

UNIVERSIDAD DE SALAMANCA
FACULTAD DE FILOLOGÍA



GRADO EN
FILOLOGÍA HISPÁNICA

Trabajo de Fin de Grado

**DISPONIBILIDAD LÉXICA EN LOS
CENTROS DE VALLADOLID, A
CORUÑA Y CÁDIZ**

Autor: REBECA CARPIO LÓPEZ
Tutor/a: Dr. /Dra. JOSÉ ANTONIO BARTOL HERNÁNDEZ

Salamanca. Curso 2018- 2019

UNIVERSIDAD DE SALAMANCA
FACULTAD DE FILOLOGÍA

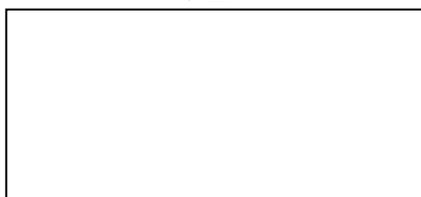
GRADO EN
FILOLOGÍA HISPÁNICA

Trabajo de Fin de Grado

**DISPONIBILIDAD LÉXICA EN LOS
CENTROS DE VALLADOLID, A
CORUÑA Y CÁDIZ**

Autor: REBECA CARPIO LÓPEZ
Tutor/a: Dr. /Dra. JOSÉ ANTONIO BARTOL HERNÁNDEZ

VºBº



Salamanca. Curso 2018-2019

DECLARACIÓN JURADA

Yo, REBECA CARPIO LÓPEZ con DNI

70896115-H DECLARO que he sido la única persona que ha realizado el presente trabajo íntegramente y que ninguno de los materiales que se adjuntan ha sido escrito o elaborado por otra persona, excepto las citas o el material identificado como perteneciente a otro.

Hago esta declaración jurada sabiendo y comprendiendo que, de comprobarse su falsedad, la calificación será negativa.

Fdo.

En Salamanca, ____ de _____ 20__

AUTORIZACIÓN PARA LA INCORPORACIÓN DEL TFG AL REPOSITORIO INSTITUCIONAL DE LA UNIVERSIDAD.

D/D^a.....con

D.N.I.....

AUTORIZO que el Trabajo de Fin de Grado titulado

".....",

sea incorporado al Repositorio Institucional de la Universidad de Salamanca en caso de que sea evaluado positivamente con una nota numérica de 9 o superior.

Fdo.

En Salamanca, ____ de _____ 20__

ÍNDICE

1. INTRODUCCIÓN.....	5
1.1. ANTECEDENTES.....	6-8
1.2. APLICACIONES DE LA DISPONIBILIDAD LÉXICA.....	8
1.2.1. Sociolingüística.....	8
1.2.2. Psicolingüística.....	8-9
1.2.3. Etnolingüística.....	9
1.2.4. Dialectología.....	9
1.3. LA DISPONIBILIDAD LÉXICA EN EL MUNDO HISPANO....	9-10
2. EDICIÓN DE RESPUESTAS.....	11-14
3. ANÁLISIS.....	14
3.1. DESCRIPCIÓN DE LA MUESTRA.....	14
3.2. DATOS CUANTITATIVOS TOTALES.....	14
3.2.1. Variable ‘tipo de centro’ público/privado.....	15-17
3.2.2. Variable sociocultural.....	17-19
3.2.3. Variable sexo.....	19-21
3.3. DATOS CUALITATIVOS.....	22-23
4. CONCLUSIÓN.....	24
5. BIBLIOGRAFÍA.....	25-26
6. APÉNDICE.....	27

1. INTRODUCCIÓN

El objetivo de este TFG es la comparación y estudio de los datos del *Léxico Disponible de España* (LDE) que nos proporciona el programa Displex, los cuales han ido obtenidos a través de una serie de encuestas realizadas a tres provincias diferentes: A Coruña, Cádiz y Valladolid, sobre el dominio léxico en el ámbito de la *comida y bebida*.

Cada una de las tres provincias que se han estudiado tiene un número de informantes diferente, ello depende del valor demográfico total y del número de matriculados en el curso de segundo de bachillerato. De esta forma, en total por cada provincia contamos con:

A Coruña: 157

Valladolid: 102

Cádiz: 157

Al realizar la investigación se han tenido en cuenta una serie de variables. Por un lado, las variables *sexo* y *nivel sociocultural familiar*, que afectan a cada informante. Y por otro, se ha tenido en cuenta la titularidad: *público/privado*; y tamaño de la población donde se encuentra: *menos de 10.000, de 10.001 a 20.000, de 20.001 a 100.000, más de 100.000*, que se refieren al centro de estudio. En cuanto a esto último, en lo que se refiere a nuestro estudio y como se verá posteriormente, solo hemos tenido en cuenta la variable público/privado, que en las provincias puede verse repartido de la siguiente manera:

	PÚBLICOS	PRIVADOS
A CORUÑA	127	30
VALLADOLID	74	28
CÁDIZ	126	31

1.1. Antecedentes

Los estudios de disponibilidad léxica surgieron en Francia a mediados del siglo XX. El primero de ellos fue realizado por Georges Gougenheim, René Michéa, Paul Rivenc y Aurélien Sauvageot, *L'élaboration du Français Élémentaire* 1956. Su principal objetivo consistía en dar a conocer el vocabulario que se debía incluir en los manuales de enseñanza del francés elemental especialmente ligado a extranjeros, es decir las palabras cotidianas o léxico temático. Para ello se recurrió primero a los recuentos léxicos basados en la frecuencia, ya que se pensaba que las palabras más frecuentes¹ (palabras atemáticas) eran las que debían enseñarse en primer lugar. Sin embargo, los investigadores se dieron cuenta de que en estos estudios faltaban palabras muy normales y de uso cotidiano puesto que estaban condicionadas por el contexto (palabras temáticas o disponibles²).

Me parece interesante destacar que el programa a partir del cual se ha trabajado, Dispolex, es una página web creada y mantenida por José Antonio Bartol H. y Natividad Hernández desde la Universidad de Salamanca.

En la propia página se define la *disponibilidad léxica* de la siguiente manera “La *Disponibilidad Léxica (DL)* es el campo de investigación que dentro de la lingüística tiene como objetivo la recogida y el posterior análisis del léxico disponible de una determinada comunidad de habla. El *léxico disponible* es el conjunto de palabras que los hablantes tienen en el lexicón mental y cuyo uso está condicionado por el tema concreto de la comunicación. Lo que se pretende es descubrir qué palabras sería capaz de usar un hablante en determinados temas de comunicación. Se diferencia del léxico básico en que este lo componen las palabras más frecuentes de una lengua con independencia del tema tratado.

¹ En contraposición con lo que sucedía con el léxico frecuente, el léxico disponible está formado principalmente por palabras con fuerte carga semántica: sustantivos, verbos y adjetivos. Bartol Hernández, J.A. (2010) “Disponibilidad léxica y selección del vocabulario”, en Rosa María Castañer y Vicente Lagüens (coords.), de moneda nunca usada. Estudios de dicados a José María Enguita Utrilla, Zaragoza, Instituto Fernando el Católico, pp. 81-107.)

² El vocablo disponible se caracteriza porque se presenta en la mente del hablante de forma inmediata y natural curando se trata de un determinado tema. Es una palabra que, sin ser necesariamente frecuente, vive potencialmente en el hablante y se actualiza en cuanto se producen ciertas asociaciones, Michéa (1953:340)

Mientras que el léxico básico está formado en su mayoría por palabras gramaticales —y las que no lo son pertenecen, en orden decreciente de frecuencia, a verbos, adjetivos y sustantivos de significado general—, en el léxico disponible abundan los sustantivos que aluden a realidades concretas.

Como señala el profesor Humberto López Morales, el léxico disponible es un léxico potencial, no actualizado; mientras que la frecuencia solo trabaja con léxico actualizado. La suma de ambos tipos de léxico constituye el léxico fundamental de una lengua.”

Los estudios de disponibilidad léxica se han centrado en dar cuenta del conocimiento que los hablantes nativos tienen sobre el léxico de su propia lengua. Aunque este sigue siendo el sujeto de encuestas más habitual, están siendo muchas hoy las investigaciones que incorporan informantes no nativos, ya sea estudiantes de español como L2, ya sea inmigrantes que están en situación de inmersión (Jing 2006, Gallego 2011, Sánchez-Saus 2011).

La disponibilidad léxica nació en el seno de la lingüística aplicada, donde los resultados más abundantes probablemente hayan sido en la enseñanza de la lengua materna y de segundas lenguas. En la recogida de datos se trabajan con listados léxicos basados en encuestas para posteriormente realizar los estudios. Puesto que las encuestas se desarrollan de forma escrita, constituyen estos materiales una vía para conocer la destreza en la escritura. Para ello, se les proporciona a los sujetos una serie de estímulos (centros de interés³) a los que se les tiene que añadir en un tiempo limitado de dos minutos el mayor número de palabras posible -sistema de listas abiertas-, esto se debe a que se suponía que cuanto mayor tiempo, más posibilidades había de que el informante anotara palabras que en realidad no usaba. El hecho de que se produzca esta prueba por escrito tiene otra razón de tipo práctico y es que, de esta manera en pocas sesiones se obtiene un material amplio y pueden realizarse de forma colectiva, a diferencia de lo que ocurre en las pruebas orales donde deberían realizarse de forma individual.

Dichas encuestas se han optado por no llevarlas a cabo a niños o adolescentes porque como indica López Morales (1999a: 28) el propósito para el que se realizan estas

³ Son dieciséis con los que trabaja el PPHDL, sin embargo ha habido a lo largo del tiempo muchas variaciones.

investigaciones, es la descripción de la norma léxica disponible de los adultos. Por lo que se han realizado sobre todo a estudiantes en etapa de educación secundaria o preuniversitaria. Las razones de esta elección no tienen que ser únicamente con la facilidad que supone este colectivo a la hora de recolectar amplias muestras de materiales de grupos homogéneos, sino sobre todo con las posibilidades que ofrece este tipo de informante como representante de la norma culta de la comunidad estudiada, ya que han alcanzado suficiente grado de maduración léxica sin estar aún expuesto a contaminación técnica que supone la inserción en el ámbito profesional (López Morales 1999:28). Además de que los sujetos a los que se les realiza las encuestas no tienen que manejar una variedad de lengua especializada como ocurriría en aquellos que tienen una profesión o estudios especializados.

1.2. Aplicaciones de la Disponibilidad Léxica

La disponibilidad léxica resulta una herramienta de gran ayuda para otras disciplinas como la dialectología, la sociolingüística, la etnolingüística, la psicolingüística o la lingüística cognitiva (López Morales 1999:19-26) ('Desarrollos teóricos y metodológicos recientes de los estudios de disponibilidad léxica', Paredes García. F. Universidad de Alcalá. P.81)

1.2.1. Sociolingüística

Esta disciplina no se planea desde una perspectiva únicamente sociolingüística, sin embargo a través de ello se obtiene información sociolingüística acerca de los encuestados como puede ser el sexo, el nivel sociocultural, el centro público o privado en el tipo de enseñanza, o la edad,⁴ información que resulta básica para el análisis de los resultados.

1.2.2. Psicolingüística

El léxico disponible es una herramienta que puede servir para conocer la organización del lexicón mental. Para poder conocer la relación que se da entre las palabras es necesario tener en cuenta el 'índice de cohesión' que se obtienen en el análisis de los centros de interés, ya

⁴ Todo ello será comentado posteriormente en el análisis de los datos obtenidos en el estudio de las encuestas realizadas en A Coruña, Cádiz y Valladolid.

que según Echeverría *et al.*, “mide el grado de coincidencia en las respuestas de los sujetos para un mismo centro” (1987.68).

La fórmula que permite determinar cuál es el índice de disponibilidad léxica de cada vocablo se expresa entre el 0 y 1, siendo el 1 la disponibilidad máxima (una palabra mencionada por todos los informantes) y el 0 la mínima. La fórmula de Chávez y Strassburger para hallar los índices de disponibilidad léxica es la que hoy se usan en programas como Dispolex.

1.2.3. Etnolingüística

Los estudios de disponibilidad léxica ofrecen información de tipo cultural sobre las zonas geográficas que se están estudiando posibilitando así una forma más sencilla de analizar los diferentes lugares.

Las primeras aportaciones que se realizaron fueron de Mackey (1971a, 1971b). Para Samper Padilla (2003, p.111) este tema no ha sido muy analizado en el mundo hispánico, sin embargo hay trabajos como el de Carcedo que resultan ser muy relevantes ya que en él se pone de manifiesto que el léxico está relacionado con los diversos factores como pueden ser culturales o económicos de las diferentes comunidades.

1.2.4. Dialectología

La relación de los estudios de la disponibilidad léxica con la dialectología reside en las pruebas asociativas. Además permite juntar un gran número de datos de una misma comunidad ya que en las encuestas siempre se obtiene léxico dialectal.

1.3. La disponibilidad léxica en el mundo hispano

En 2006 el profesor de la Universidad de Salamanca José Antonio Bartol Hernández, en su artículo “La disponibilidad léxica”, publicado en *La Revista Española de Lingüística* (RSEL), señaló que hasta el momento solo eran tres los bloques de investigación realizados. El

primero de ellos es el Proyecto Panhispánico de léxico disponible⁵, dirigido por el prof. H. López Morales. Como objetivo tenía la publicación del léxico disponible común del mundo hispánico, así como el de cada uno de los países que forman la comunidad hispánica. Información de dicho proyecto puede ser encontrada en la página web de Dispolex.

En segundo lugar son los estudios que no se ajustan a las características necesarias que presentan el Proyecto Panhispánico. Y, en tercer lugar, tenemos los estudios sobre la disponibilidad léxica de estudiantes de español como segunda lengua cuyo objetivo es el comprobar la aplicación de estos estudios a la enseñanza del léxico de una L/2.

Además de Dispolex, existen otros corpus como el de PRESEA, un gran corpus oral cuya intención es convertirse en uno de los de mayor tamaño en español. En el 2006 en el artículo

“Información básica sobre el Proyecto para el estudio sociolingüístico del español de España y de América” (Francisco Moreno Fernández, 2006) se menciona que para el futuro se pretendía que “a través de la página se pudiera acceder a una muestra muy significativa de los materiales y a las voces originales de los informantes, almacenadas en archivos de sonido con alineamiento automático de voz y de texto, así como la búsqueda de secuencias de caracteres en todo el corpus oral y a sus concordancias”.

Para entender la eclosión de la disponibilidad léxica en España y en América resulta imprescindible remitirse a la figura y la obra de Humberto López Morales, no solo [...] constituye el arranque de los estudios de disponibilidad, sino que dio lugar al Proyecto Panhispánico de Estudio sobre la Disponibilidad Léxica (PPHDL) [...] cuyo fin es la elaboración de un diccionario léxico disponible del español. [...] Puede verse información completa y detallada sobre los objetivos del PPHDL en la página anteriormente mencionada www.dispolex.com. (‘Desarrollos teóricos y metodológicos recientes de los estudios de disponibilidad léxica’, Paredes García. F. Universidad de Alcalá. P.79)

2. EDICIÓN DE RESPUESTAS

Para la elaboración de los léxicos disponibles se emplean los datos obtenidos en pruebas asociativas controladas, que constituyen un modo artificial de conseguir que afloren a la superficie, a través de losetas escritas o grabadas, las palabras de las que dispone, para uso inmediato, un determinado hablante o un grupo específico de ellos. (Bellón Fernández, Samper Padilla y Hernández, “El proyecto de estudio de la disponibilidad léxica en español”, Universidad de Las Palmas de Gran Canaria, 2003: 47)

Según López Morales, estas pruebas son “las únicas que hacen posible, en condiciones experimentales, que se realicen en la actuación lingüística las unidades léxicas con poca estabilidad estadística” (1999a: 32)

Los centros de interés⁶ se han usado en todas los estudios de disponibilidad léxica, como se ha mencionado anteriormente en la introducción, se trata de una serie de estímulos o temas en los que se intenta provocar una respuesta en los sujetos. El nuestro, Alimentos y bebidas, constituye el quinto campo de estudio en los estudios realizados de disponibilidad léxica, ello se debe a su grado de productividad, siendo además un campo de “validez universal, por su presencia contraste en prácticamente todas las investigaciones previas al proyecto” (Bellón Fernández, Samper Padilla y Hernández, ‘El proyecto de estudio de la disponibilidad léxica en español’, Universidad de Las Palmas de Gran Canaria 2003: 55).

El tipo de léxico que prima en los listados son los sustantivos y de forma denotativa los de carácter concreto.

El proyecto Panhispánico cuenta con unos criterios de edición de materiales propios. Para este estudio que se está realizando se han seguido criterios a la hora de editar las encuestas como puede ser:

- a) **Corrección ortográfica:** aunque los sujetos encuestados forman parte de la misma media de edad y etapa educativa, no quita que encontremos errores ortográficos en las encuestas, esto puede deberse a que los sujetos no usen de manera habitual estas palabras por vía escrita sino en comunicación oral. Ejemplo de esto mismo podrían ser tales como: trúa por trufa, Mohama por Mojama, biofutas por bifrutas, recezas por cerezas, cuscus

⁶ Deben ser siempre los mismos centros de interés (dieciséis correspondientes al Proyecto Panhispánico) en todos los estudios del proyecto hispánico para poder realizar las comparaciones entre los mismos estudios.

por cuscús, cucharón por cucharón, gintonic y gintónico por gin-tonic ya que así es como parece en el Diccionario de la Lengua Española. Mozzarella por mozzarella ya que también de esta forma es como la que aparece en el DLE.

- b) **La unificación ortográfica.** Me he visto con la obligación de ante algunas cuantas palabras, elegir el modo de escritura de las mismas. Es el caso de variantes como abas/haba/ **habas**; abellanas/ avellanas/ **avellana**; **absenta**/ asenta; **albóndigas**/albóndiga/ albondigas/ albondigas; **alcohol**/alchol/alcol/alcoholes; **aceituna**/acituna; **atún**/atun; **azúcar**/azucar; bacalao/bacalado; **bistec**/bisté/vistec/bisteck/bistes/bitec/bistes/bíter/pistec; bodka/bozca/bozka/**vodka**/vozka/vozca; **bogavante**/bogabante; **bollo**/bolos/boyo; **brandy**/brandi; brecol/brécoli/brocolí/brocolí/**brócoli**; buei/**buey**; burbon/**bourbon**; cerbeza/**cerveza**; cafe/**café**; canalones/**canelones**; **bizcocho**/ bizcochoz/vizcocho; **chocolate**/chococate; coña/coñá/**coñac**; **desayuno**/desalluno; **empanada**/empañada; epaguetis/espaguetis/espagettis/espagheti/espaghetis/espaghetis/espagueti/**espaguetis**/espagueti/espaguetis/espaguetis/espaguetis/spageti/spagettis/spagheti/spaghetis/spaguet his/spaguetis/spagueti/spaguetis/spagueti. **Escarola**/escarona; **espárragos**/esparragos; **güisqui**/ güisquicola/ whishey/ whiskie/ whisky/ wiskhy/ wiski/ wisky/ wisqui; **cruasán**/croissant/croisant/croissants/crusán/curasan.
- c) Dentro de este apartado me parece que es importante destacar que no solo se ha procedido a la unificación ortográfica, sino que también se han unificado plurales y singulares. Es decir, se han **neutralizado las variantes flexivas**. El criterio que he seguido ha sido tratando de agrupar bajo los plurales las formas contables: **alcachofas**/alcachofa; almeja/**almejas**; almendra/**almendras**; aceituna/**aceitunas**; acelga/**acelgas**; ajo/**ajos**; alubia/**alubias**; anchoa/**anchoas**; angula/**angulas**; berberecho/**berberechos**; berza/**berzas**; bombón/**bombones**; calamar/**calamares**; camarón/**camarones**; caramelo/**caramelos**; castaña/**castañas**; chicharro/**chicharros**; chuchería/**chucherías**; **ciruelas**/ciruela; croqueta/**croquetas**; dulce/**dulces**; espinaca/**espinacas**.

Elegidas por la frecuencia de aparición son algunas como **berenjena**/berenjenas; **bocadillo**/bocadillos; cacahuete/**cacahuetes**; **ajo**/ajos; **carne**/carnes; **cebolla**/cebollas; **cena**/cenar; **centollo**/centollos; cereal/**cereales**; champiñón/**champiñones**; **chícharo**/chícharos; **chorizo**/chorizos; **chuleta**/chuletas; **churro**/churros; **cóctel**/cócteles; conserva/**conservas**; costillas/**costilla**; **embutido**/embutidos.

- d) Al tratarse de listas lematizadas, encontramos **derivados, sinónimos, y compuestos de palabras** que significan lo mismo aunque se han dejado por separado como lemas diferentes, es decir, no se han unificado bajo un mismo lema. Tenemos casos como chorizo/chorizo a la rioja/chorizo criollo; batido/batido de coco/ batido de fresa, etc. arroz/arroz con leche/arroz a la cubana/ arroz con habichuelas, etc. Conservas/conservas en lata. Ensalada/ensaladilla/ensaladilla rusa. Ejemplos de sinónimos serían tales como patata/papa; plátano/banana o jugo/zumo.
- e) Se han usado los **paréntesis** para aquellos lexemas que unas veces aparecen en su forma completa y en otras aparece de forma reducida como por ejemplo vino; (vino) tinto, vino blanco; (vino) fino; batido/ batido (de) chocolate; isotónica/ (bebida) isotónica; comida/comida(s); alimentar/alimentar(se); Fanta/fanta (de) naranja;
- f) En cuanto a los **nombres de marcas comerciales**, lo que se ha preferido hacer es admitir la forma de aquellos lexemas que están lexicalizados, es decir aquellos en los que en una situación comunicativa sustituyen la marca pueda sustituir al producto designado. Por ejemplo es lo que ha sucedido en Coca cola/Coca-Cola/Cocacola/**Coca-Cola**/cocacola/Cocal-cola/caca-cola/cola con bebida de cola; champagne/**champán**/champan/ chamapng/chanpaña; Colacao/**Cola Cao**/colacao/Colacao/ colacao/cacao. Fanta de limón/naranja frente a Kas limón/naranja

Cabe destacar que en el estudio de estos listados se han encontrado palabras que llaman la atención por no estar relacionadas de forma directa con los alimentos o bebidas. Explicaciones para este tipo de palabras como adelgazar, atragantar, calentar, cocer, comedor, comer, comprar, cuchara, cuchillo, dieta, digestión, engordar, engullir, estiércol, exprimir, entre muchas otras. Ello puede deberse a que el sujeto que está respondiendo a las pruebas hace algún tipo de asociación entre comida/bebida con este tipo de palabras bien sea por problemas alimentarios como podría ser el caso de la aparición de palabras como engordar,

dieta, ansiedad o vomitar. O, por asociaciones con instrumentos de la cocina lo que explicarían los casos de cuchillo, cuchara, tenedor, comedor, merienda.

3. ANÁLISIS DE DATOS

Se va a proceder al análisis de los datos obtenidos en las encuestas realizadas a estas tres provincias y sus respectivos centros públicos y privados. De forma posterior al análisis de cada uno de los subapartados se llevará a cabo una pequeña conclusión de todo lo realizado en el trabajo.

3.1. Descripción de la muestra

Es un campo temático con una gran variedad de entradas referidas a comidas, bebidas, utensilios y platos típicos de algunas de las provincias que se están analizando como por ejemplo A Coruña (véase comentado esto en el partido 3.4).

3.2. Datos cuantitativos totales

En la investigación hemos encontrado entre todos los resultados de las encuestas realizadas en las provincias de Valladolid, Cádiz y A Coruña (como se puede ver en el apéndice) que en el campo que nos compete -Comidas y bebidas-, el número de palabras totales a las que los informantes contestaron de manera escrita es de 11.375. De la misma forma tenemos el promedio de palabras diferentes que se han encontrado en el estudio comparativo de dichas encuestas con un total de 811. En cuanto al número de palabras por informante tenemos 27.34, que posteriormente serán analizados por provincias. Y, en el índice de cohesión un 0.03.

Los condicionantes extralingüísticos que veremos a continuación dan datos muy importantes para el análisis de los resultados obtenidos en las encuestas.

Atenderemos a tres tipos de variables: centro de estudio (público/privado), nivel sociocultural y sexo.

3.2.1 . Variable ‘tipo de centro’ públicos/privados

Esta variable no fue introducida desde los principios de los estudios sobre disponibilidad léxica, sino que se introdujo al comienzo de los estudios en el español. El primero en incorporar esta variable a sus estudios fue H. López Morales en el *Léxico disponible de Puerto Rico* (1999a). En España los primeros en mostrar interés por esta variable con independencia de la de nivel sociocultural fueron los investigadores de disponibilidad léxica grancanarios.

Lo primero que se comentan serán los resultados de las encuestas que tenemos por cada una de las provincias y, a continuación los analizaremos haciendo comparación entre los privados correspondientes a las tres provincias y de la misma forma se procederá con los públicos.

El dato más importante que vamos a comentar será el promedio de palabras por informante de cada provincia. Esto es lo que nos permite ver si los alumnos de una determinada provincia dan de media más respuestas que otra, es decir, tienen más dominio léxico.

En primer lugar, en la provincia de **Cádiz** tenemos:

CÁDIZ	P. TOTALES	P. DIFERENTES	P. INFORMANTE	ÍNDICE DE COHESIÓN
PRIVADO	823	186	26.55	0.14
PÚBLICO	3321	358	26.36	0.07

Lo que se pretende mostrar con esta tabla es en primer lugar la recogida de palabras totales que tenemos de la provincia de Cádiz, a continuación nos muestra la cantidad de palabras diferentes que se añadieron, sin embargo como se ha dicho anteriormente el dato más relevante es el tercero, el promedio de palabras por informante. En la provincia de Cádiz contamos con que entre los centros públicos y privados la diferencia de media de palabras por informante apenas es significativa, por lo que se podría comentar que el dominio léxico entre ambos centros no hay diferencia.

Sin embargo, en la provincia de **A Coruña** los resultados son diferentes a la anterior:

A CORUÑA	P. TOTALES	P.DIFERENTES	P.INFORMANTE	ÍNDICE DE COHESIÓN
PRIVADO	1082	319	36.07	0.11
PÚBLICO	3504	515	27.59	0.05

Como en el anterior en primer lugar tenemos en porcentaje de las palabras totales recogidas en las encuestas realizadas, seguida de las palabras diferentes anotadas.

Haciendo hincapié en lo que más nos interesa, vemos que hay diferencia entre el promedio de palabras por informante entre los centros públicos y privados, es decir que el dominio léxico de públicos y privados en A Coruña. Estos datos responden a los resultados habituales, ya que normalmente los informantes de centros privados tienen más dominio léxico que los públicos.

En cuanto a lo que se refiere a la provincia de **Valladolid** tenemos los siguientes resultados:

VALLADOLID	P.TOTALES	P.DIFERENTES	P.INFORMANTE	ÍNDICE DE COHESIÓN
PRIVADO	725	224	25.89	0.12
PÚBLICO	1920	353	25.95	0.07

Yendo de nuevo al dato que nos interesa, vemos que al igual que lo que ocurría en Cádiz, los resultados recogidos sobre el promedio de palabras por informante en los centros privados y públicos de Valladolid apenas varían. Esto es que, a diferencia de lo que suele suceder que como se ha dicho anteriormente lo más habitual es que los centros privados tengan más dominio léxico, aquí no solo es que apenas haya diferencia entre ellos, sino que el centro público destaca, aunque sea mínimamente, sobre el privado.

A continuación como se ha dicho antes, se va a proceder a comparar los resultados entre las ciudades. Para empezar se analizarán los resultados obtenidos entre los centros privados. Aquí lo que podemos ver en los datos que se han recogido en estas tablas -datos pertenecientes a Dispollex- es que hay una gran diferencia entre el dominio léxico de los

privados de A Coruña siendo de 36.07 con respecto a Valladolid (25.89) y Cádiz (26.55), pues entre estos dos la diferencia no es muy significativa, siendo el promedio de palabras de Cádiz un poco más bajo.

De la misma manera, en el dominio léxico en los centros públicos sigue destacando A Coruña pero con una diferencia poco significativa respecto a las otras dos provincias:

A Coruña: 27.59

Cádiz: 26.36

Valladolid: 25.95

De manera general se podría concluir con esta variable comentando que está claro que en la media es mayor el promedio de palabras por informante en los centros privados, es decir tienen un mayor dominio léxico.

3.2.2. Niveles socioculturales

Es una de las variables más importantes en el estudio de las investigaciones sociolingüísticas. La importancia de esta variable fue manifestada desde los primeros estudios realizados por H. López Morales.

Como se señala en el artículo “Proyecto de estudio de la disponibilidad léxica en español” (Bellón, Samper Padilla y Hernández, 2003) la incorporación de los informantes a cada nivel sociocultural se produce de forma posterior al momento en el que ellos mismos han dado datos sobre las unidades familiares a las que pertenecen. Así mismo, se nos dice también en este artículo que los estudios que se han hecho en España han seguido siempre las propuestas realizadas por Samper y Hernández Cabrera (1997:230) para el estudio del léxico disponible de Gran Canaria. Dividían y establecían por niveles socioculturales teniendo cuenta dos parámetros: el nivel educativo y ocupación profesional de los padres.

Nuestros informantes pertenecen a tres niveles socioculturales (alto, medio y bajo). El número de informantes por cada nivel han quedado recogidos en la siguiente tabla:

	CÁDIZ	VALLADOLID	A CORUÑA	TOTAL INFORMANTES
NIVEL ALTO	32	20	41	93
NIVEL MEDIO	65	46	94	205
BAJO	60	36	22	118

(El total de informantes en las tres provincias sumando los privados y los públicos⁷)

Se ha recogido en esta tabla los índices cuantitativos totales sobre el promedio de palabras en general de los tres niveles socioculturales:

NIVEL	P. TOTALES	P. DIFERENTES	P. INFORMANTE	ÍNDICE DE COHESIÓN
BAJO	3292	443	27.90	0.06
MEDIO	5625	628	27.44	0.04
ALTO	2458	423	26.43	0.06

Por lo que podemos ver el promedio de palabras por informante, los alumnos cuyos padres tienen un nivel sociocultural bajo no difiere en exceso con los de nivel sociocultural medio y sin embargo se diferencian ambos del de nivel sociocultural alto porque vemos que el promedio de palabras por informante es ligeramente más bajo. Un dato sorprendente puesto que normalmente asociamos que un nivel sociocultural alto va ligado a un mayor nivel educativo con lo que el dominio léxico debería de ser mayor.

Es interesante señalar que los resultados son diferentes analizando las provincias por separado puesto que en el caso de Cádiz y Valladolid el promedio de palabras por informante

⁷ En el nivel medio alto en los informales de colegios privados de Valladolid no había ninguno.

Del nivel medio bajo en los resultados buscados en Displex tampoco aparece ninguno en Cádiz ni A Coruña privados. En Valladolid no aparecen resultados ni en públicos ni en privados.

Del nivel bajo no hay informantes recogidos en Displex en Cádiz, A Coruña y Valladolid privados.

de nivel sociocultural alto es más bajo en los centros privados que públicos, contrario es en lo que ocurre en A Coruña.

En el nivel sociocultural medio destaca que el promedio de palabras por informante es mayor en los privados de Cádiz y A Coruña y sin embargo es más bajo en el de Valladolid.

Por último, en el nivel sociocultural bajo el número de palabras por informante es mayor en los centros privados que públicos de las tres provincias.

Para terminar con esta variable me pareció que podría ser interesante analizar los niveles socioculturales en función del sexo. Al realizar el estudio y comparar los datos, los resultados que se obtuvieron fueron los siguientes:

En el nivel sociocultural bajo, el promedio de palabras por informante para los hombres es de 28.06 y en cambio para las mujeres es de 27.77 por lo que vemos que los hombres del nivel bajo tienen un mayor dominio léxico.

En cambio, en el nivel sociocultural medio se nota ya un cambio puesto que para los hombres el promedio es de 27.18 y para las mujeres 27.65, es por ello que las mujeres del nivel sociocultural medio tienen un mayor dominio léxico que los hombres aunque entre estos datos hay una gran diferencia.

Por último, en el nivel sociocultural alto son las mujeres las que tienen un mayor dominio léxico (27.18) frente a los hombres (23.49). Diferencia que por otro lado es bastante notable, ya que en los dos niveles socioculturales anteriores la diferencia entre los promedios no era muy grande, a diferencia de lo que se está mostrando aquí.

3.2.3. Variable sexo

Esta variable está presente en prácticamente todas las investigaciones que se han realizado en los diferentes trabajos desde los años 50 ya que constituye una gran fuente de información sociolingüística. La presencia constante de la variable 'sexo, salvo excepciones generalmente limitadas a ciertos centros de interés donde se observan divergencias significativas, pone de manifiesto, en la mayoría de los trabajos, una gran homogeneidad en las respuestas aportadas por hombres y mujeres (Bellón Fernández, Samper Padilla y Hernández, 'El proyecto de estudio de la disponibilidad léxica en español', Universidad de Las Palmas de Gran Canaria 2003. P 71)

Los datos acumulados hasta la fecha demuestran de manera evidente que las mujeres son más conscientes de la valoración que su comunidad hace de los fenómenos del lenguaje, y apoyan aquellos que obtienen más alto estatus en la evaluación social. (...) El habla de las mujeres no sólo es diferente al habla de los hombres sino que es mejor socialmente hablando. Así como se espera que la conducta social de las mujeres sea más correcta, también su habla debe serlo (López Morales, 1989:125-126).

Analizando los datos que nos ofrece Displex sobre las encuestas en cuanto al número de informantes por cada provincia en función del sexo, tenemos los siguientes resultados:

HOMBRES	CÁDIZ	VALLADOLID	A CORUÑA
PRIVADOS	14	12	13
PÚBLICOS	58	33	54

MUJERES	CÁDIZ	VALLADOLID	A CORUÑA
PRIVADOS	17	16	17
PÚBLICOS	68	41	73

Lo que se puede en este centro de interés es la notable y sobre todo destacable presencia general del sexo femenino en las encuestas. Si bien, en cuanto a la riqueza léxica también destaca en sexo femenino en las tres provincias, pauta general en todos los estudios de disponibilidad. Estos resultados prueban lo que encontramos en el artículo *Proyecto de estudio de la disponibilidad léxica en español*, donde encontramos que estos mismos resultados aparecen ya en la *Disponibilidad léxica en estudiantes chilenos*, Valencia y Echeverría (1999a) y en los estudios de Samper Hernández (2000, 2001) sobre el léxico en español de estudiantes extranjeros. De la misma manera, en los estudios de A. González (1997) en Cádiz ya teníamos estos resultados donde destacaba el sexo femenino en las aportaciones léxicas en campos como el que aquí se está tratando ‘Alimentos y bebidas’. Esto que s esa dicho, analizando en la ‘variable sexo’ por provincias el promedio de palabras que contestó cada informante lo vemos en las siguientes tablas:

MUJERES	CÁDIZ	A CORUÑA	VALLADOLID
PRIVADO	25.88	37.65	26.00
PÚBLICO	26.28	28.68	26.71

HOMBRES	CÁDIZ	A CORUÑA	VALLADOLID
PRIVADO	27.36	34.00	25.75
PÚBLICO	26.45	26.11	25.00

En el análisis de los datos obtenidos que se están mostrando da pie a comentar un dato interesante, pues en el promedio de palabras por informante en la provincia de Cádiz podemos ver que la media es más baja en las mujeres que en los hombres, resultado que se diferencia totalmente de la generalidad comentada anteriormente y de las otras dos provincias donde los resultados son los esperables, las mujeres tiene mayor disponibilidad léxica que los hombres. Un dato por el que quizá se explique que las mujeres tengan mayor disponibilidad léxica que los hombres es que ellas tienen más conciencia de lo prestigioso.

Me gustaría comentar también de manera un poco más general los resultados que obtenemos analizando a la vez las variables sexo y nivel sociocultural tanto en los hombre como en las mujeres.

3.3. Datos cualitativos

Se procederá en primer lugar en este apartado a analizar el índice de disponibilidad léxica de forma general entre todos los resultados de las encuestas. Posteriormente se comparará la disponibilidad por provincias.

Cuando una palabra es más disponible que otra significa que dicha palabra será incluida por el informante antes en su lista dentro del centro de interés pertinente, en este caso *Comidas y bebidas*. No es lo mismo disponibilidad que frecuencia, es decir, el hecho de que una palabra se mencione en un lugar de la lista no va relacionado con el índice de aparición de una voz.

La lista de palabras que nos proporcionan el programa DispoLex en este centro de interés viene ordenada según el índice de disponibilidad de las palabras, es decir, según la posibilidad de aparición de dichas palabras. En primer lugar tenemos que en los resultados generales la palabra con mayor índice de disponibilidad es *agua* teniendo un 0.70182 es decir que el 89.904% de los informantes mencionaron en primer lugar esta palabra. Va seguida de Coca-Cola teniendo un 0.56128 y mencionada por un 67.548% de los informantes. Como ya se ha dicho el índice de disponibilidad no va ligado al índice de aparición, un ejemplo de ello podría ser que la palabra *judías* tiene un índice de aparición de 11.538% mayor que la palabra *batido* 9.856% y sin embargo tiene una menor disponibilidad (0.05356) que la palabra *batidos* (9.856).

En la lista general la última palabra de la misma es *higos chumbos* con un índice de disponibilidad de 0.00029 y un índice de aparición de 0.240% lo que supone que esta palabra fue mencionada por apenas un informante. Solamente las 62 primeras palabras fueron mencionadas por más del 10% de los informantes de todas las provincias.

En cuanto a los listados por provincias, se van a analizar los centros públicos y privados de las tres provincias comprando los resultados de las mismas.

En primer lugar los resultados que obtenemos de los listados de los centros privados de Cádiz es que las cinco primeras palabras más mencionadas son: Coca-Cola (con una disponibilidad de 0.84406), agua, güisqui, patatas y lentejas. En los centros privados de A Coruña tenemos ya un cambio en la palabra con mayor disponibilidad, pues son: agua (con una disponibilidad de 0.73188), Coca-Cola, pescado, carne y galletas, coincidente con la palabra más disponible de los privados de Valladolid (con una disponibilidad de 0.69414): agua, carne, pescado, zumo y leche.

Para los centros públicos con respecto a los privados no cambia la primera palabra de la lista en ninguna de las tres provincias. Se podría comentar como dato diferente entre ellas que las palabras auga [g], raxo[g], pinto [g], viño [g], zorza [g], quente [g], cebola [g], castañeta [g], leite [g], augardente [g], caña de herbas [g], ollomol [g], guiso de choupa [g], millo [g], longueirón [g], meiga [g] solo han sido mencionadas en A Coruña. En la lista de Cádiz teneos palabras como pinchito, papa, percebes, pijota que no parecen en las demás, estas diferencias probablemente se deban a que son alimentos o palabras propias de la provincia donde se han señalado.

4. CONCLUSIÓN

En este trabajo de fin de grado se ha tratado de analizar de la mejor manera posible la disponibilidad léxica en los alumnos de segundo de bachillerato de las provincias de Cádiz, Valladolid y A Coruña. Los datos que se han obtenido han resultado ser los introducidos en la página DispoLex de forma posterior a la realización de la práctica de las encuestas de los centros privados y públicos de las tres provincias. A través de esta página se nos ha facilitado los resultados de las encuestas de forma que se ha podido analizar cualitativamente y cuantitativamente.

A grandes rasgos podemos señalar que los alumnos de los centros privados contestan a un mayor promedio de palabras en el tiempo establecido, es decir, tienen un mayor dominio léxico, aunque dependiendo de la provincia este dato es más elevado o menos con respecto a los públicos. En cuanto al nivel sociocultural fue interesante destacar que el mayor promedio de palabras por informante resultó ser para el nivel sociocultural bajo. Un dato que se encuentra expuesto en este trabajo y que coincide con los resultados en el resto de trabajos sobre disponibilidad es que, respecto a la ‘variable sexo’ son las mujeres las que presenta un mayor dominio léxico. Cabe destacar que en este trabajo no se ha contemplado la ‘variable edad’ ya que en los estudios de disponibilidad léxica el objetivo consiste en analizar la disponibilidad de los alumnos de segundo de bachillerato con lo que todos tienen la misma edad.

Por último, para los datos cualitativos es obvio que dependiendo de la provincia que se analice van a aparecer una serie de palabras bien sean comidas o platos típicos de dicha comunidad.

5. BIBLIOGRAFÍA

- Bartol Hernández, José Antonio, 2006 “La Disponibilidad léxica”, *Revista Española de Lingüística (RSEL)*. España: Universidad de Salamanca.
- Bartol Hernández, José Antonio “Disponibilidad léxica y selección de vocabulario”. España: Universidad de Salamanca.
- Bartol Hernández, José Antonio y Natividad Hernández. DispoLex/ investigación léxica: <http://212.128.132.42/dispolexspain/bancodatos/loguser.jsp>
- Echeverría Max S., María Olivia Herrera, Patricio Moreno y Francisco Pradenas. 1987. “Disponibilidad léxica en Educación Media”. *Revista de Lingüística Teórica y Aplicada*.
- Ferreira, Roberto y Max. S. Echeverría. “Redes semánticas en el léxico disponible de inglés L1 e inglés LE”
- Gallego Gallego, Diego J. 2011. Disponibilidad léxica en estudiantes de español como lengua extranjera en la Comunidad de Madrid. Aspectos y primeros resultados, en a. M. Cestero, I. Molina y F. Paredes (comps). Universidad de Alcalá.
- Herrero de la Fuente, Mercedes. *Estudio de disponibilidad de los hablantes de herencia alumnos del CEHI*. Madrid: Universidad de Nebrija.
- Jing, Lin, 2006. *El estudio de disponibilidad léxica de los estudiantes chinos de español como lengua extranjera*. Memoria de Máster inédita. Universidad de Alcalá.
- Lopez Morales. H. 1989. *Sociolingüística*, Madrid, Gredos.
- Lopez Morales. H. 1999 . *Léxico disponible de Puerto Rico*. Madrid: Arco Libros.

- Mackey, William Francis. 1971a. *Le vocabulaire disponible du français*, 2 tomos. París-Bruxelles-Montréal:Didier.
- Martínez Segura, Gricelda Elizabeth. *Disponibilidad léxica de las/los estudiantes de II de bachillerato académico, sección "A" del instituto 1º de Mayo de 1954, una propuesta para la enseñanza del vocabulario*. Tesis doctoral. Universidad Pedagógica Nacional Francisco Morán
- Michéa, René, 1953: "Mots fréquents e mots disponibles. Un aspect nouveau de la statistique du langage", *Les Langues modernes*.
- Moreno Fernández, Francisco, 2006 "Información básica sobre «el proyecto para el estudio sociolingüístico del español de España y de América» - PRESEA (1996-2010)", *Revista Española de Lingüística (RSEL)*. España: Universidad de Alcalá de Henares.
- Palapanidi, Kiriakí, 2012 "La aplicación de la disponibilidad léxica a la didáctica del léxico de LE". *Revista Nebrija de Lingüística Aplicada a la Enseñanza de las Lenguas*.
- Paredes García "Desarrollos teóricos y metodológicos recientes de los estudios de disponibilidad léxica" *Revista Nebrija*. España: Universidad de Alcalá de Henares. p 78-91.
- Samper Padilla, José Antonio y Clara Eugenia Hernández Cabrera. 1997. "El estudio de la disponibilidad léxica en Gran Canaria: datos iniciales y variación sociolingüística". Universidad de la Laguna.
- Samper Padilla, José Antonio, Juan José Bellón Hernández y Marta Samper Hernández, 2003 "El Proyecto de estudio de la disponibilidad Léxica en español". España: Universidad de Las Palmas de Gran Canaria.
- Sánchez-Saus Laserna, M. 2011. *Bases semánticas para el estudio de los centos de interés del léxico disponible de informantes extranjeros en las universidades andaluzas*. Tesis doctoral inédita. Univeridad de Cádiz.

6. APÉNDICE

Índices Cuantitativos	
-----------------------	--

Nº	Centro Interés	PT	PD	PI	IC
	5Comidas y bebidas	11375	811	27.34	0.03

·Disponibilidad General

Nº	Palabra	Disponibilidad	Frecuencia relativa	% Aparición	Frecuencia acumulada
1	agua	0.70182	3.293 %	89.904 %	3.293 %
2	Coca-Cola	0.56128	2.474 %	67.548 %	5.767 %
3	patatas	0.32203	1.875 %	51.202 %	7.642 %
4	güisqui	0.31733	1.849 %	50.481 %	9.491 %
5	pan	0.31123	1.805 %	49.279 %	11.296 %
6	carne	0.30986	1.743 %	47.596 %	13.039 %
7	pescado	0.27785	1.682 %	45.913 %	14.721 %
8	zumو	0.27015	1.541 %	42.067 %	16.262 %
9	lentejas	0.26167	1.638 %	44.712 %	17.900 %
10	Fanta	0.24561	1.400 %	38.221 %	19.300 %