166.2.04gyl

Herdan, G. (1966). *The advanced theory of language as a choice and chance.* Berlin: Springer.

Hirsh, D., & Nation, P. (1992). What Vocabulary Size is Needed to Read Unsimplified Texts for Pleasure? *Reading in a Foreign Language*.

Hu, H.-C., & Nation, P. (2000). Unknown Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language*, *13*(1), 403–430.

Hulstijn, J. (2007). The Shaky Ground beneath the CEFR : Quantitative and Qualitative Dimensions of Language Proficiency and Qualitative The Shaky Ground Beneath the CEFR : Quantitative Dimensions of Language Proficiency1. *The Modern Language Journal*, *91*(4), 663–667. doi: 10.1111/j.1540-4781.2007.00627_5.x

Hulstijn, J. (2014). The Common European Framework of Reference for Languages: A challenge for applied linguistics. *ITL - International Journal of Applied Linguistics*, *165*(1), 3–18. doi: doi: 10.1075/itl.165.1.01hul

Iwashita, N. (2010). Features of oral proficiency in task performance by EFL and JFL learners. *Selected Proceedings of the 2008 Second Language Research Forum*, (January 2010), 32–47.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, *29*(1), 24–49. doi: 10.1093/applin/amm017

Jarvis, S. (2002). Short texts , best- tting curves and new measures of lexical diversity. *Language Testing*, *19*(1), 57–84. doi: doi: 10.1191/0265532202lt220oa

Jarvis, S. (2013). Capturing the Diversity in Lexical Diversity. *Language Learning*, *63*(March), 87–106. doi: 10.1111/j.1467-9922.2012.00739.x

Johnson, W. (1944). Studies in language behavior: I. A program of research. *Psychological Monographs, 56*, 1-15.

Kim, H.-Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, *38*(1), 52. doi: 10.5395/rde.2013.38.1.52

Koizumi, R. (2005). Predicting Speaking Ability From Vocabulary Knowledge. *The Japan Language Testing Association*, *7*, 1–20.

Koizumi, R., & In'nami, Y. (2013). Vocabulary Knowledge and Speaking Proficiency among Second Language Learners from Novice to Intermediate Levels. *Journal of Language Teaching and Research*, *4*(5), 900–913. doi: 10.4304/jltr.4.5.900-913

Kremmel, B. (2016). Word Families and Frequency Bands in Vocabulary Tests: Challenging Conventions. *TESOL Quarterly*, *50*(4), 976–987. doi: 10.1002/tesq.329

Kremmel, B., & Schmitt, N. (2016). Interpreting Vocabulary Test Scores: What Do Various Item Formats Tell Us About Learners' Ability to Employ Words? *Language Assessment Quarterly*, *13*(4), 377–392. doi: 10.1080/15434303.2016.1237516

Kučera, H., & Francis, W. N. (1967). *A Computational Analysis of Present-Day American English.* Providence, Rhode Island: Brown University Press.

Kuiken, F., Vedder, I., & Gilabert, R. (2010). Communicative adequacy and linguistic complexity in L2 writing. *Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research. Eurosla Monographs 1.*, 81–100. Retrieved from http://eurosla.org/monographs/EM01/EM01tot.pdf#page=81

Kusseling, F., & Decoo, W. (2010). Europe and language learning: the challenges of comparable assessment. *34th European Studies Conference: selected proceedings,* 41-53. Omaha, Nebraska: University of Nebraska.

Kusseling, F., & Lonsdale, D. (2013). A Corpus-Based Assessment of French CEFR Lexical Content. *Canadian Modern Language Review/ La Revue Canadienne Des Langues Vivantes*, *69*(4), 436–461. doi: 10.3138/cmlr.1726.436

Kyle, K., & Crossley, S. (2015). Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application. *TESOL Quarterly*, *49*(4), 757–786. doi: 10.1002/tesq.194

Kyle, K., Crossley, S., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior Research Methods*, *50*(3), 1030–1046. doi: 10.3758/s13428-017-0924-4

Lamote, B., Desmet, P., & Janssen, R. (2014). *Frans in de balans.* Antwerpen: Garant Uitgevers.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special Language: From Humans Thinking to Thinking Machines* (pp. 316-323). Philadelphia: Multilingual Matters.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In J. L. Arnaud & H. Bejoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 126-132). London: Macmillan.

Laufer, B. (1998). The Development of Passive and Active Vocabulary in a Second Language : Same or Different ? *Applied Linguistics*, *19*(2), 255–271. doi: 10.1093/applin/19.2.255

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adapativeness. *Language Learning*, *54*(3), 399–436. doi: doi: 10.1111/j.0023-8333.2004.00260.x

Laufer, B., & Levitzky-Aviad, T. (2018). Loanword proportion in vocabulary size tests. *ITL - International Journal of Applied Linguistics*, *169*(1), 95–114. doi: 10.1075/itl.00008.lau

Laufer, B., & McLean, S. (2016). Loanwords and Vocabulary Size Test Scores: A Case of Different Estimates for Different L1 Learners. *Language Assessment Quarterly*, *13*(3), 202–217. doi: 10.1080/15434303.2016.1210611

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*, 307–322. doi: 10.1093/applin/16.3.307

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33–51. doi: 10.1177/026553229901600103

Laufer, B., & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? In S. Foster-Cohen & A. Nizegorodcew, EUROSLA yearbook (pp. 7-28). Amsterdam and Philadelphia: John Benjamins. https://doi.org/10.1075/eurosla.1.05lau

Laufer, B., & Paribakht, S. (1998). The relationship between passive and active vocabularies: effects of language learning context. *Language Learning*, *48*(3), 365–391. doi: 10.1111/0023-8333.00046

Laufer, B., & Ravenhorst-Kalovski, G. (2010). Lexical threshold revisited : Lexical text coverage , learners ' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15–30.

Lenhard, W. & Lenhard, A. (2016). *Calculation of Effect Sizes.* Dettelbach, Germany: Psychometrica.

Linck, J., & Cunnings, I. (2015). The Utility and Application of Mixed-Effects Models in Second Language Research. *Language Learning*, *65*(S1), 185–207. doi: 10.1111/lang.12117

Lindqvist, C. (2010). La richesse lexicale dans la production orale de l'apprenant avancé de français. *Canadian Modern Language Review/ La Revue Canadienne Des Langues Vivantes*, *66*(3 (March)), 393–420. doi: 10.3138/cmlr.66.3.393

Lindqvist, C., Bardel, C., & Gudmundson, A. (2011). Lexical richness in the advanced learner's oral production of French and Italian L2. *IRAL - International Review of Applied Linguistics in Language Teaching*, *49*(3), 221–240. doi: 10.1515/iral.2011.013

Little, D. (2007). The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy. *The Modern Language Journal*, *91*, 645–655. doi: 10.1111/j.1540-4781.2007. 00627_2.x

Long, M. H. and Richards, J. C. (2007). Series Editors' Preface. In: H. Daller, J. Milton, and J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press, pp xii–xiii.

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

Lorenzo-Dus, N., & Meara, P. (2005). Examiner support strategies and test-taker vocabulary. *IRAL - International Review of Applied Linguistics in Language Teachings*, *43*(3), 239–258. doi: 10.1515/iral.2005.43.3.239

Lu, X. (2012). The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives. *Modern Language Journal*, *96*(2), 190–208. doi: 10.1111/j.1540-4781.2011.01232_1.x

Maas, H. (1972). Zusammenhang zwischen Wortschatzumfang und Länge eines Textes. *Zeitschrift für Literaturwissenschaft und Linguistik, 8*, 73-79.

MacWhinney, B. (2000a). The CHILDES Project: Tools for Analyzing Talk. 3rd Edition. Part 1: The CHAT transcription format. Mahwah, NJ: Lawrence Erlbaum Associates. Doi: 10.21415/3mhn-0z89

MacWhinney, B. (2000b). The CHILDES Project: Tools for Analyzing Talk. 3rd Edition. Part 2: The CLAN program. Mahwah, NJ: Lawrence Erlbaum Associates. doi: 10.21415/T5G10R

Malvern, D. & Richards, B. (1997). A new measure of lexical diversity. In A. Ryan, & A. Wray (Eds.) *Evolving models of language*: papers from the annual meeting of the British Association for Applied Linguistics held at the University of Wales, Swansea, September 1996. Clevedon, UK: Multilingual Matters, 58–71.

Malvern, D., & Richards, B. (2002). Investigating accomodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*(1), 85–104. doi: 10.1191/0265532202lt221oa

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical Diversity and Language Development*. *Assessment*. doi: 10.1057/9780230511804

Matthews, J. (2018). Vocabulary for listening: Emerging evidence for high and mid-frequency vocabulary knowledge. *System*, *72*, 23–36. doi: 10.1016/j.system.2017.10.005

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System, 52*,1-13. doi: 10.1177/0265532216676851

McCarthy, P. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)* (Unpublished doctoral dissertation). University of Memphis, Memphis.

McCarthy, P., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing*, *24*(4), 459–488. doi: 10.1177/0265532207080767

McCarthy, P., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392. doi: 10.3758/BRM.42.2.381

McCarthy, P., Watanabe, S., & Lamkin, T. A. (2012). The Gramulator: A tool to identify differential linguistic features of correlative text types. In P. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing: Identification, investigation, and resolution* (pp. 312–333). Hershey, PA: IGI Global.

McLean, S. (2017). Evidence for the Adoption of the Flemma as an Appropriate Word Counting Unit. *Applied Linguistics*, 1–24. doi: 10.1093/applin/amw050

McLean, S., & Kramer, B. (2015). The Creation of a New Vocabulary Levels Test. *Shiken*, *19*(2), 1–32. Retrieved from http://teval.jalt.org/sites/teval.jalt.org/files/Shiken_19-02_Complete-1.pdf#page=26

Meara, P. (1990). A note on passive vocabulary. *Second Language Research*, *6*(2), 150–154. doi: 10.1177/026765839000600204

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Competence and Performance in Language Learning* (pp.35-53). Cambridge: Cambridge University Press.

Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics, 26,* 32–47. doi: 10.1093/applin/amh037

Meara, P., & Jones, G. (1990). *Eurocentres Vocabulary Size Tests 10KA*. Zurich: Eurocentres Learning Service.

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

Meara, P., & Miralpeix, I. (2016). *Tools for researching vocabulary.* Bristol, UK: Multilingual Matters.

Mecartty, F. (2000). Lexical and grammatical knowledge in reading and listening comprehension by foreign language learners of Spanish. *Applied Language Learning, 11*, 323–348.

Milton, J. (2006). Language Lite? Learning French Vocabulary in School. *Journal of French Language Studies*, *16*(02), 187. doi: 10.1017/S0959269506002420

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol, Buffalo, Toronto: Multilingual Matters. doi: 10.1080/09571736.2012.677623

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. *Eurosla Monograph Series 1*, 211–232. Retrieved from http://www.eurosla.org/monographs/EM01/EM01home.php

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. *Eurosla Monograph Series 2*, 57–78. Retrieved from http://www.eurosla.org/monographs/EM02/EM02home.php

Milton, J., & Alexiou, T. (2009). Vocabulary size and the Common European Framework of Reference for Languages. In B. Richards, H. Daller, D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary Studies in First and Second Language Acquisition: The Interface Between Theory and Application* (pp. 194–211). Palgrave Macmillan. doi: 10.1057/9780230242258

Milton, J., & Donzelli, G. (2013). 21 The lexicon. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge Handbook of Second Language Acquisition* (pp. 441–460). Cambridge University Press.

Milton, J., & Hopkins, N. (2005). *Aural lex.* Swansea, UK: Swansea University.

Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In M. D. M. Chacón-Beltrán, R.; Abello-Contesse, C.; Torreblanca-Lopéz (Ed.), *Insights into non-native vocabulary teaching and learning.* (pp. 83–98). Bristol, U.K.: Multilingual Matters.

Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *IRAL - International Review of Applied Linguistics in Language Teaching*, *56*(1), 1–24. doi: 10.1515/iral-2017-0016

Mochida, K., & Harrington, M. (2006). The yes/no test as a measure of receptive vocabulary knowledge. *Language Testing, 23*,73–98. doi:10.1191/02655322 06lt321oa

Nation, P. (1983). Testing and teaching vocabulary. *Guidelines, 5*(1), 12-25. Retrieved from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/1983-Testing-and-teaching.pdf

Nation, P. (1990). *Teaching and learning vocabulary.* Rowley, Mass: Newbury House.

Nation, P. (2001). *Learning Vocabulary in Another Language.*Cambridge: Cambridge University Press. doi: 10.1093/elt/56.1.91

Nation, P. (2006). How Large a Vocabulary Is Needed for Reading and Listening? *The Canadian Modern Language Review / La Revue Canadienne Des Langues Vivantes*, *63*(1), 59–81. doi: 10.1353/cml.2006.0049

Nation, P. (2012). *The BNC/COCA word family lists.* Retrieved from https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf

Nation, P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139524759

Nation, P. (2016). *Making and using word lists for language learning and testing.* Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/z.208

Nguyen, L. T. C., & Nation, P. (2011). A Bilingual Vocabulary Size Test of English for Vietnamese Learners. *RELC Journal*, *42*(1), 86–99. doi: 10.1177/0033688210390264

Noreillie, A., Kestemont, B., Heylen, K., Desmet, P., & Peters, E. (2018). Vocabulary knowledge and listening comprehension at an intermediate level in English and French as foreign languages An approximate replication study of Stæhr (2009). *ITL - International Journal of Applied Linguistics*, *169*(1), 212–231. doi: doi: 10.1075/itl.00013.nor

Ovtcharov, V., Cobb, T., & Halter, R. (2006). La richesse lexicale des productions orales : mesure fiable du niveau de compétence langagière. *Canadian Modern Language Review/ La Revue Canadienne Des Langues Vivantes*, *63*, 107–125. doi: 10.3138/cmlr.63.1.107

Pauwels, P. (2016, July). Lexical development in/through writing. Paper presented at the Vocabulary Special Interest Group of the British Association for Applied Linguistics, Nottingham, UK.

Peters, E., Noreillie, A.S., Heylen, K., Bulté, B., & Desmet P. (2019). The impact of instruction and out-of-school exposure to FL input on learners' vocabulary knowledge in two languages. *Language Learning*. doi; 10.1111/lang.12351

Peters, E., Velghe, T., & Rompaey, T. Van. (2019). The VocabLab tests The development of an English and French vocabulary test. *ITL - International Journal of Applied Linguistics*, *170*(1), 53–78. doi: 10.1075/itl.17029.pet

Plonsky, L. (2015). *Advancing quantitative methods in second language research*. Retrieved from https://books.google.com/books?id=jA0XCgAAQBAJ&pgis=1

Plonsky, L., & Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. *Modern Language Journal*, 100(2), 538–553. doi: 10.1111/modl.12335

Porte, G. (2012). Introduction. In G. Porte (Ed.), *Replication research in applied linguistics* (pp.1–17). Cambridge: Cambridge University Press.

Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review/ La Revue Canadienne Des Langues Vivantes*, *56*(2), 282–307. doi: 10.3138/cmlr.56.2.282

Qian, D. (2000). Validating the role of depth of vocabulary knowledge in assessing reading for basic comprehension [TOEFL 2000 Research Report]. Princeton, NJ: Educational Testing Service.

Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513-536. doi: 10.1111/1467-9922.00193

Read, J . (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal, 19*, 12 –25.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, *10*(3), 355–371. doi: 10.1177/026553229301000308

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

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined?. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language. Selection, acquisition and testing* (pp. 209-227). Amsterdam/Philadelphia: Benjamins.

Revesz, A., Ekiert, M., & Torgersen, E. N. (2016). The Effects of Complexity, Accuracy, and Fluency on Communicative Adequacy in Oral Task Performance. *Applied Linguistics*, *37*(6), 828–848. doi: 10.1093/applin/amu069

Richards, B., & Malvern, D. (2000, September). *Measuring vocabulary richness in teenage learners of French* . Paper presented at the Britisch Educational Research Association Conference, Cardiff University.

Richards, J., & Schmidt, R. (2002). *Longman Dictionary of language teaching and applied linguistics* (3rd ed.). London: Taylor & Francis Ltd.

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

Schmitt, N. (2008). *Review article: Instructed second language vocabulary learning*. *Language Teaching Research* (Vol. 12). doi: 10.1177/1362168808089921

Schmitt, N. (2010). *Researching vocabulary: a vocabulary research manual*. *Language Teaching*. doi: 10.1017/CBO9781107415324.004

Schmitt, N. (2014). Size and Depth of Vocabulary Knowledge: What the Research Shows. *Language Learning*, *64*(4), 913–951. doi: 10.1111/lang.12077

Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2017). How much vocabulary is needed to use English ? Replication of. *Language Teaching*, *50*(2), 212–226. doi: 10.1017/S0261444815000075

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. doi: 10.1111/j.1540-4781.2011.01146.x

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, *47*(4), 484–503. doi: 10.1017/S0261444812000018

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. doi: 10.1191/026553201668475857

Schmitt, N., & Zimmerman, C. (2002). Derivative word forms: What do learners know?. *TESOL Quarterly*, *36*(2), 145-171. doi: 10.2307/3588328

Schoonen, R., van Gelderen, A., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First Language and Second Language Writing: The Role of Linguistic Knowledge, Speed of Processing, and Metacognitive Knowledge. *Language Learning*, *53*(1), 165–202. doi: 10.1111/1467-9922.00213

Schoonen, R., Van Gelderen, A., Stoel, R. D., Hulstijn, J., & De Glopper, K. (2011). Modeling the Development of L1 and EFL Writing Proficiency of Secondary School Students. *Language Learning*, *61*(1), 31–79. doi: 10.1111/j.1467-9922.2010.00590.x

Segers, J. (2015). *Tester le vocabulaire français des élèves flamands. Une analyse de deux variantes d'un test réceptif* (Unpublished Bachelor's paper). KU Leuven: Antwerpen.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, *36*(2), 139–152. doi: 10.1080/09571730802389975

Stæhr, L. S. (2009). *Vocabulary Knowledge and Advanced Listening Comprehension in English As a Foreign Language*. *Studies in Second Language Acquisition* (Vol. 31). doi: 10.1017/S0272263109990039

Teng, F. (2014). Assessing the depth and breadth of vocabulary knowledge with listening comprehension. *Passa*, *48*(2), 29–56.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's wordbook of 30,000 words.* New York: Teachers College, Columbia University.

Tidball, F., & Treffers-Daller, J. (2008). Analysing lexical richness in French learner language: What frequency lists and teacher judgements can tell us about basic and advanced words. *Journal of French Language Studies*, *18*(Special Issue 03), 299–313. doi: 10.1017/S0959269508003463

Treffers-Daller, J. (2013). Measuring lexical diversity among L2 learners of French. An exploration of the validity of D, MTLD and HD-D as measures of language ability. In S. Jarvis & M. Daller (Eds.), *Vocabulary knowledge: human ratings and automated measures* (pp.79–103). Philadelphia/Amsterdam: John Benjamins.

Treffers-Daller, J., Daller, M., Malvern, D., Richards, B., Meara, P., & Milton, J. (2008). Introduction: Special issue on knowledge and use of the lexicon in French as a second language. *Journal of French Language Studies*, *18*(03), 269–276. doi: 10.1017/s095926950800344x

Treffers-Daller, J., Parslow, P., & Williams, S. (2016). Back to Basics: How Measures of Lexical Diversity Can Help Discriminate between CEFR Levels. *Applied Linguistics*, 1–27. doi: 10.1093/applin/amw009

Uchihara, T., & Clenton, J. (2018). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research.* Advance online publication. doi: 10.1177/1362168818799371

Uchihara, T., & Saito, K. (2016). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *Language Learning Journal*, *44*(3), 1–15. doi: 10.1080/09571736.2016.1191527

van Zeeland, H., & Schmitt, N. (2013). Lexical Coverage in L1 and L2 Listening Comprehension: The Same or Different from Reading Comprehension? *Applied Linguistics*, *34*(4), 457–479. doi: 10.1093/applin/ams074

Verlinde, S., & Selva, T. (2001). Corpus-based versus intuition-based lexicography: defining a word list for a French learners' dictionary. *Proceedings of the Corpus Linguistics 2001 Conference*, 594–598. Retrieved from http://www.kuleuven.ac.be/grelep/publicat/verlinde.pdf

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, *17*(1), 65–83. doi: 10.1191/026553200676636328

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1 / L2 acquisition and frequency of input. *Applied Psycholinguistics*, *22*, 217–234.

Vlaamse Overheid. (2013). Peiling Frans luisteren en spreken.

Vrancx, S. (2016). *Franse woordenschattoetsen : een vergelijking tussen een eentalige en een tweetalige versie* (Unpublished Master's thesis). KU Leuven: Antwerpen.

Wang, S. (2015). An Empirical Study on the Role of Vocabulary Knowledge in EFL Listening Comprehension. *Theory and Practice in Language Studies*, *5*(5), 989–995. doi: http://dx.doi.org/10.17507/tpls.0505.14

Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System*, *65*, 139–150. doi: 10.1016/j.system.2016.12.013

Waring, R. (1997). A comparison of the receptive and productive vocabulary sizes of some second language learners. *Immaculata (Notre Dame Seishin University, Okayama), 1*, 53-68.

Webb, S. (2008). Receptive and Productive Vocabulary Sizes of L2 Learners. *Studies in Second Language Acquisition*, *30*(01), 79–95. doi: 10.1017/S0272263108080042

Webb, S., & Rodgers, H. (2009a). Vocabulary Demands of Television Programs. *Language Learning*, *59*(2), 335–366. doi: 10.1111/j.1467-9922.2009.00509.x

Webb, S., & Rodgers, M. P. H. (2009b). The Lexical Coverage of Movies. *Applied Linguistics*, *30*(3), 407–427. doi: 10.1093/applin/amp010

Webb, S., & Sasao, Y. (2013). New directions in vocabulary testing. *RELC Journal*, *44*(3), 263–277. doi: 10.1177/0033688213500582

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test. *ITL - International Journal of Applied Linguistics*, *168*(1), 33–69. doi: 10.1075/itl.168.1.02web

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

Wisniewski, K. (2017). Empirical Learner Language and the Levels of the Common European Framework. *Language Learning*, *67*(S1), 232–253. doi: 10.1111/lang.12223

Wu, T. (1993). An accurate computation of the hypergeometric distribution function. *ACM Trans. Math. Softw.*, *19*(1), 33–43. doi: http://doi.acm.org/10.1145/151271.151274

Xue, G., & Nation, P. (1984). A university word list. *Language Learning and Communication, 3*(2), 215–229.

Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied Linguistics, 31*(2), 236–259. doi: 10.1093/applin/amp024

Appendices

Appendix 1 Global scale of the proficiency levels (Council of Europe, 2001, p. 24)

	C2	Can understand with ease virtually everything heard or read. Can summarize information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
Proficient user	C1	Can understand a wide range of demanding, longer texts, and recognize implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organizational patterns, connectors and cohesive devices.
	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialization. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
Independent user	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.

Basic user	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g., very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Appendix 2 French DELF listening comprehension test (Study 1)

Partie 1

COMPRÉHENSION DE L'ORAL

25 points

■

■ Consignes

Vous allez entendre trois documents sonores, correspondant à des situations différentes.
Pour le premier et le deuxième document, vous aurez :
- 30 secondes pour lire les questions ;
- une première écoute, puis 30 secondes de pause pour commencer à répondre aux questions ;
- une deuxième écoute, puis 1 minute de pause pour compléter vos réponses.
Répondez aux questions, en cochant (☒) la bonne réponse, ou en écrivant l'information demandée.

6 points

Document 1

1. Quel est le problème de la mère ? *1 point*
   - ☐ Elle a une très grave maladie.
   - ☐ Elle oublie beaucoup de choses.
   - ☐ Elle se sent très seule.

2. La mère est étonnée parce que sa fille : *1 point*
   - ☐ est rentrée de voyage.
   - ☐ lui apporte des médicaments.
   - ☐ a préparé un gâteau.

3. Qu'a fait la mère pour l'anniversaire de Philippe ? *1 point*
   - ☐ Un gâteau.
   - ☐ Un repas.
   - ☐ Rien.

4. La mère est allée à l'anniversaire de Philippe : *1 point*
   - ☐ Vrai
   - ☐ Faux
   - ☐ On ne sait pas

5. Quel âge a la mère ? *1 point*
   - ☐ 65 ans
   - ☐ 57 ans
   - ☐ 75 ans

6. La mère va partir une semaine avec sa fille : *1 point*
   ☐ Vrai ☐ Faux ☐ On ne sait pas

Document 2

6 points

1. Ce document est : *1 point*
   - ▯ une publicité.
   - ▯ un témoignage.
   - ▯ une annonce.

1. Cet homme a des problèmes pour : *1 point*
   - ▯ écrire.
   - ▯ parler.
   - ▯ chanter.

2. Il se souvient des numéros de téléphone de : *1 point*
   - ▯ tous ses amis.
   - ▯ tous ses enfants.
   - ▯ toute sa famille.

3. Il a des souvenirs précis des événements : *1 point*
   - ▯ très récents.
   - ▯ très anciens.
   - ▯ anciens et récents.

5. Que pense-t-il de cette faculté ? *1 point*
   - ▯ Il trouve cela très pratique.
   - ▯ Il trouve cela plutôt gênant.
   - ▯ Il en est fier.

6. Sa « maladie » peut être soignée. *1 point*
   ▯ Vrai ▯ Faux ▯ On ne sait pas

Document 3

13 points

Vous allez entendre un document sonore. Vous aurez tout d'abord 1 minute pour lire les questions, puis vous entendrez deux fois l'enregistrement avec une pause de 3 minutes entre les deux écoutes. Après la deuxième écoute, vous aurez encore 2 minutes pour compléter vos réponses.
Répondez aux questions en cochant la bonne réponse, ou en écrivant l'information demandée.

1. Quel est le sujet de l'interview ? *1 point*
   - ▯ Le théâtre
   - ▯ La mémoire
   - ▯ Le cirque

2. Sur scène l'acteur interprète : *1 point*
   - ▯ des dizaines de personnages.
   - ▯ un seul personnage.
   - ▯ des centaines de personnages.

3. Quelle est la durée de ce spectacle ? *1 point*
   - ▯ Une heure.
   - ▯ Deux heures.
   - ▯ Trois heures.

4. Il se souvient de ses textes : *1 point*
   - ▯ grâce à son corps.
   - ▯ sans travailler.
   - ▯ en dansant

5. Quel auteur est cité ? *1 point*
   - ☐ Molière
   - ☐ Anouilh
   - ☐ Proust

6. Pour se souvenir de ses textes, il doit absolument : *1 point*
   - ☐ les entendre.
   - ☐ les visualiser.
   - ☐ les lire.

7. A quoi compare-t-il la mémoire ? (donnez au moins un exemple) *2 points*

   .....

8. Selon l'acteur, quel est le principal ennemi de la mémoire? *1 point*

   .....

9. Peut-il revenir en arrière lorsqu'il dit un texte ? *1 point*

   ☐ Oui ☐ Non ☐ On ne sait pas

10. Qu'arrive-t-il au dresseur qui a peur ? *1 point*

    .....

11. Qu'arrive-t-il à l'acteur qui a peur ? *2 points*

    ....

Appendix 3 French monolingual vocabulary knowledge test (Study 1)

1. **certain**
0 ne plus en danger
0 pas encore employé
0 ce qui est assuré, vrai
0 préparé à faire quelque chose
0 je ne connais pas la réponse

2. **mort**
0 qui ne vit plus
0 qui dit ce qu'il pense
0 qui ne perd pas son calme
0 qui ne prend pas de risques
0 je ne connais pas la réponse

3. **empêcher**
0 rendre impossible
0 disparaître sous l'eau
0 donner la faute à quelqu'un
0 ne pas donner le résultat espéré
0 je ne connais pas la réponse

4. **peuple (m)**
0 situation où il n'y a pas de guerre
0 groupe de personnes qui vivent ensemble
0 développement vers une situation meilleure
0 volonté de faire ou de voir quelque chose se produire
0 je ne connais pas la réponse

5. **beau**
0 qui a peu d'argent
0 ce que l'œil apprécie
0 pas habituel, anormal
0 avec une influence importante
0 je ne connais pas la réponse

6. **quitter**
0 être entraîné vers le sol
0 avoir droit à quelque chose
0 laisser quelqu'un en partant
0 se composer de plusieurs parties
0 je ne connais pas la réponse

7. **nuit (f)**
0 période de l'année
0 partie d'un jour sans lumière
0 le fait de se passer après le moment prévu
0 la force, l'importance de quelqu'un ou de quelque chose
0 je ne connais pas la réponse

8. **lutte (f)**
0 combat, conflit
0 sentiment triste
0 le fait de tuer volontairement une personne
0 quelque chose que l'on veut réaliser mais qui est difficile
0 je ne connais pas la réponse

9. **accès (m)**
0 sentiment de désir
0 la façon dont quelqu'un réagit
0 la possibilité d'entrer quelque part
0 période donnée à quelqu'un pour faire quelque chose
0 je ne connais pas la réponse

10. **avantage (m)**
0 travail qu'il faut faire
0 petit texte qu'une personne envoie à une autre
0 le fait de donner et de recevoir ensuite quelque chose
0 ce qui donne plus de chances de réaliser quelque chose
0 je ne connais pas la réponse

11. **environnement (m)**
0 nombre de kilos
0 partie inférieure
0 espace naturel où vit l'homme
0 proportion exprimée en pourcentage

0 je ne connais pas la réponse

12. **effectuer**
0 créer une idée, un objet
0 rassembler quelque chose
0 faire une action, une activité
0 faire mieux que quelqu'un d'autre
0 je ne connais pas la réponse

13. **lourd**
0 qui a beaucoup de poids
0 qui est dur pour les autres
0 qui ressemble très bien à quelque chose
0 qui regroupe beaucoup d'éléments, grand
0 je ne connais pas la réponse

14. **lumière (f)**
0 ce que donne le soleil ou une lampe
0 grand nombre de personnes à un lieu
0 situation de quelqu'un qui ne travaille pas
0 condition de l'eau qui permet de faire du ski
0 je ne connais pas la réponse

15. **sang (m)**
0 fait de tomber
0 grande masse d'eau
0 liquide rouge dans le corps
0 activité de prendre des poissons
0 je ne connais pas la réponse

16. **prétendre**
0 affirmer, dire
0 donner à quelqu'un tout ce qu'il faut
0 demander de bien garder quelque chose
0 être sûr que quelque chose de négatif va se produire
0 je ne connais pas la réponse

17. **promesse (f)**
0 problème de santé
0 engagement de faire quelque chose
0 fait de mettre brusquement fin à quelque chose
0 ce que l'on peut utiliser pour améliorer une situation
0 je ne connais pas la réponse

18. **rare**
0 capable de quelque chose
0 avec une dimension limitée
0 qu'on ne trouve pas beaucoup
0 qui ne change pas d'idées, qui tient à quelque chose
0 je ne connais pas la réponse

19. **lendemain (m)**
0 le jour après
0 ligne sur une surface
0 où deux lignes se rencontrent
0 action de placer dans une nouvelle situation
0 je ne connais pas la réponse

20. **ciel (m)**
0 ensemble de maisons
0 route pour aller d'un lieu à un autre
0 espace énorme dans lequel se trouve le soleil
0 ensemble des déplacements des moyens de transport
0 je ne connais pas la réponse

21. **suspendre**
0 rendre inutilisable
0 exploser en morceaux
0 arrêter provisoirement
0 ne pas savoir quelque chose
0 je ne connais pas la réponse

22. **épreuve (f)**
0 situation difficile
0 marque de respect
0 ensemble de flammes
0 manière d'exprimer l'imagination
0 je ne connais pas la réponse

23. **écrivain (m)**
0 responsable politique d'une ville
0 personne qui fait de la littérature

0 où arrivent et partent les bateaux
0 façon selon laquelle quelque chose peut se faire
0 je ne connais pas la réponse

24. **boîte (f)**
0 meuble pour s'asseoir
0 programme radio ou télé
0 somme d'argent qu'il faut payer à l'état
0 objet dans lequel on peut mettre quelque chose
0 je ne connais pas la réponse

25. **veiller**
0 faire durer
0 utiliser pour justifier une action
0 être confronté à une nouvelle situation
0 rester debout la nuit pour s'occuper de quelqu'un
0 je ne connais pas la réponse

26. **goût (m)**
0 morceau long et fin
0 période de cent ans
0 terre entourée d'eau
0 impression laissée dans la bouche
0 je ne connais pas la réponse

27. **oreille (f)**
0 partie arrière du corps
0 partie du corps pour entendre
0 partie du corps qui permet de marcher
0 partie du membre supérieur du corps humain
0 je ne connais pas la réponse

28. **poche (f)**
0 le fait de gagner
0 le fait de ne pas avoir réussi
0 ouverture dans un pantalon où on peut mettre quelque chose
0 quelqu'un qui subit les conséquences négatives de quelque chose
0 je ne connais pas la réponse

29. **cerveau (m)**
0 le fait de ne pas être là
0 action de commencer le combat
0 somme d'argent qu'on reçoit pour son travail
0 organe qui gère les mouvements et les pensées
0 je ne connais pas la réponse

30. **inquiétant**
0 très connu
0 qui cause de la crainte
0 avec un rapport aux paysans
0 qui a des instruments pour lutter
0 je ne connais pas la réponse

31. **séjour (m)**
0 le fait de résider temporairement quelque part
0 le fait d'abandonner ce à quoi on tient beaucoup
0 le fait que l'on choisit des représentants politiques
0 le fait d'avoir perdu un match, une compétition, une guerre
0 je ne connais pas la réponse

32. **censé**
0 qui est supposé
0 qui se produit tout à coup
0 qui vient avant quelque chose d'autre
0 qui ne correspond pas à ce qu'on attendait
0 je ne connais pas la réponse

33. **fuite (f)**
0 action de se tuer soi-même
0 action de s'écarter d'un danger
0 action de courir après quelqu'un
0 action de donner à plusieurs personnes
0 je ne connais pas la réponse

34. **vœu (m)**
0 sentiment d'être trompé dans son attente
0 grand désir de voir quelque chose se réaliser

0 sentiment d'une personne qui est extrêmement en colère contre quelqu'un
0 le fait de poser une série de questions à un groupe pour connaître leur opinion
0 je ne connais pas la réponse

35. b**ranche (f)**
0 partie d'un arbre
0 ce qui a été mis ensemble
0 surface qui permet de visualiser des images
0 pièce élastique qui reprend sa forme d'origine après avoir été contractée
0 je ne connais pas la réponse

36. **songer**
0 faire pénétrer dans quelque chose
0 bouger dans tous les sens plusieurs fois
0 garder à peine le contact avec une surface
0 se représenter quelque chose dans son esprit
0 je ne connais pas la réponse

37. **enceinte**
0 agréable à voir
0 qui n'entend rien
0 agréable à toucher
0 qui attend un enfant
0 je ne connais pas la réponse

38. **contrainte (f)**
0 caractère menaçant
0 ce qui limite ce que l'on peut faire
0 fait d'être le propriétaire de quelque chose
0 mesure que l'on prend pour éviter un danger ou quelque chose de désagréable
0 je ne connais pas la réponse

39. **répandre**
0 partager
0 diminuer la vitesse
0 apparaître par surprise
0 mettre partout autour de soi
0 je ne connais pas la réponse

40. **huile (f)**
0 ce qui coule des yeux
0 de l'eau qui tombe du ciel
0 liquide blanc donné par la vache
0 liquide tiré de plantes, de pétrole
0 je ne connais pas la réponse

41. **guide (m)**
0 personne à qui l'on est opposé
0 personne qui montre le chemin
0 personne qui crée une organisation, une publication
0 quelqu'un qui paie une autre personne pour travailler pour lui
0 je ne connais pas la réponse

42. **creuser**
0 mettre des graines dans la terre
0 placer sous une pierre au cimetière
0 réaliser un trou, un tunnel dans la terre
0 déformer quelque chose en le comprimant
0 je ne connais pas la réponse

43. **couche (f)**
0 petit élément d'un corps
0 petite hauteur de terrain arrondie
0 somme payée, pour devenir membre
0 quantité d'une matière d'une certaine épaisseur
0 je ne connais pas la réponse

44. **accueil (m)**
0 vue sur la nature
0 raison qui cache la vraie raison
0 ce que l'on voit de quelque chose
0 le fait de recevoir une personne qui vient vous visiter
0 je ne connais pas la réponse

45. **lent**
0 pas très vite
0 stupide, imbécile
0 sans eau ni autre liquide
0 le fait de mériter du respect
0 je ne connais pas la réponse

46. **gare (f)**
0 où s'arrêtent les trains
0 lieu vers lequel on se rend
0 passage par où on peut sortir
0 grand meuble de présentation
0 je ne connais pas la réponse

47. **provenance (f)**
0 couple qui se sépare
0 chose difficile réussie
0 ensemble coordonné de mesures
0 origine d'un objet, source d'information
0 je ne connais pas la réponse

48. **similaire**
0 où il n'y a pas de violence
0 partagé par plusieurs personnes
0 associé à quelqu'un d'autre pour l'aider
0 avec à peu près les mêmes caractéristiques
0 je ne connais pas la réponse

49. **ravir**
0 obliger, forcer
0 plaire beaucoup
0 consacrer à quelque chose
0 mettre à la disposition de quelqu'un
0 je ne connais pas la réponse

50. **acquisition (f)**
0 ce que l'on a obtenu
0 personne qui dirige une entreprise
0 le fait que des personnes se comprennent bien
0 indication d'un nombre important de personnes, de choses
0 je ne connais pas la réponse

51. **distinct**
0 différent
0 qui est déterminé
0 imposé par un règlement
0 qui est basé sur de bons arguments
0 je ne connais pas la réponse

52. **couloir (m)**
0 le repas du midi
0 avec les couleurs d'un pays
0 espace avec des plantes, près d'une maison
0 passage pour aller d'une chambre à une autre
0 je ne connais pas la réponse

53. **étoile (f)**
0 grande plante
0 morceau fin coupé
0 objet au ciel qui émet de la lumière
0 matière dure et transparente où la lumière peut passer
0 je ne sais pas la réponse

54. **trahir**
0 rendre plus pénible
0 rendre quelqu'un fatigué
0 ne pas être fidèle à quelqu'un ou à quelque chose
0 croire qu'une personne a fait quelque chose qu'elle ne pouvait pas faire
0 je ne connais pas la réponse

55. **temporaire**
0 non définitif
0 qui rapporte à la ville
0 qui appartient à une commune
0 dit de quelqu'un qui regarde bien ce qui se passe
0 je ne connais pas la réponse

56. **robe (f)**
0 vent très fort
0 vêtement de femme
0 grand feu non contrôlé
0 inquiétude, émotion dans laquelle se trouve quelqu'un
0 je ne connais pas la réponse

57. **toit (m)**
0 partie supérieure d'un bâtiment
0 la façon dont se répand le son, la lumière
0 modification du déroulement de quelque chose
0 tache sombre produite par un objet, une personne
0 je ne connais pas la réponse

58. **envoi (m)**
0 ce qu'on risque de perdre
0 la moitié du jour qui est passée
0 action de faire arriver une lettre
0 personne qui regarde quelque chose (un film, un concert)
0 je ne connais pas la réponse

59. **deviner**
0 accourir rapidement
0 faire tout pour réussir quelque chose
0 obtenir quelque chose pour quelqu'un
0 arriver à trouver quelque chose en posant des questions
0 je ne connais pas la réponse

60. **sagesse (f)**
0 souffrance physique ou morale
0 ce qui recouvre le corps d'une personne
0 l'importance, l'intensité de quelque chose
0 caractéristique d'une personne qui agit intelligemment
0 je ne connais pas la réponse

61. **déroulement (m)**
0 maison
0 infraction, crime
0 développement d'une action
0 petite quantité peu importante
0 je ne connais pas la réponse

62. **ancêtre (m)**
0 type de métal
0 système de vote
0 personne qui appartient à la même profession
0 un des grands-parents, des arrière-grands-parents
0 je ne connais pas la réponse

63. **hostile**
0 de l'ennemi
0 qui a du plaisir à causer de la douleur
0 qui donne suffisamment de garanties
0 que l'on peut facilement obtenir, comprendre
0 je ne connais pas la réponse

64. **trésor (m)**
0 investissement
0 argent qui reste
0 ce qui représente une grande valeur
0 somme d'argent qu'il faut payer si on n'a pas respecté la loi
0 je ne connais pas la réponse

65. **mépris (m)**
0 désaccord
0 opinion que l'on se fait d'avance
0 sentiment que le comportement de quelqu'un n'a aucune valeur
0 avoir de la compréhension pour la souffrance, les peines de quelqu'un
0 je ne connais pas la réponse

66. **consécutif**
0 plus tard
0 qui peut être blessé
0 qui suit immédiatement
0 qui exprime un ordre absolu
0 je ne connais pas la réponse

67. **coutume (f)**
0 usage établi, habitude
0 terrain où l'on enterre les morts
0 importance d'une action, personne

0 date à laquelle il faut payer quelque chose
0 je ne connais pas la réponse

68. **abriter**
0 répandre, étendre
0 adapter parfaitement quelque chose
0 faire disparaître un sentiment de malaise
0 protéger quelqu'un ou quelque chose en le mettant en sécurité
0 je ne connais pas la réponse

69. **queue (f)**
0 file de personnes
0 avec qui on a de bons rapports
0 attaque inattendue et violente
0 rapport sexuel imposé à quelqu'un sans sa permission
0 je ne connais pas la réponse

70. **chantier (m)**
0 structure de base
0 de l'eau entourée de terre de tous côtés
0 où l'on attend le train, le métro, un bateau
0 terrain où l'on fait des travaux ou de construction
0 je ne connais pas la réponse

71. **inconvénient (m)**
0 élément avantageux
0 aspect négatif qui dérange
0 action de rendre de l'argent
0 point faible dans une argumentation, dans un comportement de quelqu'un
0 je ne connais pas la réponse

72. **emprunt (m)**
0 demande officielle
0 destruction, dommages
0 information visible sur par exemple une étiquette
0 somme d'argent accordée qu'il faut rendre par après
0 je ne connais pas la réponse

73. **hâte (f)**
0 mouvement de la mer
0 état de celui qui est pressé
0 perte matérielle ou morale
0 fort sentiment d'inquiétude
0 je ne connais pas la réponse

74. **ennuyer**
0 faire cesser le feu
0 rassembler la poussière
0 craindre vivement quelque chose
0 causer du souci ou un sentiment de fatigue
0 je ne connais pas la réponse

75. **creux**
0 peu de largeur
0 sans vêtements
0 vide à l'intérieur
0 d'aspect massif, qui ne coule pas
0 je ne connais pas la réponse

76. **complice**
0 mêlé à une affaire
0 qui n'existait pas avant
0 qui convient dans les circonstances actuelles
0 se dit des choses qui viennent l'une après l'autre
0 je ne connais pas la réponse

77. **raide**
0 peu de volume
0 très droit, sans souplesse
0 avec un sentiment de gêne
0 dit d'une personne qui possède des terres
0 je ne connais pas la réponse

78. **singulier**
0 terrible, grave
0 sur le point de se produire
0 mauvais, peu sympathique
0 bizarre, étrange, inhabituel
0 je ne connais pas la réponse

79. **approprié**
0 certain, sûr
0 qui est ajouté
0 adapté, qui convient

0 juste, qui ne favorise ni défavorise personne
0 je ne connais pas la réponse

80. **ange (m)**
0 désaccord, dispute assez violente
0 reconnaître à quelqu'un ses fautes
0 être céleste entre Dieu et l'homme
0 symbole géométrique du christianisme
0 je ne connais pas la réponse

81. **pâte (f)**
0 très petite particule de matière
0 produit alimentaire à base de farine
0 grosse masse de fines gouttes d'eau
0 matière fine qui recouvre certaines plages
0 je ne connais pas la réponse

82. **logiciel (m)**
0 programme d'ordinateur
0 objet sur lequel on met des aliments
0 objet rond, en forme de cercle sur lequel roule une voiture
0 quelque chose d'embêtant, de lourd que l'on doit supporter
0 je ne connais pas la réponse

83. **couteau (m)**
0 petit enfant
0 mauvais rêve
0 instrument pour couper
0 un groupe de personnes
0 je ne connais pas la réponse

84. **décéder**
0 mourir
0 encourager
0 faire un trou dans une surface
0 avoir de l'horreur pour quelqu'un ou quelque chose
0 je ne connais pas la réponse

85. **prospérité (f)**
0 fait d'être en prison
0 heureux état, heureuse situation
0 situation d'une personne qui dort
0 petit groupe de personnes qui font quelque chose ensemble
0 je ne connais pas la réponse

86. **poing (m)**
0 main fermée
0 partie centrale du corps
0 partie latérale du visage
0 partie du corps qui relie la tête au reste du corps
0 je ne connais pas la réponse

87. **rallier**
0 continuer à exister
0 arriver quelque part
0 regrouper, rassembler
0 au moyen d'une machine, mettre un dessin, un texte sur du papier
0 je ne connais pas la réponse

88. **défaillance (f)**
0 vêtement
0 petit texte littéraire en vers
0 le fait d'arrêter de fonctionner
0 visage d'une personne ou d'un animal
0 je ne connais pas la réponse

89. **indemnisation (f)**
0 colère très forte
0 animal qui miaule
0 ensemble des personnes qui produisent
0 le fait de compenser quelqu'un de ses pertes, de ses frais
0 je ne connais pas la réponse

90. **harceler**
0 se presser
0 prendre et réunir
0 soumettre quelqu'un à de petites attaques, provoquer
0 se soumettre à la volonté de quelqu'un, à un règlement
0 je ne connais pas la réponse

91. **s'envoler**
0 se dit d'un avion qui part
0 couper les cheveux près de la peau
0 mettre ses espoirs dans quelque chose
0 mal utiliser quelque chose et ne pas en profiter
0 je ne connais pas la réponse

92. **muet**
0 qui ne peut pas parler
0 extrêmement modeste
0 qui n'est pas dur, qui est souple
0 qui excite la haine, qui est méchant
0 je ne connais pas la réponse

93. **incidence (f)**
0 amélioration d'une situation
0 le fait d'avoir des réserves, des doutes
0 conséquence plus ou moins directe de quelque chose
0 attitude d'une personne qui ne s'occupe pas de ce qui se passe autour d'elle
0 je ne connais pas la réponse

94. **omettre**
0 oublier de faire quelque chose
0 passer un fer chaud sur du linge
0 retirer un objet d'où il se trouve
0 maintenir une action dans des limites
0 je ne connais pas la réponse

95. **montre (f)**
0 règle que l'on doit respecter
0 discours très positif sur quelqu'un
0 développement de quelque chose
0 petit appareil servant à donner l'heure
0 je ne connais pas la réponse

96. **distorsion (f)**
0 indication de la hauteur
0 condition médicale, blessure des muscles
0 action de faire le bien, de donner aux pauvres
0 l'eau de la rivière qui coule sur les champs, sur les terrains
0 je ne connais pas la réponse

97. **gai**
0 bref, court, sans détails
0 que l'on fait sans peine, facile
0 qui prête à rire, qui est amusant
0 d'une lumière faible, très blanche, en parlant du visage
0 je ne connais pas la réponse

98. **mordre**
0 préparer un aliment
0 rendre immobile, invariable
0 changer tout à coup profondément
0 serrer quelque chose entre les dents
0 je ne connais pas la réponse

99. **vigoureux**
0 courageux
0 dur, brutal
0 plein de santé
0 qui n'est pas ce qu'il paraît être
0 je ne connais pas la réponse

100. **bannir**
0 dire à quelqu'un de quitter un lieu
0 rendre quelqu'un heureux, comblé
0 rendre quelque chose moins stable, affaiblir
0 être imposé à quelqu'un en tant que charge
0 je ne connais pas la réponse

101. **parenthèse (f)**
0 remarque accessoire
0 trace que laisse un bateau derrière lui
0 trace laissée par une substance, saleté

0 état de celui qui est sous le pouvoir d'une autorité contre laquelle il a lutté
0 je ne connais pas la réponse

102. **avalanche (f)**
0 sorte de fil qui pousse sur le corps
0 grande masse de neige qui tombe des montagnes
0 plate-forme dans un escalier au niveau des étages
0 temps violent avec du tonnerre, de la pluie, des éclairs
0 je ne connais pas la réponse

103. **jet (m)**
0 grand sac
0 fait d'admettre un échec
0 action de lancer quelque chose
0 objet dont on se sert pour encourager les chevaux
0 je ne connais pas la réponse

104. **barreau (m)**
0 tumulte dans la foule
0 barre qui sert de fermeture
0 plante dont les graines servent à nourrir
0 somme d'argent qu'on paie pour pouvoir habiter dans une maison
0 je ne connais pas la réponse

105. **nouer**
0 trouver la mort dans l'eau
0 rester à la surface d'un liquide
0 donner une position supérieure
0 établir, former, se dit des relations
0 je ne connais pas la réponse

106. **bénévole**
0 relatif au mariage
0 qui n'a rien de remarquable
0 faire quelque chose sans être payé
0 relatif à la marine de guerre ou aux bateaux de commerce
0 je ne connais pas la réponse

107. **signataire (m)**
0 une très grande quantité
0 personne qui enseigne dans une école primaire
0 personne qui a mis son nom sur un document officiel
0 plusieurs feuilles de papier, sur lesquelles on peut écrire
0 je ne connais pas la réponse

108. **amer**
0 relatif au train
0 originaire du pays où il vit
0 qui marque un échange de même valeur
0 qui produit au goût une sensation souvent désagréable
0 je ne connais pas la réponse

109. **réconcilier**
0 faire perdre du temps à quelqu'un
0 heurter le pied contre quelque chose
0 rétablir entre des personnes des relations amicales
0 séduire quelqu'un par des présents ou des promesses
0 je ne connais pas la réponse

110. **appliqué**
0 type de saveur
0 qui est clair, facile à saisir
0 qui est ponctuel, qui fait tout avec soin
0 se dit d'un processus qu'on ne peut pas arrêter, ni inverser
0 je ne connais pas la réponse

111. **tremblement (m)**
0 compensation
0 objet pour diriger une voiture
0 colonne de plusieurs véhicules
0 mouvement incontrôlé des mains
0 je ne connais pas la réponse

112. **plaie (f)**
0 blessure, due à un accident
0 température trop élevée du corps
0 sentiment exagéré de sa propre valeur
0 substance blanche que l'on peut extraire de l'eau de mer
0 je ne connais pas la réponse

113. **boucher (m)**
0 religieux qui vit en communauté
0 personne avec des pouvoirs magiques
0 personne qui dirige un groupe de religieux
0 personne qui prépare et vend de la viande
0 je ne connais pas la réponse

114. **foie (m)**
0 partie du corps, organe
0 partie supérieure de la tête
0 phénomène naturel, éclat de lumière
0 construction en fils de fer pour tenir enfermés des animaux
0 je ne connais pas la réponse

115. **puits (m)**
0 trou profond dans le sol
0 étage partiellement enterré
0 le fait de remettre à plus tard
0 organe nécessaire pour respirer
0 je ne connais pas la réponse

116. **case (f)**
0 grand animal sauvage, dangereux
0 petit espace délimité sur un formulaire
0 partie du corps, partie inférieure du cou
0 un type d'arbre qu'on cultive pour ses fruits
0 je ne connais pas la réponse

117. **détérioration (f)**
0 température égale ou inférieure à zéro
0 action par laquelle une chose a perdu de ses qualités
0 une personne qui prend la fonction d'une autre personne
0 du matériel, des personnes qui rendent un groupe plus fort
0 je ne connais pas la réponse

118. **moyenne (f)**
0 aspect différent d'une chose
0 surface sur laquelle figure quelque chose
0 petite boule, par exemple remplie d'air, de gaz
0 valeur qui indique le milieu entre plusieurs autres
0 je ne connais pas la réponse

119. **malin**
0 intelligent
0 pas exagéré
0 horrible, terrible
0 destiné à brûler pour chauffer
0 je ne connais pas la réponse

120. **s'écrouler**
0 rire, s'amuser
0 préciser quelque chose
0 arriver par accident, par surprise
0 tomber en pièces et perdre toute valeur
0 je ne connais pas la réponse

Appendix 4 Bilingual receptive vocabulary knowledge test (Study 2)

1. **certain**
0 veilig
0 nieuw
0 zeker
0 bereid
0 ik weet het niet

2. **mort**
0 dood
0 rechtuit
0 rustig
0 voorzichtig
0 ik weet het niet

3. **coup** (m)
0 deeltje
0 mening
0 product
0 slag
0 ik weet het niet

4. **empêcher**
0 verhinderen
0 zinken
0 beschuldigen
0 mislukken
0 ik weet het niet

5. **peuple** (m)
0 vrede
0 volk
0 vooruitgang
0 hoop
0 ik weet het niet

6. **beau**
0 arm
0 mooi
0 gek
0 machtig
0 ik weet het niet

7. **prêt**
0 veilig
0 werkloos
0 zeker
0 bereid
0 ik weet het niet

8. **mouvement** (m)
0 beweging
0 begin
0 afstand
0 glimlach
0 ik weet het niet

9. **quitter**
0 vallen
0 verdienen
0 verlaten
0 bestaan uit
0 ik weet het niet

10. **tomber**
0 achterlaten
0 recht hebben op
0 uit meerdere delen bestaan
0 vallen
0 ik weet het niet

11. **paix** (f)
0 vrede
0 familie
0 verbetering
0 wilskracht
0 ik weet het niet

12. **nuit** (f)
0 zomer
0 nacht
0 vertraging
0 kracht
0 ik weet het niet

13. **ignorer**
0 breken
0 uiteenbarsten
0 opschorten
0 niet weten
0 ik weet het niet

14. **pauvre**
0 arm
0 mooi
0 ongewoon
0 invloedrijk
0 ik weet het niet

15. **taux** (m)
0 gewicht
0 omgeving
0 onderste gedeelte
0 percentage
0 ik weet het niet

16. **espoir** (m)
0 volk
0 hoop
0 vrede
0 vooruitgang
0 ik weet het niet

17. **faible**
0 genoeg
0 goed
0 tijdbesparend
0 zwak
0 ik weet het niet

18. **avis** (m)
0 mening
0 slag
0 merk
0 deel
0 ik weet het niet

19. **lutte** (f)
0 gevecht
0 verdriet
0 moord
0 uitdaging
0 ik weet het niet

20. **satisfaire**
0 bevestigen, zeggen
0 bewaren
0 doemdenken
0 tevredenstellen
0 ik weet het niet

21. **échange** (m)
0 briefje
0 hulpmiddel
0 opdracht
0 ruil
0 ik weet het niet

22. **absence** (f)
0 afwezigheid
0 aanval
0 loon
0 hersenen
0 ik weet het niet

23. **cher**
0 dichtbij
0 duur
0 risicovol
0 waar
0 ik weet het niet

24. **accès** (m)
0 verlangen
0 reactie
0 toegang
0 termijn
0 ik weet het niet

25. **proche**
0 dichtbij
0 echt
0 risicovol
0 duur
0 ik weet het niet

26. **avantage** (m)
0 taak
0 bericht
0 ruil
0 voordeel
0 ik weet het niet

27. **nécessité** (f)
0 gunst
0 droom
0 noodzaak
0 blijdschap
0 ik weet het niet

28. **échapper**
0 ontsnappen
0 remmen
0 nalaten
0 bouwen
0 ik weet het niet

29. **environnement** (m)
0 gewicht
0 onderkant
0 milieu
0 percentage
0 ik weet het niet

30. **effectuer**
0 uitvinden
0 verzamelen
0 uitvoeren
0 verslaan
0 ik weet het niet

31. **ferme** (f)
0 boerderij
0 geboorte
0 honger
0 hulpvaardigheid
0 ik weet het niet

32. **échec** (m)
0 broekzak
0 slachtoffer
0 mislukking
0 overwinning
0 ik weet het niet

33. **poids** (m)
0 gewicht
0 onderkant
0 omgeving
0 percentage
0 ik weet het niet

34. **honneur** (m)
0 eer
0 fantasie
0 moeilijkheid
0 vlammenzee
0 ik weet het niet

35. **prétendre**
0 beweren
0 bevredigen
0 toevertrouwen
0 vrezen
0 ik weet het niet

36. **retraite** (f)
0 afwezigheid
0 ontstaan
0 pensioen
0 verkoop
0 ik weet het niet

37. **fil** (m)
0 eeuw
0 draad
0 eiland
0 smaak
0 ik weet het niet

38. **puissant**
0 gek
0 arm
0 mooi
0 machtig
0 ik weet het niet

39. **rare**
0 vatbaar
0 nauw
0 zeldzaam
0 trouw
0 ik weet het niet

40. **envie** (f)
0 deadline
0 reactie
0 toegankelijkheid
0 zin
0 ik weet het niet

41. **lendemain** (m)
0 de volgende dag
0 streep
0 hoek
0 boot
0 ik weet het niet

42. **retard** (m)
0 kracht, belang
0 nacht
0 periode
0 vertraging
0 ik weet het niet

43. **port** (m)
0 auteur/schrijver
0 burgemeester
0 haven
0 methode
0 ik weet het niet

44. **rêve** (m)
0 kern
0 geluk
0 droom
0 waardering
0 ik weet het niet

45. **trait** (m)
0 kruising
0 overgang
0 streep/lijn
0 volgende dag
0 ik weet het niet

46. **distance** (f)
0 afstand
0 begin
0 glimlach
0 verplaatsing
0 ik weet het niet

47. **foyer** (m)
0 mislukking
0 netwerk
0 punt, plaats
0 tehuis
0 ik weet het niet

48. **suspendre**
0 vernielen
0 uiteenbarsten
0 schorsen
0 niet weten
0 ik weet het niet

49. **sévère**
0 zwaar
0 streng
0 gelijkaardig
0 omvangrijk
0 ik weet het niet

50. **chute** (f)
0 val
0 zee
0 bloed
0 visvangst
0 ik weet het niet

51. **veiller**
0 aanhouden
0 funderen
0 aansnijden
0 waken
0 ik weet het niet

52. **goût** (m)
0 draad
0 eeuw
0 eiland
0 smaak
0 ik weet het niet

53. **trace** (f)
0 deurkruk
0 merkteken
0 stilte
0 vis
0 ik weet het niet

54. **foule** (f)
0 licht
0 menigte
0 sneeuw
0 werkloosheid
0 ik weet het niet

55. **poche** (f)
0 overwinning
0 mislukking
0 broekzak
0 slachtoffer
0 ik weet het niet

56. **éclater**
0 schorsen
0 vernielen
0 ontploffen
0 niet weten
0 ik weet het niet

57. **bonheur** (m)
0 geluk
0 geloof
0 gevoel
0 bank
0 ik weet het niet

58. **plonger**
0 verhinderen
0 duiken
0 beschuldigen
0 mislukken
0 ik weet het niet

59. **cerveau** (m)
0 afwezigheid
0 aanval
0 inkomen
0 hersenen
0 ik weet het niet

60. **inquiétant**
0 beroemd
0 verontrustend
0 agrarisch
0 gewapend
0 ik weet het niet

61. **séjour** (m)
0 verblijf
0 opoffering
0 verkiezingen
0 nederlaag
0 ik weet het niet

62. **censé**
0 geacht
0 plotseling
0 voorafgaand
0 ontgoochelend
0 ik weet het niet

63. **fuite** (f)
0 zelfmoord
0 vlucht
0 achtervolging
0 verdeling
0 ik weet het niet

64. **vœu** (m)
0 teleurstelling
0 wens
0 haat
0 opinieonderzoek
0 ik weet het niet

65. **branche** (f)
0 tak
0 mengeling
0 scherm
0 veer
0 ik weet het niet

66. **songer**
0 inslaan
0 schudden
0 glijden
0 peinzen
0 ik weet het niet

67. **enceinte**
0 mooi
0 doof
0 zacht
0 zwanger
0 ik weet het niet

68. **contrainte** (f)
0 ernst
0 beperking
0 bezit
0 voorzorgsmaatregel
0 ik weet het niet

69. **répandre**
0 verdelen
0 vertragen
0 opduiken
0 verspreiden
0 ik weet het niet

70. **huile** (f)
0 traan
0 regen
0 melk
0 olie
0 ik weet het niet

71. **guide** (m)
0 tegenstander
0 gids
0 stichter
0 werkgever
0 ik weet het niet

72. **creuser**
0 zaaien
0 begraven
0 graven
0 pletten
0 ik weet het niet

73. **couche** (f)
0 cel
0 heuvel
0 lidgeld
0 laag
0 ik weet het niet

74. **accueil** (m)
0 boom
0 smoes
0 uiterlijk
0 onthaal
0 ik weet het niet

75. **lent**
0 traag
0 stom
0 droog
0 waardig
0 ik weet het niet

76. **gare** (f)
0 station
0 bestemming
0 uitgang
0 schap
0 ik weet het niet

77. **provenance** (f)
0 scheiding
0 prestatie
0 pakket maatregelen
0 herkomst
0 ik weet het niet

78. **similaire**
0 vredig
0 gemeenschappelijk
0 toegevoegd
0 gelijkaardig
0 ik weet het niet

79. **ravir**
0 dwingen
0 verrukken
0 wijden aan
0 ter beschikking stellen
0 ik weet het niet

80. **acquisition** (f)
0 aanschaf
0 ondernemer
0 verstandhouding
0 toevloed
0 ik weet het niet

81. **distinct**
0 verschillend
0 zeker
0 verplicht
0 gegrond
0 ik weet het niet

82. **couloir** (m)
0 lunch
0 vlag
0 tuin
0 gang
0 ik weet het niet

83. **étoile** (f)
0 landschap
0 schijf
0 ster
0 glas
0 ik weet het niet

84. **trahir**
0 verergeren
0 uitputten
0 verraden
0 verdenken
0 ik weet het niet

85. **temporaire**
0 tijdelijk
0 stedelijk
0 gemeentelijk
0 aandachtig
0 ik weet het niet

86. **robe** (f)
0 storm
0 jurk
0 brand
0 verwardheid
0 ik weet het niet

87. **toit** (m)
0 dak
0 golf
0 keerpunt
0 vlek
0 ik weet het niet

88. **envoi** (m)
0 inzet
0 zending
0 middag
0 toeschouwer
0 ik weet het niet

89. **deviner**
0 toesnellen
0 zijn uiterste best doen
0 volharden
0 raden
0 ik weet het niet

90. **sagesse** (f)
0 pijn
0 huid
0 omvang
0 wijsheid
0 ik weet het niet

91. **déroulement** (m)
0 huis
0 misdrijf
0 verloop
0 handvol
0 ik weet het niet

92. **ancêtre** (m)
0 staal
0 verkiezing
0 collega
0 voorouder
0 ik weet het niet

93. **hostile**
0 vijandig
0 wreed
0 geloofwaardig
0 toegankelijk
0 ik weet het niet

94. **trésor** (m)
0 investering
0 overschot
0 schat
0 geldboete
0 ik weet het niet

95. **mépris** (m)
0 meningsverschil
0 vooroordeel
0 minachting
0 medelijden
0 ik weet het niet

96. **consécutif**
0 later
0 kwetsbaar
0 opeenvolgend
0 gebiedend
0 ik weet het niet

97. **coutume** (f)
0 gewoonte
0 kerkhof
0 omvang
0 vervaldag
0 ik weet het niet

98. **abriter**
0 verspreiden
0 aanpassen
0 bedaren
0 beschermen
0 ik weet het niet

99. **queue** (f)
0 rij mensen
0 vriend
0 aanval
0 verkrachting
0 ik weet het niet

100. **chantier** (m)
0 stramien
0 meer
0 perron
0 werf
0 ik weet het niet

101. **inconvénient** (m)
0 troef
0 ongemak
0 terugbetaling
0 fout
0 ik weet het niet

102. **emprunt** (m)
0 verzoek
0 schade
0 vermelding
0 lening
0 ik weet het niet

103. **hâte** (f)
0 tij
0 haast
0 schade
0 angst
0 ik weet het niet

104. **ennuyer**
0 uitdoven
0 vegen
0 vrezen
0 vervelen
0 ik weet het niet

105. **creux**
0 dun
0 naakt
0 hol
0 dik
0 ik weet het niet

106. **complice**
0 medeplichtig
0 onuitgegeven
0 gepast
0 opeenvolgend
0 ik weet het niet

107. **raide**
0 dun
0 stijf
0 beschaamd
0 machtig
0 ik weet het niet

108. **singulier**
0 vreselijk
0 nakend
0 gemeen
0 vreemd
0 ik weet het niet

109. **approprié**
0 zeker
0 toegevoegd
0 geschikt
0 rechtvaardig
0 ik weet het niet

110. **ange** (m)
0 ruzie
0 bekentenis
0 engel
0 kruis
0 ik weet het niet

111. **pâte** (f)
0 stofje
0 pasta
0 wolk
0 zand
0 ik weet het niet

112. **logiciel** (m)
0 software
0 bord
0 wiel
0 last
0 ik weet het niet

113. **couteau** (m)
0 jong kind
0 nachtmerrie
0 mes
0 stam
0 ik weet het niet

114. **décéder**
0 sterven
0 aanmoedigen
0 boren
0 haten
0 ik weet het niet

115. **prospérité** (f)
0 gevangenschap
0 voorspoed
0 slaap
0 kern
0 ik weet het niet

116. **poing** (m)
0 vuist
0 buik
0 wang
0 nek
0 ik weet het niet

117. **rallier**
0 voortbestaan
0 aankomen
0 verenigen
0 afdrukken
0 ik weet het niet

118. **défaillance** (f)
0 hemd
0 gedicht
0 defect
0 muil
0 ik weet het niet

119. **indemnisation** (f)
0 woede
0 kat
0 arbeidskrachten
0 schadevergoeding
0 ik weet het niet

120. **harceler**
0 zich haasten
0 verzamelen
0 pesten
0 onderwerpen
0 ik weet het niet

Appendix 5 Bilingual productive test (0-3,000) (Pilot study 2)

1. Philippe est ce_ _ _ _ _ que les résultats des deux matchs suivants seront décisifs. (zeker)
2. Mon grand-père est m _ _ _ à la suite d'une grave maladie. (dood)
3. Le maire a interdit la manifestation de ce mouvement d'extrême droite. Ainsi, il voulait e_ _ _ _ _ _ _ des problèmes entre des manifestants de gauche et de droite. (verhinderen)
4. Dans une démocratie, c'est le p_ _ _ _ _ qui doit choisir son gouvernement. (volk)
5. Elle avait l'air sympa avec cette b_ _ _ _ robe. (mooi)
6. La p_ _ _ _ de la Bastille a été un évènement important dans l'histoire de la France. (verovering)
7. Olivier a q_ _ _ _ _ sa femme pour une autre. (verlaten)
8. Le pôle Sud, c'est l'e_ _ _ _ _ _ le plus froid de la planète. (plaats)
9. Il n'a pas pu p_ _ _ _ _ _ _ _ _ _ à Roland Garros à cause d'une blessure. (deelnemen)
10. La l_ _ _ _ contre la criminalité est une priorité du nouveau maire. (strijd)
11. Cette entreprise veut c_ _ _ _ _ _ _ _ _ la tour la plus haute d'Europe. (bouwen)
12. L'a_ _ _ _ _ _ _ de pouvoir acheter des livres et dvd en ligne, c'est qu'il ne faut plus se déplacer. (voordeel)
13. Je peux te demander une f_ _ _ _ _? Je cherche quelqu'un pour traduire ce texte en anglais. (gunst)
14. Tu peux m'aider? Cette boîte est trop l_ _ _ _ _ pour la porter moi seul. (zwaar)
15. Le chasseur avait touché le lapin : il y avait du s_ _ _ partout. (bloed)
16. Il p_ _ _ _ _ _ avoir serré la main au Président des États-Unis, mais je ne le crois pas. (beweren)
17. Je n'ai plus confiance en lui parce qu'il ne tient jamais ses p_ _ _ _ _ _ _ _ . Il ne fait jamais ce qu'il a dit. (beloftes)
18. Le tigre est devenu un animal r_ _ _ . (zeldzaam)
19. Dans la soirée du 20 juin, la rock star assistera à une cérémonie à Paris, et puis le

l_ _ _ _ _ _ _ _, elle rendra visite à des enfants malades. (‘s anderendaags)

20. Juste après la n_ _ _ _ _ _ _ _ de sa fille, il a mis les premières photos sur Instagram. (geboorte)
21. Sa f_ _ en Dieu aide le croyant à surmonter toutes les difficultés. (geloof)
22. J'ai froid. Tu peux mettre du b_ _ _ sur le feu? (hout)
23. Ça se voit qu’il est heureux. Il a toujours un grand s_ _ _ _ _ _ . (glimlach)
24. Il a beaucoup étudié, mais je ne sais pas si ce sera s_ _ _ _ _ _ _ _ pour réussir. (voldoende)
25. - Tu as lu *La carte et le territoire*? - Ah oui, Michel Houellebecq est un grand é_ _ _ _ _ _ _ . (schrijver)
26. Un mode de vie plus s_ _ _ permettrait de prévenir beaucoup de problèmes physiques. (gezond)
27. Donne-lui une b _ _ _ _ de bonbons afin de le féliciter pour sa promotion. (doos)
28. Son appartement témoigne d'un bon g_ _ _ . J'aime bien les couleurs fraîches et les meubles design. (smaak)
29. Il a découvert une pièce de deux euros dans la p_ _ _ _ de son pantalon. (zak)
30. L’âge agit sur le c_ _ _ _ _ _. Voilà pourquoi il faut rester actif aussi au niveau intellectuel : lire des livres, apprendre une nouvelle langue, faire des sudokus, etc. (hersenen)
31. Il fait trop chaud pour se mettre au soleil. Il faut s’installer à l’o _ _ _ _ (schaduw).
32. Après son s_ _ _ _ _ aux soins intensifs, le prince poursuit sa revalidation dans un autre service. (verblijf)
33. Les c_ _ _ _ _ _ _ _ _ _ religieuses appartiennent à la sphère privée et tout le monde devrai avoir le droit de s'exprimer librement. (overtuigingen)
34. Nous vous offrons un lit d_ _ _ et confortable et un petit-déjeuner buffet soigné. (zacht)

35. Depuis ses origines, notre produit fonde sa légitimité sur son expertise du soin des p_ _ _ _ sensibles car rien n'est plus fragile que la p_ _ _ du bébé qui ne tolère pas la moindre allergie. (huid)
36. On a dû couper cette b_ _ _ _ _ _ du grand arbre pour éviter qu'elle ne tombe sur notre maison. (tak)
37. Tu as aussi entendu la r_ _ _ _ _ selon laquelle le président aurait une affaire avec une actrice ? (gerucht)
38. Le représentant de cette organisation a demandé de mieux r_ _ _ _ _ _ _ les bénéfices entre les populations africaines. (verdelen)
39. Vraiment ? Elle est e _ _ _ _ _ _ _ depuis cinq mois ? Je ne savais pas qu'elle attendait un enfant. (zwanger)
40. Au cours de la journée, le temps devrait rester s_ _ (droog) dans le sud du pays. Il y aura éventuellement de la pluie dans le nord.
41. Après avoir infecté les PC, les virus se r_ _ _ _ _ _ _ _ maintenant aussi sur les smartphones et les tablettes. (verspreiden)
42. J'adore les pâtes avec de l'h_ _ _ _ d'olive. (olie)
43. En voyant son cadavre, elle a poussé des c_ _ _ de terreur. (schreeuwen)
44. Pourquoi les chiens c_ _ _ _ _ _ _ -ils des trous? (graven)
45. Avant de passer à la seconde c_ _ _ _ _, il faut être sûr que la peinture est sèche. (laag)
46. L'é_ _ _ _ _ relie le bras avec le reste du corps. (schouder)
47. De l'eau c_ _ _ _ le long du mur de ma chambre! Qu'est-ce qu'il faut faire? (stromen)
48. Si vous avez encore des questions, adressez-vous à l'a_ _ _ _ _ _ où on vous aidera volontiers. (onthaal)
49. Je voudrais savoir quelle est la voiture la plus rapide et quelle est la plus l_ _ _ _. (traag)
50. Nous sommes de plus en plus nombreux à être attentifs à la p_ _ _ _ _ _ _ _ _ des produits et à préférer ceux de chez nous. (herkomst)
51. J'ai appris que vous serez bientôt père et j'en suis r_ _ _ . (verheugd)
52. Google a réalisé une nouvelle a_ _ _ _ _ _ _ _ _ _ en achetant Boston Dynamics. (aanwinst)
53. Il nous a appris que le rêve et la réalité sont deux choses d_ _ _ _ _ _ _ _ _ . (verschillend)

54. Vous prenez l'escalier jusqu'au deuxième étage. Ensuite, vous arrivez dans le
    c_ _ _ _ _ _ et vous frappez à la dernière porte sur votre droite où se trouve son bureau. (gang)
55. Quel est votre l_ _ _ _ _ préféré pendant votre temps libre ? (hobby)
56. Le jour après le lancement du nouvel iPhone 6s, il s'est p_ _ _ _ _ _ _ _ à la Fnac. (heeft gehaast/haastte)
57. Le procureur a décidé la c_ _ _ _ _ _ de ce dossier de fraude. Il y a trop peu de preuves pour pouvoir condamner cet industriel. (afsluiting)
58. Prise d'émotion, elle a laissé libre cours aux l_ _ _ _ _ . (tranen)
59. Elle a décidé de mettre sa plus jolie r_ _ _ pour aller au mariage de sa sœur. (kleed)
60. La *s*_ _ _ _ _ _ comprend le savoir et la vertu de quelqu'un. (wijsheid)

Appendix 6 Bilingual productive test (0-2,000) (Study 2)

1. Philippe est ce_ _ _ _ _ (zeker) que les résultats des deux matchs suivants seront décisifs.
2. Mon grand-père est m _ _ _ (dood) à la suite d'une grave maladie.
3. Son agresseur a continué à lui porter des c_ _ _ _ (trap) de pied.
4. Le maire a interdit la manifestation de ce mouvement d'extrême droite. Ainsi, il voulait e_ _ _ _ _ _ _ (verhinderen) des problèmes entre des manifestants de gauche et de droite.
5. Le d _ _ _ _ (begin) du film est passionnant : dans la première scène le héros est attaqué par son plus grand ennemi.
6. Dans une démocratie, c'est le p_ _ _ _ _ (volk) qui doit choisir son gouvernement.
7. Le général j _ _ _ (oordelen) qu'il n'est pas nécessaire d'envoyer des troupes supplémentaires au front. Son adjudant n'est pas d'accord, mais n'ose pas protester.
8. Il a été condamné à une p _ _ _ _ (straf) de prison de 5 ans.
9. La p_ _ _ _ (verovering) de la Bastille a été un évènement important dans l'histoire de la France.
10. Ils sont contents que leur fils soit revenu sain et s _ _ _ (veilig) de la guerre.
11. Olivier a q_ _ _ _ _ (verlaten) sa femme pour une autre.
12. L'o _ _ _ _ _ _ _ (doel) de cette organisation est de protéger les droits des enfants.
13. La jeune femme est t_ _ _ _ _ (gevallen) du wagon et s'est retrouvée en dessous du train.
14. Pour l'achat d'un nouveau vélo, je vous conseille ce magasin. Il y a un l _ _ _ _ (ruime) choix de vélos pour toute la famille.
15. Il a mal aux p _ _ _ _ (voeten) après cette promenade de trois heures dans la montagne.
16. Le pôle Sud, c'est l'e_ _ _ _ _ _ (plaats) le plus froid de la planète.
17. La l_ _ _ _ (strijd) contre la criminalité est une priorité du nouveau maire.
18. Toutes les entreprises veulent s_ _ _ _ _ _ _ _ _ (tevredenstellen) les besoins de leurs clients.

19. La Belgique et les États-Unis ont conclu un accord qui permettra l'é_ _ _ _ _ _ (uitwisseling) automatique des informations financières entre les deux pays. De cette façon, il sera plus difficile de frauder.
20. Un pyromane est obsédé par le f _ _ (vuur).
21. Il m'a envoyé un m _ _ _ _ _ _ (bericht) pour me dire qu'il ne peut pas venir ce soir.
22. Cette entreprise veut c_ _ _ _ _ _ _ _ _ (bouwen) la tour la plus haute d'Europe.
23. Désolé, je n'ai pas a_ _ _ _ (toegang) à ces données. Je ne peux pas vous aider.
24. Ce pays n'a plus de r_ _ _ _ _ _ _ _ _ (bronnen) naturelles et doit donc importer son énergie.
25. On pouvait lui c _ _ _ _ _ _ (toevertrouwen) n'importe quel secret: il est très discret.
26. L'a_ _ _ _ _ _ _ (voordeel) de pouvoir acheter des livres et dvd en ligne, c'est qu'il ne faut plus se déplacer.
27. Le champion olympique, hospitalisé depuis trois mois et demi après un grave accident de ski, fait de petits p_ _ _ _ _ _ (vooruitgang) encourageants.
28. La Syrie s'est engagée à d_ _ _ _ _ _ _ (vernietigen) son arsenal d'armes chimiques d'ici le 30 juin. Après cette date, aucune arme chimique ne devrait encore se trouver dans une caserne militaire.
29. Dans son *e* _ _ _ _ _ _ _ _ _ _ _ _ _ (omgeving) *naturel*, le *chat sauvage* s'adapte à de nombreux habitats : la savane, la forêt et la steppe.
30. Je peux te demander une f_ _ _ _ _(gunst) ? Je cherche quelqu'un pour traduire ce texte en anglais.
31. Il aime travailler dans les champs et il adore les vaches. Cela m'étonnerait pas qu'il ait plus tard sa propre f_ _ _ _ (boerderij).
32. Tu peux m'aider ? Cette boîte est trop l_ _ _ _ _ (zwaar) pour la porter moi seul.
33. - Quelle est ton é_ _ _ _ _ _ _ (uitzending) préférée à la télé? - Ah, j'adore *The Voice*.
34. Le chasseur avait touché le lapin : il y avait du s_ _ _ (bloed) partout.
35. Le président de ce parti politique d'extrême droite est a_ _ _ _ _ (beschuldigd) d'antisémitisme.

36. Il p_ _ _ _ _ _ (beweren) avoir serré la main au Président des États-Unis, mais je ne le crois pas.
37. Je n'ai plus confiance en lui parce qu'il ne tient jamais ses p_ _ _ _ _ _ _ _ . Il ne fait jamais ce qu'il a dit. (beloftes)
38. L'économie chinoise est une des plus p _ _ _ _ _ _ _ _ _ (machtig) du monde.
39. Le tigre est devenu un animal r_ _ _ (zeldzaam).
40. Avec le beau temps qu'il fait en ce moment, j'ai e_ _ _ _ (zin) d'une glace.
41. J'aimerais passer mes vacances sur une î_ _ (eiland) exotique quelque part dans l'océan Pacifique.
42. Dans la soirée du 20 juin, la rock star assistera à une cérémonie à Paris, et puis le l_ _ _ _ _ _ _ _ ('s anderendaags), elle rendra visite à des enfants malades.
43. À la suite d'un accident, le train entre Bruxelles et Anvers aura un r_ _ _ _ _ (vertraging) d'environ 50 minutes. Veuillez nous en excuser.
44. Sa f_ _ (geloof) en Dieu aide le croyant à surmonter toutes les difficultés.
45. Après sa v_ _ _ _ _ _ _ (overwinning) contre Zulte Waregem, Anderlecht prend option sur le titre.
46. D'après cet article, le botox change les t_ _ _ _ _ (trekken) du visage.
47. J'ai froid. Tu peux mettre du b_ _ _ (hout) sur le feu?
48. La Cour d'assises l'a condamné à 25 ans pour le m _ _ _ _ _ _ (moord) de sa fille de 7 mois.
49. Il a beaucoup étudié, mais je ne sais pas si ce sera s_ _ _ _ _ _ _ _ (voldoende) pour réussir.
50. Cet homme politique de droite r _ _ _ _ _ _ _ (verwijten) à ce journal de l'avoir comparé à Hitler.
51. Il est temps de revenir sur la semaine de 35 heures pour redonner de la v _ _ _ _ _ _ (kracht).
52. Un mode de vie plus s_ _ _ (gezond) permettrait de prévenir beaucoup de problèmes physiques.
53. Les habitants de cette petite île vivent de la p_ _ _ _ (visvangst). Ils vendent tout ce que la mer leur donne.

54. Il y a une relation é_ _ _ _ _ _ (nauw) entre le tabac et le cancer du poumon.
55. L'avion chinois a disparu et n'a laissé aucune t_ _ _ _ (spoor).
56. Les cafés et restaurants ont beaucoup d'argent en c _ _ _ _ _ (kas) le samedi soir.
57. Il a découvert une pièce de deux euros dans la p_ _ _ _ (zak) de son pantalon.
58. Parfois je me demande qui a in _ _ _ _ _ (bedacht) le concept de pop-up store, tu sais, ces magasins à court terme.
59. Nous avons des moments de j_ _ _ (blijdschap), même s'ils sont rares et que nous devons faire face à beaucoup de difficultés, explique Ali, habitant de la capitale syrienne.
60. On va manger bientôt ? J'ai déjà f _ _ _ (honger).

Appendix 7 Speaking tasks (Studies 2 & 3)

Sujet 1

Vous êtes allé(e) à la montagne et en rentrant vous vous rendez compte que vous êtes malade. Vous vous rendez chez le médecin et vous lui décrivez vos symptômes.

Vous pouvez utiliser les images ci-dessous pour vous aider.

Sujet 2

Vous vous présentez à une agence pour l'emploi pour obtenir un job d'étudiant.

Il y a deux propositions :

- un job d'été dans une boulangerie

- un job d'étudiant dans un supermarché

L'employé vous demande de vous présenter (études, expérience, motivation) et vous essayez de le convaincre.

Appendix 8 Flash card hotel reservation (Pilot study 1)

Appendix 9 Flash card job interview (Pilot study 1)

Leerling A

U bent de verantwoordelijke van de lokale supermarkt (Carrefour, Delhaize, Colruyt, etc.) in Doornik (Tournai). U interviewt een student, omdat hij/zij als jobstudent(e) in uw winkel wil werken.

Speel de dialoog met de volgende aanwijzingen:

- Ontvang de student en stel uzelf voor
- Zeg de student dat het om een job als XXX gaat (zelf in te vullen)

- Stel vragen over de volgende onderwerpen:
  - Opleiding
  - Vorige werkervaring(en)
  - Niveau talen: in het bijzonder Frans en Engels
  - Vervoersmiddel om naar het werk te komen
  - E-mailadres en telefoonnummer

- Vraag wanneer de student beschikbaar is tijdens de grote vakantie
- Vraag naar de motivatie van de student
- Stel het volgende loon voor: 10 € per uur
- Vraag of de student nog vragen heeft
- Bedank de student en neem afscheid

Leerling B

U bent een student die gaat solliciteren voor een vakantiejob als kassier(ster) bij de lokale supermarkt (Carrefour, Delhaize, Colruyt, etc.) in Doornik (Tournai).

Speel de dialoog met de volgende aanwijzingen:

- Begroet de verantwoordelijke en stel uzelf voor (naam en leeftijd)
- Opleiding: uw studierichting (bv. Wetenschappen-wiskunde, Moderne talen, …)
- Vorige werkervaring: XXX (zelf in te vullen)
- Niveau talen: XXX (zelf in te vullen) niveau voor Frans, Engels, Nederlands, …
- In bezit van rijbewijs
    - Vervoersmiddel: auto
- E-mailadres: NAAM @hotmail.com & Telefoonnummer: 0476 / 23 36 57
- U bent beschikbaar van 1 juli tot 25 juli en van 16 augustus tot 31 augustus
- Motivatie: Frans oefenen, … (zelf aan te vullen)
- U gaat akkoord met het loon
- U hebt geen verdere vragen

- Bedank de verantwoordelijke en neem afscheid

Appendix 10 Flash card doctor's visit (Pilot study 1)

Rol: arts

Je bent huisarts en iemand komt op consultatie. Vraag naar de klachten en levensstijl van je patiënt en vertel hem wat hij moet doen om beter te worden. De conversatie mag voorbereid worden met behulp van onderstaande afbeeldingen en je eigen verbeelding. **Probeer de situatie zo realistisch mogelijk na te bootsen.**

Rol: patiënt

Je gaat naar je huisarts met een aantal klachten. Leg uit wat er scheelt en vraag wat je moet doen om zo snel mogelijk beter te worden. De conversatie mag voorbereid worden met behulp van onderstaande afbeeldingen en je eigen verbeelding. **Probeer de situatie zo realistisch mogelijk na te bootsen.**

Appendix 11 Questionnaire (Pilot study 1)

NAAM EN VOORNAAM: _____

KLAS: _____

GEBOORTEDATUM: _____

Vragenlijst: contact met het Frans. Omcirkel je antwoord/ vul in.

Door deze vragenlijst in te vullen, stem ik ermee toe dat de informatie die ik geef, gebruikt mag worden voor onderzoek en ik begrijp dat deze alleen op anonieme basis gebruikt zal worden.

1. Frans is mijn moedertaal.

Ja
Nee

2. Thuis spreek ik…

- met mijn moeder: _____
- met mijn vader: _____
- met broers/ zussen: _____

3. Ik heb familieleden die Frans spreken.

Ja: ouder(s)/ grootouder(s)/ andere:_____
Nee

Indien ja: ik spreek zeer vaak / vaak / soms / zelden met hen in het Frans.

4. Ik heb vrienden/kennissen die Frans spreken.

Ja
Nee

Indien ja: ik spreek zeer vaak / vaak / soms / zelden met hen in het Frans.

5. Ik heb ooit in een Franstalig gebied gewoond.

Ja, gedurende _____
Nee

6. Ik vind Frans een aangename taal.

Ja
Nee
Neutraal

7. Ik vind Frans een aangenaam vak.

Ja
Nee
Neutraal

8. Buiten school spreek ik Frans

Op talenkamp
Op reis
In jeugdbewegingen/ op mijn vakantiejob
Met familie of vrienden
Tijdens bijlessen
In geen enkele situatie
Andere: _____

9. Ik vind mijn niveau Frans…

Zeer goed
Goed
Gemiddeld
Slecht
Zeer slecht

Appendix 12 Speaking tasks (Pilot study 2)

Doctor's visit

Visite chez le médecin

Vous êtes allé(e) à la montagne et de retour vous vous rendez compte que vous avez pris froid : vous avez mal à la tête, à la gorge… Vous vous rendez chez le médecin et vous lui décrivez vos symptômes.

Job interview

Entretien d'embauche

Vous vous présentez à une agence pour l'emploi pour obtenir un stage dans un commerce. L'employé vous demande de vous présenter (études, expérience, motivation) et vous essayez de le convaincre.

Appendix 13 Holistic scale used by the raters (Study 3)

Evaluation holistique

Vous avez écouté différents dialogues et vous avez complété les évaluations analytiques, accordez maintenant un score holistique à tous les dialogues basé sur les détails ci-dessous. Vous serez amenés à utiliser une échelle graduelle allant de 1 (minimum) à 5 (maximum). Comme pour les évaluations analytiques, la distance entre chaque grade (p.ex. 1-2, 3-4, 4-5) doit être considérée comme égale.

UN SCORE DE 5 : Un dialogue dans cette catégorie fait preuve d'une maîtrise claire et cohérente du vocabulaire français, même s'il peut y avoir encore quelques erreurs mineures. Dans un dialogue type, le locuteur devrait utiliser les catégories conceptuelles appropriées (tant concrètes qu'abstraites), montrer une cohérence claire entre les mots, des connexions lexico-sémantiques, et le dialogue type est lexicalement varié (assez de mots pour expliquer des éléments complexes en détail). En général, le dialogue montre la maîtrise de la langue, par un vocabulaire varié, approprié et adéquat, le tout utilisé de manière naturelle et avec fluidité.

UN SCORE DE 4 : Un dialogue dans cette catégorie fait preuve d'une maîtrise raisonnablement cohérente du vocabulaire français, même s'il y a des erreurs occasionnelles ou des égarements par rapport à la qualité lexicale. Le dialogue présente une utilisation appropriée des catégories conceptuelles (tant concrètes qu'abstraites), de la cohérence entre les mots, des connexions lexico-sémantiques et une diversité lexicale qui permet des discussions d'éléments complexes, mais pas systématiquement. En général, le dialogue fait preuve d'une utilisation appropriée et précise du vocabulaire et il semble fluide et approprié.

UN SCORE DE 3 : Un dialogue dans cette catégorie fait preuve d'une maîtrise lexicale adéquate, même s'il y aura des égarements ou des méprises. Le dialogue montre une certaine utilisation appropriée des catégories conceptuelles (dont des concepts abstraits, mais pour la plupart des concepts concrets), de la cohérence entre les mots, des connexions lexico-sémantiques et de la diversité lexicale. En général, le dialogue utilise un vocabulaire approprié et précis, mais fait preuve d'une maîtrise inconsistante.

UN SCORE DE 2 : Un dialogue dans cette catégorie montre une maîtrise lexicale en voie de développement, mais montre UNE OU PLUSIEURS faiblesses dans les catégories conceptuelles, les connexions lexico-sémantiques, la cohésion entre les mots et la diversité lexicale. Cependant, le discours dans le dialogue est généralement structuré. Le dialogue est marqué essentiellement par un lexique restreint ou un choix de mots inapproprié, un manque de variation et consiste surtout en des mots concrets. Les inexactitudes lexicales dans le dialogue nuisent à la compréhension.

UN SCORE DE 1 : Un dialogue dans cette catégorie fait preuve d'une maîtrise lexicale restreinte et montre DEUX OU PLUS de faiblesses dans les catégories conceptuelles, les connexions lexico-sémantiques, la cohésion entre les mots et la diversité lexicale. Le discours est très peu cohérent et le locuteur a du mal à s'exprimer facilement. Des répétitions ainsi que l'emploi de phrases apprises par cœur sont fréquents. En gros, le dialogue fait preuve d'un vocabulaire limité, des choix de mots incorrects et montre de nombreux problèmes lexicaux, qui rendent la compréhension difficile.

Score holistique basé sur les rubriques ci-dessus (1-5): ___

Appendix 14 English PET listening comprehension test (Study 1)

Listening • Part 1

Questions 1–7

There are seven questions in this part.
For each question, choose the correct answer (**A**, **B** or **C**).

Example: Where did the girl and her family go on holiday?



(A) B C

---

1 Which cyclist won the race?



A B C

2 Why couldn't the girl go to photography club yesterday?



A B C

3 Where will the boy and his friend practise basketball?

A

B

C

4 How did the girl get to school?

A

B

C

5 Which music poster does the boy prefer?

A

B

C

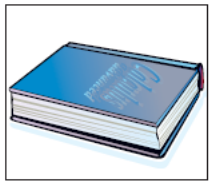
6 Where will they go after the cinema?

A

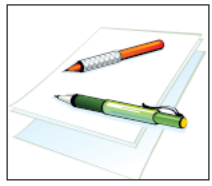
B

C

7 What will Sam take on the school museum trip tomorrow?



A



B

C

Listening • Part 2

Questions 8–13

You will hear an interview with a student called Sarah Mercer, who is planning to become a weather forecaster. For each question, choose the correct answer **A**, **B** or **C**.

---

8 Why did Sarah first become interested in the weather?

- **A** She studied weather in school science lessons. ☐
- **B** She had experience of bad weather where she lived. ☐
- **C** She saw programmes about weather on TV. ☐

9 Who encouraged Sarah's interest in the weather at home?

- **A** her mother ☐
- **B** her father ☐
- **C** her grandfather ☐

10 Pupils at Sarah's school club

- **A** produced a book about the weather. ☐
- **B** provided information to the school for projects. ☐
- **C** set up equipment for studying the weather. ☐

11 What mistake did Sarah's family make during a sailing trip?

- **A** They didn't recognise signs of bad weather coming. ☐
- **B** They hadn't made preparations for bad weather. ☐
- **C** They failed to check weather forecasts regularly. ☐

12 Sarah thinks in future she'd like to

- **A** work in the area of sport. ☐
- **B** be on TV weather programmes. ☐
- **C** travel abroad for her job. ☐

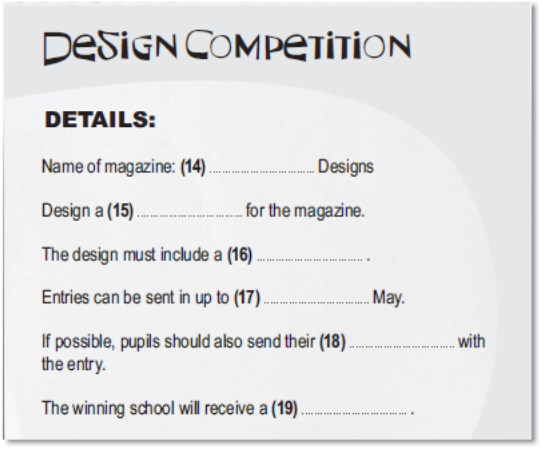
13 What kind of weather does Sarah like best?

- **A** when the sun is shining ☐
- **B** when it's raining hard ☐
- **C** when there's fog ☐

Listening • Part 3

Questions 14–19

You will hear an art teacher talking to a class about a design competition run by an online magazine. For each question, fill in the missing information in the numbered space.



DESIGN COMPETITION

DETAILS:

Name of magazine: **(14)** ..... Designs

Design a **(15)** ..... for the magazine.

The design must include a **(16)** ..... .

Entries can be sent in up to **(17)** ..... May.

If possible, pupils should also send their **(18)** ..... with the entry.

The winning school will receive a **(19)** ..... .

Listening • Part 4

Questions 20–25

Look at the six sentences for this part.
You will hear a boy, Mark, and a girl, Anna, talking about winter sports.
Decide if each sentence is correct or incorrect.
If it is correct, choose the letter **A** for **YES**. If it is not correct, choose the letter **B** for **NO**.

		YES	NO
20	Mark was disappointed at the amount of snow at his holiday centre.	**A**	**B**
21	Anna is surprised at how quickly Mark learnt to snowboard.	**A**	**B**
22	Mark thinks it's important to be fit in order to snowboard well.	**A**	**B**
23	Anna is shocked that some people don't protect themselves better when snowboarding.	**A**	**B**
24	Mark is keen to learn more advanced snowboarding in future.	**A**	**B**
25	Anna and Mark both prefer doing snowboarding to other winter sports.	**A**	**B**

Appendix 15 English VocabLab monolingual vocabulary knowledge test (Study 1)

1. **talk**:
0 to speak
0 to remove something
0 to give in return for money
0 to start an important activity
0 I don't know the answer

2. **lose**:
0 to put down
0 to enjoy or want
0 to continue to exist
0 to no longer have something
0 I don't know the answer

3. **real**:
0 true
0 having a lot of money
0 suffering from a disease
0 happening every week, month, …
0 I don't know the answer

4. **remember**:
0 to not forget
0 to get something
0 to speak for another person
0 to say that you will not do something
0 I don't know the answer

5. **building**:
0 something you can drink
0 the head, arms and legs of a person
0 a structure with walls, such as a house
0 something you use to keep things together in your hands
0 I don't know the answer

6. **society**:
0 a feeling
0 how fast something moves
0 a very large group of people
0 a small amount of something
0 I don't know the answer

7. **church**:
0 something you can sit on
0 a building for religious activities
0 the part of your body that holds your heart
0 something you use to hold your coffee or tea
0 I don't know the answer

8. **deal**:
0 to use a car
0 to do business
0 to form a group
0 to make something work again
0 I don't know the answer

9. **rise**:
0 to give food
0 to look at words in a book
0 to increase or move upwards
0 to know something you have seen before
0 I don't know the answer

10. **animal**:
0 a series of lessons
0 a living thing, such as a dog
0 a place where you buy things
0 something that you use to travel
0 I don't know the answer

11. **benefit**:
0 a good effect
0 a part of an organization
0 the time when a baby is born
0 a relationship or connection between people
0 I don't know the answer

12. **seat**:
0 something you can sit on
0 something that you can hear
0 the body part that covers a person or animal
0 a thing with words or pictures giving information

0 I don't know the answer

13. **affect**:
0 to accept
0 to state in public
0 to change something
0 to believe that something is true
0 I don't know the answer

14. **customer**:
0 a person who buys things
0 the action of choosing between things
0 an area of land such as Belgium, France or England
0 something that you write or say to express your opinion
0 I don't know the answer

15. **track**:
0 a large plant
0 a road or way
0 a group of soldiers
0 the process of testing
0 I don't know the answer

16. **feature**:
0 an important part or quality
0 a set of papers or documents
0 a group of people who control a country
0 the process in which something changes or grows
0 I don't know the answer

17. **attempt**:
0 an effort to do something
0 the right to use something
0 a record of your money with a bank
0 a request or question when you need help
0 I don't know the answer

18. **equipment**:
0 a person who changes texts
0 a group of tools or machines
0 someone who hates and fights you
0 a dangerous situation that happens unexpectedly
0 I don't know the answer

19. **female**:
0 last
0 very large
0 not difficult
0 relating to women
0 I don't know the answer

20. **wave**:
0 a young boy or girl
0 a space in the wall with glass
0 a line of water that moves across the sea
0 the part of your body that you use to see
0 I don't know the answer

21. **shut**:
0 to go
0 to close
0 to change
0 to make part of a group
0 I don't know the answer

22. **sweet**:
0 certain
0 not in danger
0 without money
0 being like sugar in your mouth
0 I don't know the answer

23. **package**:
0 an area with water
0 a feeling of being happy
0 a box with things to send to someone
0 money that you make by selling things
0 I don't know the answer

24. **loan**:
0 the person that is in control
0 a large amount or number of something
0 a group of teams playing against each other
0 the money that you receive but have to give back
0 I don't know the answer

25. **traffic**:
0 a small city
0 a subject that you write or speak about
0 the material of the ground in which plants grow
0 the group of moving machines, cars, ships and trains
0 I don't know the answer

26. **guide**:
0 a table that you use when you work
0 something that you hope will happen
0 the money that you have to give back to someone
0 a person or book that gives information about a place
0 I don't know the answer

27. **manner**:
0 a small number of people or things
0 something that is wrong or incorrect
0 the process or action of changing position
0 the way or style in which something is done
0 I don't know the answer

28. **confirm**:
0 to go up
0 to do something wrong
0 to have something inside
0 to say that something is true
0 I don't know the answer

29. **confidence**:
0 a difficult job
0 the reason why something happens
0 the belief that you do things well or right
0 the power to decide what someone has to do
0 I don't know the answer

30. **grey**:
0 very big or large
0 long between the edges
0 between black and white
0 living in a place since being born
0 I don't know the answer

31. **taste**:
0 a large group of families
0 a statement about something serious
0 the type of food or things that you like
0 the belief that you can trust someone or something
0 I don't know the answer

32. **guilty**:
0 covered in water
0 lighter than normal
0 nice or good-looking
0 when you have done something wrong
0 I don't know the answer

33. **prayer**:
0 someone who starts a business
0 a large room in a public building
0 something that covers a surface
0 the words that you speak to God
0 I don't know the answer

34. **suspect**:
0 to give
0 to remove
0 to move from side to side
0 to believe that something is true
0 I don't know the answer

35. **violent**:
0 causing pain
0 easy to attack
0 worth a lot of money
0 continuous over a period of time
0 I don't know the answer

36. **wedding**:
0 the process of giving things to each person
0 the event in which two people get married
0 the action of travelling from one place to another
0 the serious responsibility that you have to deal with
0 I don't know the answer

37. **lake**:
0 a large area of water
0 one of two different things
0 the possibility of something
0 when you do not have enough money
0 I don't know the answer

38. **bother**:
0 to change in form
0 to cause problems for someone
0 to put something or someone in the ground
0 to do something special for an important event
0 I don't know the answer

39. **reply**:
0 to give an answer to a question
0 to build something large and complex
0 to change something so that it works better
0 to make someone believe that something is true
0 I don't know the answer

40. **exposure**:
0 the time when you stop working
0 the event of becoming less or worse
0 the state of being in a situation that might affect you
0 the feeling of pleasure when you get something you want
0 I don't know the answer

41. **amazing**:
0 very good
0 not very important
0 including many details
0 behaving in an angry way
0 I don't know the answer

42. **impose**:
0 to breathe in through your nose
0 to move quickly across a surface
0 to force people to accept something
0 to put something in a place where people can see it
0 I don't know the answer

43. **mood**:
0 the way you feel
0 a sound you do not like
0 someone who leads a city
0 something to cover your face
0 I don't know the answer

44. **toy**:
0 a young person
0 something to play with
0 a tall (part of a) building
0 a piece of soft and thin paper
0 I don't know the answer

45. **anxiety**:
0 the feeling that you are good at something
0 the feeling of respect for something important
0 the feeling that makes you worry about something
0 the feeling that makes you want to hurt other people
0 I don't know the answer

46. **exhibition**:
0 a set of laws
0 a large company
0 a public show with interesting things
0 an amount of money that is kept in case of future loss
0 I don't know the answer

47. **giant**:
0 happy
0 very large
0 responsible
0 excellent and important
0 I don't know the answer

48. **restore**:
0 to stop doing
0 to fight against
0 to bring back to an earlier situation
0 to give the enthusiasm to do something
0 I don't know the answer

49. **hip**:
0 something you do often
0 an area of land higher than the land around it
0 the place where good people go after they die
0 the part which connects your legs to your body
0 I don't know the answer

50. **draft**:
0 an object used for cutting food
0 a group of small and unclean pieces
0 a combination of two or more elements
0 a plan or drawing that still needs changes
0 I don't know the answer

51. **ceiling**:
0 a plan
0 the surface above you in a room
0 a crime, something that you do wrong
0 the activity of making sure that people follow the rules
0 I don't know the answer

52. **bunch**:
0 a dollar
0 a yellow food
0 a group of things or people
0 a metal thing shot from a gun
0 I don't know the answer

53. **resort**:
0 a member of your family
0 the time when you stop working
0 a place that you go to for a holiday
0 something that tells you how to prepare food
0 I don't know the answer

54. **tight**:
0 with fear
0 most important
0 very close together
0 in the same situation
0 I don't know the answer

55. **justify**:
0 to use
0 to leave
0 to win control
0 to give a reason for something you have done
0 I don't know the answer

56. **insight**:
0 a list at the back of a book
0 a particular event or thing
0 the skill to understand something
0 something that makes you want to do a job
0 I don't know the answer

57. **shelter**:
0 a change
0 a piece of wood against a wall
0 the hard part of an egg or nut
0 a building that protects people or things
0 I don't know the answer

58. **rub**:
0 to hit something
0 to not agree to an offer
0 to touch someone with your lips
0 to move your hands over an object
0 I don't know the answer

59. **can**:
0 an accident or noise
0 a plant grown for food
0 a group of people who make decisions
0 something made from metal that contains drink
0 I don't know the answer

60. **steady**:
0 not smart
0 firm and strong
0 not resembling a circle
0 relating to feelings or religion
0 I don't know the answer

61. **slight**:
0 strong
0 small in size
0 allowed by law
0 without clothes
0 I don't know the answer

62. **flavour**:
0 the particular taste of food
0 something coming from a fire
0 when water covers something
0 the soft part of the body of a person or animal
0 I don't know the answer

63. **sequence**:
0 a good quality
0 a series of events
0 the person in charge
0 a new business or activity
0 I don't know the answer

64. **bite**:
0 to break suddenly
0 to use your teeth to eat
0 to clean or touch quickly
0 to act in a particular way
0 I don't know the answer

65. **cope**:
0 to fall down
0 to put together parts
0 to deal with a situation
0 to decide that something will not happen
0 I don't know the answer

66. **mortgage**:
0 the right to make your own decisions
0 the things that you want to throw away
0 something that you receive from someone after they die
0 an agreement in which you receive money from the bank
0 I don't know the answer

67. **acid**:
0 a group of people
0 a chemical substance
0 a result that you hope to achieve
0 an arrangement to meet at a particular time
0 I don't know the answer

68. **ugly**:
0 enough
0 very strange
0 not beautiful
0 for a short period of time
0 I don't know the answer

69. **charity**:
0 the quality of being honest
0 the activity of giving money
0 the state of understanding something
0 the process of entering a country with many people
0 I don't know the answer

70. **blanket**:
0 the flat part of a knife
0 a wooden object to hit a ball
0 a part of land near to the sea
0 a thick cover to keep warm in bed
0 I don't know the answer

71. **patch**:
0 a hole in the ground
0 a piece of something
0 the level or degree of something
0 an animal that is kept in the home

0 I don't know the answer

72. **coal**:
0 a hard, black substance
0 a stick that produces light
0 a type of cloth made from a plant
0 a line or shape that is not straight
0 I don't know the answer

73. **peer**:
0 to ask
0 to remove a cover
0 to look with difficulty
0 to express a feeling of respect
0 I don't know the answer

74. **upset**:
0 to disappear
0 to experience
0 to move away
0 to make feel angry
0 I don't know the answer

75. **surgeon**:
0 a central part
0 a type of animal without legs
0 a doctor who cuts open your body
0 something that makes you feel nervous
0 I don't know the answer

76. **continuous**:
0 not formal
0 without stopping
0 causing discussion
0 wanting to find out something
0 I don't know the answer

77. **flour**:
0 the difference
0 the range of a subject
0 a type of white powder used in food
0 something that you get when you win
0 I don't know the answer

78. **closet**:
0 a fight
0 a group of animals
0 someone who works in an office
0 a place where you keep your clothes
0 I don't know the answer

79. **trap**:
0 to not notice
0 to become soft
0 to make changes
0 to keep in one place
0 I don't know the answer

80. **nail**:
0 a thin stream
0 the soft earth
0 one complete turn around
0 a small, thin piece of metal
0 I don't know the answer

81. **inevitable**:
0 certain to happen
0 with nothing added
0 naturally existing in a place
0 experienced in a very strong way
0 I don't know the answer

82. **herb**:
0 a type of plant
0 a large hole in a mountain
0 the front part of your body
0 the brother of your mother or father
0 I don't know the answer

83. **curtain**:
0 a part of a plant
0 a place in the ground
0 a beautiful way of moving
0 a piece of cloth that covers a window
0 I don't know the answer

84. **limb**:
0 a unit
0 a type of food
0 an arm or a leg
0 the first room in a large building
0 I don't know the answer

85. **endure**:
0 to do
0 to suffer
0 to go above a limit
0 to put your name on the participant list
0 I don't know the answer

86. **turkey**:
0 a large bird
0 an area of land between hills
0 a part of your body that hurts
0 the short finger at the side of your hand
0 I don't know the answer

87. **resume**:
0 to start again
0 to make stronger
0 to give up and leave
0 to make pain go away
0 I don't know the answer

88. **bold**:
0 basic
0 angry
0 not afraid of danger
0 between red and blue
0 I don't know the answer

89. **reluctant**:
0 likely to be true
0 chosen by chance
0 sold directly to the public
0 not willing to do something
0 I don't know the answer

90. **flip**:
0 to turn over quickly
0 to stay on the surface
0 to send out light suddenly
0 to put one part on another part
0 I don't know the answer

91. **rage**:
0 a strong feeling of being angry
0 a place surrounded on all sides by bars
0 an idea or guess about something that is not known
0 a situation in which there is not enough of something
0 I don't know the answer

92. **precious**:
0 earlier
0 not modern
0 of great value or expensive
0 said or thought to be, but there is no proof
0 I don't know the answer

93. **rear**:
0 hard, not moving easily
0 at the back of something
0 not spiritual, not religious
0 almost straight up and down
0 I don't know the answer

94. **jaw**:
0 the lower part of your face
0 a hard hat that protects your head
0 the hair that grows on the lower part of a man's face
0 the part of piece of clothing that goes around your neck
0 I don't know the answer

95. **glory**:
0 admiration and praise
0 the right to control a country
0 the fact that you are responsible
0 the state of controlling your emotions
0 I don't know the answer

96. **condemn**:
0 to develop
0 to come together
0 to criticize something
0 to make more beautiful
0 I don't know the answer

97. **shrimp**:
0 a small sea animal
0 a flying insect with colourful wings
0 an animal kept by farmers for its meat
0 a very large wild animal with a very long nose
0 I don't know the answer

98. **bull**:
0 a male cow
0 an offer to pay money
0 a large size or mass of something
0 a tool that is used to hold and carry water
0 I don't know the answer

99. **object**:
0 to use for your benefit
0 to give someone a place to sit
0 to express opposition or dislike
0 to pull something heavy with a lot of effort
0 I don't know the answer

100. **scramble**:
0 to understand
0 to fall while walking
0 to react or to answer
0 to move using your hands and feet
0 I don't know the answer

101. **dough**:
0 a type of plant
0 dry grass that is used as food for animals
0 something made from glass to keep food
0 a mixture of ingredients that is baked to make cookies or pizza
0 I don't know the answer

102. **shy**:
0 stupid, not serious
0 sad because you are not with other people
0 supporting one person or opinion more than others
0 nervous and scared in the company of other people
0 I don't know the answer

103. **scent**:
0 a natural smell
0 an unhappy feeling
0 something that is the same
0 an agreement or permission
0 I don't know the answer

104. **stack**:
0 a pile of things
0 a flat circular object
0 the beginning of the day
0 a piece of stone that covers a roof
0 I don't know the answer

105. **fierce**:
0 near each other
0 private or personal
0 involving very strong feelings of hate or anger
0 belonging to a place with a local government of its own
0 I don't know the answer

106. **aisle**:
0 a strong feeling of sadness
0 a passage between rows or seats
0 a personal quality that attracts people to you
0 a small group of soldiers who do a particular job
0 I don't know the answer

107. **chunk**:
0 the bones of the head
0 a large piece of something
0 a ball that has the map of the world on it
0 an object that produces light and fits into a lamp
0 I don't know the answer

108. **hurricane**:
0 a violent storm with very strong winds
0 the part of your body that cleans your blood
0 a body temperature that is higher than normal
0 one of the many soft, light things that cover a bird's body
0 I don't know the answer

109. **embarrassed**:
0 never happened before
0 far away from other places
0 when you cannot use part of your body or brain

0 feeling nervous and stupid in front of other people
0 I don't know the answer

110. **cheer**:
0 to make healthy again
0 to shout with happiness
0 to reply to something you disagree with
0 to use your teeth to bite food into small pieces
0 I don't know the answer

111. **scratch**:
0 to make changes
0 to rub with your nails
0 to move something quickly
0 to put into water or tea for a short time
0 I don't know the answer

112. **neat**:
0 shocking
0 being the only one of a type
0 carefully arranged and looking nice
0 having a short distance from the top to the bottom
0 I don't know the answer

113. **threshold**:
0 a piece of land
0 the moment when an airplane lands
0 a step or level at which something starts
0 the activity of moving something to another place
0 I don't know the answer

114. **cruel**:
0 basic, essential
0 heavy or crowded
0 very unkind and causing suffering
0 done because you choose to do it
0 I don't know the answer

115. **cottage**:
0 a type of small house
0 the study and use of numbers
0 an area of water where ships stay
0 a natural substance in the earth such as gold
0 I don't know the answer

116. **retreat**:
0 the feeling of respect
0 a private and safe place
0 a book in which you write down meetings
0 the activity of taking something
0 I don't know the answer

117. **contemplate**:
0 to say that something cannot be done
0 to explain something clearly and exactly
0 to force something out from your mouth
0 to spend time thinking over a future action
0 I don't know the answer

118. **arrow**:
0 a sign that points in a direction
0 a sudden and quick spreading of a disease
0 a situation where you work for someone without pay
0 a type of smooth, cold stone used in buildings and statues
0 I don't know the answer

119. **ribbon**:
0 a type of large tree
0 a long piece of material used to tie things together
0 a small animal with long ears and large front teeth
0 the thick and soft hair that covers the body of an animal
0 I don't know the answer

120. **ridge**:
0 a type of heavy wet soil
0 a large, soft, comfortable seat
0 the long and narrow top of a group of mountains
0 a claim that someone has done something wrong

0 I don’t know the answer

Nederlandse samenvatting

Het *Europees Referentiekader voor Talen* (ERK) is wellicht het belangrijkste beleidsdocument van de laatste jaren voor het vreemdetalenonderwijs in Europa. Het taalneutrale ERK beschrijft talige competenties voor alle Europese talen en dit binnen zes taalvaardigheidsniveaus (van A1 tot C2) op basis van descriptoren. Ondanks de populariteit van het ERK zijn de taalspecifieke invullingen empirisch onvoldoende gevalideerd omdat zij voornamelijk gebaseerd zijn op de meer intuïtieve beoordeling van taaldocenten. Bovendien is het verband tussen de talige competenties enerzijds en concrete taaltaken anderzijds nauwelijks onderzocht. Ten slotte ontbreekt het het ERK aan een eenduidig theoretisch kader op basis waarvan taalvaardigheid kan geconceptualiseerd worden. De vraag die men zich dus kan stellen is of het op dit moment wel mogelijk is om op een betrouwbare en valide manier een ERK-niveau aan een taaltaak en/of een vreemdetaalleerder toe te kennen.

Voor dit onderzoek focussen we concreet op het schoolvak Frans en het B1 niveau, dat de verwachtte beheersingsgraad is op het einde van het secundair onderwijs. Daarbij is vooral het beheersen van de mondelinge taalvaardigheid cruciaal. We focussen hier dan ook expliciet op luister- en spreekvaardigheid.

Na de algemene inleiding en afbakening van het onderwerp (hoofdstuk 1), het theoretische (hoofdstuk 2) en het methodologische kader (hoofdstuk 3), behandelen de volgende hoofdstukken de drie empirische studies die tijdens dit doctoraatsonderzoek uitgevoerd werden.

Meer concreet onderzoekt dit proefschrift de relatie tussen woordenschatkennis en luistervaardigheid enerzijds, en tussen woordenschatkennis en woordenschatgebruik in spreekvaardigheid anderzijds. Aangezien woordenschat een sleutelfactor is bij het voorspellen van iemands taalvaardigheid, staat lexicon centraal in de drie empirische studies.

In de eerste studie (hoofdstuk 4) gaan we na hoe sterk de relatie is tussen woordenschatkennis en luistervaardigheid voor Engels en Frans op het intermediair niveau (B1) (N = 199 Engels; N = 330 Frans). Verder gaan we na hoeveel woordenschat nodig is voor 'adequate' luistervaardigheid op het B1-niveau. Deze analyse is een replicatiestudie van Staehr (2009). Dit gebeurde aan de hand van een receptieve woordenschattest en een gestandaardiseerde luistervaardigheidstest (PET voor Engels, DELF voor Frans). De resultaten tonen aan dat er een sterke correlatie bestaat tussen de woordenschatkennis van leerlingen zowel voor Frans als voor Engels. Verder tonen de resultaten aan dat voor Engels ongeveer 1.000 woorden nodig zijn om tot adequaat luisterbegrip te komen, terwijl dat er voor Frans 2.000 zijn.

De tweede studie (hoofdstuk 5) onderzoekt hoe de relatie tussen een receptieve en productieve woordenschattoets en het taalgebruik van vreemdetaalleerders eruit ziet bij spreekvaardigheid (N = 51). Hiervoor is een leerdercorpus, dat bestaat uit mondelinge taaltaken uitgevoerd door vreemdetaalleerders Frans, aangelegd en volledig geannoteerd. Hun taalgebruik werd gemeten aan de hand van (1) het aantal woorden, verschillende woorden (types) en lemma's, (2) het lexicaal profiel van de gebruikte woorden en (3) de lexicale diversiteit van de gebruikte woorden. De resultaten tonen aan dat taalleerders die een grotere woordenschatkennis hebben, zowel receptief als productief, meer (verschillende) woorden gebruiken. Verder blijkt uit deze studie dat een productieve woordenschattoets een betere voorspeller is van woordenschatgebruik in gesproken taal dan een receptieve woordenschattoets.

In studie 3 (hoofdstuk 6) gaan we dieper in op woordenschatgebruik en spreekvaardigheid. Vertrekkend van het leerdercorpus, waaraan ook dezelfde taken uitgevoerd door moedertaalsprekers toegevoegd werden, en de beoordeling van de beschikbare taaltaken door experten, wordt gekeken welke lexicale parameters het woordenschatgebruik

van zowel moedertaalsprekers als vreemdetaalleerders Frans in twee interactieve spreektaken kunnen voorspellen (*N* = 27 L1, *N* = 51 L2). Die lexicale parameters die in globo relevant gebleken zijn, zijn het aantal woorden (tokens, types en lemma's), het frequentieprofiel van de gebruikte woorden en de lexicale diversiteit (D, HD-D). Verder wordt het verschil in woordenschatgebruik door moedertaalsprekers en vreemdetaalleerders onderzocht alsook de relatie tussen scores op een receptieve en productieve woordenschattoets, die vreemdetaalleerders aflegden, en de score die experten toekenden aan hun mondelinge output in twee interactieve spreektaken. Volgens de resultaten heeft de lexicale diversiteit van de gesproken output de grootste voorspellende waarde voor het inschatten van het bereikte taalvaardigheidsniveau, gevolgd door het aantal types en het aantal laagfrequente lemma's die een spreker gebruikt. Uit de resultaten komt ook naar voor dat zowel moedertaalsprekers als vreemdetaalleerders zich in deze situaties beperken tot de 1.000 meest frequente woorden. Moedertaalsprekers gebruiken echter iets meer laagfrequente woorden dan vreemdetaalleerders en hun output is lexicaal gezien ook meer gevarieerd.

Het laatste hoofdstuk van dit doctoraatsonderzoek (hoofdstuk 7) vat de belangrijkste resultaten samen en stelt een aantal pistes voor met betrekking tot toekomstig onderzoek. De studies in dit doctoraat dragen bij tot de empirische validering van de mondelinge taalcomponent (luisteren en spreken) voor Frans op het B1 niveau van het ERK. Bovendien wordt op deze manier bijgedragen aan de theorievorming in taalverwerving Frans door zowel receptieve als productieve woordenschattoetsen te gebruiken, op interactieve taaltaken te focussen in plaats van op monologale, en door Frans te onderzoeken bij leerlingen uit het secundair onderwijs.

Tenslotte is hiermee een opstap gemaakt naar het uittekenen van de noodzakelijke componenten voor het adequaat voorspellen van het beheersingsniveau van vreemdetaalleerders.