

Ingrid Schiefloe Lange

El tamaño del vocabulario pasivo de estudiantes de español en la universidad noruega

Bacheloroppgave i Spansk

Veileder: Gro Nygård

Juni 2023



NTNU

Kunnskap for en bedre verden

Ingrid Schiefloe Lange

El tamaño del vocabulario pasivo de estudiantes de español en la universidad noruega

Bacheloroppgave i Spansk
Veileder: Gro Nygård
Juni 2023

Norges teknisk-naturvitenskapelige universitet
Det humanistiske fakultet
Institutt for språk og litteratur



Kunnskap for en bedre verden

Resumen

Este trabajo es una investigación del vocabulario pasivo de 17 estudiantes de cursos del primer año de español como lengua extranjera en una universidad noruega. Los participantes hicieron una encuesta de información general y una prueba de reconocimiento léxico (X-Lex) de 120 palabras. 100 de ellas representaron las 5000 palabras más frecuentes del español, y 20 eran pseudopalabras. Basado en los resultados de la prueba calculamos una aproximación de los vocabularios pasivos de los participantes, y el vocabulario promedio. Basado en sus respuestas del cuestionario de información general observamos en qué coincidieron los estudiantes con vocabularios más grandes del promedio, para encontrar qué factores pueden influir en el tamaño del vocabulario pasivo. Además, hicimos una comparación con los resultados de otra investigación de reconocimiento léxico de estudiantes del español, el trabajo de Fairclough (2011). Nuestros resultados muestran que hay gran variación en el tamaño del vocabulario de los estudiantes, pero que el promedio es de 2588 palabras. También indican que la exposición extramural de la lengua parece influir bastante en el vocabulario, y que los estudiantes obtuvieron resultados similares a los de otras investigaciones.

Sammendrag

Dette arbeidet er en undersøkelse av det passive ordforrådet til 17 studenter av emner fra det første året ved spansk som fremmedspråk på et norsk universitet. Deltakerne gjennomførte en spørreundersøkelse som inneholdt spørsmål om bakgrunnsinformasjon, og en prøve som testet evnen til å gjenkjenne ord (X-Lex) som besto av 120 ord. 100 av disse representerte de 5000 hyppigste ordene i spansk, og 20 var falske ord. Basert på resultatene av prøven, kalkulerte vi en tilnærming av de passive ordforrådene til deltakerne og det gjennomsnittlige ordforrådet. Basert på svarene deres i spørreskjemaet om bakgrunnsinformasjon fant vi også ut hva studentene med større enn gjennomsnittlig ordforråd hadde til felles, for å finne ut hvilke faktorer som kan påvirke størrelsen på det passive ordforrådet. I tillegg sammenlignet vi resultatene våre med resultater fra en annen undersøkelse av ordgjenkjenning blant spanskstudenter, arbeidet til Fairclough (2011). Resultatene våre viser at det er stor variasjon i størrelsen på studentenes ordforråd, men at gjennomsnittet er 2588 ord. Resultatene viser også at ekstramural språkkontakt ser ut til å ha en betydelig innflytelse på ordforrådet, og at studentene hadde resultater som ligner de fra andre undersøkelser.

Índice

1. Introducción	3
2. Teoría	4
2.1 La prueba X-Lex.....	4
2.2 La distinción entre el vocabulario receptivo y el vocabulario pasivo	4
2.3 ¿Cuántas palabras hay que saber para utilizar una L2?	6
2.4 El tamaño de vocabulario y los niveles del MCER	8
2.5 La exposición extramural en el aprendizaje de lenguas	10
3. Método	11
3.1 Los participantes.....	11
3.2 La encuesta	11
3.3 El cálculo de la puntuación.....	13
4. Resultados	14
4.1 El vocabulario.....	14
4.2 Las pseudopalabras.....	15
4.3 Correlaciones entre el vocabulario y otros factores.....	16
4.4 Comparación con los resultados de Fairclough (2011)	18
5. Conclusiones	20
Referencias bibliográficas	22
Apéndices	24
La lista de palabras	24
El cuestionario de información general (<i>Bakgrunnsinformasjon</i>).....	25

1. Introducción

El vocabulario es una parte fundamental de cualquier lengua. Es lo que nos permite comunicar con otros y expresarnos de una manera coherente y precisa. Sin un vocabulario suficiente, no es posible enunciar lo que se quiere decir con exactitud y eficacia. Además, hay que entender una parte considerable de una lengua para poder entender el lenguaje escrito y hablado. Por eso, el vocabulario tiene un papel importante en todas las destrezas de un idioma. El objetivo principal de esta investigación es encontrar cuántas palabras conocen los estudiantes de español en una universidad noruega. En concreto, encontraremos el tamaño de su vocabulario pasivo basándonos en su conocimiento de las 5000 palabras más frecuentes del español. Los participantes son estudiantes de dos asignaturas universitarias de lengua española como lengua extranjera. Estas asignaturas normalmente se hacen durante el primer año de un grado de tres años, o como un programa independiente de un año. También hay estudiantes que hacen estas asignaturas como cursos individuales dentro de otro programa de estudio. En el presente trabajo queremos responder a las siguientes preguntas de investigación:

1. ¿Cuál es el tamaño del vocabulario pasivo de los estudiantes?
2. ¿Qué conexiones hay entre el tamaño de vocabulario y otros factores, basado en qué características tienen en común los estudiantes que tienen un vocabulario más grande o pequeño?
3. ¿Cómo se comparan nuestros resultados con los de otras investigaciones?

Primero, en la parte de teoría, explicaremos el formato de prueba llamado X-Lex, y la definición del vocabulario pasivo en nuestra investigación y cómo se distingue del vocabulario receptivo. Además, veremos qué algunas investigaciones anteriores han encontrado en cuanto a qué tamaño de vocabulario se necesita para utilizar una segunda lengua, cómo este corresponde a los niveles del Marco común europeo de referencia para las lenguas, y cómo la exposición extramural influye en el aprendizaje de lenguas extranjeras. A continuación, en la sección de método, describiremos quiénes son los participantes, además de los contenidos de nuestra encuesta y cómo hemos elegido calcular la puntuación de la prueba X-Lex, y por qué hemos elegido ese método. Seguidamente presentamos nuestros resultados: el tamaño del vocabulario de los participantes, algunas cosas que destacan en cuanto a la selección de pseudopalabras en la prueba, y qué conexiones vemos entre el tamaño de vocabulario y las respuestas en la encuesta de información general. También habrá una comparación con una investigación anterior de estudiantes de lengua española, de Fairclough (2011). Al final habrá una conclusión donde resumimos las respuestas a nuestras preguntas de investigación.

2. Teoría

2.1 La prueba X-Lex

Para coleccionar los datos para la investigación, hicimos una encuesta en forma de una prueba de reconocimiento léxico llamada X-Lex. Este tipo de prueba mide la capacidad de reconocer las palabras más comunes de una lengua y evitar seleccionar palabras falsas, o pseudopalabras. Los resultados son una buena medida de la competencia general de un idioma (Milton, 2010; Fairclough, 2011), y la prueba se ha utilizado para investigaciones de otras lenguas extranjeras, como el francés, el inglés, y el griego (Lindqvist, 2016; Milton, 2008; Milton & Alexiou, 2009). Aunque la utilizamos en esta investigación para medir el vocabulario pasivo, la prueba X-Lex es principalmente una prueba de reconocimiento léxico. Por eso es importante recordar que, en realidad, medimos el conocimiento pasivo de las palabras más frecuentes, y calculamos un vocabulario aproximado basado en esto. Además, como la prueba se limite a las 5000 palabras más frecuentes, no necesariamente dará una medida exacta del vocabulario del estudiante, ya que no incluye las palabras menos frecuentes. Fairclough (2011) realizó una investigación del reconocimiento léxico de estudiantes del español en los Estados Unidos, que demuestra que los resultados de una prueba de vocabulario basada en las 5000 palabras más frecuentes, como la prueba X-Lex, es una manera práctica de examinar la capacidad general de una lengua. Esto se debe a que los resultados correlacionan con los de otras pruebas de vocabulario, y el vocabulario, a su vez, se ha relacionado con otras destrezas como la comprensión escrita y oral (p. 275). En su versión de la prueba, Fairclough pidió a los participantes que seleccionaran las palabras que “pueden explicar a un amigo” (p. 279). Esto corresponde a un conocimiento receptivo de una palabra. En nuestra versión, en cambio, queremos investigar el vocabulario pasivo. Aunque este tipo de vocabulario está relacionado con el vocabulario receptivo, hemos decidido establecer una distinción entre los dos, lo que haremos en la siguiente sección, para que no haya dudas en cuanto a la diferencia entre los dos términos.

2.2 La distinción entre el vocabulario receptivo y el vocabulario pasivo

Cuando se habla del vocabulario, es importante aclarar de qué tipo se habla, como hay varios tipos de vocabulario y distintas maneras de clasificarlos. Nation (2012) hace una distinción principal entre lo receptivo y lo productivo, en que lo receptivo tiene que ver con comprender o entender lo que leemos o escuchamos, y lo productivo con producir lengua por escribir o hablar. Sin embargo, no son definiciones perfectas, ya que existen rasgos productivos en las

habilidades receptivas, por ejemplo, se puede decir que producimos significados cuando leemos y escuchamos. Su descripción básica de los dos es que el vocabulario receptivo se relaciona con sacar significados por leer y escuchar la lengua, y el productivo con expresar significados por escribir y hablar (pp. 46-47). Más detalladamente, saber una palabra en el sentido receptivo quiere decir que se puede reconocerla al escucharla o leerla además de reconocer su composición morfológica, saber que la palabra señala un significado particular y si se ha utilizado correctamente en el contexto (pp. 48-50). En la prueba X-Lex no se puede comprobar estos últimos dos puntos, como las palabras aparecen en una lista sin ningún contexto. Los términos pasivo y receptivo a veces se usan como sinónimos, por ejemplo, en Laufer (1998) que distingue entre un “conocimiento receptivo (pasivo)” (p. 257) y dos tipos de conocimiento productivo de palabras. Ya que a veces se utilizan como sinónimos en otros textos, es difícil encontrar comparaciones y distinciones entre los dos, pero en esta investigación queremos diferenciar entre los dos términos. En relación con la distinción entre el vocabulario receptivo y el pasivo también cabe mencionar los términos amplitud de vocabulario (*vocabulary breadth*) y profundidad de vocabulario (*vocabulary depth*). La amplitud del vocabulario es el número de palabras que una persona conoce sin importar en qué medida las conozca. Esto contrasta con la profundidad del vocabulario, que está relacionada con el conocimiento de cómo funcionan las palabras y sus varios posibles significados (Milton, 2010, pp. 218-219). Otra cuestión es cómo se comparan los tamaños de los vocabularios receptivos y pasivos.

Laufer (1998) en su investigación del desarrollo de vocabulario en lenguas extranjeras ha descubierto que el vocabulario receptivo es más grande que el activo, que sugiere que el pasivo, según nuestra definición, será aún más grande ya que el umbral para la inclusión de palabras en el vocabulario pasivo es más bajo. También se ha encontrado que entre los estudiantes de niveles más altos la diferencia entre el vocabulario receptivo y el activo es incluso más grande. Esto quiere decir que el número de palabras que el estudiante puede utilizar correctamente es más pequeño que el número de palabras conocidas receptiva y pasivamente. En resumen, podemos definir el vocabulario pasivo como un aspecto del vocabulario que se relaciona con la amplitud del vocabulario, y sus requisitos para conocer una palabra son más sencillos que los del vocabulario receptivo. Se trata de un conocimiento pasivo de una palabra, en que se reconoce su forma, pero no necesariamente su sentido. Por eso suponemos que el pasivo es la clasificación de vocabulario más grande. No obstante, no podemos saber con seguridad la magnitud de la diferencia de tamaños.

Para medir el conocimiento receptivo de palabras de maneras más precisas se puede utilizar otras formas de pruebas más complejas, que ponen las palabras en un contexto para comprobar si el participante sabe utilizarlas correctamente (por ejemplo, el *Vocabulary Levels Test* descrito en Nation, 2012, pp. 35-36). Sin embargo, también hay investigaciones que utilizan X-Lex para medir el conocimiento receptivo. En estos casos, si se utiliza solo la prueba de reconocimiento, no hay forma de verificar con total seguridad que los participantes realmente entienden las palabras que seleccionan. Para solucionar esto se puede, por ejemplo, pedir a los participantes que traduzcan algunas de las palabras de la prueba después de acabarla (Mochida & Harrington, 2006). Sin embargo, esto añadiría complejidad a la prueba, y no debe ser necesario ya que la inclusión de las pseudopalabras sirve para determinar si los participantes han respondido con sinceridad. Además, la formulación de la tarea o las instrucciones de la prueba también influirán en las respuestas de los participantes, y pueden determinar la distinción entre el conocimiento receptivo y el pasivo. Como ya hemos visto, Fairclough (2011) en su investigación pidió a los participantes que seleccionaran las palabras que podían entender en al menos uno de sus sentidos, que sería un conocimiento receptivo. En nuestra encuesta, en cambio, hemos optado por pedir a los participantes que marquen las palabras que simplemente pueden reconocer como palabras españolas. Creemos que esto corresponde más con un conocimiento pasivo, y además simplificará el cálculo de la puntuación y del vocabulario, como volveremos a ver cuándo tratamos cómo calcular las puntuaciones en el apartado 3.3. En resumen, para medir el vocabulario pasivo no ponemos las palabras en un contexto, porque no nos importa si los participantes pueden entender la palabra o no. Además, para distinguir entre el conocimiento receptivo y pasivo de una palabra, podemos decir que el pasivo es menos complejo que la comprensión receptiva. No importa el significado ni el uso de la palabra, solo la capacidad de reconocer su forma fuera de un contexto. Teniendo estos factores en mente, el formato de la prueba X-Lex es ideal para medir el conocimiento pasivo del vocabulario de una manera sencilla.

2.3 ¿Cuántas palabras hay que saber para utilizar una L2?

Hay varias investigaciones que han intentado hallar cuántas palabras se necesita entender o poder utilizar para usar una segunda lengua (L2) efectivamente. Un número recurrente es el conocimiento de un 95 % de un texto para comprensión general (Laufer, 1989; Milton y Alexiou, 2009; Nation 2012). Se ha encontrado que si se conoce el 95 % de las palabras de un texto, se puede comprender un texto escrito con éxito. La misma cobertura también parece

bastar para la comprensión oral. Un uso independiente de una lengua exige una cobertura del 98 %, pero con la asistencia de un diccionario el 95 % será suficiente (Robles-García, 2022, pp. 966-967). Nation (2012) también menciona una cobertura del 98 % como un objetivo para estudiantes para poder leer independiente, además hay investigaciones que sugieren que este es un número manejable de palabras desconocidas para poder leer un texto (Hu & Nation, 2000; Schmitt et al., 2011; van Zeeland & Schmitt, 2012, citado en Nation, 2012, p. 14). En cuanto a qué número de palabras se debe saber para alcanzar este porcentaje en español, Davies (2005) ha encontrado que esto varía según el género del texto. Es decir, un vocabulario de cierto tamaño cubriría distintos porcentajes de diferentes géneros de texto. Con sólo las 1000 palabras más frecuentes, se tiene capacidad de reconocer un 76-79,6 % de un texto escrito, pero aprender las palabras de las siguientes bandas de frecuencia no aumenta la cobertura mucho. Añadir las siguientes 1000 palabras sólo cubriría un 6-8 % más del texto. La tercera banda de 1000 palabras añade unos 3-4 % más, lo que nos lleva a que un vocabulario que consiste en las 3000 palabras más comunes del español cubre un 88,2 % de un texto no ficción, un 89,6 % de un texto de ficción, y un 94 % de un texto oral, ya que hay más repetición de ciertas palabras en el lenguaje hablado (p. 109).

Además, Robles-García (2022) ha encontrado que el aprendizaje de las 3000 palabras más frecuentes se hace principalmente durante los primeros semestres de la carrera universitaria. El aprendizaje de estas palabras de frecuencia alta disminuye en el segundo año y durante las clases de niveles más altos (p. 972). En general, los estudiantes del español como lengua extranjera en su investigación faltan conocimiento de las 3000 palabras más frecuentes independientemente de su nivel y cuánto tiempo llevan estudiando la lengua. Más en concreto, los estudiantes en los niveles más altos solamente reconocieron un 74,5 % de las 3000 palabras más frecuentes, que está lejos de la cobertura del 95 % necesario para la comprensión (p. 975). Stæhr (2008), una investigación de estudiantes daneses del inglés como L2, encontró que la habilidad de leer es la que depende más del tamaño del vocabulario. Además, encontró que los estudiantes que no dominaron las 2000 palabras más frecuentes tuvieron peor comprensión de lectura (p. 147). Otro resultado fue que el conocimiento receptivo de las 2000 palabras más frecuentes era importante para obtener resultados mejores en pruebas de lectura, escuchar, y escribir (p. 149). En general, las 2000 o 3000 palabras más frecuentes son especialmente importantes para estudiantes de lenguas extranjeras para poder utilizar su L2, pero los estudiantes, también en niveles universitarios más altos, faltan conocimiento de ellas, como ha

concluido Robles-García (2022) y otras investigaciones (Laufer, 2000; Milton, 2006; Stæhr, 2008; Webb & Chang, 2012).

2.4 El tamaño de vocabulario y los niveles del MCER

Aunque la versión actual del Marco común europeo de referencia para las lenguas (MCER) no menciona números específicos en cuanto al tamaño de vocabulario, su volumen complementario (Consejo de Europa, 2020) contiene descripciones detalladas sobre cómo el estudiante debe poder utilizar la lengua en los distintos niveles, en varias categorías de uso. El volumen complementario menciona el conocimiento de palabras de frecuencia alta, y un lenguaje cotidiano de uso muy frecuente, en cuanto a la habilidad de leer por placer. Cuando se habla de leer por placer, se trata de “[...] textos escritos y signados, tanto de ficción como de no ficción [...]” (p. 70). Hemos elegido esta categoría como ejemplo porque corresponde a las categorías de textos que ya hemos visto en las investigaciones de Davies (2005) y Robles-García (2022). En cuanto a los niveles del MCER, en esta sección las competencias recomendadas para el Nivel B1 son que el estudiante comprende la información principal de textos de carácter divulgativo, y poemas y letras de canciones que empleen un lenguaje sencillo. Además, puede entender narraciones, guías y artículos de revistas que empleen un lenguaje cotidiano de uso muy frecuente. Igualmente, puede seguir el argumento de varios tipos de historias escritas en un lenguaje cotidiano de uso muy frecuente, siempre que pueda hacer un uso frecuente del diccionario (Consejo de Europa, 2020, p. 71). La descripción del nivel B2 es parecida, pero el estudiante debe tener un elevado grado de autonomía, y todavía se espera que tome su tiempo y use un diccionario (p. 71).

No hay números oficiales de cuántas palabras se debe conocer en los distintos niveles del MCER, pero investigaciones como Milton (2010) han intentado llenar este vacío y encontrar qué tamaños de vocabulario son típicos entre los estudiantes de cada nivel del MCER. Para encontrar estos números, se han comparado las puntuaciones de X-Lex y los niveles del MCER de estudiantes de lenguas extranjeras. El resultado fue que los números varían entre lenguas, es decir, el tamaño estimado del vocabulario depende de qué lengua se examina y también en cómo se ha seleccionado las palabras que componen la encuesta. Por ejemplo, para obtener una cobertura de un 80 % de un texto, se necesitan más palabras en inglés que en francés, y aún más en griego (Milton & Alexiou, 2009, p. 205). Asimismo, hay variación entre el tamaño del vocabulario para cada nivel del MCER entre la misma lengua. El cuadro 1 contiene las

puntuaciones promedias de X-Lex y los niveles correspondientes del MCER de estudiantes de varias lenguas extranjeras en distintos países. Los niveles del MCER de los estudiantes fueron determinados por sus profesores (Milton, 2010, p. 225). Además, podemos calcular un tamaño de vocabulario promedio a través de estas lenguas para encontrar un número general de palabras conocidas para cada nivel del MCER, lo que muestra el cuadro 2.

MCER	Inglés en Hungría	Inglés en Grecia	Francés en Grecia	Francés en España	Griego en Grecia
A1		1477	1125	894	1492
A2		2156	1756	1700	2237
B1	3135	3263	2422	2194	3338
B2	3668	3304	2630	2450	4012
C1	4340	3690	3212	2675	
C2		4068	3525	3721	

Cuadro 1, vocabulario promedio para lenguas extranjeras en varios países según las puntuaciones de X-Lex. *Griego en Grecia* se refiere al griego como lengua extranjera (Milton, 2010, p. 225, datos originalmente de Milton & Alexiou, 2009)

Nivel de MCER	Vocabulario promedio (X-Lex)	Vocabulario mínimo	Vocabulario máximo
A1	1247	894	1492
A2	1962	1700	2237
B1	2870	2194	3338
B2	3213	2450	4012
C1	3479	2675	4340
C2	3771	3525	4068

Cuadro 2, números promedios, mínimos, y máximos de todos los estudiantes de lenguas extranjeras del cuadro 1

Los resultados de este cálculo muestran que el tamaño de vocabulario no necesariamente puede traducirse directamente al nivel de CEFR. Por ejemplo, en el nivel C2, aunque el promedio es más grande que el de C1, el máximo es menor. Además, es posible que dos estudiantes que tienen el mismo tamaño de vocabulario no estén en el mismo nivel porque sus competencias en otras áreas de la lengua son distintas. Naturalmente, lo mismo ocurre a la inversa, que dos estudiantes que tienen el nivel B1, por ejemplo, pueden tener muy distintos tamaños de

vocabulario. Estos números deben ser una representación válida como aproximaciones de tamaños de vocabulario para varias lenguas, aunque no sean tan precisos como los rangos para lenguas específicas. Adicionalmente, en cuanto al nivel C2 no debe haber un límite máximo para el tamaño del vocabulario, el número máximo aquí simplemente refleja la puntuación más alta entre los estudiantes de este nivel.

2.5 La exposición extramural en el aprendizaje de lenguas

Aprender una lengua no siempre es un proceso que tiene lugar exclusivamente dentro de un aula u otros ámbitos de aprendizaje. El estudiante puede entrar en contacto con una L2 en su tiempo libre, ya sea deliberadamente o no. Este contacto puede ayudar en la adquisición de lengua o vocabulario, y se llama exposición extramural. Sundqvist & Sylvén (2016) define la exposición extramural como el contacto con el idioma fuera de los muros, que quiere decir los encuentros con la lengua que un estudiante tiene separado de clases de lengua. Este contacto con la lengua no es iniciado por un profesor u otro tipo de pedagogo, sino más bien por el estudiante mismo, o un amigo o pariente. Algunas actividades mencionadas como ejemplos de exposición extramural son mirar películas, series, u otras formas de videos, escuchar música, leer textos, jugar a videojuegos, o utilizar la lengua de otras maneras solo o acompañado, en la vida real o a través de internet. La exposición extramural de lengua incluye una amplia variedad de actividades, e incorpora varias destrezas esenciales de lengua (pp. 6-7). En cuanto a los efectos de la exposición extramural, se ha encontrado que influye positivamente en el vocabulario de estudiantes del inglés como L2, y aunque su relación con competencia general de lengua sea más complicada, hay una relación positiva entre la exposición extramural y el aprendizaje exitoso de lengua (Sundqvist, 2009; Sundqvist & Sylvén, 2016). Asimismo, estudiantes del inglés como L2 de niveles más altos utilizan más frecuentemente varias formas de exposición extramural, como mirar televisión y películas y escuchar música en la lengua extranjera (Griffiths, 2008, p. 90). El aprendizaje extramural es relevante en este trabajo porque nuestra encuesta incluye preguntas sobre cómo los estudiantes utilizan lenguas extranjeras fuera de sus estudios en la universidad, y por eso la exposición extramural es uno de los aspectos que vamos a ver cómo influye en el vocabulario de los estudiantes.

3. Método

3.1 Los participantes

Hicimos la recolección de datos durante dos clases universitarias de diferentes asignaturas españolas. Había varios alumnos presentes en ambas clases a los que se indicó que no participaran si ya habían contestado a la encuesta. Además, hubo un número pequeño de participantes que rellenaron la encuesta fuera de las clases. En total hubo 18 participantes, 8 de la primera clase, 7 de la segunda clase, y 3 que participaron fuera de las clases de manera digital. 17 de los participantes tenían el noruego como lengua materna, y entre ellos 2 también tenían otra lengua materna, 1 de ellos el español. En el análisis de los resultados hemos excluido los resultados de la persona con el español como lengua materna. Las asignaturas que realizan los participantes son del segundo semestre del primer año del programa de estudios, y la mayoría de los participantes estuvieron en su primer año de estudios de español en la universidad. También hubo algunos participantes que ya estaban en su segundo o tercer año, por consiguiente, hay una disparidad entre cuánto tiempo llevan estudiando la lengua. También puede variar cuándo empezaron a aprender el idioma durante la carrera y cuántos años de escolarización tienen en total. Por eso suponemos que habrá grandes variaciones en los niveles lingüísticos entre los estudiantes, dado que no hay requisitos o un examen de admisión para los estudios. No obstante, hay recomendaciones de qué competencia se debe tener antes de empezar el programa de estudios. Al inscribirse en el curso del primer semestre se aconseja que los estudiantes tengan un nivel de A2 o que han aprendido español en el bachillerato, y después de haber terminado los cursos del primer semestre, los estudiantes deben tener el nivel B1, y el B2 al terminar el segundo semestre.

3.2 La encuesta

La prueba se hizo en forma de encuesta digital en *Nettskjema*, una herramienta de encuesta en línea desarrollada por la Universidad de Oslo. La prueba consistió en dos partes. La primera parte de la prueba era un cuestionario de información general, con preguntas sobre las experiencias previas del participante con el español, el inglés, y otras lenguas que pudiera conocer. Preguntamos a los participantes sobre qué programa de estudios hacen, su lengua materna, si tienen amigos o familia con que hablan otras lenguas que la noruega, su edad al empezar a aprender el inglés y el español, y cómo describen su propia capacidad en el inglés, el español, y en otras lenguas conocidas en una escala de 1 a 5. Además, preguntamos sobre la

frecuencia estimada con que han escuchado o leído inglés y español fuera de la escuela o universidad semanalmente los últimos 12 meses, si han vivido en el extranjero y cuánto tiempo, y al final si tienen o han tenido dificultades de leer o escribir, por ejemplo dislexia, u otros diagnósticos que pueden influir en el aprendizaje de lengua o vocabulario. Para estas últimas preguntas hemos recibido permiso de NSD (*Norsk senter for forskningsdata*) para recoger información sensible.

La segunda parte era la prueba X-Lex, que presentó a los participantes una lista de 120 palabras. Entre ellas, hay 100 palabras recogidas de una lista de las 5000 palabras españolas más frecuentes. Se ha dividido esta lista en cinco categorías de frecuencia de 1000 palabras cada una, y seleccionado 20 palabras de cada de estas categorías. Además, hay 20 pseudopalabras, que son palabras falsas que parecen ser palabras reales a partir de su morfología, pero no tienen ningún sentido ni serían utilizadas por un hablante nativo. También hemos verificado con hablantes nativos que estas pseudopalabras no son demasiadas parecidas a palabras reales. Las palabras que forman la lista vienen de Fairclough (2011), que las ha recogido de *A Frequency Dictionary of Spanish: Core Vocabulary for Learners* de Mark Davies (2006). Es un diccionario de frecuencia del español que está hecho de un corpus del español escrito además de transcripciones de lengua hablada. Contiene un balance de textos de España (el 43 %) y América Latina (el 57 %), y prácticamente todos los textos son de 1970-2000, la mayoría de la década de 1990. (pp. 2-3). Para elegir las palabras que componen la lista, se eliminaron los 114 primeros elementos del diccionario porque eran elementos funcionales con poco contenido semántico, como determinantes. A continuación, se seleccionó una palabra aleatoriamente de cada 40, produciendo 24 palabras para representar cada banda de frecuencia. Además, se incluyeron los cognados ya que pueden también ayudar o confundir en el reconocimiento de una palabra (Fairclough, 2011, pp. 278-279). Un cognado es una palabra que es similar tanto en su forma como en su significado en dos o más lenguas (Lindqvist, 2020, p. 149). Para nuestra versión de la encuesta, hemos utilizado sólo 20 de las 24 palabras para cada categoría de frecuencia. Una lista completa de las palabras que formaron parte de la prueba de reconocimiento léxico y el cuestionario de información general se pueden encontrar como apéndices al final del texto.

3.3 El cálculo de la puntuación

Hay varios métodos que se pueden emplear para calcular una puntuación para la prueba X-Lex, y hay algunas variables que se deben tener en cuenta al elegir el método de calcular la puntuación. No solamente el conocimiento que tiene el participante de una palabra influye en sus respuestas en la prueba. Como explica Huibregtse et al. (2002), un aspecto que desempeña un papel importante es la adivinanza. Una persona puede dudar de si conoce o no la palabra en cuestión, o puede no estar segura de su significado concreto. En estos casos, generalmente, los participantes no adivinan aleatoriamente. Huibregtse et al. lo llama adivinanza sofisticada (*Sophisticated guessing*, p. 230). Otra cosa que hay que tener en cuenta es que cada persona tiene su propio estilo de adivinar. Cuando hay duda, algunos participantes tienden a responder *sí* y otros *no*. Esto se llama *response style*, o estilo de respuesta. Por ejemplo, los que tienden a responder *no* siempre que no tienen confianza total, tienen un estilo de respuesta conservativo (pp. 230-239). En Huibregtse et al., también se discute y compara distintas maneras de calcular la puntuación para una prueba de vocabulario con palabras reales y pseudopalabras. Estos métodos se basan en el *hit rate* (el porcentaje de palabras reales seleccionadas) y el *false alarm rate* (el porcentaje de pseudopalabras seleccionadas), y resultan en una puntuación entre 0 y 1, donde 1 es la mejor. Varios de estos métodos intentan tomar en cuenta los efectos de la adivinanza y el estilo de respuesta para obtener resultados lo más precisos posibles, pero el método más sencillo y directo es el llamado *h-f*, o El número de respuestas correctas (*The number of correct responses, h-f* en lo sucesivo), en que simplemente se calcula el porcentaje de palabras reales seleccionadas y se resta el porcentaje de pseudopalabras seleccionadas (p. 231).

En nuestra versión de la encuesta hemos quitado el requisito de entender las palabras que forman parte de la prueba, y el participante solo tiene que indicar si reconoce la palabra o no. Esto debe más o menos eliminar el elemento de adivinanza, o al menos reducir su importancia, que crea menos posibilidades de duda. Además, pedir a los estudiantes que simplemente seleccionen palabras que reconocen debe eliminar el problema de que se puede interpretar de distintas maneras lo que quiere decir saber una palabra, como discute Laufer & Paribakht (1998, p. 366). Como no hay gran diferencia entre los resultados de los varios métodos de calcular la puntuación (Mochida & Harrington, 2006), y corregir para la adivinanza no es muy importante en nuestra investigación, hemos optado por utilizar el método más sencillo, *h-f*. En la práctica, esto da los mismos resultados que el método utilizado por Milton (2008) y su versión francesa de X-Lex, en que cada palabra real que ha elegido el participante le da 50 puntos, y cada

pseudopalabra elegida le da -250 puntos (p. 340). Con este número se puede aproximar el tamaño del vocabulario del participante, con una puntuación máxima de 5000 palabras. Este método da el mismo resultado que multiplicar el resultado de $h-f$ con 5000, el número total de palabras en las bandas de frecuencias representadas en la prueba, y podemos utilizar cualquiera de los dos para calcular el tamaño del vocabulario pasivo.

4. Resultados

Los resultados de la encuesta se organizaron automáticamente en la página web de *Nettskjema*, y también fueron descargados en un documento Excel. En este documento codificamos los datos para facilitar los cálculos.

4.1 El vocabulario

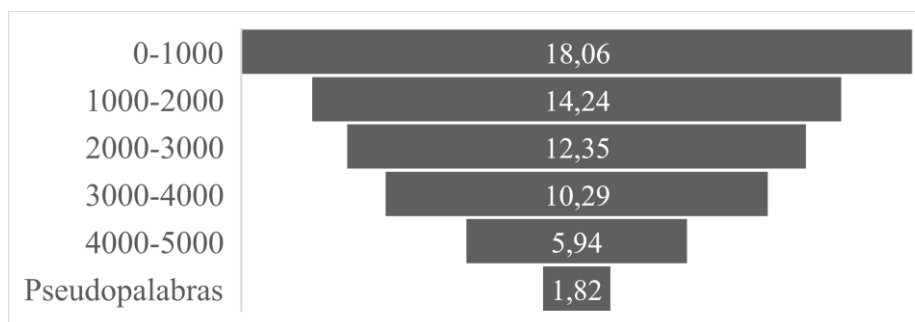
Los resultados del cálculo de la puntuación muestran que hay bastante variación en el tamaño del vocabulario pasivo de los participantes. El cuadro 3 muestra la puntuación mínima, máxima y promedia, y la desviación típica de los 17 participantes incluidos en los cálculos. Esta puntuación equivale al tamaño del vocabulario pasivo. El tamaño promedio del vocabulario pasivo de los participantes fue 2588 palabras. Si comparamos esta puntuación promedia con los números de Milton (2010) (en el apartado 2.4, cuadro 2), vemos que el participante promedio estará alrededor del nivel B1, lo que corresponde a lo que se espera de los alumnos, como descrito en el apartado 3.1.

Vocabulario mínimo	Vocabulario máximo	Vocabulario promedio	Desviación típica
1400	3900	2588	690

Cuadro 3, el vocabulario pasivo mínimo, máximo, y promedio de los 17 estudiantes que no tenían el español como lengua materna

Adicionalmente, podemos ver el número promedio de palabras reconocidas en cada categoría de frecuencia en el cuadro 4. Hay una disminución gradual del número promedio de palabras que reconocieron los participantes cuando bajamos en la frecuencia de las palabras. De las 1000 más comunes, reconocieron 18,06 palabras en promedio, y al bajar en frecuencia hay una disminución gradual hasta solamente 5,94 palabras reconocidas en promedio para la quinta banda de frecuencia. Comparado con los resultados de Robles-García (2022) que vimos en el apartado 2.3, parece que nuestros participantes igualmente, en general, no conocieron muy bien

las palabras de la segunda y tercera banda de frecuencia. Esto puede implicar que no conocen suficientes palabras para obtener una comprensión adecuada de la lengua, según las investigaciones que vimos en el apartado 2.3.



Cuadro 4, números promedios de palabras seleccionadas de cada banda de frecuencia, y de las pseudopalabras

4.2 Las pseudopalabras

Lo más común entre los participantes fue seleccionar solamente una pseudopalabra (6), y el número promedio de pseudopalabras seleccionadas fue bastante bajo, de 1,82. Además, hubo 5 participantes que no seleccionaron ninguna pseudopalabra. También cabe mencionar que las únicas palabras que no fueron seleccionadas por ningún estudiante todas fueron pseudopalabras. Sin embargo, hay algunas de las palabras falsas que destacan por ser seleccionadas muy frecuentemente. Las pseudopalabras más seleccionadas fueron *festivar* (6 veces), *calorar* (5), *gustamente* (5) y *reprendizaje* (5). Entre estas palabras, *festivar* y *calorar* fueron seleccionadas más por los estudiantes de la primera clase, mientras que *gustamente* y *reprendizaje* fueron seleccionadas más o menos uniformemente entre todos los participantes. Esto puede deberse al contenido de la clase durante la cual pedimos al primer grupo de estudiantes que hicieran la encuesta. El tema de esta clase eran la morfología y las familias léxicas. Por eso, es posible que la información adquirida sobre este tema inmediatamente antes de la prueba influyera en qué palabras los estudiantes consideraban que podrían ser españolas. Además, la pseudopalabra *festivar* es bastante parecida al verbo noruego *å feste* (ir a fiestas), y posiblemente fue interpretada incorrectamente como un cognado de ello. También es interesante que el único participante que no tiene el noruego como una de sus lenguas maternas no ha seleccionado esta palabra, pero esto es un tamaño de muestra demasiado pequeño para sacar conclusiones sobre esta cuestión. La teoría de que esta palabra en particular fue seleccionada tanto por su posible interpretación como un cognado se apoya en una investigación anterior que encontró que los

cognados de la lengua materna, y también de otras lenguas segundas, pueden influir en el conocimiento de palabras en una lengua extranjera. Es posible que los participantes hayan seleccionado palabras que, aunque no reconocieran como palabras españolas, las reconocieron como cognados de palabras de otras lenguas conocidas (Lindqvist, 2020, p. 159).

Algunas de las funciones de las pseudopalabras en la prueba es ajustar la puntuación y hacer correcciones para la adivinanza, y otras investigaciones que utilizan la prueba X-Lex tienden a excluir las respuestas con un 20 % o más de pseudopalabras seleccionadas, porque puede indicar que el participante haya recurrido demasiado a la adivinanza en lugar de responder sinceramente (por ejemplo, en Milton, 2006; Milton, 2008). En nuestra investigación hubo 3 participantes que seleccionaron 4 pseudopalabras, y 2 que seleccionaron 5. Entonces, en total hubo 5 participantes que seleccionaron un 20 % o más de las pseudopalabras. Hemos elegido no excluir a los participantes que seleccionaron más de un 20 % de las pseudopalabras, ya que las cuatro pseudopalabras más elegidas fueron seleccionadas tanto por los participantes, y si no fueron parte de la prueba, ningún participante tendría más de un 15 % de las pseudopalabras seleccionadas. Otra particularidad de las pseudopalabras, como se discute en Huibregtse et al. (2002), es que los participantes no pueden seleccionar una pseudopalabra como una palabra real con certeza. En otras palabras, es improbable que un participante creyera con seguridad que ha reconocido una pseudopalabra, y sólo puede tener confianza en no reconocerla (p. 241). En cuanto a estas cuatro palabras en particular, no sabemos con qué certeza los estudiantes las seleccionaron, pero una forma posible de hacerlo habría sido utilizar una herramienta que nos permitiera ver cuánto tiempo tardaron los participantes antes de seleccionar o rechazar cada palabra. Esto podría indicar con qué confianza seleccionaron cada palabra, y qué certeza tuvieron de las diferentes pseudopalabras.

4.3 Correlaciones entre el vocabulario y otros factores

Para determinar cuáles de los factores sobre los que preguntamos a los estudiantes durante la encuesta pueden influir en el vocabulario, podemos separar las puntuaciones de los estudiantes según lo que respondieron en la encuesta de información general. Así podemos comparar las puntuaciones promedias entre estos grupos y ver cómo se diferencian. Por ejemplo, los estudiantes que han respondido que tienen una competencia mejor del español también tienen un vocabulario más grande (cuadro 5), al igual que los que han respondido que utilizan el español fuera de la escuela con más frecuencia (cuadro 6).

Nivel medio (n = 5)	Independiente (n = 10)	Competente y fluido (n = 2)
2060	2645	3625

Cuadro 5, vocabulario promedio según cómo describen los estudiantes su propia competencia

Menos de 1 hora por semana (n = 7)	1-3 horas por semana (n = 6)	Más de 3 horas por semana (n = 4)
2086	2867	3050

Cuadro 6, vocabulario promedio según el uso del español en el tiempo libre

También podemos separar los participantes en dos grupos, de vocabulario más grande y pequeño del promedio. Esto produce un grupo de 9 estudiantes con un vocabulario más grande del promedio, y otro grupo de 8 estudiantes con un vocabulario más pequeño. Los que tienen un vocabulario más grande del promedio (>2588) tienen en común que la mayoría (5) utilizan otra lengua con familia/amigos. Entre ellos, 3 personas dicen que el español es una de las lenguas que hablan con ellos. Casi todos (8) empezaron a aprender español en el octavo grado de la escuela (edad 13-14), y 1 persona en el primer año del bachillerato (edad 16-17). 3 de ellos también empezaron a aprender español de una edad menor de la edad promedio (12,18). Todos los participantes que median su competencia como competente (1) o con fluidez (1), pertenecen a esta categoría. Entre los demás, la mayoría (6) han dicho que tienen un nivel independiente, y 1 persona nivel medio. 3 de los estudiantes en este grupo utilizan el español más de 3 horas fuera de la escuela semanalmente, 5 lo usan 1-3 horas, y sólo 1 persona ha dicho menos que 1 hora. En cuanto a estancias en el extranjero, 4 personas han vivido en un país hispanohablante más de 3 meses. Por otra parte, los participantes en este grupo seleccionaron más pseudopalabras, también en total y en promedio.

Entre los 8 estudiantes que tienen un vocabulario más pequeño del promedio (<2588), la mayoría (7) no utilizan otra lengua con familia/amigos. Además, la única persona que lo hace no ha mencionado utilizar el español. Como los participantes del otro grupo, casi todos (7) empezaron a aprender el español en el octavo grado, pero también hubo 1 persona que no había aprendido el español antes de la universidad. Casi todos (6) empezaron a aprender el español de una edad más alta del promedio, a edades entre 13 y 19 años, y todos evaluaron su propia habilidad como independiente (4) o media (4). La mayoría de los estudiantes que componen este grupo (6) respondieron que utilizan el español menos de 1 hora semanalmente, y solamente

1 persona lo utiliza más de 3 horas, y 1 persona entre 1 y 3 horas. La mayoría (6) no han estado más de tres semanas en el extranjero, y 2 personas han estado en un país hispanohablante. Además, comparado con los del otro grupo, estos participantes seleccionaron menos pseudopalabras.

Basado en estas características comunes, parece que los aspectos que pueden influir más en el vocabulario pasivo son cuánto tiempo se utiliza la lengua fuera de la escuela, incluido el uso de la lengua con familia o con amigos, y también la edad que se tiene cuando se empieza a aprender la lengua. Este último punto es lógico, ya que empezar más temprano significará que han llevado más tiempo utilizando o estudiando la lengua. En cuanto a los otros dos, pueden considerarse formas de exposición extramural que, como vimos en el apartado 2.5, tiene una correlación positiva con el aprendizaje exitoso de una lengua extranjera. Parece que cuando durante la escuela se empieza a aprender la lengua posiblemente influye menos, pero casi todos los participantes empezaron a aprender la lengua en octavo grado, el primer año de la escuela secundaria (*ungdomsskolen*). Entonces no hay suficientes datos para llegar a una conclusión, pero el hecho de que los que aprenden la lengua fuera de la escuela más temprano tiene un nivel más alto, sugiere que aprenderlo más temprano en la escuela tendrá el mismo efecto. Otro aspecto que se puede considerar es cómo la competencia de otras lenguas extranjeras puede influir la competencia de una L2. Investigaciones como Molnár (2010) han encontrado que el conocimiento de una L2 puede influir y ayudar con el vocabulario de una L3. Sin embargo, todos nuestros participantes respondieron que utilizaban frecuentemente (más de 3 horas en promedio semanalmente) el inglés en su tiempo libre, y todos igualmente respondieron que tienen un nivel competente o fluido de inglés. Además, el número de estudiantes que utilizan otras lenguas que el español o el inglés era muy pequeño. Por consiguiente, no podemos ver cómo el conocimiento de otra lengua ha influido en nuestros resultados.

4.4 Comparación con los resultados de Fairclough (2011)

Como mencionado cuando la introdujimos en el apartado 2.1, en Fairclough (2011) se ha realizado una investigación sobre el reconocimiento léxico de estudiantes del español en los Estados Unidos. El objetivo era encontrar si una prueba de reconocimiento léxico es un formato válido para medir la competencia general del español. Hubo dos categorías de participantes en la investigación, estudiantes de lengua española de herencia y estudiantes de español como L2. Los participantes que fueron expuestos al español antes de empezar la escuela se consideraban

estudiantes de lengua de herencia (p. 277). En este apartado comparamos los resultados de los participantes de Fairclough en la prueba de reconocimiento léxico con los de nuestros participantes. El cuadro 7 muestra el porcentaje promedio de palabras elegidas por los participantes de cada banda de frecuencia (*hit rate*), para los estudiantes de español como lengua de herencia (HLL), español como L2 (SLL), y el promedio de todos los participantes de Fairclough combinado, comparado con los estudiantes de nuestra investigación. No hemos incluido el grupo de control de Fairclough, que consistió en estudiantes graduados que obtuvieron resultados casi perfectos en todas las bandas.

Banda de frecuencia	Fairclough, HLL (n = 183)	Fairclough, SLL (n = 147)	Fairclough, todos (n = 330)	Nuestros participantes (n = 17)
1	.96	.73	.86	.90
2	.92	.56	.76	.71
3	.86	.39	.65	.62
4	.80	.34	.60	.51
5	.78	.31	.57	.30

Cuadro 7, el *hit rate* para cada banda de frecuencia. HLL = estudiantes del español como lengua de herencia (*heritage language learners*), SLL = estudiantes del español como segunda lengua (*second language learners*), copiado de Fairclough (2011)

Vemos que nuestros participantes, comparado con todos de Fairclough, en promedio tuvieron una puntuación más alta sólo cuando se trata de las palabras de la primera banda de frecuencia, y puntuaciones un poco más bajas en la mayoría de las demás bandas de frecuencia. Los números de las otras bandas no son muy distintos de los de todos los participantes de Fairclough, salvo el último, en que los estudiantes de español como lengua de herencia tiene una puntuación alta que mejora el promedio del grupo mezclado, y nuestros estudiantes tienen una puntuación más parecida a los estudiantes de español como L2. En otras palabras, los resultados de nuestros participantes generalmente se parecen más al grupo mezclado que a cualquiera de los dos grupos. Entre nuestros participantes solamente hay 2 personas que serían considerados estudiantes del español como lengua de herencia según la definición que ha utilizado Fairclough. Esto es un número muy pequeño para hacer conclusiones definitivas, pero podemos mencionar que ellos, comparado con el promedio, obtuvieron puntuaciones más altas, seleccionaron casi el doble de cantidad de palabras para las dos bandas de frecuencias más bajas, y también seleccionaron más pseudopalabras. Otra cosa que encontró Fairclough fue que los estudiantes de español como lengua de herencia seleccionaron más palabras reales, pero

también más pseudopalabras (p. 289), que es parecido a nuestros resultados donde los participantes con vocabularios más grandes también seleccionaron más pseudopalabras. Fairclough sugiere que esto puede explicarse porque los estudiantes de lengua de herencia tienen un sentido mejor para cuales palabras pueden ser reales, basado en las reglas morfológicas y fonológicas (p. 289). Esto puede significar que lo mismo sea verdad para los estudiantes con mayor competencia lingüística, que explicaría la selección más alta de pseudopalabras por ellos en nuestra investigación.

5. Conclusiones

Para concluir este trabajo, volvemos a las preguntas de investigación para resumir nuestros descubrimientos.

1. ¿Cuál es el tamaño del vocabulario pasivo de los estudiantes?

Basado en las puntuaciones de los 17 participantes en la prueba X-Lex, hemos encontrado que hubo una gran variación en su vocabulario pasivo. El vocabulario mínimo fue 1400, el máximo fue 3900, y el promedio fue 2588 entre los 17 estudiantes. Esto no fue un resultado inesperado, como los estudiantes de estos cursos de español pueden tener muy diferentes experiencias previas con la lengua. Esto también se verifica en investigaciones que intentan encontrar conexiones entre el tamaño del vocabulario y los niveles del Marco común europeo de referencia para las lenguas, que han encontrado que puede haber gran variación entre estudiantes en el mismo nivel. Nuestros métodos también tienen algunos límites que debemos mencionar. Hemos calculado un vocabulario pasivo aproximado de los participantes, que solamente es un aspecto del conocimiento léxico, la amplitud. No nos dice mucho sobre la habilidad de los estudiantes de utilizar sus vocabularios efectivamente en usos activos. Además, X-Lex es una prueba de reconocimiento léxico que mide la capacidad de reconocer las 5000 palabras más comunes. Basado en los resultados de esta prueba hemos calculado los vocabularios pasivos de los participantes. Pero, no nos dice nada sobre el conocimiento de las palabras más allá de las 5000 palabras más frecuentes. En realidad, sería más preciso decir que hemos calculado su conocimiento pasivo de las 5000 palabras más frecuentes del español. Ciertamente hay algunas palabras menos frecuentes que los estudiantes conocen, pero es poco probable que los estudiantes conozcan muchas palabras de las frecuencias más allá de la quinta banda de frecuencia basado en la disminución que vimos en el reconocimiento de palabras de las frecuencias más bajas. Sin embargo, puede haber vocabulario específico de baja frecuencia

que conozcan bien debido a su uso en contextos universitarios, que no se ha tenido en cuenta en esta investigación.

2. ¿Qué conexiones hay entre el tamaño de vocabulario y otros factores, basado en qué características tienen en común los estudiantes que tienen un vocabulario más grande o pequeño?

Para responder a nuestra segunda pregunta de investigación separamos a los participantes en dos grupos según si su vocabulario estaba por encima o por debajo del vocabulario promedio. Esto mostró que la mayoría de los que tuvieron un vocabulario más grande utilizan más el español con familia o amigos, y también pasan más tiempo utilizando la lengua receptivamente en su tiempo libre. Además, en algunos casos también llevan más años aprendiendo la lengua. Parece que uno de los factores que principalmente influye en el tamaño del vocabulario pasivo es la exposición extramural del español, que igualmente es el caso en investigaciones de los efectos de la exposición extramural en otras lenguas extranjeras (Sundqvist, 2009; Sundqvist & Sylvén, 2016).

3. ¿Cómo se comparan los resultados con los de otras investigaciones?

Descubrimos que el vocabulario promedio de los participantes de nuestra investigación equivale aproximadamente a un nivel B1 del MCER, comparado con previas investigaciones que comparan el tamaño del vocabulario receptivo con otras pruebas (Milton, 2010). Este nivel corresponde a lo que se espera que los estudiantes tengan al entrar en cursos universitarios de este nivel. Sin embargo, nuestros resultados también pueden indicar que muchos de los estudiantes no tengan un vocabulario suficiente para utilizar la lengua efectivamente, en comparación con las conclusiones de investigaciones anteriores como las de Robles-García (2022) y Nation (2012). En cuanto al número promedio de palabras seleccionadas por cada banda de frecuencia (*hit rate*), nuestros resultados son parecidos a los de Fairclough (2011). Los resultados de nuestros participantes fueron especialmente similares a los del grupo mezclado de estudiantes de español como lengua extranjera y lengua de herencia, salvo la quinta banda de frecuencia. En ella, nuestros participantes tuvieron un promedio bastante inferior, pero muy cercano a los estudiantes de segunda lengua de Fairclough. Además, hay una semejanza en el patrón de que los estudiantes de mayor competencia seleccionaron más pseudopalabras en la prueba X-Lex.

Referencias bibliográficas

- Consejo de Europa (2020). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación. Volumen complementario*. Servicio de publicaciones del Consejo de Europa.
- Davies, M. (2005). Vocabulary Range and Text Coverage: Insights from the Forthcoming Routledge Frequency Dictionary of Spanish. En D. Eddington (Ed.), *Selected Proceedings of the 7th Hispanic Linguistics Symposium* (pp. 106-115). Cascadilla Proceedings Project.
- Davies, M. (2006). *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*. Routledge.
- Fairclough, M. (2011). Testing the lexical recognition task with Spanish/English bilinguals in the United States. *Language Testing*, 28(2)(2), 273-297.
<https://doi.org/https://doi.org/10.1177/0265532210393151>
- Griffiths, C. (2008). Strategies and good language learners. En *Lessons from Good Language Learners*. Cambridge University Press.
- Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language* 23(1), 403-430.
- Huibregtse, I., Admiraal, W., & Meara, P. (2002). Scores on a yes-no vocabulary test: correction for guessing and response style. *Language Testing*, 19(3), 227-245.
<https://doi.org/10.1191/0265532202lt229oa>
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? En C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316-323). Multilingual Matters.
- Laufer, B. (1998). The Development of Passive and Active Vocabulary in a Second Language: Same or Different? *Applied Linguistics*, 19(2), 255-271.
- Laufer, B. (2000). Task effect on instructed vocabulary learning: The hypothesis of 'involvement'. *Selected Papers from AILA '99 Tokyo*, 47-62. (Waseda University Press)
- Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: effects of language learning context. *Language Learning*, 48(3), 365-391.
<https://doi.org/Doi 10.1111/0023-8333.00046>
- Lindqvist, C. (2016). Le développement de la taille du vocabulaire en français L2 chez les élèves suédophones. *Synergies Pays Scandinaves*, 11-12, 151-161.

- Lindqvist, C. (2020). First and second language cognate effects in third language vocabulary size estimates. In *First language influences on multilingual lexicons* (pp. 149-161). Routledge.
- Milton, J. (2006). Language lite? Learning French vocabulary in school. *Journal of French Language Studies*, 16, 187-205.
- Milton, J. (2008). French vocabulary breadth among learners in the British school and university system: comparing knowledge over time. *French Language Studies*, 18, 333-348. <https://doi.org/10.1017/S0959269508003487>
- Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. *Eurosla Monographs Series*, 1, 211-232.
- Milton, J., & Alexiou, T. (2009). Vocabulary Size and the Common European Framework of Reference for Languages. En B. Richards, Daller, M.H., Malvern, D.D., Meara, P., Milton, J., Treffers-Daller, J. (Ed.), *Vocabulary Studies in First and Second Language Acquisition* (pp. 194-211). https://doi.org/https://doi.org/10.1057/9780230242258_12
- Mochida, A., & Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing*, 23(1), 73-98. <https://doi.org/https://doi.org/10.1191/0265532206lt321oa>
- Molnár, T. (2010). Cognate Recognition and L3 Vocabulary Acquisition. *Acta Universitatis Sapientiae. Scientific Journal of Sapientia University*, 2(2), 337-349.
- Nation, I. S. P. (2012). *Learning Vocabulary in Another Language* (2 ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139524759>
- Robles-García, P. (2022). Receptive vocabulary knowledge in L2 learners of Spanish: The role of high-frequency words. *Foreign Language Annals*, 55(4), 963-984. <https://doi.org/10.1111/flan.12630>
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, 95(1), 26-43.
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36, 139-152.
- Sundqvist, P. (2009). *Extramural English matters: Out-of-school English and its impact on Swedish ninth graders' oral proficiency and vocabulary*. [Tesis]. Karlstad University, Karlstad.
- Sundqvist, P., & Sylvén, L. K. (2016). *Extramural English in Teaching and Learning*. <https://doi.org/https://doi.org/10.1057/978-1-137-46048-6>

- van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), 457-479.
- Webb, S. A., & Chang, A. C.-S. (2012). Second Language Vocabulary Growth. *RELC Journal*, 43(11), 113-126.

Apéndices

La lista de palabras

La lista de palabras que forman parte de la prueba, categorizadas por sus bandas de frecuencia. Las palabras son las mismas que ha utilizado Fairclough (2011), pero hemos utilizado sólo las veinte primeras de cada categoría para reducir el número total de palabras de la prueba. (20 de cada categoría para un total de 120).

0-1000	1000-2000	2000-3000	3000-4000	4000-5000	Pseudopalabras
vivir	matrimonio	colgar	encendido	caridad	disalejante
perder	acontecimiento	aceite	cazar	indudable	abucar
mil	ajeno	selva	cemento	cuidadoso	churruminos
calle	tren	valioso	enfermero	vejez	egüero
morir	silla	cuadrado	descargar	poblar	robonente
pagar	fuego	copa	reemplazar	ida	alcanaban
comer	pegar	ruso	prisa	mármol	aprisar
aprender	verdaderamente	judío	lástima	parroquia	reprendizaje
esfuerzo	seguido	desarrollado	pauta	cima	canupia
lleno	beber	amable	cocinar	conejo	paujón
escuela	rendir	pelota	despacio	apurar	confúnfuro
joven	amarillo	despacho	reparto	abanico	festivar
lanzar	mezcla	trozo	martes	encabezar	benefocio
mar	cocina	remontar	incierto	dichoso	camindante
precio	playa	mediodía	puño	desayunar	recelto
junto	tiro	húmedo	liso	burro	gustamente
carne	suave	madurez	estirar	caldo	releerizante
debajo	desprender	novio	prescindir	bicho	billetar
mío	encerrar	enterrar	cuna	adentrar	alabazante
derecho	dulce	convivencia	reposo	buey	calorar

El cuestionario de información general (*Bakgrunnsinformasjon*)

1) I forbindelse med hvilket studieprogram studerer du spansk?

- Spansk årsstudium/1. år bachelor
- 2./3. år bachelor
- Europastudier
- Spansk som enkeltemne/annet

2) Hva er morsmålet ditt?

Lærte du to eller flere språk fra fødselen av, oppgir du begge/alle språkene.

--

2) Har du familie/nære venner som du snakker et annet språk enn norsk med?

- Nei
- Ja – spesifiser hvem og hvilke(t) språk dere bruker:

Språk	Hvem snakker du det med?				
	Foreldre/fore-satte, søsken	Besteforeldre, tanter/onkler, søskenbarn	Fjernere slektninger	Venner jeg treffer ofte	Andre venner og bekjente

3) Hvor gammel var du da du begynte å lære *spansk*? _____

4) Når begynte du med *spansk* på skolen (klassetrinn)? _____

5) Hvor gammel var du da du begynte å lære *engelsk*? _____

7) Hvilke språk kan du og hvordan vil du vurdere ferdighetene dine i disse språkene?

Hvis du kan flere språk enn engelsk og spansk, fyll inn disse selv.

SPRÅK	Grunnleggende Jeg kan kjenne igjen og bruke enkle ord og uttrykk.	Middels Jeg kan bruke vanlige ord og enkle setninger, og kan kommunisere på grunnleggende nivå.	Selvstendig Jeg kan forstå hovedinnholdet i standard tale og snakke/skrive sammenhengende om kjente tema.	Kompetent Jeg kan snakke ganske flytende og forstå sammenhengende tale om vanskelige, kjente tema. Jeg kan skrive detaljert om mange tema.	Flytende Jeg kan uttrykke meg flytende, spontant og presist. Jeg forstår uten problemer alt jeg leser og hører, og kan bruke språket effektivt i sosiale og faglige sammenhenger.
Engelsk					
Spansk					

8) I løpet av det siste året, hvor ofte har du hørt *spansk* utenom skolen?

For eksempel gjennom film, YouTube, TV, dataspill, bøker, musikk eller lignende?

- Aldri
- Av og til, men mindre enn 1 time i uka i gjennomsnitt
- 1-3 timer i uka i gjennomsnitt
- Mer enn 3 timer i uka i gjennomsnitt

9) I løpet av det siste året, hvor ofte har du hørt *engelsk* utenom skolen?

For eksempel gjennom film, YouTube, TV, dataspill, bøker, musikk eller lignende?

- Aldri
- Av og til, men mindre enn 1 time i uka i gjennomsnitt
- 1-3 timer i uka i gjennomsnitt
- Mer enn 3 timer i uka i gjennomsnitt

10) Har du bodd i utlandet mer enn tre måneder?

- Ja Nei

Hvis ja, hvor bodde du, hvor lenge var du der, og hvor gammel var du?

11) Har du dysleksi eller lese- og skrivevansker?

- Ja Nei

12) Har du, eller har du hatt, diagnoser som kan påvirke språklæring?

(f.eks. spesifikke språkvansker, autisme, sterkt nedsatt hørsel e.l.)

- Ja Nei

