



Universidad Autónoma de Querétaro  
Facultad de Lenguas y Letras

Maestría en lingüística

Desarrollo de la riqueza léxica en estudiantes universitarios de francés L2

Tesis

Como parte de los requisitos para obtener el grado de Maestro en Lingüística

Presenta

Etienne Luis García Greiner

Dirigido por:

Dra. Adelina Velázquez Herrera

Dr. Ignacio Rodríguez Sánchez

Secretario

Dr. Antonio Reyes Pérez

Vocal

Dra. Karina Hess Zimmermann

Suplente

Dra. Gloria Nelida Avecilla Ramírez

Suplente

Centro Universitario, 29 de mayo del 2025, Querétaro, Qro. México

La presente obra está bajo la licencia:  
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

### Usted es libre de:

**Compartir** — copiar y redistribuir el material en cualquier medio o formato

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

### Bajo los siguientes términos:



**Atribución** — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



**NoComercial** — Usted no puede hacer uso del material con [propósitos comerciales](#).



**SinDerivadas** — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

**No hay restricciones adicionales** — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

### Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.

## **Dedicatorias y agradecimientos**

Este trabajo está dedicado, en primer lugar, a todas las personas que me acompañaron a lo largo de mi recorrido en la maestría. Gracias a su apoyo y aliento constante, pude enfrentar cada reto sabiendo que no caminaba solo.

En especial, agradezco a mis sinodales. No solo fueron mis profesores que esclarecieron dudas y proporcionaron valioso material para este proyecto, sino que también demostraron una paciencia (¡y vaya que la demostraron!) y una guía inquebrantable durante todo el proceso. Sus clases siempre inspiraron curiosidad y entusiasmo, alentándome a profundizar más.

Mi gratitud para mi compañera y futura doctora en lingüística, Elisabetta, cuya ayuda fue invaluable. Su generosidad al compartir conocimientos, su tiempo y su amistad.

Mis padres han sido el pilar fundamental de este recorrido: su apoyo incondicional, su dedicación constante y su confianza en mi capacidad me impulsaron a no rendirme ante cada obstáculo. Su ejemplo de esfuerzo y perseverancia me ha recordado siempre la importancia de seguir adelante, incluso en momentos de incertidumbre.

A mi doctora Diana, cuyo amor y respaldo me sostuvieron a lo largo de cada etapa. Tu compañía, comprensión y ánimo en esta etapa importante de mi vida.

## Índice

Abreviaturas y Siglas.....	5
Resumen en Español.....	6
Abstract (English).....	7
Abstract (Français).....	8
1. Introducción.....	9
1.2 Planteamiento del Problema.....	12
1.3 Preguntas de Investigación.....	13
1.4 Hipótesis.....	14
1.5 Objetivos.....	14
2. Marco Teórico.....	16
2.1 Conceptos Fundamentales.....	17
2.1.1 <i>Definición del Concepto de Palabra:</i> .....	17
2.1.2 <i>El Tipo y el Caso</i> .....	18
2.1.3 <i>Las Frecuencias de Uso</i> .....	19
2.2 Adquisición y Desarrollo del Vocabulario en una Segunda Lengua.....	23
2.2.1 <i>Los Factores Determinantes en Adquisición de Vocabulario en L2</i> .....	23
2.2.2 <i>La Variabilidad en la Adquisición de Vocabulario en L2</i> .....	27
2.2.3 <i>El Valor de las Palabras Frecuentes en el Aprendizaje de un Idioma</i> .....	28
2.3 Consideraciones en los Estudios Sobre Léxico.....	32
2.3.1 <i>Las Pruebas para Medir el Léxico</i> .....	32
2.3.2 <i>El Rol del Vocabulario en la Transmisión de Temas e Intenciones</i> .....	34
2.4 El Vocabulario Técnico y sus Categorías.....	36
3. Antecedentes.....	39
3.1 La Medición de la Riqueza Léxica por Conteo de Palabras y Uso de Formulas.....	39
3.1.1 <i>La Medición de la Riqueza Léxica en L1</i> .....	39

3.1.2 Las Mediciones de la Riqueza Léxica en L2.....	41
3.2 La Medición de Riqueza Léxica Mediante Pruebas.....	43
3.2.1 El Uso de Pruebas para Medición de Riqueza Léxica.....	44
3.2.2 Adaptación de VLT para Francés como L2.....	48
3.3 Observaciones Finales.....	50
4. Metodología.....	52
4.1 El Diseño de la Prueba.....	52
4.2 Pilotaje.....	54
4.3 Los Participantes.....	56
4.4 La Aplicación de la Prueba.....	57
4.5 El Procedimiento de Análisis.....	58
5. Resultados.....	60
5.1 Los Resultados Generales.....	60
5.2 El Análisis Estadístico.....	62
5.3 El Análisis Estadístico para Bandas de Frecuencia.....	64
5.4 Las Observaciones Finales Entre Bandas de Frecuencia.....	65
6. Discusión.....	67
6.1 La Creación de una Prueba VLT Adaptada para Estudiantes Mexicanos....	71
6.2 Las Limitaciones del Estudio.....	72
6.3 Las Perspectivas Futuras.....	73
7. Conclusiones.....	76
8. Bibliografía.....	78
9. Anexo.....	84
9.1 Formato de la Prueba Utilizada para el Análisis de Esta Tesis.....	84
9.2 Formato de Carta de Consentimiento.....	113
9.3 Link de Acceso a Datos de la Prueba.....	115

## **Índice de tablas**

Tabla 1: Media y desviación estándar para cada grupo de semestre.....	60
Tabla 2: Cálculo ANOVA entre los grados de licenciatura.....	63
Tabla 3: Post Hoc por grado académico.....	64
Tabla 4: Anova con corrección de esfericidad Greenhouse-Geisser.....	65

## **Índice de figuras**

Figura 1: Ejemplo de los reactivos de la prueba, Banda 1K.....	53
Figura 2: Puntajes de los grupos en la prueba, con su media y su desviación estándar.....	61
Figura 3: Resultados de cada banda de frecuencia por el frado de licenciatura....	62

## **Abreviaturas y Siglas**

**ANOVA:** Análisis de Varianza (Analysis of Variance).

**AWL:** Academic Word List (Lista de Palabras Académicas).

**DVK:** Depth of Vocabulary Knowledge (Profundidad del Conocimiento de Vocabulario).

**GCSE:** General Certificate of Secondary Education (Certificado General de Educación Secundaria, utilizado en el sistema educativo del Reino Unido).

**G\_Lex:** Prueba de evaluación del conocimiento productivo de vocabulario en diversos contextos.

**L2:** Segunda Lengua.

**Lex30:** Prueba de vocabulario productivo basada en la generación de palabras relacionadas con cargas semánticas.

**LLMF:** Licenciatura en Lenguas Modernas en Francés.

**MCER:** Marco Común Europeo de Referencia para las Lenguas (Common European Framework of Reference for Languages - CEFR en sus siglas en inglés)

**MTLD:** Measure of Textual Lexical Diversity (Medida de la Diversidad Léxica Textual).

**PLVT:** Productive Vocabulary Levels Test.

**TESOL:** Teachers of English to Speakers of Other Languages.

**TOEFL:** Test of English as a Foreign Language (Examen de Inglés como Lengua Extranjera).

**TOEIC:** Test of English for International Communication (Examen de Inglés para la Comunicación Internacional).

**TTR:** Type-Token Ratio (Relación entre Tipos y Tokens).

**TTV:** Test de la Taille de Vocabulaire (VLT en francés).

**UVLT:** Updated Vocabulary Levels Test.

**VLT:** Vocabulary Levels Test.



## **Resumen en Español**

Esta tesis tuvo como propósito analizar el desarrollo de la riqueza léxica en estudiantes mexicanos inscritos en una Licenciatura en Lenguas Modernas en Francés (LLMF). El estudio tuvo como objetivo examinar si la exposición continua al francés como segunda lengua influye en la adquisición de vocabulario a lo largo de los distintos semestres de la carrera.

Para llevar a cabo este análisis, se diseñó una prueba específica basada en instrumentos validados, como el *Vocabulary Levels Test* (VLT) y el *Test de la Taille de Vocabulaire* (TTV). Esta prueba se aplicó a estudiantes de distintos semestres (2°, 4°, 6° y 8°) con el propósito de medir la riqueza léxica, con especial atención a las bandas de frecuencia léxica.

La investigación hipotetizó que la riqueza léxica aumentaría conforme los estudiantes avanzaban en su formación académica, aunque también consideró la posibilidad de una estabilización en los semestres más avanzados. Se emplearon análisis estadísticos, como ANOVA y pruebas Post Hoc, para identificar diferencias significativas entre los distintos grupos de estudiantes y evaluar si mejoró la adquisición léxica en función del grado académico de los participantes.

Asimismo, se esperaba que los estudiantes manejaran mejor el vocabulario de alta frecuencia, mientras que el dominio del vocabulario de menor frecuencia representara un desafío mayor. Estos hallazgos podrían proporcionar información valiosa para adaptar y mejorar en un futuro las estrategias pedagógicas en la enseñanza del francés como L2.

## **Palabras Clave**

Riqueza léxica, adquisición de vocabulario, francés como segunda lengua.

## **Abstract (English)**

This thesis aimed to analyze the development of lexical richness in Mexican students enrolled in a Bachelor's program in Modern Languages in French (LLMF). This study aims to examine whether ongoing exposure to French as a second language influences vocabulary acquisition throughout the various semesters of the program.

To carry out this analysis, a specific test was developed based on validated instruments such as The *Vocabulary Levels Test* (VLT) and the *Test de la Taille de Vocabulaire* (TTV). This test was administered to students from different semesters (2nd, 4th, 6th, and 8th) to measure their lexical richness, with particular attention paid to lexical frequency bands.

The study hypothesized that lexical richness would increase as students advanced in their academic studies, while also considering the possibility of leveling off at higher semesters. Statistical analyses, including ANOVA and Post Hoc tests, were employed to identify significant differences among the different groups of students and to evaluate whether their lexical acquisition improved according to academic level.

It was likewise expected that students would have a stronger command of high-frequency vocabulary, whereas mastering lower-frequency vocabulary would pose a greater challenge. These findings may provide valuable insights to adapt and enhance pedagogical strategies in teaching French as an L2 in the future.

## **Keywords**

Lexical richness, vocabulary acquisition, French as second language.

## **Abstract (Français)**

Cette thèse visait à analyser le développement de la richesse lexicale chez des étudiants mexicains inscrits en Licence en Langues Modernes option Français (LLMF). L'étude a cherché à déterminer si une exposition continue au français comme langue seconde (L2) influence l'acquisition du vocabulaire au fil des différents semestres du programme.

Pour réaliser cette analyse, un test spécifique a été élaboré à partir d'instruments validés tels que le *Vocabulary Levels Test* (VLT) et le Test de la *Taille de Vocabulaire* (TTV). Ce test a été administré à des étudiants de plusieurs semestres (2e, 4e, 6e et 8e) afin de mesurer leur richesse lexicale, en accordant une attention particulière aux bandes de fréquence lexicale.

La recherche a émis l'hypothèse que la richesse lexicale augmenterait à mesure que les étudiants progressent dans leur formation, tout en envisageant la possibilité d'une stabilisation aux semestres les plus avancés. Des analyses statistiques, incluant ANOVA et des tests Post Hoc, ont été menées pour identifier des différences significatives entre les divers groupes d'étudiants et pour évaluer si leur acquisition lexicale s'est améliorée en fonction du niveau académique.

On s'attendait également à ce que les étudiants maîtrisent mieux le vocabulaire de haute fréquence, tandis que l'acquisition d'un lexique de fréquence plus faible représenterait un défi plus important. Ces résultats pourraient fournir des informations précieuses pour adapter et améliorer, à l'avenir, les stratégies pédagogiques dans l'enseignement du français en tant que L2.

## **Mots-clés**

Richesse lexicale, acquisition de vocabulaire, français langue seconde.

## 1. Introducción

El desarrollo del vocabulario es uno de los fundamentos en el proceso de adquisición de una segunda lengua (L2), sobre el que se edifican las competencias lingüísticas, tales como la comprensión auditiva, lectora, la expresión oral y escrita. La riqueza léxica es entendida como la variedad del vocabulario, la velocidad de acceso, la adecuación y la complejidad en el uso de las palabras que un estudiante es capaz de manejar (Daller, Milton y Treffers-Daller, 2007) y ha sido un objeto de estudio recurrente en la investigación sobre el aprendizaje de lenguas extranjeras. Sin embargo, el crecimiento léxico no es un fenómeno lineal ni uniforme, pues está condicionado por múltiples factores: la exposición al idioma, la frecuencia de uso del vocabulario, las estrategias pedagógicas empleadas por el docente y las técnicas de aprendizaje del estudiante durante su proceso de adquisición del vocabulario.

El presente trabajo constituye un estudio transversal, de enfoque cuantitativo y de amplia cobertura, pues abarca casi la totalidad de los estudiantes de la Licenciatura en Lenguas Modernas en Francés (para futuras referencias: LLMF) de la Universidad Autónoma de Querétaro, México. Esta aproximación ofrece un panorama general de la adquisición del vocabulario de francés como segunda lengua en distintos semestres.

En la LLMF, la lengua extranjera meta no se practica de manera habitual fuera del aula, por lo que la adquisición se ve influenciada por la falta de contacto directo y continuo con el idioma. Es importante considerar que los materiales empleados en la licenciatura, textos literarios, artículos científicos y académicos, tienden a promover la adquisición de un lenguaje de baja frecuencia y especializado, poco utilizado en la comunicación cotidiana. Aunque estos textos, en materias como lingüística, literatura, traducción, docencia, entre otras,

contribuyen al desarrollo de la riqueza léxica de los estudiantes al introducir terminología específica de diversas áreas, surge la cuestión de si esta especialización podría dejar lagunas en el dominio del vocabulario de alta frecuencia, esencial para una comunicación efectiva en contextos generales. Al respecto, se ha demostrado que el vocabulario de alta frecuencia, al ser la base de la comunicación en cualquier idioma, se adquiere más rápidamente en las primeras etapas del aprendizaje (Batista y Horst, 2016; Milton, 2006; Webb y Nation, 2017). Por lo tanto, en esta investigación se analizará no tanto el 'uso' del vocabulario, sino la capacidad de los estudiantes para reconocer el significado de las palabras.

Esta investigación se concentró en la observación de la riqueza léxica en estudiantes en formación como especialistas del francés L2, en función de su grado y la frecuencia del vocabulario. El objetivo central fue evaluar cómo la exposición continua al francés como L2 influye en la adquisición de vocabulario a lo largo de los distintos semestres del programa. Este análisis fue llevado a cabo mediante la implementación de un instrumento de evaluación diseñado específicamente para este estudio, el cual fue elaborado con base en dos pruebas léxicas reconocidas en el campo de la lingüística aplicada: el *Vocabulary Levels Test* (VLT) (Webb, Sasao y Ballance, 2017) y el *Test de la Taille de Vocabulaire* (TTV) (Batista y Horst, 2016). Estas herramientas permiten medir el conocimiento léxico de los estudiantes en función de diferentes bandas de frecuencia léxica y proporciona una visión clara de su dominio del vocabulario tanto de alta como de baja frecuencia.

Un aspecto central de la investigación es examinar cómo los estudiantes reconocen el vocabulario de alta frecuencia en contraste con el de baja frecuencia. La teoría sugiere que los estudiantes tienden a adquirir más fácilmente el

vocabulario de alta frecuencia<sup>1</sup>, dado que estas palabras son más comunes. Sin embargo, debemos considerar que el desafío en la adquisición radica en reconocer el significado básico de palabras que, a pesar de su baja frecuencia en el uso cotidiano, son importantes para el dominio del idioma. Al vocabulario especializado se le considera por lo general de baja frecuencia y representa un mayor desafío para los estudiantes, ya que se encuentra con menor regularidad y requiere un mayor esfuerzo cognitivo para su adquisición (Webb y Nation, 2017; Milton 2009).

Este estudio ofrece una contribución al campo del desarrollo léxico en francés como segunda lengua (L2), ya que aborda de manera detallada la adquisición de vocabulario en estudiantes universitarios en distintos semestres. A diferencia de investigaciones previas como la de Webb, Sasao y Ballance (2017) y Batista y Horst (2016), que se han centrado mayormente en estudios generales<sup>2</sup> sobre léxico, este trabajo proporciona un análisis detallado de cada una de las bandas de frecuencia léxica (1K, 2K, 3K, 5K y 10K) en diferentes semestres de formación académica. Lo que propone este estudio es un desglose a detalle de la adquisición léxica por bandas de frecuencia, lo cual permite detectar patrones y necesidades específicas en el proceso de aprendizaje del vocabulario. Esto no solo proporciona una visión más detallada de la evolución del léxico, sino que permite comprender en qué etapas y bandas específicas los estudiantes requieren mayor apoyo pedagógico. La investigación permite ver si los estudiantes consolidan el vocabulario básico en los primeros semestres, o si presentan una ralentización en el vocabulario de baja frecuencia, representada por la banda 10K. Esta perspectiva segmentada ofrece datos específicos que podrían ayudar a

---

<sup>1</sup> Más adelante en la sección de *Marco Teórico* (p. 16) de esta tesis, se observará a detalle que existe un sistema para diferenciar las palabras por su frecuencia de uso.

<sup>2</sup> Al decir “generales” esto se refiere a que se hace un análisis general de todas las bandas de un grupo, sin embargo en esta tesis se hace un análisis separado por cada una de estas bandas.

identificar áreas de mejora en los programas de enseñanza del francés y optimizar las estrategias de adquisición del vocabulario.

## **1.2 Planteamiento del Problema**

El aprendizaje de una lengua extranjera enfrenta una serie de retos derivados de factores tanto internos como externos al proceso de enseñanza-aprendizaje. A partir de la experiencia docente se ha observado que, en el caso específico del francés como segunda lengua (L2), el vocabulario aprendido en el aula por los estudiantes de la LLMF no corresponde del todo con el que es usado en la vida cotidiana, sino que este se centra principalmente en vocabulario especializado dirigido de acuerdo a sus materias.

A pesar del creciente interés académico sobre el desarrollo léxico en estudiantes de francés L2, existen pocas investigaciones específicas que analicen cómo evoluciona la riqueza léxica del francés en universidades hispanohablantes. Además, los estudios disponibles frecuentemente presentan enfoques globales, sin considerar en detalle el nivel de frecuencia léxica (alta o baja). De igual manera, se pretende identificar patrones específicos de adquisición en función de la frecuencia de las palabras. En este sentido, el presente estudio propone un análisis segmentado según bandas de frecuencia léxica para obtener una comprensión más precisa del desarrollo del vocabulario en estudiantes de la LLMF.

### **1.3 Preguntas de Investigación**

#### ***Pregunta General***

¿Qué nivel de riqueza léxica (tamaño de vocabulario) presentan los estudiantes de la LLMF de la Universidad Autónoma de Querétaro?

#### ***Preguntas Específicas***

¿Existen diferencias significativas en la riqueza léxica entre los estudiantes de distintos semestres de la LLMF?

¿Cómo se relaciona la frecuencia léxica (alta o baja) en el nivel de adquisición de vocabulario en estudiantes de la LLMF?

¿Qué patrones existen en la adquisición de vocabulario de alta y baja frecuencia a lo largo de los diferentes semestres?

### **1.4 Hipótesis**

Se encontrarán diferencias estadísticamente significativas en la riqueza léxica entre estudiantes de distintos semestres, de modo que los estudiantes de semestres avanzados mostrarán mayor riqueza léxica que los de semestres iniciales.

También se espera que el vocabulario de alta frecuencia sea adquirido en mayor medida que el de baja frecuencia.

### **1.5 Objetivos**

#### **General**

Analizar la riqueza léxica en estudiantes de la LLMF.



### **Objetivos Específicos**

1. **Medir la riqueza léxica** de los estudiantes de la LLMF en cada uno de los semestres del programa, mediante pruebas de vocabulario.
2. **Comparar los niveles de riqueza léxica** entre estudiantes de diferentes semestres para determinar si existen diferencias significativas en la adquisición de vocabulario conforme a su grado.
3. **Evaluar la relación** entre las palabras de alta y baja frecuencia y la adquisición léxica de los estudiantes e identificar si existen patrones en el aprendizaje de vocabulario de mayor o menor frecuencia.

## 2. Marco Teórico

En este apartado se abordan los principales conceptos teóricos que sustentan el análisis de la adquisición léxica en estudiantes de francés como segundo idioma (en futuras referencias: L2) . Se exponen las definiciones fundamentales de *palabra*, *familia léxica*, *lema* y las distinciones entre *tipo* y *caso*, conceptos clave para comprender cómo se mide la riqueza léxica. En el último apartado de la sección 2.1 se presenta como la frecuencia de las palabras juega un papel fundamental en el aprendizaje del vocabulario. Autores como Meara (1980) y Laufer y Nation (1995) han desarrollado modelos que organizan las palabras en bandas de frecuencia según su recurrencia en el uso cotidiano. Este enfoque permite evaluar el conocimiento léxico de los estudiantes, desde las palabras de uso más frecuente a las menos frecuentes. Esto otorga una idea más clara de cómo el vocabulario se adquiere más rápidamente debido a su alta frecuencia, por lo que es más probable que sea aprendido fácilmente en las primeras etapas, mientras que las palabras de baja frecuencia requieren mayor esfuerzo y exposición (Webb y Nation, 2017).

En la sección 2.2, se hace una observación de cómo la riqueza léxica se desarrolla en términos receptivos y productivos, resalta la importancia de las palabras como cognados y su impacto en la adquisición léxica (Daulton, 2008).

Por otro lado, se examina la variabilidad en la adquisición léxica entre estudiantes de distintos niveles; se observa que el aprendizaje del vocabulario no sigue una trayectoria lineal, ya que progresa rápidamente en las etapas iniciales con palabras de alta frecuencia, pero se desacelera en niveles más avanzados con términos menos frecuentes.

Asimismo, se ha sugerido que parte de esta variación puede derivar de las necesidades contextuales e inmediatas de los estudiantes; sin embargo, el objetivo principal de esta tesis se orienta hacia una comprensión profunda del significado, más allá de la mera acumulación de léxico (Agustín-Llach, 2022).

Finalmente, se analiza el papel del vocabulario técnico y especializado, y se describe su impacto en la enseñanza de lenguas y el aprendizaje incidental. Particularmente, se enfatiza la importancia de medir y priorizar la frecuencia léxica como herramienta clave para optimizar el aprendizaje, especialmente en contextos educativos como el de los estudiantes de la LLMF, en los que se observa una interacción constante entre el lenguaje oral y escrito.

## **2.1 Conceptos Fundamentales**

### ***2.1.1 Definición del Concepto de Palabra:***

El término “palabra” suele considerarse el concepto más básico en lingüística. De manera práctica, muchas veces se concibe como una secuencia de elementos impresos en la página; sin embargo, Ferreiro y colaboradores (1996) señalan la dificultad de establecer límites precisos debido a la ausencia de una definición universalmente aceptada. Esta dificultad para definir claramente qué es una palabra se relaciona directamente con la medición del vocabulario, ya que según como se delimite, se obtienen resultados distintos en los estudios de riqueza léxica (Daller, Milton y Treffers-Daller, 2007).

Algunos autores proponen diversos criterios para determinar esos límites, pues dependen de cómo se agrupen o clasifiquen las palabras al momento de medir la riqueza léxica. Una posibilidad consiste en concebir la palabra como parte de una *familia léxica*, esto significa que incluye una raíz común y todas sus formas

derivadas (Daller, Milton y Treffers-Daller, 2007). Por ejemplo, la palabra *educar* abarcaría *educación*, *educador* y *educado*. Este planteamiento sugiere que el conocimiento de la raíz permite inferir el significado de sus derivaciones, y resulta útil para evaluar la riqueza léxica, porque mide la habilidad de reconocer y utilizar variaciones morfológicas.

Por otro lado, algunos investigadores proponen medir las palabras a partir del *lema*, entendido como una unidad compuesta por una forma base y sus variantes flexivas más frecuentes (Daller, Milton & Treffers-Daller, 2007). En este caso, *hablar* abarcaría *hablé*, *hablaste* y *hablamos*, pero no incluiría *hablante* o *hablador*. Este criterio facilita el análisis detallado de la forma base y sus variaciones, lo que ofrece comparaciones más precisas entre diferentes lenguas y contextos lingüísticos. Además, no se limita a memorizar una definición aislada, sino que incorpora el conocimiento de las flexiones y su uso apropiado, por lo que esta forma de delimitación de palabras es la que se utilizará en este estudio.

### **2.1.2 El Tipo y el Caso**

Dos conceptos con los que se tiene que estar familiarizado al momento de involucrarse con el análisis de léxico son los de *tipo* (*type* en inglés) y *caso* (*token* en inglés). Diversos autores han utilizado el concepto de *tipo* para designar cada forma única de una palabra tal como aparece en un texto, sin importar sus inflexiones o derivaciones (Webb y Nation, 2017). Por ejemplo, *correr* y *corriendo* se considerarían *tipos* diferentes. En estudios de frecuencia de palabras y de riqueza léxica es uno de los conceptos básicos y se usa muy comúnmente (Daller, Milton y Treffers-Daller, 2007, p.13), ya que permite contar cada forma específica de una palabra que aparece en un corpus textual.

Por otro lado, *caso* se refiere a cada palabra en un texto. Es decir, un *caso* es cualquier secuencia de caracteres que se considera una palabra o un símbolo separado dentro de un corpus de datos. Los *casos* incluyen palabras, números, signos de puntuación, y otros elementos que se identifican como unidades separadas en un análisis textual.

Un ejemplo para explicar la diferencia entre *tipo* y *caso* se puede observar en la oración: "Un zorro corre por un bosque", en la que hay seis *casos*: *Un*, *zorro*, *corre*, *por*, *un*, *bosque*. En este caso, cada palabra se cuenta como un *caso* individual (Daller, Milton y Treffers-Daller, 2007). Incluso si la frase se repitiera en un corpus, cada aparición de las palabras se contaría nuevamente como un *caso* distinto. Mientras que un *tipo* es cada instancia individual de una palabra en un texto (Daller, Milton y Treffers-Daller, 2007). En el ejemplo anterior, aunque existen seis *casos*, se tienen solo cinco *tipos*, ya que la palabra "*un*" aparece dos veces.

Aunque en esta tesis no se observe ni se haga conteos de tipos ni casos su función va más allá de una simple definición, ya que están ligados a la medición de la riqueza léxica, eje central de la tesis.

### **2.1.3 Las Frecuencias de Uso**

Laufer y Nation (1995) desarrollaron una herramienta para analizar cómo se distribuyen las palabras en un texto de acuerdo con su frecuencia de aparición. De acuerdo a este método, las palabras se clasifican en distintas "bandas de frecuencia". Dichas bandas son agrupaciones de palabras que se organizan en conjuntos de 1000 palabras cada uno, y se ordenan según su frecuencia relativa en el lenguaje. Por ejemplo, la primera banda de frecuencia, también conocida como banda 1K, comprende las 1000 palabras más comunes en un determinado idioma. Esto se refiere a las palabras que encontramos con mayor frecuencia en

una amplia variedad de textos y que desempeñan un papel fundamental en la comunicación cotidiana, Las palabras más frecuentes conforman la mayor parte del léxico utilizado en los textos académicos, cubriendo alrededor del 80 %, mientras que las palabras menos frecuentes solo contribuyen con cerca del 5 % del total (Karim, Karim y Chandio, 2019) . Ejemplos de palabras que se encuentran en esta banda son los artículos "el", "la", "los", los pronombres personales "yo", "tú", "él", y los verbos más utilizados como "ser", "estar" y "tener".

A medida que se avanza en las bandas de frecuencia, como la banda 2K, nos encontramos con palabras menos frecuentes pero aún comunes en el lenguaje. Estas son utilizadas con menos regularidad en comparación con las de la banda 1K, pero se consideran dentro de una categoría fundamental para la comunicación. En la banda 2K podemos encontrar palabras como "amigo", "trabajo", "casa" y una variedad de sustantivos, adjetivos y verbos que son ampliamente utilizados en diferentes contextos.

Las bandas posteriores, como la 3K y la 4K, comprenden palabras de frecuencia moderada que se encuentran en un rango intermedio de uso. En esta y otras investigaciones, la banda 4K suele no emplearse de forma directa en las pruebas de medición de léxico, debido a su similitud en porcentaje de frecuencia y aparición con la banda 3K, motivo por el cual generalmente se mide solo la 3K dentro de este rango intermedio, esto se puede observar en pruebas como la de Batista y Horst (2016). Este enfoque busca simplificar el análisis, centrándose en los extremos del espectro de frecuencia. Asimismo, en las pruebas de vocabulario se omiten las bandas intermedias entre la 5K y la 10K para facilitar el estudio de los niveles léxicos más básicos y avanzados en L2.

La creencia generalizada de que las palabras de baja frecuencia se limitan exclusivamente a términos técnicos no es necesariamente precisa (Webb y

Nation, 2017). De hecho, existe un grupo de palabras que pueden encontrarse en publicaciones altamente especializadas y, sin embargo, se consideran de alta frecuencia, por citar un ejemplo tenemos: *programa* o *enchufe*, que son términos de ingeniería. Sin embargo, estos términos son usados frecuentemente por personas fuera de esa área de estudios.

Existe un conjunto de palabras conocidas como vocabulario académico, frecuentes en diversas disciplinas académicas, pero poco comunes en textos generales. Estas palabras, también denominadas sub-vocabulario técnico, no sólo contienen términos puramente especializados, sino que también forman parte en el uso del vocabulario técnico en general (Nation, 2000). Coxhead (2000) establece que las palabras académicas son las que frecuentemente presentan desafíos para los estudiantes, ya que no suelen ser ampliamente conocidas ni enseñadas en cursos basados en contenidos cotidianos. Snow y Uccelli (2009) coinciden en que el vocabulario académico comprende tanto términos específicos de contenido como vocabulario general empleado en contextos educativos para adquirir conocimiento nuevo, describir ideas abstractas y desarrollar la comprensión conceptual de los estudiantes. De forma similar, Zwiers (2008) señala que el vocabulario académico, al carecer del apoyo extralingüístico presente en el lenguaje cotidiano, plantea dificultades significativas al ser más complejo y abstracto. Además, dicho vocabulario es la base para la comprensión y producción de textos académicos debido a sus funciones cognitivas específicas y su uso en estructuras discursivas propias del ámbito escolar (Snow y Uccelli, 2009). Por otra parte, estas palabras suelen pasar desapercibidas como vocabulario especializado en comparación con el resto del vocabulario técnico, lo que las hace menos visibles en la enseñanza explícita y, por lo tanto, potencialmente más problemáticas para los estudiantes.

Coxhead (2000) indica que las palabras académicas representan un porcentaje reducido de aparición en los textos, y estas pueden resultar complicadas para los lectores de L2<sup>3</sup>. Entre ellas se encuentran términos como *cita* (textual), *aislar* y *conferir*<sup>4</sup>. Esto se debe a que estas palabras no son ampliamente conocidas, no se enseñan con frecuencia y suelen ser menos visibles que el vocabulario técnico, lo que representa un desafío adicional para los estudiantes de idiomas. Por lo tanto, es importante reconocer el impacto que el conocimiento de este tipo de vocabulario puede tener en la comprensión de textos académicos y conferencias. Sin embargo, estudios como los de Coxhead (2000) y Nation (2001) han demostrado que, debido a su baja frecuencia y características particulares, el vocabulario académico requiere instrumentos de medición específicos. Por ello, en esta investigación nos centraremos en el léxico general, definido según las bandas de frecuencia mencionadas previamente. Por otra parte, el vocabulario académico se encuentra distribuido en diversas bandas de frecuencia, lo cual dificulta su evaluación mediante los instrumentos diseñados para medir el vocabulario general por bandas específicas. No obstante, es importante mencionar que, aun cuando no se mida en este estudio, dicho vocabulario académico es el que se encuentra con mayor presencia en los alumnos de la LLMF.

<sup>3</sup> Coxhead (2000) menciona que el vocabulario académico representa aproximadamente el 10% del total en los textos académicos en inglés. No obstante, es importante remarcar que no menciona si el fenómeno ocurre de manera similar en los textos académicos en español. Zwiers (2008) añade otra visión al señalar que la dificultad proviene también de su carácter abstracto y su uso en contextos escolares específicos sin apoyo extralingüístico.

<sup>4</sup> Considere un contexto en el que el hablante se limita al uso de términos cotidianos. En este sentido, Coxhead (2000) reflexiona sobre la situación de los hablantes cuya lengua materna (L1) no proviene de las lenguas romances. Se puede apreciar que una parte significativa de los términos en el *Academic Word List* tiene raíces grecolatinas, lo que podría representar un desafío adicional para este grupo de aprendices.



## **2.2 Adquisición y Desarrollo del Vocabulario en una Segunda Lengua**

### ***2.2.1 Los Factores Determinantes en Adquisición de Vocabulario en L2***

El aprendizaje de vocabulario se ve influido por una serie de diferentes factores, entre los que se encuentran la frecuencia y la diversidad léxica, el contexto de exposición y las bases de conocimiento previas del aprendiz. Los estudios sobre frecuencia de palabras han demostrado que el aprendizaje de vocabulario de alta frecuencia, aquel que comprende las palabras más comunes, suele ser prioritario y tiene el mayor impacto en la capacidad de comprensión general de palabras. En francés, por ejemplo, las primeras 2000 palabras más frecuentes abarcan tanto el léxico cotidiano como el académico, esto alcanza aproximadamente el 85% de cobertura léxica de textos en francés (Cobb y Horst, 2004).

En el ámbito de la adquisición de un segundo idioma, la familiaridad con las palabras y sus orígenes juegan un papel importante. Las palabras similares o prestadas, conocidas como cognados, son un punto de partida significativo (Webb y Nation, 2017). Estas palabras, que mantienen una forma y un significado similares en dos idiomas, representan una ventaja notable para los aprendices. Por ejemplo, un hablante nativo de español encontraría poco desafío al aprender la palabra francesa “intéressant” debido a la similitud que tiene con la palabra “interesante” en español (Daulton, 2008). Este fenómeno no es aislado; un gran porcentaje de palabras en español, al igual que el francés, tiene su origen en idiomas como el latín o el griego, por lo que facilita el aprendizaje entre estos idiomas. En el inglés se calcula que aproximadamente la mitad de las 3,000 familias de palabras más frecuentes y de este idioma han sido adoptadas de alguna forma en otros idiomas como en el japonés (Daulton, 2008). Estos

llamados préstamos, una vez asimilados, toman significados y usos adaptados a los usos locales.

Una ventaja adicional de los préstamos es que muchas palabras provenientes del latín y del griego contienen componentes, como raíces o afijos, que modifican de manera similar su significado en idiomas como el francés o el español. Por ejemplo, saber que el prefijo 'in' denota negación en ambos idiomas resulta útil, como se evidencia en palabras tales como 'incapable' en francés o 'innecesario' en español. Esta comprensión de las partes de las palabras ayuda a los estudiantes a reducir la carga de aprendizaje a largo plazo, ya que les permite analizar palabras desconocidas y deducir sus posibles significados.

Nation (2001) indica que, al explorar las definiciones y la etimología (a través de diccionarios o glosarios), los aprendices pueden construir conexiones semánticas que facilitan la memorización y el uso adecuado en diferentes contextos, pero en este método no siempre muestra siempre la forma real en la que se usa por los nativo-hablantes. Un ejemplo de ello en francés es el verbo «trouver». Generalmente, un diccionario traduce este verbo como «encontrar». No obstante, en situaciones cotidianas, los hablantes nativos le dan otro uso, como en la frase «Je trouve ça intéressant», equivalente en español a «Me parece interesante», en este caso se puede apreciar como el significado original del verbo se pierde para ser cambiado por otro.

La adquisición de vocabulario en L2 va más allá del reconocimiento superficial y la memorización de palabras aisladas; implica el uso contextual, la comprensión de las relaciones semánticas y la habilidad para emplear el léxico en situaciones comunicativas reales (Webb y Nation, 2017). Según Nation (2001) este aprendizaje se divide en conocimiento receptivo (esencial para la comprensión auditiva y lectora) y conocimiento productivo (imprescindible para la

expresión oral y escrita). En las primeras etapas, el vocabulario activo, es decir, el que se utiliza en la producción, es más limitado que el pasivo, que comprende las palabras que el estudiante puede reconocer y entender. Ambos tipos resultan importantes para alcanzar una competencia lingüística completa en L2 (Laufer, 1998).

En esta tesis se da relevancia al estudio del vocabulario a través del análisis de bandas de frecuencia léxica, consideradas aquí como una herramienta que ayuda a evaluar el desarrollo léxico y, en el futuro, podría guiar la instrucción de acuerdo con las necesidades lingüísticas específicas de los estudiantes. Además, autores como Laufer y Nation (1995) sostienen que la producción de palabras menos frecuentes es un claro indicador de progreso en la adquisición de una L2, pues refleja la ampliación del repertorio hacia términos técnicos y especializados. En este marco, instrumentos como el Test de la *Taille du Vocabulaire* (TTV) y su homólogo en inglés, el *Vocabulary Levels Test* (VLT), permiten medir la amplitud léxica en función de la frecuencia de uso, esto facilita el análisis de la riqueza del vocabulario. Dichos recursos evalúan tanto el vocabulario de alta frecuencia, esencial para la fluidez básica, como el de baja frecuencia, que enriquece la precisión y sofisticación en el uso del idioma.

Además, la enseñanza explícita del vocabulario complementa el aprendizaje incidental, entendido como la adquisición de nuevas palabras como resultado de actividades centradas en el significado (por ejemplo, lectura o escucha) y sin la intención explícita de memorizar vocabulario. Aunque el léxico puede adquirirse mediante la exposición continua a la lectura y las clases en la licenciatura, se ha demostrado que la retención a largo plazo es más efectiva cuando se incorpora instrucción sistemática (Agustín-Llach, 2022). En este sentido, Lewis (1993) enfatiza que el aprendizaje de vocabulario no debe

depender únicamente de la exposición, sino que requiere un enfoque estructurado que facilite su consolidación. Asimismo, destaca que la enseñanza de unidades léxicas completas (*chunks*) favorece la retención y el uso natural del idioma, al permitir a los estudiantes reconocer y emplear combinaciones predecibles de palabras en distintos contextos. Por ello, tanto la instrucción explícita como el uso de estrategias intencionales resultan efectivas para fortalecer la adquisición léxica y mejorar la competencia lingüística en una segunda lengua (Lewis, 1993; Agustín-Llach, 2022). Además, la inclusión de listas de frecuencia y la selección de vocabulario relevante para contextos específicos ayudan a los estudiantes a priorizar términos de mayor utilidad en su aprendizaje y comunicación cotidiana.

Un aspecto aparte en la adquisición de vocabulario es la importancia del análisis morfológico, que abarca el estudio de prefijos, sufijos y raíces. La capacidad de descomponer palabras en sus componentes morfológicos ofrece una ventaja al aprender vocabulario especializado o técnico, sobre todo en contextos académicos, en los que la precisión léxica resulta una de las bases para el dominio del lenguaje (Webb, Sasao y Ballance, 2017). Además, este conocimiento favorece el desarrollo de estrategias metalingüísticas, mediante las cuales el estudiante puede transferir habilidades de deducción léxica de su lengua materna a la L2.

Finalmente, la comprensión profunda de una palabra incluye no solo su significado, sino también sus asociaciones semánticas, colocaciones y connotaciones contextuales. Treffers-Daller (2022) y Qian (2002) hacen la observación de que los estudiantes avanzados de L2 a menudo presentan una mayor riqueza léxica y capacidad para emplear vocabulario preciso en distintos contextos debido a estas capacidades mencionadas. Este nivel de conocimiento

se desarrolla después de mucho tiempo y requiere de una exposición continua a situaciones de uso real del idioma.

### ***2.2.2 La Variabilidad en la Adquisición de Vocabulario en L2***

Para este estudio es importante comprender el ritmo y el alcance del crecimiento del vocabulario entre los estudiantes de L2 en diferentes períodos. Aunque la investigación en esta área aún no ha alcanzado un consenso metodológico, los estudios ya existentes concuerdan en que la adquisición de vocabulario en un segundo idioma no sigue una trayectoria lineal, sino que se desacelera a medida que avanza el proceso de aprendizaje. En las etapas iniciales, el aprendizaje de palabras frecuentes ocurre con mayor rapidez debido a su presencia constante y valor comunicativo. Sin embargo, en niveles más altos, la incorporación de términos menos frecuentes se vuelve más lenta y requiere tanto una exposición prolongada como estrategias de aprendizaje más deliberadas. (Webb y Nation, 2017; Laufer, 1998; Schmitt, 1997). Esta variabilidad sugiere la necesidad de emplear métodos de evaluación que combinen análisis por bandas de frecuencia con enfoques cualitativos para obtener una visión más precisa del desarrollo léxico.

Existe una diversidad entre los métodos utilizados para contar y medir el vocabulario. Estos métodos incluyen contar por lemas, familias de palabras o tipos de palabras (Daller, Milton y Treffers-Daller, 2007), cada uno ofrece una lente diferente a través de la cual se evalúa el crecimiento léxico. Esta variación en las unidades de medida ha generado controversias y desafíos a la hora de generalizar los hallazgos en diferentes contextos de aprendizaje (Webb y Nation, 2017),

debido a que muchos autores deciden hacer sus mediciones con herramientas y cálculos diferentes<sup>5</sup>.

A pesar de lo mencionado, un resultado documentado en la investigación sobre adquisición de vocabulario en L2 es que en las primeras etapas del aprendizaje, los estudiantes incorporan con relativa rapidez palabras de alta frecuencia, ya que estas aparecen con mayor regularidad en el input lingüístico y están directamente relacionadas con las necesidades comunicativas básicas (Webb & Nation, 2017). Sin embargo, conforme avanzan en su formación, la cantidad de vocabulario nuevo que adquieren disminuye debido a varios factores interrelacionados. En primer lugar, las palabras menos frecuentes requieren mayor número de exposiciones para ser adquiridas, lo que incrementa la carga cognitiva del aprendizaje (Laufer, 1998). En segundo lugar, a medida que los estudiantes desarrollan una base léxica más sólida, la redundancia en el input aumenta, es decir, encuentran con mayor frecuencia palabras que ya conocen en lugar de términos nuevos, lo que limita la oportunidad de expansión léxica (Schmitt, 1997). Además, la producción y retención de vocabulario de baja frecuencia está condicionada por su menor utilidad en la comunicación cotidiana, lo que puede llevar a su olvido si no es reforzado con estrategias explícitas de aprendizaje (Milton, 2009). Por lo tanto, la disminución en la tasa de crecimiento léxico en estudiantes avanzados no solo responde a la frecuencia de exposición, sino también a las limitaciones en la memoria, la competencia discursiva y la necesidad de mayor procesamiento para consolidar palabras en contextos de uso natural.

### ***2.2.3 El Valor de las Palabras Frecuentes en el Aprendizaje de un Idioma***

<sup>5</sup> Esto se puede observar en la sección de *Antecedentes*, donde se reportan varios tipos de cálculos de riqueza léxica y pruebas para medir el vocabulario.

En algunos métodos de enseñanza se piensa que, para lograr dominar un idioma de manera eficiente, hay que entender el papel de las palabras frecuentes en el proceso de aprendizaje. Es común ver que en los manuales de L2 se tengan como prioritarias palabras específicas de alta frecuencia, por lo que se puede considerar que las palabras frecuentes son más valiosas en el aprendizaje de un idioma. En este sentido, la relevancia de medir su frecuencia y su clasificación en categorías puede potenciar el desarrollo lingüístico de los estudiantes. Un ejemplo de esta distinción se puede observar en exámenes como el DELF<sup>6</sup>. En el nivel A1, se evalúa el conocimiento de palabras de alta frecuencia, como vocabulario relacionado con el hogar y verbos de uso cotidiano. En contraste, en el DALF, se espera que el examinado posea un dominio de palabras de baja frecuencia, generalmente asociadas con un lenguaje más especializado y un bagaje cultural amplio.

Las palabras frecuentes son aquellas que aparecen con mayor regularidad en la lengua, tanto en el habla como en la escritura. Aprender estas palabras es la base para la construcción de oraciones y la comprensión de textos cotidianos (Webb y Nation, 2017). Por ejemplo, el conocimiento de palabras como “esto”, “agua” o “caminar” es más útil que vocabulario menos frecuente como “fluctuante”, “escalada” o “desfibrilador”. Es evidente que las palabras frecuentes son más valiosas para los estudiantes porque resulta más probable que se necesiten para la comunicación cotidiana.

Medir la frecuencia de las palabras resulta una de las bases para priorizar el vocabulario que conviene aprender en las primeras etapas, ya que la mayoría de

<sup>6</sup> El **DELF** (Diplôme d'Études en Langue Française) y el **DALF** (Diplôme Approfondi de Langue Française) son certificaciones oficiales de dominio del idioma francés, otorgadas por el Ministerio de Educación de Francia. El DELF evalúa competencias en niveles básicos e intermedios (A1-B2), mientras que el DALF está dirigido a niveles avanzados (C1-C2) según el Marco Común Europeo de Referencia para las Lenguas (MCER).

la comunicación se sostiene sobre un conjunto reducido de términos de alta frecuencia (Nation, 2001). Gracias a este criterio, se han diseñado actividades y materiales ajustados a las necesidades y al nivel de los estudiantes, que favorecen un aprendizaje más eficaz. Al centrarse en palabras que aparecen con más regularidad en contextos reales, los alumnos adquieren con rapidez una base léxica que les permite abarcar un gran porcentaje de la comunicación.

La importancia de aprender palabras según su frecuencia se destaca aún más al examinar cómo se distribuye el vocabulario en textos reales. La *Ley de Zipf* establece que las palabras en un idioma siguen un patrón de frecuencia decreciente: unas pocas palabras muy utilizadas constituyen la mayor parte del lenguaje cotidiano, mientras que el resto de las palabras aparecen con menor frecuencia (Zipf, 1949).

Este patrón que propone la *Ley de Zipf* no solo se observa en el lenguaje, sino también en otros fenómenos naturales y sistemas complejos, como la distribución del tamaño de las ciudades o la frecuencia de uso de ciertos patrones en redes sociales. En lingüística, esta ley explica que el lenguaje humano está organizado de manera eficiente para minimizar el esfuerzo cognitivo y de producción. Las palabras más frecuentes tienden a ser cortas y versátiles, mientras que las menos frecuentes suelen ser más largas y especializadas, esto facilita una comunicación efectiva con un número limitado de palabras esenciales (Zipf, 1949).

La ventaja de los estudiantes que cursan una licenciatura en francés es que tienen la oportunidad de escuchar a sus profesores y adquirir un vocabulario basado en el lenguaje oral; por otro lado, el lenguaje escrito se encuentra en la exposición cotidiana a la lectura y escritura. Las lecturas en francés realizadas en sus cursos enriquecen su conocimiento léxico, permiten la comprensión de



términos técnicos, expresiones idiomáticas y el uso contextual de las palabras en diferentes géneros textuales.

Un factor de adquisición de nuevo léxico proviene del proceso de lectura; esta aproximación se encuentra en algunos programas educativos. La investigación ha mostrado que las posibilidades de que un hablante nativo de un idioma reconozca el significado de una nueva palabra cuando la encuentra casualmente durante la lectura son extremadamente bajas, apenas un 0,07% (Cobb y Horst, 2004) . Aunque este ritmo es suficiente para que los hablantes nativos desarrollen un vocabulario completo en francés a lo largo de su vida, los estudiantes de L2 a menudo no se sumergen en la lectura extensa. De hecho, por lo general los documentos de dichos estudiantes se encuentran alejados de la lengua hablada en Francia (Cobb y Horst, 2004).

A diferencia de los hablantes nativos, que suelen encontrar palabras nuevas rodeadas de vocabulario familiar, los aprendices de L2 a menudo tropiezan con varios términos desconocidos simultáneamente dentro de un mismo pasaje, lo cual dificulta aún más la comprensión y la inferencia del significado (Cobb y Horst, 2004; Laufer, 2003). Se ha mostrado que se necesita una proporción mínima de palabras conocidas y desconocidas de al menos 20:1 para lograr una comprensión confiable y la adquisición de nuevas palabras (Cobb y Horst, 2004), lo que significa que al menos el 95% de las palabras en el contexto deben ser conocidas (Cobb y Horst, 2004; Laufer, 2003). El cumplir con esta condición es menos probable cuando los estudiantes de L2 leen textos no simplificados, lo que representa un obstáculo significativo para la adquisición natural de vocabulario.

Para comprender cómo progresa un estudiante típico hacia la comprensión del 95% de las palabras en un texto promedio, es fundamental considerar cómo

ocurre normalmente la adquisición de vocabulario. Los estudiantes tienden a aprender vocabulario en L2 en un orden aproximado de frecuencia. Las primeras 1,000 familias de palabras más comunes, que representan aproximadamente el 75% de las palabras que se usan en el lenguaje hablado y en diversos textos escritos, suelen ser las primeras en adquirirse. Luego, se adquieren las siguientes 1,000 familias de palabras más frecuentes, que son menos comunes en el habla pero más comunes en la escritura, y contribuyen a otro 5% de las palabras que se encuentran con regularidad en los textos. Esto significa que con el conocimiento de sólo 2,000 familias de palabras, un estudiante ya comprende aproximadamente el 80% de las palabras en un texto promedio (Cobb y Horst, 2004).

Sin embargo, el camino desde el 80% hasta el 95% de comprensión no es sencillo y no sigue una progresión lineal, la adquisición se comporta de forma exponencialmente decreciente. Con esto se quiere decir que las familias de K1 son usadas cientos de veces más que una banda alejada como la K5.

Para concluir esta sección, la riqueza léxica se define como una dimensión del conocimiento lingüístico en lengua. Se refiere a la variedad del léxico que un hablante o escritor es capaz de utilizar en situaciones comunicativas específicas, ya sea en la producción oral o escrita (Laufer y Nation, 1995; Milton, 2009; Daller, Milton y Treffers-Daller, 2007). En esta investigación, el concepto de riqueza léxica es el núcleo central que nos sirve para examinar los patrones de adquisición léxica en estudiantes universitarios de francés L2, en función de su grado universitario.

## **2.3 Consideraciones en los Estudios Sobre Léxico**

### ***2.3.1 Las Pruebas para Medir el Léxico***

Desde el siglo pasado se han creado varias pruebas con el objetivo de evaluar el conocimiento de estudiantes de L2. Uno de los enfoques más conocidos

es el Marco Común Europeo de Referencia para las Lenguas (MCER), el cual busca definir las destrezas lingüísticas de los estudiantes en una escala de niveles de diferentes lenguas que va desde un A1, nivel básico, hasta un C2, el nivel más alto de competencia. Sin embargo, estas pruebas a menudo sólo se centran en medir las capacidades del estudiante para desenvolverse en situaciones concretas, sin tener en cuenta realmente la profundidad o precisión del vocabulario.

Se han concebido pruebas que específicamente han sido desarrolladas para evaluar la competencia léxica en un segundo idioma, tales como las pruebas de tipo *YES/NO*. Algunos estudios, como el de Meara y Buxton (1987) en inglés L1 o Milton (2006) en francés como L2, han empleado este tipo de pruebas para evaluar la riqueza léxica de los estudiantes. No obstante, es importante destacar que estas pruebas, aunque útiles hasta cierto grado, no profundizan en algunos aspectos de los que se hablarán a continuación.

Si bien la prueba *YES/NO* ha mostrado ser útil para evaluar la competencia comunicativa general de una lengua, su enfoque amplio puede obviar la relevancia del vocabulario en el *dominio efectivo*<sup>7</sup>: la habilidad de usar las palabras con precisión y adecuación contextual (Nation, 2001). Por ejemplo, se puede aprender la palabra “Mirabel” que puede ser tomado solo como el nombre de una persona, pero también implica una fruta específica de la región de Lorena, Francia, por lo tanto, una persona puede reconocer la palabra, pero no saber realmente qué es o si existe algún otro significado (Rodríguez, 2022). El vocabulario desempeña un papel fundamental en el dominio de un idioma, y tener un conocimiento profundo y

<sup>7</sup> La capacidad real y funcional que tiene un hablante para utilizar el vocabulario de una lengua de manera adecuada en contextos comunicativos diversos, más allá del conocimiento pasivo o teórico de las palabras (Daller, Milton y Treffers-Daller, 2007).

preciso de las palabras y expresiones es esencial para lograr una comunicación efectiva.

Si se toma en cuenta lo mencionado anteriormente, resulta necesario desarrollar métodos de evaluación con un mayor énfasis en el conocimiento del vocabulario. En este estudio, sin embargo, sólo se evalúa la capacidad del estudiante para reconocer y emplear palabras de ciertas bandas de frecuencia, sin abarcar la medición de múltiples significados, sinónimos o antónimos ni su aplicación en contextos diversos.

### ***2.3.2 El Rol del Vocabulario en la Transmisión de Temas e Intenciones***

Nation (2000) explica que el vocabulario usado puede denotar la intención del autor y el tipo de documento. De esta forma, cuando los escritores emplean un vocabulario particular relacionado con un campo o tema específico, dichas palabras sirven como una señal para los lectores sobre el enfoque central del texto. Por ejemplo, en un artículo científico sobre el cambio climático, el uso frecuente de palabras como "calentamiento global", "emisiones de carbono" e "impacto ambiental" transmite claramente el tema a los lectores. Por otra parte, el vocabulario muestra la formalidad del texto al incorporar léxico académico y etimologías greco-latinas. Esto demuestra la intención del escritor de presentar la información de manera académica y enfocar a una audiencia en específico, que al mismo tiempo espera cierto nivel de discurso.

En el plano oral, Nation (2000) también señala que la selección léxica durante una presentación o conferencia puede guiar la percepción de la audiencia sobre la seriedad o el objetivo de la exposición. Cuando el emisor opta por terminología especializada o tecnicismos en su discurso oral, refuerza tanto su autoridad en la materia como limita la audiencia que realmente podrá seguirlo. Del

mismo modo, el uso de vocabulario coloquial o más cercano a la lengua cotidiana, incluso en ámbitos formales, puede generar empatía y facilitar la comprensión para un público diverso, sin que esto implique perder exactitud en los conceptos.

Asimismo, algunos textos reflejan una naturaleza técnica a través de la inclusión de terminología especializada y detallada, como ocurre en campos científicos o académicos. El uso de esta terminología específica dentro de un área particular no solo demuestra el nivel de conocimiento técnico necesario para comprender el tema, sino que también estructura y delimita el discurso dentro de dicho campo. Por ejemplo, en un artículo de investigación médica, la aparición de términos como *neurotransmisor*, *placebo* y *vesícula* ilustra claramente la tecnicidad y profundidad del contenido (Nation, 2000).

Del mismo modo que la terminología técnica puede delimitar el conocimiento en campos específicos, el vocabulario también puede ser un reflejo de la posición ideológica del comunicador. Las palabras seleccionadas por el escritor no solo transmiten información, sino también sus creencias, valores y posturas políticas o sociales (Fairclough, 1989). Por ejemplo, el uso de términos como “progreso”, “reforma” o “libertad” en un artículo político puede evidenciar la orientación ideológica del autor, ya que estas palabras llevan connotaciones que trascienden su significado literal.

Este matiz ideológico se vuelve palpable también en la oralidad a través de la entonación, la velocidad de habla y la manera en que el emisor recalca ciertas palabras (Fairclough, 1989). Un ponente, por ejemplo, podría enfatizar repetidamente palabras clave en un discurso político o emplear un campo semántico concreto como “democracia” o “justicia” para subrayar su postura y persuadir a la audiencia de forma más directa que en un texto escrito.

Por lo que se ha visto, las elecciones de vocabulario pueden reflejar la intención del escritor de hacer que el texto sea accesible o inaccesible para ciertos lectores. Los escritores pueden decidir usar un vocabulario que atienda a una audiencia específica o se dirija a un grupo de lectores que posea conocimientos especializados. Esta selección deliberada de vocabulario y la inclusión de definiciones dentro del texto pueden facilitar la comprensión de lectores específicos o crear una barrera para aquellos que no están familiarizados con el tema (Fairclough, 1989; Nation, 2000).

## **2.4 El Vocabulario Técnico y sus Categorías**

Nation (2000) plantea que, para estudiar el vocabulario, no basta con elaborar listas que muestren qué palabras son más frecuentes. Además, propone clasificar las palabras según los lugares o ámbitos en los que tienden a aparecer con mayor frecuencia; por ejemplo, si se usan sobre todo en contextos académicos, técnicos o en la vida cotidiana. De este modo, podemos entender no solo cuántas veces aparece una palabra, sino también en qué clases de textos o situaciones es más probable encontrarla.

La primera categoría del vocabulario técnico incluye aquellas palabras que rara vez se utilizan fuera de un campo específico. Estas palabras son técnicas y especializadas, su significado está delimitado a un ámbito específico. Para poner un ejemplo, en el campo de la lingüística, se encuentran términos como "morfema", "lema" y "fonema", que son palabras que se utilizan principalmente dentro de la disciplina y tienen un significado preciso y técnico. Estas palabras son críticas al momento de hablar en la lingüística y es poco probable que se encuentren en conversaciones cotidianas o en otras áreas de estudio sin relación con dicha disciplina .

La segunda categoría propuesta por Nation (2000) abarca palabras que se usan tanto dentro como fuera del campo, pero que adquieren significados diferentes según el contexto. En lingüística, encontramos términos como "sentido", "tono" y "raíz". Estas palabras tienen significados específicos cuando se utilizan en el ámbito de la lingüística, pero también se emplean en otros contextos con connotaciones distintas. Por ejemplo, mientras en el campo lingüístico «raíz» hace referencia a la base morfológica de una palabra, es decir, la unidad mínima portadora de significado a partir de la cual se forman otras palabras mediante afijos, en otros contextos como la botánica se refiere al órgano subterráneo de las plantas, y en matemáticas al valor que, al multiplicarse por sí mismo, da como resultado un número determinado. Este tipo de vocabulario, que se sitúa entre lo general y lo técnico, recibe el nombre de vocabulario semitécnico. (Nation, 2000).

La tercera categoría incluye palabras que se usan tanto dentro como fuera del campo, pero que son préstamos de otros campos o disciplinas. A diferencia del campo anterior, el origen de estos términos es más marcado, estas palabras son tomadas prestadas y adaptadas para su uso en un ámbito específico. En lingüística, encontramos palabras como "frecuencia", "rango" y "longitud", que tienen su origen en el campo de la física, pero adquieren un significado específico dentro del estudio del lenguaje.

Finalmente, la cuarta categoría abarca palabras que son frecuentes dentro del un campo de estudio, pero que no son especializadas ni técnicas. Es un léxico de uso general y se utiliza tanto dentro como fuera del campo de estudio. En el ámbito de la lingüística, encontramos palabras como "decir", "forma" y "sonido", que son términos usados comúnmente en el estudio formal del lenguaje básico pero que también se emplean en conversaciones cotidianas.

Por lo tanto, si se toma en consideración lo expuesto en este capítulo, la presente tesis se centrará en el análisis del desarrollo de la riqueza léxica en estudiantes de laLLMF; además, se examinará el impacto de la frecuencia del vocabulario en su adquisición y se evaluarán las diferencias en el crecimiento léxico entre los distintos semestres.



### **3. Antecedentes**

En este capítulo se expondrán algunos estudios previos que han tenido como objetivo medir la riqueza léxica, específicamente en la lengua escrita, con investigaciones de finales del siglo XX e inicios del siglo XXI. Para lograr esto, se observarán los distintos enfoques que se han abordado, se partirá de lo más general a lo más específico. En primer lugar, se revisarán aquellos estudios cuyo propósito ha sido medir la riqueza léxica tanto en L1 como en L2, mediante el conteo de palabras y el uso de diferentes fórmulas para evaluarla. Finalmente, se presentarán los estudios más relevantes para esta investigación, los cuales emplean pruebas de medición de léxico, primero con los de L1 y por último los enfocados en L2 y que utilizan metodologías similares a la usada en esta tesis.

En las últimas décadas, el incremento de las capacidades tecnológicas ha permitido analizar de manera más exhaustiva los elementos básicos del léxico y sus relaciones con la adquisición y la lengua. A fin de contextualizar la pertinencia de esta tesis, a continuación se presentan las siguientes investigaciones.

#### **3.1 La Medición de la Riqueza Léxica por Conteo de Palabras y Uso de Formulas.**

##### ***3.1.1 La Medición de la Riqueza Léxica en L1***

El primer estudio a observar es el de Laimí Álvarez (2019), quien se centró en analizar la riqueza léxica en la lengua materna de dos grupos de estudiantes cubanos de diferentes escuelas con dos diferentes niveles socioeconómicos. Para ello, se analizaron 20 textos escritos por estudiantes de 11 años y otros 20 textos escritos por alumnos de 16 años (pre-universitarios) de cada una de las dos

escuelas. Con el propósito de calcular los índices de riqueza léxica, se emplearon dos cálculos distintos: el de López Morales (1982) y el de Lidia Cuba (2006). El autor utiliza estas dos fórmulas con el fin de mostrar las grandes diferencias de datos que arrojan ambos cálculos.

La fórmula de López Morales (1982), basada en el calculo de *relación tipo/caso* usada por Johnson (1944), que mide la proporción de palabras diferentes (vocablos) en un texto en relación con el total de palabras. Esta fórmula incluye todas las palabras y proporciona un índice de diversidad léxica general. En contraste, la fórmula que emplea Lidia Cuba (2006) se enfoca en el porcentaje de vocablos de contenido nocional (PVN), es decir, aquellas palabras que tienen contenido semántico significativo (sustantivos, verbos, adjetivos, etc.), las palabras funcionales como artículos, preposiciones y pronombres son excluidas.

Los resultados del estudio de Laimi Álvarez (2019) brindan dos hallazgos: la influencia del factor socioeconómico en el nivel léxico de los estudiantes y la gran diferencia de porcentajes obtenidos al usar dos fórmulas distintas para calcular el índice de riqueza léxica. En la muestra de estudiantes de 11 años, se observa un índice de riqueza léxica del 40.39% en la escuela privada y del 37.36% en la escuela pública al utilizar la fórmula de López Morales. En contraste, al utilizar la fórmula de Lidia Cuba, se obtiene un índice del 63.12% en la escuela privada y del 62.58% en la escuela pública. En los estudiantes de 16 años, la primera fórmula arroja un índice de riqueza léxica del 43.20% en la escuela privada y del 41.57% en la escuela pública, mientras que la segunda fórmula da como resultado un índice del 66.21% en la escuela pública y del 65.63% en la escuela privada. En todos los resultados se puede observar una diferencia mayor a 20 %.

Es importante señalar esta diferencia porque muestra cómo los cambios en los cálculos pueden modificar los valores que reflejan la riqueza léxica, y permite ilustrar las variaciones en los resultados de acuerdo con el enfoque aplicado.

### ***3.1.2 Las Mediciones de la Riqueza Léxica en L2***

El estudio de Eva Thue Vold (2022) analizó la riqueza léxica de estudiantes noruegos de 16 a 18 años en dos escuelas diferentes, quienes ya habían recibido 335 horas de clases de francés como L2. Para este análisis, se emplearon dos medidas de riqueza léxica: la sofisticación léxica, evaluada a través de las bandas de frecuencia (K) de las palabras empleadas en los textos, y la diversidad léxica, medida mediante la puntuación D. Esta última mide la relación entre el número de palabras diferentes (types) y el total de palabras producidas (tokens), junto con el conteo de familias léxicas en las primeras bandas de frecuencia.

El estudio incluía pruebas escritas, en las que los estudiantes redactaron textos de tres tipos en francés (narrativos, argumentativos e informativos<sup>8</sup>) sobre un tema específico. Dichas producciones fueron analizadas con un software de análisis léxico para obtener los valores de K y D. Además, se emplearon grabaciones de audio para evaluar la riqueza léxica en la producción oral de los estudiantes, quienes hablaron sobre un tema dado en francés. Estas grabaciones también se transcribieron y se analizaron con el mismo software. Los datos se recopilaban a lo largo de seis meses, en cerca de 60 horas totales.

Los resultados del estudio de Eva Thue Vold (2022) revelaron que los estudiantes noruegos de francés como L2 muestran un aumento limitado en la diversidad léxica y dependían de vocabulario de alta frecuencia, sin evidencias

<sup>8</sup>Los estudiantes podían elegir entre varias tareas de escritura, por ejemplo: describir una visita al Festival de Cannes, reflexionar sobre la importancia del cine en la sociedad francesa, o continuar una historia relacionada con un robo.

claras de progreso significativo en vocabulario menos común. Estos hallazgos sugieren que, en contextos con exposición limitada al idioma, la adquisición de vocabulario avanza con mayor lentitud y requiere medidas pedagógicas más especializadas que extiendan el aprendizaje más allá del lenguaje empleado con frecuencia en el salón de clases.

En el mismo estudio, Vold (2022) discutió sobre la adquisición de léxico en L2 y el impacto de la pedagogía en el índice de riqueza léxica. Resaltó la relevancia de realizar un análisis cuidadoso de la muestra, pues factores como el tipo y nivel de curso, la elaboración de materiales, las expectativas y la experiencia del docente, junto con las experiencias previas de los estudiantes, influyen en la forma en que se desarrolla el aprendizaje del idioma.

Para evitar variables indeseadas al medir la riqueza léxica, es importante seleccionar la herramienta de medición adecuada, pues esta elección influye significativamente en los resultados y en la interpretación de los datos. Esto se observa en un estudio realizado por Treffers-Daller (2013), quien analizó la riqueza léxica en hablantes de inglés que aprenden francés (L2). Participaron 21 estudiantes de primer año de universidad (de licenciatura, mencionados como "nivel 1") y 20 estudiantes de último año (nivel 3) de estudios en francés, además de un grupo control compuesto por 23 hablantes nativos de francés. Se utilizaron dos tiras cómicas de las historias de padre e hijo de Plaue y se solicitó a cada estudiante narrar la historia con sus propias palabras; se les concedió la oportunidad de practicar antes de la grabación.

Los resultados muestran que, en promedio, los estudiantes produjeron 96 tipos lematizados<sup>9</sup> y 325 tokens por historia. Se observó que los datos lematizados tendían a obtener puntajes más bajos en comparación con los datos no

<sup>9</sup>Refiriéndose al proceso de reducir una palabra a su forma base o raíz, conocida como "lema".

lematizados. Sin embargo, en la transformación logarítmica propuesta por MAAS<sup>10</sup>, no se apreció un efecto relevante de la lematización. La medida MTLD mostró una mayor variación según la longitud del texto, en contraste con otras medidas analizadas. En los textos que registran una extensión de 200 a 666 palabras, la medida MAAS presenta mejoras significativas. Se eligió el cálculo de Herdan's C como el mejor predictor de la riqueza léxica del grupo, seguido por la D de Guiraud, mientras que MAAS y MTLD resultaron menos acertadas. TTR fue la menos eficaz. Treffers-Daller (2013) sostiene que, al realizar un análisis mediante conteo bruto de palabras<sup>11</sup>, los resultados pueden verse afectados por múltiples factores que generen variables indeseadas, lo cual arrojaría datos inexactos para una investigación.

La investigación de Treffers-Daller (2011) hace énfasis en la importancia de seleccionar herramientas precisas para medir la riqueza léxica en estudiantes de francés como L2. Se subraya que la lematización es ideal en idiomas altamente flexivos como el francés — motivo por el cual esta tesis opta por un método mediante pruebas de vocabulario—, pues permite obtener resultados más fiables y evita inflaciones en las métricas de vocabulario.

### **3.2 La Medición de Riqueza Léxica Mediante Pruebas**

En el apartado anterior se mostró cómo algunos autores emplearon el conteo bruto de palabras para medir la riqueza léxica. No obstante, se observó que no existe consenso respecto a las fórmulas y métodos utilizados. Por ello, en el este apartado se abordará cómo diversos autores han diseñado pruebas

<sup>10</sup>Para futuras referencias: D hace referencia a “Guiraud's Root Type”, MTLD a “Measure of Textual Lexical Diversity”, MAAS a “Moving-Average Type-Token Ratio”, y HD-D a “Herdan's C”, TTR “Type-Token Ratio”. Todos estos son métodos usados por investigadores para hacer cálculos de riqueza léxica. Estos cálculos no se profundizarán en este estudio debido a que se desvían de la investigación principal.

<sup>11</sup> El concepto de conteo bruto se refiere al conteo manual de tipos y casos en un texto.

específicas para evaluar la riqueza léxica, así como la aplicación y validación de dichas pruebas en distintos contextos.

### **3.2.1 El Uso de Pruebas para Medición de Riqueza Léxica**

Para evaluar la eficacia y validez de las pruebas de vocabulario, Meara y Buxton (1987) llevaron a cabo un experimento que incluyó a 100 estudiantes de Reino Unido. Se les aplicó una prueba reconocida (Cambridge First Certificate Examination) y una prueba que se le conoce como *test YES/NO*, es decir, una prueba de vocabulario en la cual se les presenta una lista de 20 palabras reales por cada rango de frecuencia (con un total de 100 palabras) y 20 palabras inventadas, con un total de 120 palabras. Se les preguntaba a los estudiantes si conocían cada palabra o no, con el fin de comparar resultados.

Este estudio demostró que el *test YES/NO* ofrece una alternativa eficaz y más práctica que las pruebas de opción múltiple para evaluar el conocimiento de vocabulario, ya que se pudo observar una alta correlación con el examen tradicional y una mejor capacidad de discriminación entre niveles de competencia.

James Milton (2006), uno de los investigadores con mayor experiencia en el análisis de la riqueza léxica, publicó el artículo *Language Lite? Learning French Vocabulary in School*. El objetivo fue investigar el aprendizaje de vocabulario en francés como lengua extranjera en estudiantes de una escuela secundaria británica, en un análisis longitudinal de 7 años. Empleó el método X\_Lex, un *test YES/NO* con listas de palabras reales y palabras inventadas.

Los resultados revelaron que los estudiantes aprendían alrededor de 170 palabras por año hasta el nivel GCSE (General Certificate of Secondary Education), y cerca de 535 palabras por año entre GCSE y el nivel “A” (Advanced Education).

level). Aun así, se detectó un progreso pequeño en los años 2 y 3, a pesar de la reducción del tiempo en clase. Lo que permite anticipar un patrón similar en la progresión de la riqueza léxica, tema central de esta tesis.

Este estudio expone varias razones por las cuales las pruebas de vocabulario ofrecen mayor precisión que un conteo simple de palabras. En primer lugar, un conteo bruto no distingue entre vocabulario de alta y baja frecuencia. En segundo lugar, ignora la complejidad léxica y contabiliza repeticiones o variaciones morfológicas que pueden inflar los resultados, mientras que un instrumento especializado evita esta distorsión al enfocarse en familias léxicas definidas. Por último, existe una marcada falta de estandarización cuando se hace este tipo de conteo.

Por su parte, Edmonds (2022) diseñó un estudio para explorar y evaluar la validez de varias pruebas que miden el conocimiento de vocabulario en un segundo idioma (L2). Se centró en la relación y la confiabilidad de estos métodos de evaluación. Los participantes se encontraban en el cuarto semestre de un programa universitario de inglés en una institución académica francesa, con un perfil de competencia avanzada. Se aplicaron cuatro pruebas de vocabulario en inglés: las pruebas VLT (Nation 1983), Lex30<sup>12</sup> (Meara y Fitzpatrick, 2000), G\_Lex<sup>13</sup> y PVLT<sup>14</sup> (Laufer y Nation 1995), cada una enfocada en evaluar áreas específicas del conocimiento léxico.

La investigación comprendió dos fases. En la primera, los participantes completaron los tests VLT y G\_Lex, junto con un cuestionario sobre sus

<sup>12</sup> Prueba que mide el vocabulario productivo en una segunda lengua mediante la generación de palabras relacionadas con estímulos semánticos.

<sup>13</sup> Variación de Lex30, es una evaluación del conocimiento productivo de vocabulario, que mide la capacidad de generar palabras en respuesta a diversos contextos o temas.

<sup>14</sup> *Productive Vocabulary Levels Test*, prueba que evalúa la capacidad de producir palabras en diferentes niveles de frecuencia en una segunda lengua.

antecedentes educativos y lingüísticos. En la segunda, presentaron el PVLT y el Lex30. Al finalizar el experimento, se confirmó que Lex30 y G\_Lex muestran una correlación notable (conocimiento productivo de vocabulario), mientras que el PVLT y el VLT se asocian con el conocimiento receptivo. Este dato aporta relevancia a las pruebas VLT, porque dicha prueba se aplicará para investigar la riqueza léxica de los participantes de esta tesis.

Hung Tan Ha (2022) examinó la correlación entre el tamaño del vocabulario receptivo y la competencia en lectura de 234 estudiantes de una universidad de Vietnam, quienes habían estudiado inglés como segundo idioma durante al menos nueve años y mostraban un nivel intermedio (B1). Se realizó una evaluación del tamaño de vocabulario (UVLT<sup>15</sup>) y un análisis de lectura del IETS<sup>16</sup>. El estudio sustentó la hipótesis acerca de la dificultad de las pruebas de vocabulario en niveles elevados. También planteó inquietudes sobre el uso de puntajes de corte<sup>17</sup> para clasificar a los estudiantes en distintos niveles léxicos, pues no reflejaban con exactitud la competencia lectora.

Ademas de la evaluación del nivel léxico del alumnado, las pruebas VLT han servido para otros propósitos. En la Universidad de Tecnología de Jiangsu, China, la prueba se aplicó para determinar si el libro de texto de inglés coincidía con el nivel de vocabulario de los estudiantes. Jiangfeng (2016) llevó a cabo un estudio con 108 alumnos de primer año, de 19 a 21 años, con el fin de juzgar la adecuación de ese material didáctico. El grupo logró un dominio considerable en el

<sup>15</sup>Updated Vocabulary Levels Test.

<sup>16</sup> International English Testing System.

<sup>17</sup> "Puntajes de corte" son criterios específicos que se usan en las pruebas para clasificar a los participantes en diferentes categorías. Aquí se establece para dividir a los participantes en grupos basados en su desempeño. Por ejemplo, en una prueba de vocabulario, un puntaje de corte podría designar un límite que separa a los estudiantes en categorías como "nivel básico", "intermedio" o "avanzado" según la cantidad de palabras que comprenden.



nivel de 2,000 palabras (81.48 %), pero obtuvo puntajes mucho más bajos en los niveles 3K, 5K y en la Academic Word List (AWL).

Qian (2002) realizó un estudio con 217 estudiantes de un programa intensivo de inglés en la Universidad de Toronto. Analizó la relación entre las puntuaciones en la profundidad y amplitud del conocimiento del vocabulario, los ítems de vocabulario del TOEFL y la comprensión lectora básica. Los participantes hablaban 19 lenguas maternas diferentes y estaban inscritos en clases de nivel intermedio o superior, según los resultados del TOEIC (Test de Inglés para la Comunicación Internacional). El grupo era diverso en género, área académica y motivos para aprender inglés. Muchos de ellos planeaban estudiar en universidades de países angloparlantes.

Los instrumentos utilizados incluyeron una versión no divulgada del subtest de comprensión de lectura del TOEFL, profundidad en el conocimiento del vocabulario o DVK (por sus siglas en inglés: *Depth of Vocabulary Knowledge*), y el VLT para evaluar el tamaño del vocabulario. El objetivo del estudio fue determinar en qué medida cada una de estas pruebas contribuía a predecir el desempeño en tareas de comprensión lectora básica. Los resultados mostraron que todas las medidas de vocabulario presentaban una alta intercorrelación con la comprensión lectora, aunque ninguna se destacó significativamente sobre las demás en términos de correlación.

En la Universidad Putera Batam (Isla Batam, Indonesia), Susanto (2017) empleó el VLT para determinar si el libro de texto *Mosaic 1 Reading* era adecuado para el nivel de vocabulario de 30 estudiantes de primer año de inglés, con edades entre 18 y 22 años. Los resultados mostraron que la mayoría de los estudiantes no habían alcanzado un dominio adecuado en los niveles de 2000 y 3000 palabras del VLT, lo que indicaba un desajuste con el material didáctico. Esta investigación

confirma la utilidad del VLT como herramienta de diagnóstico y ajuste en contextos educativos.

Una investigación particularmente relevante para esta tesis es la realizada por Webb, Sasao y Ballance (2017), quienes desarrollaron y validaron dos versiones actualizadas del *Vocabulary Levels Test* (VLT) para evaluar el conocimiento léxico en los niveles de frecuencia de 1000, 2000, 3000, 4000 y 5000 palabras. Para elaborar su test, inicialmente se crearon 331 *clusters* (grupos de ítems), de los cuales se seleccionaron 234 considerados de alta calidad. Cada nivel del test contiene 10 *clusters*: cinco de sustantivos, tres de verbos y dos de adjetivos. Asimismo, se emplearon 1463 participantes de diversos orígenes y otros 250 participantes adicionales para evaluar las nuevas formas del test.

Los resultados de la investigación indicaron que ambas versiones del VLT (Form A y Form B) presentan una dificultad de ítems similar, según los análisis estadísticos (t-tests y mapas de Rasch). Además, la validación realizada mediante el marco teórico de Messick, que abarca aspectos como contenido, estructura, generalizabilidad y factores externos, confirmó su fiabilidad. En conclusión, los investigadores demostraron que estas versiones del VLT cumplen con la función propuesta.

### **3.2.2 Adaptación de VLT para Francés como L2**

Hasta ahora, en esta tesis, la mayoría de pruebas VLT se enfocan en aprendices de inglés como L2. Sin embargo, esta investigación buscó crear y validar una nueva herramienta para medir el tamaño de vocabulario receptivo en estudiantes de francés como segunda lengua (L2). Este test fue desarrollado por

Batista y Horst (2016) y se denomina TTV (*Test de la Taille du Vocabulaire*<sup>18</sup>), con base en trabajos de corpus de vanguardia en francés.

El TTV se basó en trabajos de corpus de vanguardia en francés con el fin de crear una medida actualizada del tamaño de vocabulario receptivo y ponerla a disposición de la comunidad de enseñanza e investigación. Se diseñó a partir del VLT y abarcó niveles de frecuencia de 2,000, 3,000 y 5,000 palabras. A diferencia del VLT, el TTV no incluyó una sección análoga a la Academic Word List (AWL), porque no existía un listado similar para el francés. Para la sección de 10,000 palabras, se recurrió a la lista de Baudot (1992) debido a la falta de palabras menos frecuentes del corpus originalmente usado.

La población piloto de la investigación incluyó a 63 estudiantes adultos inmigrantes con diversas lenguas maternas, quienes estaban inscritos en un programa gubernamental en una escuela de Montreal. El tiempo promedio para responder la prueba TTV fue de 30 a 35 minutos.

Antes de pasar al estudio principal, se llevaron a cabo entrevistas individuales con algunos participantes de la población piloto para evaluar su conocimiento real de las palabras que habían respondido correctamente en el TTV. Estas entrevistas revelaron resultados mixtos, con algunos participantes que respondieron correctamente en la prueba no lo hicieron en las entrevistas, y otros que explicaron correctamente las palabras en las entrevistas pero no en la prueba.

En la versión final, participaron 175 estudiantes adultos inmigrantes inscritos en cursos intensivos de francés en la misma escuela de Montreal. Se dividieron en cuatro grupos de competencia (principiante, bajo-intermedio, alto-intermedio o avanzado), con base en una prueba de colocación. Representaron 39 países y 21 idiomas diferentes.

<sup>18</sup> Test del tamaño del vocabulario.

En el análisis de datos, se calcularon las medias de las secciones del TTV y se efectuaron pruebas de ANOVA de una vía para evaluar las diferencias entre grupos. También se calculó el coeficiente alfa de Cronbach para evaluar la confiabilidad interna de las cuatro secciones del test. Finalmente, se estimó el tamaño de vocabulario general de cada participante y se aplicó porcentajes a la cantidad de palabras en cada sección del TTV.

Los resultados indican que los participantes obtuvieron puntajes más altos en los niveles de menor frecuencia (2K y 3K) y niveles más bajos en 5K y 10K. Esto refleja un conocimiento más firme de palabras de alta frecuencia y una menor familiaridad con el léxico menos común.

Al final, el TTV se validó satisfactoriamente como instrumento para medir el tamaño de vocabulario receptivo en estudiantes de francés L2. Por lo que es una valiosa herramienta de investigación en el ámbito del francés L2, sobre todo para niveles de frecuencia superiores a 5,000 palabras.

### **3.3 Observaciones Finales**

El análisis de investigaciones previas sobre la riqueza léxica en L2 es importante para comprender la evolución de este campo de estudio y la manera de medir el tamaño y la profundidad del vocabulario en estudiantes de L1 y L2, sobre todo en inglés y francés. Los estudios revisados (Álvarez, 2019; Eva Thue Vold, 2022; Treffers-Daller, 2013; Qian, 2002; Meara y Buxton, 1987; entre otros) abordan esta complejidad desde perspectivas diversas y hacen uso de medidas como la diversidad léxica, la sofisticación léxica, la lematización, además de pruebas de tipo VLT y alternativas prácticas como el test de YES/NO. Las investigaciones han mostrado la importancia de adaptar las pruebas a las particularidades del idioma y del contexto de los aprendices, lo que refuerza la

necesidad de herramientas personalizadas para idiomas distintos del inglés, como el *Test de la Taille du Vocabulaire*. Esta última prueba se ajusta de manera óptima a los objetivos de esta tesis. Su estructura facilita una evaluación más precisa del vocabulario receptivo, en especial en los niveles de frecuencia que resultan decisivos para estudiantes universitarios que aprenden una L2 en un contexto académico. Este tipo de test no solo permite medir con mayor exactitud el tamaño del vocabulario, sino que también minimiza las ambigüedades que surgen con métodos menos estructurados o menos específicos, y esto permite una interpretación más confiable de los resultados.

## 4. Metodología

### 4.1 El Diseño de la Prueba

Para analizar la riqueza léxica de los participantes en nuestra investigación, se optó por diseñar una prueba con base en *The Updated Vocabulary Levels Test* (Webb, Sasao y Ballance, 2017) y el *Test de la Taille du Vocabulaire* (Batista y Horst, 2016). Estas pruebas proporcionan un patrón base a partir del que se puede generar una nueva prueba que esté adaptada al perfil de los estudiantes mexicanos de licenciatura. La decisión de utilizar ambos modelos se debe a su validez y fiabilidad como herramientas de medición de la riqueza léxica en L2. Estas pruebas han sido utilizadas en estudios previos debido a su capacidad para evaluar la amplitud del vocabulario en distintos niveles de frecuencia, lo que proporciona datos precisos y comparables con los estudios internacionales.

Para la selección de las palabras incluidas en la prueba, se procedió a la descarga de un vocabulario basado en frecuencias léxicas. Con este propósito, se accedió al listado léxico de New, Pallier y Brysbaert (2004), denominado *Lexique 3.83*, cuya base se procesó mediante lematización para reducir la inflación de conteos provocada por la flexión morfológica. A partir de esta tabla de frecuencias lematizada, se seleccionaron 30 palabras para cada una de las siguientes bandas de frecuencia: 1K, 2K, 3K, 5K y 10K. Se evitó la inclusión de cognados y palabras con polisemia, con el objetivo de evitar sesgos.

La prueba, fue creada en Google Forms (véase en Anexo) y está organizada en categorías según las bandas de frecuencia léxica (1K, 2K, 3K, 5K y 10K), con 30 palabras asignadas a cada una de ellas, 150 palabras en total. Se acomodaron por conjuntos de tres palabras a las que deben asociarse seis

definiciones posibles así como se muestra en la *figura 1*. Todas las palabras y definiciones están redactadas en francés. La particularidad del diseño se encuentra en que las definiciones de las palabras correspondientes a la banda 1K se redactaron exclusivamente con las 500 palabras más frecuentes de esa misma banda, mientras que para las bandas superiores (2K a 10K), las definiciones se formularon con vocabulario de alta frecuencia proveniente de bandas anteriores. Este criterio tuvo como propósito mantener la dificultad controlada, sin interferir con la evaluación de la banda correspondiente.

**Figura 1**

*Ejemplo de los reactivos de la prueba, banda 1K*

	Eau très froide.	Lieu avec beaucoup d'arbres.	Accord entre des personnes.	Chute d'eau rapide.	Son indésirable.	État de calme.
<b>25. Bruit</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>26. Glace</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>27. Forêt</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Los reactivos se agrupan en categorías según el tipo de palabra (sustantivos, verbos y adjetivos). Esto tiene como objetivo reducir la posibilidad de que los participantes deduzcan respuestas basándose en diferencias morfológicas entre las categorías gramaticales.

## 4.2 Pilotaje

Se realizó un pilotaje de la prueba en noviembre de 2023 con la participación de 26 personas en total: 16 estudiantes del Centro de Enseñanza de Lenguas y Culturas (CELYC) de la Universidad Autónoma de Querétaro, de entre 18 y 30 años de edad y con distintos niveles de francés; 2 estudiantes de una escuela privada; 4 profesores de francés, y 4 hablantes nativo-hablantes con grado de bachillerato. El objetivo fue identificar posibles mejoras en la prueba, detectar eventuales problemas durante su aplicación y evaluar su confiabilidad.

Esta fase de la evaluación se llevó a cabo en el centro de cómputo del CELYC en el campus Aeropuerto de la Universidad Autónoma de Querétaro. A cada participante se le asignó una computadora individual y se le indicó completar la prueba de manera autónoma, sin recibir ningún tipo de asistencia externa, humana o tecnológica. Cualquier inconveniente o duda durante la realización se registró en una bitácora. Al finalizar la prueba se realizaron algunas preguntas de diagnóstico, tales como:

1. ¿El ejemplo inicial explica con claridad cómo se debe realizar la prueba?
2. ¿Hay alguna parte de la prueba que fuera confusa?
3. ¿Qué piensas de la duración de la prueba?
4. Comentarios sobre la prueba.

Conforme a estas preguntas, los estudiantes realizaron las siguientes observaciones, las cuales fueron atendidas: algunos señalaron la necesidad de especificar al inicio de la prueba, que cada reactivo cuenta con una única respuesta correcta. En términos generales, todos los participantes comprendieron claramente los reactivos. Asimismo, la mayoría de los estudiantes consideró que



la duración de la prueba era adecuada; no obstante, una minoría opinó que podría resultar extensa. Se reportaron también fallos en algunas computadoras, las cuales, en ocasiones, borraban las respuestas ya ingresadas por los estudiantes. Como consecuencia, se implementó un sistema para corregir este inconveniente. Se anotó la hora en que se inició la prueba, la hora del fin de la prueba, el grado y edad del estudiante, de esta manera, se podía reanudar la prueba sin perder los datos<sup>19</sup>.

Con base en los resultados generados en la prueba piloto, por medio del programa *Jamovi* 2.3 se realizó una prueba de confiabilidad con el alfa de Cronbach, dió un resultado de 0.992<sup>20</sup>. Esta cifra indica que la prueba es altamente confiable. Por lo tanto, no se realizaron modificaciones a la prueba por este motivo específico, aunque sí se introdujeron ajustes por otras razones que se explican más adelante.

Una vez completado el pilotaje se realizaron correcciones como:

- Reactivos que pueden ser confundidos
- Definiciones que fueron mejoradas para su comprensión
- Preguntas para obtener datos sociodemográficos y sobre el conocimiento de la L2

<sup>19</sup> Para ilustrar la importancia de este procedimiento, se recuperaron los datos de un estudiante cuya prueba se cerró antes de completarse. Se registró la hora en que se cerró la prueba y la hora en que se reabrió, esto permitió que el estudiante respondiera únicamente las preguntas pendientes. Gracias a esto, fue posible conservar sus datos sin necesidad de descartarlos.

<sup>20</sup> Este valor indica un nivel de fiabilidad excelente, ya que, según los estándares estadísticos, un alfa de Cronbach superior a 0.9 se considera altamente confiable.

### 4.3 Los Participantes

El estudio contó con la participación de 40 estudiantes de la LLMF, provenientes de diferentes semestres de la carrera. A los estudiantes se les permite ingresar a la licenciatura sin conocimientos previos de francés; sin embargo, al momento de graduarse, tienen como requisito obtener un nivel B2 en el idioma mediante la presentación de un examen DELF. En términos de carga académica, los estudiantes tienen aproximadamente entre 22 y 29 horas de clase a la semana, de las cuales entre el 35 % y el 80 % se imparten en francés<sup>21</sup>.

En cuanto a la distribución por semestre, participaron 7 estudiantes de 2º semestre, 9 de 4º semestre, 11 de 6º semestre y 13 de 8º semestre. En términos de género, la muestra estuvo conformada mayoritariamente por mujeres, con un total de 30 participantes, mientras que los 10 restantes fueron hombres. Solo dos participantes, un estudiante de 4º semestre y otro de 8º, contaban con un certificado DELF B2. Además, 15 de los 40 estudiantes indicaron no tomar cursos adicionales de francés fuera de su formación curricular. También cabe remarcar que una de las participantes de 8º semestre había vivido 5 años en Francia.

Respecto a la edad de los participantes, la distribución fue la siguiente: un estudiante de 18 años, cinco de 19 años, seis de 20 años, once de 21 años, seis de 22 años, seis de 23 años, dos de 24 años, dos de 26 años y uno de 29 años. Finalmente, un participante fue eliminado del estudio debido a errores técnicos durante la aplicación de la prueba. A causa de fallas en la red, se reinició su evaluación, y para evitar sesgos derivados de un posible entrenamiento previo, se

<sup>21</sup> Datos obtenidos por medio de la coordinación de la LLMF.

decidió excluir sus resultados, por lo que la muestra final quedó en 40 estudiantes.

En el marco de los principios éticos que guían la investigación científica (UAQ, 2010), se puso atención en asegurar el consentimiento informado de todos los participantes. Conforme los estándares de bioética, se proporcionó a cada participante una hoja de consentimiento detallada (Véase en *Anexo*). En esta, se especificaron los objetivos del estudio, los procedimientos involucrados y el papel que los participantes desempeñarían en la investigación. Después de una explicación completa del proceso y sus posibles implicaciones, los participantes firmaron este documento, indicaron su acuerdo para colaborar voluntaria y anónimamente. Los datos se recopilaron de forma anónima, lo que evitó cualquier posibilidad de identificación personal. Asimismo, toda la información fue almacenada en un entorno seguro y de acceso restringido, para asegurar su protección contra el uso no autorizado. Este protocolo no solo asegura el respeto y la autonomía de los individuos, sino que también garantiza la transparencia y la responsabilidad en el tratamiento de los datos obtenidos.

#### **4.4 La Aplicación de la Prueba**

La ejecución de la prueba se llevó a cabo a inicios de marzo 2024, en las instalaciones del centro de cómputo del CELYC de la Universidad Autónoma de Querétaro, Campus Aeropuerto. El examen se administró a cada uno de los grupos que conforman la LLMF, prohibiéndoles el uso de dispositivos móviles y la utilización de herramientas que pudieran facilitar la resolución del mismo.

Para llevar a cabo la prueba, el formulario fue cargado previamente en cada computadora antes del ingreso de los participantes. Se explicó a los participantes en qué consistía la prueba. También se les informó cómo serían utilizados sus

datos y que estos serían registrados de forma anónima. Por último, se les indicó el momento de inicio de la prueba y se estableció un límite de tiempo de 30 minutos.

#### **4.5 El Procedimiento de Análisis**

Una vez recopiladas las respuestas de los 40 participantes, el análisis de los datos se llevó a cabo en varias etapas, se utilizaron herramientas estadísticas y software especializado para el análisis.

Primeramente, los resultados de las pruebas de vocabulario fueron exportados desde *Google Forms* en formato CSV, lo que permitió estructurar las respuestas de manera que facilitaran su análisis posterior. Los datos de cada participante fueron codificados de acuerdo con las categorías de frecuencia léxica (1k, 2k, 3k, 5k y 10k), para una identificación clara de los aciertos y errores en cada una de las bandas de frecuencia.

Después, para evaluar la riqueza léxica de cada participante, se sumó el puntaje de respuestas correctas en cada banda de frecuencia. Este cálculo permitió determinar cuántas palabras conocían los estudiantes en cada nivel, también identificar las tendencias entre los diferentes semestres en cuanto al conocimiento de las palabras de alta y baja frecuencia, es decir, si los grupos tenían un comportamiento similar o diferente. A su vez, permitió hacer la observación de si los estudiantes de semestres más avanzados tenían un dominio más amplio y diverso del vocabulario. Se calcularon las medias y las desviaciones estándar para cada semestre, lo que facilitó la comparación cuantitativa entre los grupos y proporcionó una medida precisa del progreso léxico a lo largo de los semestres.

A continuación, se realizaron pruebas estadísticas para determinar la existencia de diferencias significativas entre los grupos. En primer lugar, se realizó un análisis ANOVA para evaluar si el dominio del vocabulario variaba significativamente entre los distintos semestres. Posteriormente, tras obtener resultados significativos en el ANOVA, se aplicó una prueba post hoc para identificar con precisión entre qué grupos existían diferencias en el dominio del vocabulario.

Se realizó una segunda prueba ANOVA con el propósito de analizar si existían diferencias significativas entre los distintos semestres y cada una de las bandas de frecuencia léxica. El objetivo de este análisis era examinar si el dominio del vocabulario variaba en cada banda de frecuencia en función del semestre cursado. Se partió de la hipótesis de que los estudiantes de octavo semestre habrían adquirido un mayor número de palabras de baja frecuencia en comparación con los semestres anteriores.

Por último, en función de las preguntas de investigación planteadas, se realizó la interpretación. Se prestó especial atención a las diferencias en la riqueza léxica entre semestres y a la relación entre la frecuencia de las palabras y su adquisición por parte de los estudiantes. Los datos obtenidos de las pruebas en *Google Forms* proporcionaron la base cuantitativa para discutir las tendencias observadas en el desarrollo léxico de los estudiantes, y esto se utilizó para contrastar con teorías previas sobre la adquisición de vocabulario en una lengua extranjera.

## 5. Resultados

### 5.1 Los Resultados Generales

Al extraer los datos de *Google Forms*, se generó una tabla con los resultados individuales de cada participante. Estos datos se procesaron para realizar los cálculos que se presentarán más adelante en esta sección. En la *Tabla 1* se muestran los resultados de la prueba de riqueza léxica, junto con dos medidas estadísticas clave para cada grupo de semestre: la media y la desviación estándar.

**Tabla 1**

*Media y desviación estándar para cada grupo de semestre*

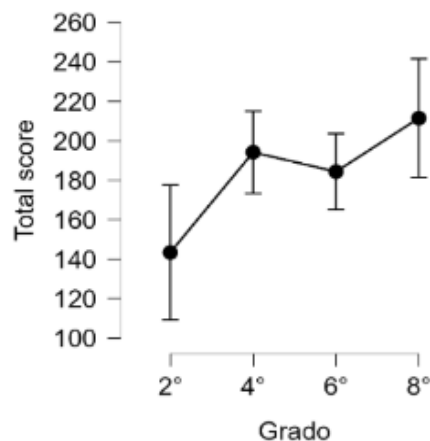
Grado	Grados de libertad	Media	Desviación Estándar
2°	6	143.42	34.15
4°	9	194.2	27.6
6°	10	184.36	27.31
8°	11	211.5	45.27

La *Tabla 1* muestra que la media del puntaje obtenido en la prueba aumenta conforme avanzan los semestres, con excepción del sexto, en el cual se observa un ligero descenso respecto a los datos del cuarto semestre. Cabe aclarar que el puntaje máximo era de 300 puntos, ya que cada acierto tenía un valor de 2 puntos.

En la *figura 2* se puede apreciar el aumento en la riqueza de léxica por semestre, en el que se ilustra la media de cada grupo con su intervalo de confianza. Esta representación permite visualizar claramente la media de cada semestre, así como la desviación estándar de cada grupo. Esto muestra un aumento de conocimiento léxico conforme al grado de la licenciatura.

**Figura 2**

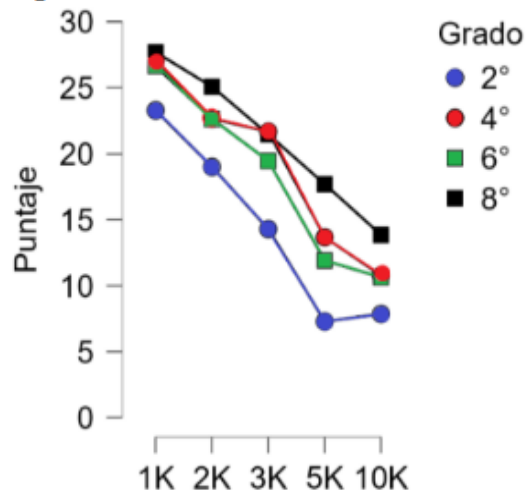
*Puntajes de los grupos en la prueba, con su media y su desviación estándar.*



Para profundizar el análisis se decidió hacer una separación de resultados por banda de frecuencia, es decir, medir que resultados obtuvo cada grupo en las bandas 1k, 2k, 3k, 5k y 10k, estos resultados se muestran en la *Figura 3*.

**Figura 3**

*Resultados de cada banda de frecuencia por el grado de licenciatura.*



Esta figura muestra que hay un aumento general en todas las bandas de frecuencia conforme al grado escolar (con excepción de 4° y 6° donde no se aprecia mucha diferencia). Los estudiantes tienen alto puntaje de las palabras que se encuentran en estas bandas de frecuencia altas y estos resultados crecen en las bandas de frecuencia más bajas.

## 5.2 El Análisis Estadístico

El primer paso fue realizar una prueba de Shapiro-Wilk para evaluar la normalidad de los datos en cada grupo correspondiente a los diferentes semestres. Ningún dato fue  $p > 0.05$ , esto mostró que los datos seguían una



distribución normal. Por lo tanto, el cumplimiento del supuesto de normalidad justifica el uso de análisis estadísticos paramétricos para la comparación entre grupos.

Para observar más a detalle los resultados obtenidos, se procedió a la realización de una ANOVA (Véase en *Tabla 2*) por medio del programa de cálculos estadísticos JASP (2024). Esto nos permitirá más adelante observar si existen diferencias significativas entre los resultados de cada grado.

**Tabla 2**

*Cálculo ANOVA entre los grados de licenciatura*

<b>variable</b>	<b>GL</b>	<b>F</b>	<b>p</b>
Grado	3	5.202	0.004*
Residuals	36		

El análisis ANOVA de una vía arrojó un resultado de  $F(3,36) = 5.2$ ,  $p = .004$ , lo cual indica que al menos un semestre presenta diferencias estadísticamente significativas en comparación con los demás en la puntuación total. Para identificar entre qué semestres se encontraban dichas diferencias, se aplicó también una prueba Post Hoc que se puede ver en la *Tabla 3*, mediante el programa JASP (2024).

**Tabla 3**

*Post Hoc por grado académico*

Grado		Diferencia de promedios	Desv Est	t	ptukey
2°	4°	-50.77	18.1	-2.8	0.04*
	6°	-40.94	17.76	-2.31	0.12
	8°	-68.07	17.47	-3.9	0.002*
4°	6°	9.84	16.05	0.61	0.93
	8°	-17.3	15.73	-1.1	0.69
6°	8°	-27.14	15.34	-1.77	0.3

Los siguientes resultados indican lo siguiente:

- Entre los semestres 2° y 4°: un valor p de Tukey de 0.039 señala una diferencia significativa en los puntajes entre ambos grupos.
- Entre los semestres 2° y 8°: se observa una diferencia estadísticamente significativa, con un valor ptukey = 0.002.
- Entre los semestres superiores (4°, 6° y 8°): los valores de ptukey superiores a 0.05 indican la ausencia de diferencias significativas en estas etapas avanzadas.

### **5.3 El Análisis Estadístico para Bandas de Frecuencia**

La siguiente etapa de la investigación era saber si había algún tipo de diferencia en las bandas de frecuencia conforme al grado escolar en el que se encontraban los estudiantes. Para esto se realizó una ANOVA de medidas repetidas *Tabla 4*.

**Tabla 4**

*Anova con corrección de esfericidad Greenhouse-Geisser*

<b>Casos</b>	<b>Suma de cuadrados</b>	<b>GL</b>	<b>F</b>	<b>p</b>
Bandas	6352.51	2.83	195.9	< .001*
Bandas * Grado	167.70	8.50	1.724	0.097

En el ANOVA de medidas repetidas, la esfericidad exige que la varianza de las diferencias entre pares de condiciones sea homogénea. Como en este estudio no se cumplía esa suposición, se aplicó la corrección de Greenhouse-Geisser para prevenir resultados sesgados. Tras dicha corrección, el análisis arrojó un valor  $p < .001$  en las bandas de frecuencia, lo que señala diferencias estadísticamente significativas según la frecuencia de las bandas léxicas.

Por otro lado, la interacción entre las bandas de frecuencia y los grados académicos no resultó ser estadísticamente significativa ( $F = 1.724$ ,  $p = .097$ ). Este resultado sugiere que la influencia de las bandas de frecuencia en el aprendizaje de vocabulario no varía significativamente entre los diferentes semestres. En otras palabras, la forma en que los estudiantes adquieren vocabulario de alta y baja frecuencia parece ser consistente a lo largo de su carrera universitaria.

#### **5.4 Las Observaciones Finales Entre Bandas de Frecuencia**

Los resultados del estudio permiten afirmar que el desarrollo de la riqueza léxica en los estudiantes de la LLMF sigue un patrón de progresión estable a lo largo de los semestres, lo cual indica que la riqueza léxica aumenta de manera continua pero relativamente uniforme a lo largo de la licenciatura.

Se puede confirmar que la adquisición de vocabulario en una segunda lengua está estrechamente vinculada con la frecuencia de las palabras. El análisis ANOVA por bandas de frecuencia (1K, 2K, 3K, 5K y 10K) revela que los estudiantes presentan un avance diferenciado según el nivel de frecuencia léxica, lo cual responde a la segunda pregunta de investigación. Se observa que las palabras de alta frecuencia se adquieren con mayor facilidad que aquellas de baja frecuencia, lo que sugiere una relación directa entre la frecuencia léxica y el grado de adquisición del vocabulario en los estudiantes de la LLMF.

Asimismo, estos resultados evidencian que la progresión de las diferentes bandas de frecuencia se mantiene relativamente homogénea en todos los semestres, lo que indica que tanto el vocabulario de alta como de baja frecuencia han sido asimilados con una consistencia similar. Este hallazgo permite identificar un patrón estable de adquisición de vocabulario entre los semestres. En este sentido, se puede afirmar que el desarrollo del léxico, tanto de alta como de baja frecuencia, sigue una trayectoria continua durante los distintos semestres de la licenciatura.

A continuación, en la sección de discusiones, se comentarán algunas de las implicaciones que tienen estos resultados.

## **6. Discusión**

Esta investigación tuvo como propósito analizar el desarrollo de la riqueza léxica en estudiantes de la LLMF de la UAQ, evaluó las diferencias en el dominio del vocabulario entre los estudiantes de distintos semestres y examinó la relación entre la frecuencia de uso del vocabulario y la adquisición del mismo. Se ha centrado en dos objetivos principales: por un lado, observar si existen diferencias significativas en la riqueza léxica entre los estudiantes de diferentes semestres, y por otro, ver cómo se relaciona la frecuencia del vocabulario con la adquisición léxica en el aprendizaje del francés como L2.

La investigación encontró diferencias significativas que indican que los estudiantes de grados más avanzados poseen mayor riqueza léxica. Este resultado coincide con las premisas de Nation (2000) y Webb y Nation (2017), quienes sostienen que el desarrollo del vocabulario está vinculado a la exposición constante al idioma y al aumento progresivo de la complejidad léxica con el tiempo. En particular, los datos indican que los estudiantes de los primeros semestres (2° y 4°) dominaban principalmente palabras de alta frecuencia (1K y 2K), mientras que aquellos de semestres avanzados (6° y 8°) comenzaron a incorporar vocabulario de menor frecuencia (3K, 5K y 10K), aunque con una progresión más lenta. Sin embargo, los resultados del sexto semestre no siguieron el patrón ascendente esperado, ya que su puntuación promedio fue inferior a la del cuarto semestre. Esta diferencia puede explicarse por el hecho de que se trata de grupos distintos, en los que intervienen múltiples factores individuales, como el

grado de exposición previa al francés, el esfuerzo personal en el estudio, o incluso experiencias fuera del entorno universitario.

Estos resultados reflejan que la adquisición léxica no se desarrolla de forma uniforme, sino que sigue un proceso progresivo en el que el tipo de vocabulario aprendido varía según el grado de exposición a la lengua. En las primeras etapas, los estudiantes incorporan con mayor facilidad palabras de alta frecuencia debido a su constante aparición en su entorno, lo que facilita su aprendizaje incidental (Lewis, 1993; Nation, 2000). No obstante, a medida que avanzan en su formación, la adquisición de vocabulario menos frecuente exige un esfuerzo cognitivo mayor y una exposición más prolongada para garantizar su consolidación, debido a que la aparición del vocabulario de baja frecuencia tiende a aparecer menos que el de alta (Milton, 2009).

En este sentido, los resultados sugieren una ralentización en el aprendizaje de vocabulario menos frecuente a medida que los estudiantes progresan en sus estudios. Este fenómeno no es atípico en la adquisición de segundas lenguas; según Webb y Nation (2017), la desaceleración en la adquisición de palabras menos frecuentes se atribuye generalmente a una exposición más limitada al vocabulario especializado. Si bien este hallazgo resulta algo intrigante, pues podría esperarse que un mayor tiempo de instrucción conlleve una expansión del léxico de baja frecuencia en la misma proporción. Lo que podría explicar este fenómeno es que los estudiantes solo reciben input especializado en palabras académicas y no están expuestos a vocabulario fuera del ámbito académico. Se sabe que los estudiantes adquieren vocabulario no necesariamente de acuerdo con la frecuencia general de las palabras, sino con la frecuencia con la que estas se utilizan en su entorno (Coxhead, 2000).

Respecto a la interacción entre los distintos semestres y las bandas de frecuencia léxica, el análisis estadístico no reveló una interacción significativa. Esto sugiere que los estudiantes adquieren vocabulario de distintas bandas de frecuencia de forma gradual. Como afirman Milton (2009) y Webb & Nation (2017), es probable que a medida que se alcanza un umbral funcional en sus tareas académicas, el crecimiento léxico se desacelere. Esta tendencia también se refleja en los estudios sobre adquisición léxica en L2, que estiman que los estudiantes pueden aprender entre 500 y 1,000 palabras nuevas por año en etapas iniciales, pero que esta tasa disminuye, conforme el vocabulario se vuelve más técnico o menos frecuente (Milton, 2006).

Es importante mencionar que, aunque este estudio observó un incremento significativo en la capacidad de reconocimiento de palabras, este incremento debe interpretarse con cautela, ya que el tipo de conocimiento evaluado corresponde únicamente a la dimensión receptiva y no garantiza un dominio productivo ni profundo del léxico. Como señalan Meara (1980), Daller, Milton y Treffers-Daller (2007), la riqueza léxica no se limita a la cantidad de palabras reconocidas, sino que incluye también la profundidad del conocimiento, la capacidad de uso, la sensibilidad morfológica y la competencia colocacional. Por tanto, los resultados del presente estudio deben entenderse dentro de los límites metodológicos del instrumento aplicado, el cual no explora aspectos productivos ni estratégicos del léxico.

A lo largo de este estudio, se observó que la exposición constante al idioma es importante para el desarrollo de la riqueza léxica. Sin embargo, los estudiantes de la LLMF, al encontrarse en un entorno no francófono, tienen pocas oportunidades para practicar el idioma fuera del aula, lo que puede haber limitado su exposición al input en francés. Al respecto, Batista y Horst (2016) han

observado que los contextos de inmersión, donde los estudiantes tienen contacto diario y continuo con el idioma objetivo, favorecen una adquisición más rápida y profunda del vocabulario. En este sentido, los estudiantes que han tenido la oportunidad de vivir en un lugar francófono, probablemente presentan un mayor progreso en su desarrollo léxico en comparación con aquellos que dependen exclusivamente del aprendizaje en el aula. Esta hipótesis obtiene respaldo al observarse los resultados individuales, en los que se mostró una variabilidad considerable entre estudiantes del mismo semestre. Tal dispersión sugiere que el desarrollo léxico no responde exclusivamente al avance académico, sino que se ve influido también por factores individuales, como el grado de exposición extracurricular, las estrategias personales de aprendizaje y el nivel de motivación (Nation, 2001; Barcroft, 2012)

Otro aporte de esta investigación es que aborda un contexto escasamente explorado en la literatura sobre adquisición de vocabulario en francés como lengua extranjera. Aunque existen numerosos estudios enfocados en la enseñanza del vocabulario en inglés o en contextos europeos, son pocos los que específicamente atienden a estudiantes universitarios mexicanos en relación con el aprendizaje del francés como L2. Los resultados obtenidos aportan datos específicos sobre el desarrollo de la riqueza léxica en estudiantes mexicanos aprendiendo francés, ampliando así las perspectivas de aplicación práctica en contextos educativos similares en Latinoamérica.

Una observación pertinente a considerar es la posibilidad de que la educación a distancia, impuesta por la pandemia, pudiera haber limitado la exposición y adquisición léxica de los estudiantes. Sin embargo, los resultados de este estudio muestran que el desarrollo léxico del grupo de 8º semestre, que cursó



la mayoría de sus estudios bajo la modalidad virtual, no parece haberse visto afectado significativamente.

### **6.1 La Creación de una Prueba VLT Adaptada para Estudiantes Mexicanos**

Una contribución de esta investigación fue la creación y validación de un nuevo *Vocabulary Levels Test* específicamente adaptado para estudiantes hispanohablantes que aprenden francés L2. Esta prueba surgió a partir de una combinación entre el VLT original en inglés propuesto por Nation (1983), así como del *Test de la Taille du Vocabulaire* (TTV) desarrollado por Batista y Horst (2016) para el francés. Esta adaptación fue realizada para esta investigación con el fin de evaluar con precisión la riqueza léxica receptiva de estudiantes universitarios de la LLMF.

La creación de esta prueba consideró cuidadosamente el contexto lingüístico de los estudiantes mexicanos, se tuvo especial precaución para seleccionar vocabulario que no incluyera palabras fácilmente reconocibles por su similitud formal con el español, con el fin de evitar respuestas influidas por esta proximidad lingüística. Esto asegura que la evaluación refleje más fielmente el conocimiento léxico real adquirido por exposición al francés y minimizar sesgos. También se realizó a la prueba un análisis de fiabilidad mediante el Alpha de Cronbach, se obtuvo un valor alto de 0.973. Este resultado muestra que la prueba posee una elevada confiabilidad.

Dada la confiabilidad estadística y la rigurosidad en su diseño metodológico, esta adaptación del VLT representa una herramienta que puede ser utilizada en futuras investigaciones. Su potencial aplicación no solo se limita a

estudiantes universitarios mexicanos, sino que puede extenderse a estudiantes hispanohablantes interesados en medir su conocimiento léxico del francés. En este sentido, este instrumento constituye un aporte para la evaluación y el monitoreo de riqueza léxica en contextos educativos similares.

## **6.2 Las Limitaciones del Estudio**

Uno de los principales puntos a destacar es que la presente tesis se apoya en una prueba de Google Forms centrada en bandas de frecuencia léxica. Esto podría no capturar la riqueza completa del uso del léxico en contextos más amplios y variados, como sería en la producción real o la comprensión profunda de la palabra. La adquisición del vocabulario implica tanto el conocimiento receptivo como productivo, así como la capacidad de usar las palabras en situaciones comunicativas diversas. En este sentido, una característica a mejorar sería contar con una evaluación del uso contextual del vocabulario, tanto en la producción oral como escrita, lo cual podría ofrecer un panorama más completo de la competencia léxica de los estudiantes.

Por otra parte, el uso de bandas de frecuencia (1k, 2k, 3k, 5k, 10k) es el núcleo de la herramienta para la medición de adquisición de vocabulario, pero también cuenta con limitaciones importantes. Las bandas de frecuencia, aunque útiles para evaluar qué tan común es un conjunto de palabras, no permiten capturar completamente el aspecto cualitativo del léxico. Un enfoque basado solo en la frecuencia puede subestimar el uso de palabras menos frecuentes pero más relevantes para contextos académicos o específicos. Por ejemplo, en estudios previos como los de Milton (2009), Laufer y Nation (1995) y Coxhead (2000), se ha señalado que el vocabulario académico y especializado no siempre se refleja adecuadamente en estas bandas. Esto plantea la cuestión sobre si el presente

estudio podría haber subestimado el valor del vocabulario académico, aspecto que resulta importante considerar dado que los estudiantes se desenvuelven en contextos educativos formales.

Este estudio representa una investigación acotada sobre el desarrollo de la riqueza léxica en estudiantes universitarios de francés como segunda lengua. Por lo tanto, se reconoce que podría alcanzar una mayor profundidad mediante la incorporación de diversos factores adicionales. La investigación podría ampliarse al observar variables como las estrategias pedagógicas específicas empleadas por los docentes, las experiencias previas de aprendizaje de los estudiantes, así como el impacto de distintos contextos de exposición al idioma francés más allá del ámbito estrictamente académico. Del mismo modo, sería pertinente contemplar factores socioculturales vinculados al entorno familiar y social de los estudiantes, que podrían influir en la adquisición del léxico. Por lo anterior, se recomienda interpretar los resultados obtenidos con cautela, reconociendo la necesidad de realizar estudios posteriores que integren dichas variables para obtener una visión más completa del fenómeno abordado en esta tesis.

### **6.3 Las Perspectivas Futuras**

A partir de los resultados obtenidos en este trabajo, surgen nuevas preguntas y líneas de investigación que podrían ser abordadas en estudios futuros, para ampliar la comprensión sobre la adquisición léxica en el aprendizaje de una segunda lengua. Un cuestionamiento que podría explorarse es si existe una correlación significativa entre la adquisición de vocabulario de alta y baja frecuencia y la competencia en habilidades productivas, como la expresión oral y escrita. En este estudio se evaluó la comprensión léxica a través de un test de Google Forms, pero no se exploró cómo este conocimiento se traduce en la

capacidad de los estudiantes para producir lenguaje en contextos comunicativos reales.

Otra línea que este estudio deja abierta para futuras investigaciones consiste en examinar cómo la enseñanza explícita del vocabulario, combinada con la exposición incidental, influye en la adquisición de palabras pertenecientes a distintas bandas de frecuencia léxica. La exploración de estas preguntas podría proporcionar información para diseñar metodologías pedagógicas más efectivas y centradas en el desarrollo integral de la competencia léxica de los estudiantes L2.

La presente investigación permitió observar el manejo léxico de los alumnos conforme a su semestre, así como las implicaciones que tiene la frecuencia léxica en el manejo de vocabulario de los estudiantes. Sin embargo, podría considerarse el realizar en un futuro un estudio de carácter longitudinal que permitiera observar qué cambios existen en el léxico de los estudiantes durante un semestre de diferencia o cómo varía este desarrollo a lo largo de la licenciatura. Estudiar la progresión de un grupo específico a lo largo de semestres consecutivos podría proporcionar una visión más detallada sobre el ritmo y las características de la adquisición léxica. Esto permitiría identificar patrones de crecimiento, no solo en términos de cantidad de vocabulario, sino también en la diversificación y el uso de vocabulario de alta y baja frecuencia.

Otra pregunta relevante se relaciona con el tipo de lenguaje evaluado en esta tesis. Dado que el contexto académico se centra en un tipo de discurso especializado, surge la duda: ¿El lenguaje de los estudiantes está centrado solamente en lo académico, más que en lo cotidiano? Se debe considerar que el entorno académico tiende a exponer a los estudiantes a un léxico especializado que no siempre se refleja en pruebas centradas en vocabulario de uso general. Si la prueba utilizada en esta investigación está más enfocada en evaluar el

conocimiento de palabras de alta frecuencia del lenguaje cotidiano, es posible que no capture adecuadamente el desarrollo léxico necesario para la competencia académica, lo cual podría plantear la necesidad de incluir instrumentos que aborden también el uso de términos más complejos y específicos del ámbito académico.

Un aspecto importante que emergió de este estudio, pero que no fue abordado en profundidad, es el impacto del contexto sociocultural en el desarrollo de la riqueza léxica de los estudiantes. Aunque la tesis se enfocó principalmente en la evaluación de la adquisición de vocabulario a través de un test en Google Forms con bandas de frecuencia, es necesario reconocer que el entorno no inmersivo en el que los estudiantes de la LLMF operan pudo haber limitado de manera significativa su desarrollo léxico. Estudios previos han destacado que la exposición al idioma en situaciones reales de uso, fuera del aula, tiene un impacto crucial en la consolidación de vocabulario y en el uso de palabras menos frecuentes y más especializadas (Batista & Horst, 2016; Milton, 2009).

## **7. Conclusiones**

La presente investigación tuvo como objetivo principal analizar el desarrollo de la riqueza léxica en estudiantes mexicanos inscritos en la LLMF. A través del análisis estadístico de la prueba de comprensión de vocabulario se ha podido observar un progreso significativo en la adquisición de vocabulario, tanto en el uso de palabras de alta y baja frecuencia, lo cual se refleja en los índices de riqueza léxica obtenidos.

Los resultados confirmaron la hipótesis inicial, mostraron que existe un aumento significativo en la riqueza léxica conforme al semestre de los estudiantes. En particular, se observó un mayor dominio del vocabulario más frecuente en todos los niveles estudiados. Además, los estudiantes de semestres más avanzados comenzaron a adquirir vocabulario de menor frecuencia (bandas 5K y 10K), aunque a un ritmo más lento que el de alta frecuencia. Esto se debe, en parte, a que no se exponen tan a menudo a dichas palabras como a las de uso más común.

En respuesta a la pregunta específica sobre los patrones existentes en la adquisición del vocabulario de alta y baja frecuencia, se observó que mientras el vocabulario de alta frecuencia presenta un desarrollo constante y rápido desde los primeros semestres, el vocabulario de baja frecuencia revela un patrón más tardío y menos pronunciado. Este resultado responde al segundo objetivo planteado en la investigación sobre evaluar la relación entre la frecuencia de las palabras y la capacidad de adquisición léxica, confirma la existencia de patrones en el

aprendizaje de vocabulario según su frecuencia. Esta tendencia podría deberse al tipo de materiales y recursos utilizados en las clases, los cuales probablemente privilegian palabras frecuentes sobre vocabulario especializado.

La falta de progresión en las bandas de baja frecuencia sugiere posibles lagunas en el manejo de textos académicos y especializados exigidos a los futuros profesionistas en lengua francesa. La ausencia de diferencias significativas en la banda 10K evidencia que los estudiantes podrían no recibir suficiente exposición al vocabulario especializado en las etapas avanzadas de su formación. Esto pone de manifiesto la necesidad de adoptar estrategias pedagógicas que favorezcan la adquisición efectiva del vocabulario de menor frecuencia.

Cabe destacar que esta tesis desarrolló una prueba confiable que ha mostrado ser adecuada para medir la riqueza léxica en estudiantes mexicanos que aprenden francés L2. La implementación de este instrumento no solo permitió identificar diferencias significativas entre los grupos de estudiantes, sino que también ofrece una herramienta que es práctica, eficaz y rápida para futuras investigaciones y evaluaciones en el ámbito de la enseñanza y aprendizaje de esta segunda lengua.

Esta tesis ha resaltado la importancia clave que tiene el desarrollo léxico en la formación de los futuros profesionales de la lengua. El dominio del vocabulario no solo refleja la competencia lingüística de un estudiante, sino que también es un componente crítico para que pueda desempeñarse con éxito tanto en contextos académicos o profesionales como en contextos sociales cotidianos.

## 8. Bibliografía

- Álvarez, L. (2019). *Índices de riqueza léxica en estudiantes de 11 y 16 años de La Habana*. Universidad de La Habana, 287, 267-286.
- Agustín-Llach, M. P. (2022). Incidental learning of L2 Spanish vocabulary. En J. Barcroft y J. Muñoz-Basols (Eds.), *Spanish vocabulary learning in meaning-oriented instruction* (pp. 44-57). Routledge.  
<https://doi.org/10.4324/9781315100364>
- Barcroft, J. y Muñoz-Basols, J. (2022). Input, tasks, and processing specificity in Spanish vocabulary learning. Routledge. En J. Barcroft y J. Muñoz-Basols (Eds.), *Spanish vocabulary learning in meaning-oriented instruction* (pp. 44-57). Routledge. <https://doi.org/10.4324/9781315100364>
- Batista, R. y Horst, M. (2016). A New Receptive Vocabulary Size Test for French. *Canadian Modern Language Review*. 72. 1-23.  
<https://doi.org/10.3138/cmlr.2820>.
- Baudot, J. (1992). *Les fréquences d'utilisation des mots en français écrit contemporain*. Montreal. Presses de l'Université de Montreal.



- Cobb, T. y Horst, M. (2004). Chapter 2. Is There Room for an Academic Word List in French. En Bogaards, P., y Laufer-Dvorkin, B. (2004). *Vocabulary in a second language: Selection, Acquisition, and Testing*. John Benjamins Publishing Company.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge University Press.
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213.\_  
<https://doi.org/10.2307/3587951>
- Daulton, F. E. (2008). *Japan's built-in lexicon of English-based loanwords*. Clevedon: Multilingual Matters LTD.
- Daller, H., Milton, J. y Treffers-Daller, J. (2007). *Modelling and Assessing Vocabulary Knowledge*. Cambridge University Press.
- Edmonds, A., Clenton, J. y Elmetaher, H. (2022). Exploring the construct validity of tests used to assess L2 productive vocabulary knowledge. *System*, 108, 102855. <https://doi.org/10.1016/j.system.2022.102855>
- Fairclough, N. (1989). *Language and Power*. London: Longman.
- Ferreiro, E., Pontecorvo, C. Ribeiro, N. y García, I. (1996). *Caperucita Roja aprende a escribir: estudios psicolingüísticos comparativos en tres lenguas*. Gedisa.
- Ha, H. T. (2022). Issues in matching learners to their reading levels using receptive vocabulary knowledge – A closer look at the vocabulary levels tests and scoring methods. *Heliyon*, 8(8), e10244  
<https://doi.org/10.1016/j.heliyon.2022.e10244>.

JASP Team (2024). JASP (Version 0.19.3)[Computer software]

Jiangfeng, L. (2016). Role of Vocabulary Levels Test (VLT) in Chinese Undergraduate Students' Comprehension of Textbooks. *Journal of Language Teaching and Research*, 7(2) <https://doi.org/10.17507/jltr.0702.16>).

Johnson, W. (1944) Studies in language behavior: I. A program of research. *Psychological Monographs* 56, 115.

Karim, S., Karim, S., y Chandio, R. (2019). Association of high frequency and low frequency words with reading comprehension of ESL undergraduates. *International Journal of English and Education*, 8(3), 105-113. <https://www.researchgate.net/publication/354985467>

Laufer, B. (2003). Vocabulary Acquisition in a Second Language: Do Learners Really Acquire Most Vocabulary by Reading? Some empirical Evidence. *The Canadian Modern Language Review*, 59.

Laufer, B., y Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, 16, 307-322. <http://dx.doi.org/10.1093/applin/16.3.307>

Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. Language Teaching Publications.

López, H. (1995). Los estudios de disponibilidad léxica: Pasado y presente. *Boletín de Filología*, 35(1), Article 1.

- Meara, P. M., y Buxton, B. (1987). *An Alternative to Multiple Choice Vocabulary Tests*. Language Testing.
- Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching and Linguistics*, 13(4), 221-246.
- Meara, P., & Fitzpatrick, T. (2000). Lex30: an improved method of assessing productive vocabulary in an L2. *System*, 28(1), 19–30.  
[https://doi.org/10.1016/s0346-251x\(99\)00058-5](https://doi.org/10.1016/s0346-251x(99)00058-5)
- Milton, J. (2006). Language Lite? Learning French Vocabulary in School. *Journal of French Language Studies*, 16, 187-205.  
<https://doi.org/10.1017/S0959269506002420>
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Multilingual Matters.
- Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, 5(1), 12–25.
- Nation, I. S. P. (2000). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- New, B., Pallier, C., Brysbaert, M., Ferrand, L. (2004). Lexique 2 : A New French Lexical Database. *Behavior Research Methods, Instruments, y Computers*, 36 (3), 516-524. [doi.org/10.3758/bf03195598](https://doi.org/10.3758/bf03195598)
- Qian, D. (2002). Investigating the Relationship Between Vocabulary Knowledge and Academic Reading Performance: An Assessment Perspective. *Language Learning*, 52(3), 513–536.

- Rodríguez, I. (2022). Testing L2 Spanish vocabulary knowledge. En J. Barcroft y J. Muñoz-Basols (Eds.), *Input, tasks, and processing specificity in Spanish vocabulary learning* (pp. 130). Routledge.  
<https://doi.org/10.4324/9781315100364>
- Susanto, A. (2017). *Assessing the Relationship between the Vocabulary Level Test (VLT) and Reading Comprehension*. Studies in English Language and Education. Batam, Indonesia.
- Schmitt, N. (1997). Vocabulary learning strategies. En N. Schmitt y M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 199-227). Cambridge University Press
- Treffers-Daller, J. (2013). Chapter 3. Measuring lexical diversity among L2 learners of French. En *Studies in bilingualism* (pp. 79-104).  
<https://doi.org/10.1075/sibil.47.05ch3>
- Tidball, F., y 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, 299-313. <https://doi.org/10.1017/S0959269508003463>
- Universidad Autónoma de Querétaro, (2010). *Manual del Comité de Ética para la Investigación*. Universidad Autónoma de Querétaro.  
<https://www.uaq.mx/finuaq/manualdelcomiteparalainvestigacion.pdf>
- Vold, E. T. (2022). Development of lexical richness among beginning learners of French as a foreign language. *Nordic Journal of Language Teaching and Learning*, 10(2), Article 2. <https://doi.org/10.46364/njltl.v10i2.1007>

Webb, S., y Nation, P. (2017). *How Vocabulary is Learned*. Oxford University Press.

Webb, S., Sasao, Y., y Ballance, O. (2017). The updated Vocabulary Levels Test. *ITL - International Journal of Applied Linguistics*, 168(1), 33-69.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Addison-Wesley.

## **9. Anexo**

### **9.1 Formato de la Prueba Utilizada para el Análisis de Esta Tesis.**

## PRUEBA LÉXICA DE FRANCÉS

Este formulario es una adaptación de las evaluaciones propuestas en "The updated Vocabulary Levels Test (Webb, Sasao & Ballance, 2017)" y "Test de la taille du vocabulaire " (Batista, 2014). Ayudará en un estudio sobre tu conocimiento léxico. Tu participación es parte de una investigación llevada a cabo en el marco de mis estudios en la Maestría en Lingüística de la UAQ . Completarlo debería requerir aproximadamente 15 minutos.

Los datos personales serán anonimizados y se garantiza la total confidencialidad de los resultados. Por favor, te solicitamos que respondas con sinceridad.

Puedes dirigirme cualquier duda o pregunta a esta dirección: [eggreiner1890@gmail.com](mailto:eggreiner1890@gmail.com)

*\* Indicates required question*

---

### Datos personales

1. Sexo: \*

*Mark only one oval.*

☐ Masculino

☐ Femenino

2. Edad \*

---

3. Grado \*

*Mark only one oval.*

☐ 2°

☐ 4°

☐ 6°

4. ¿Cuentas con algún diploma DELF? \*

*Mark only one oval.*

- ☐ Sí
- ☐ No
- ☐ Other: \_\_\_\_\_

5. ¿Has estudiado francés antes de la licenciatura? (si no ignora esta pregunta).

*Mark only one oval per row.*

	3 a 5 horas semanales	6 a 10 horas semanales
<b>1 bimestre</b>	<input type="radio"/>	<input type="radio"/>
<b>1 cuatrimestre</b>	<input type="radio"/>	<input type="radio"/>
<b>1 semestre</b>	<input type="radio"/>	<input type="radio"/>
<b>1 año</b>	<input type="radio"/>	<input type="radio"/>
<b>1 año y medio</b>	<input type="radio"/>	<input type="radio"/>
<b>2 o más años</b>	<input type="radio"/>	<input type="radio"/>
<b>Otro (informar al investigador)</b>	<input type="radio"/>	<input type="radio"/>

*Skip to question 6*

Ejemplo de cómo responder:



6. *Mark only one oval per row.*

[illegible]

## Serie K1

7. *Mark only one oval per row.*

[illegible]

8.

Mark only one oval per row.

[illegible]

9.

Mark only one oval per row.

[illegible]



12. *Mark only one oval per row.*

[illegible]

13. *Mark only one oval per row.*

[illegible]

14.

Mark only one oval per row.

[illegible]

15.

Mark only one oval per row.

[illegible]

16.

Mark only one oval per row.

[illegible]

2K

17.

Mark only one oval per row.

[illegible]

18.

Mark only one oval per row.

[illegible]

19.

Mark only one oval per row.

[illegible]















32.

Mark only one oval per row.

[illegible]

33.

Mark only one oval per row.

[illegible]



























56.

*Mark only one oval per row.*

	Agréable à l'oreille, harmonieux.	Philosophie qui nie l'existence de sens.	Qui n'est pas dur.	Ton, style déclamatoire.	Incapable de parler.	Présent partout en même temps.
28. <b>Emphase</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
29. <b>Mou</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
30. <b>Muet</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

---

This content is neither created nor endorsed by Google.

Google Forms

## **9.2 Formato de Carta de Consentimiento.**

### **Carta de consentimiento**

El propósito de este documento es darle a conocer toda la información necesaria para que usted pueda decidir libremente si desea participar en la investigación para la realización del trabajo de tesis de maestría y que a continuación se describe en forma resumida:

**Título del estudio:** Desarrollo de la riqueza léxica de estudiantes universitarios de francés L2

**Resumen:** Esta investigación se enfoca en el estudio del desarrollo de la competencia léxica en estudiantes mexicanos que aprenden francés como segunda lengua (L2), específicamente dentro de la Licenciatura en Lenguas Modernas en Francés (LLMF). El estudio busca entender cómo evoluciona el vocabulario de estos estudiantes a lo largo de la licenciatura.

#### **Al respecto expongo que:**

He sido informado(a) sobre el objetivo del estudio a desarrollar. Los datos obtenidos en la investigación en la que participaré serán absolutamente confidenciales, no aparecerá mi nombre ni mis datos personales en libros, revistas y otros medios de publicidad derivados de la investigación ya descrita. Toda mi información será utilizada exclusivamente para la investigación.

He leído el documento, entiendo las declaraciones contenidas en él y la necesidad de hacer constar mi consentimiento, para lo cual lo firmo libre y voluntariamente.

Yo (nombre completo) \_\_\_\_\_  
consiento en participar en la investigación denominada: *"Desarrollo de la riqueza  
léxica en estudiantes universitarios de francés L2"*.

**Fecha:** \_\_\_\_\_ **Firma:** \_\_\_\_\_

**Investigador responsable:**  
Etienne Luis García Greiner

### **9.3 Link de Acceso a Datos de la Prueba**

[https://drive.google.com/file/d/1kvxtJiOqyld81c1piIGM\\_dJiHpaOPcM0/view?  
usp=drive\\_link](https://drive.google.com/file/d/1kvxtJiOqyld81c1piIGM_dJiHpaOPcM0/view?usp=drive_link)