

# EVOLUCIÓN DIACRÓNICA DEL LÉXICO DISPONIBLE DE ESTUDIANTES SINOHABLANTES DE ELE

HIDALGO GALLARDO, Matías | Lector MAEC-AECID de la Universidad de Estudios Internacionales de Sichuan (CHINA)

Fecha de recepción: 15 de febrero de 2017 | Fecha de aceptación: 24 de marzo de 2017 | Fecha de publicación: 7 de abril de 2017

## RESUMEN

En este artículo presentamos un análisis diacrónico del léxico disponible de estudiantes sinohablantes de español como lengua extranjera (ELE). Nuestro objetivo es evaluar la evolución de su competencia léxica y localizar posibles diferencias intermuestrales. Para ello, hemos cotejado cuantitativa y cualitativamente los resultados de una investigación de 2006 conformada por 263 universitarios chinos con los de un estudio propio reciente compuesto por una muestra de 400 participantes también chinos. Dicho cotejo incluye un contraste de once campos nocionales o *centros de interés* que nos ha permitido determinar, por un lado, el rendimiento léxico de los informantes y, por otro, la estabilidad de su léxico más disponible. Los resultados revelan un alto grado de estabilidad léxica y una mejora de la productividad de este tipo de informantes. Toda esta información, junto con otras fuentes, puede servir de base para futuras investigaciones sobre la mejora de la selección léxica de materiales didácticos.

**PALABRAS CLAVE:** léxico disponible, español como lengua extranjera (ELE), estudiantes sinohablantes, análisis diacrónico.

## ABSTRACT

In this paper we present a diachronic comparison of the available lexicon of two samples of Chinese-speaking learners of Spanish as a Foreign Language (SFL). Our goal is to assess the evolution of their lexical competence and locate possible inter-sample differences. The quantitative and qualitative collation of eleven notional fields or *centers of interest* has allowed us to determine, on the one hand, participants' lexical performance and, on the other hand, the stability of their available lexicon. The results show a high degree of lexical stability and an improvement of the productivity of this kind of participants. All this information, together with other sources, can constitute a basis for future researches regarding the improvement of the vocabulary selection of didactic materials.

**KEYWORDS:** available lexicon, Spanish as a Foreign Language (SFL), Chinese-speaking learners, diachronic analysis.

## BIODATA

Matías Hidalgo Gallardo es doctorando en Arte y Humanidades por la Universidad de Jaén, con especialidad en disponibilidad léxica. Es Lector MAEC-AECID en la Universidad de Estudios Internacionales de Sichuan (Chongqing, China) y ha colaborado como profesor y examinador DELE en los Institutos Cervantes de Pekín y Nápoles.

## INTRODUCCIÓN

Es cierto que durante muchos años el léxico ha estado supeditado a la gramática. Sin embargo, hoy día ya no se cuestiona su posición nuclear en los procesos de enseñanza/aprendizaje de lenguas. Muestras de ello son las numerosas investigaciones que se preocupan por la competencia léxica, los enfoques o metodologías que se centran en él, como el *enfoque léxico* (Lewis, 1993) o el *lexical priming* (Hoey, 2005), o los recientes manuales que se basan en ellos. Dentro de estas preocupaciones por la enseñanza del léxico sobresale el interés por la selección del mismo, es decir, por elegir qué vocabulario debe enseñarse en función del nivel, con una especial atención a los niveles principiantes. Esta realidad es —con características particulares— aplicable también al contexto chino, donde el léxico y la gramática son los elementos fundamentales, aunque no debemos olvidar que en muchas ocasiones su enseñanza se enmarca en enfoques poco comunicativos o demasiado estructuralistas.

Uno de los campos que pretenden dar cuenta de esta selección del vocabulario es la *disponibilidad léxica* (Gougenheim *et al.*, 1954), que, a través de una metodología propia, recopila información y elabora diccionarios o listados del léxico disponible de hablantes nativos, el cual, según López Morales (1984: 62), puede definirse como «el caudal léxico utilizable en una situación comunicativa dada», es decir, aquellas unidades léxicas de carga semántica concreta —y de carácter nocional— que solo se actualizan si lo permite el tema del discurso. No obstante, esta metodología también se ha empleado en informantes extranjeros, un ámbito que nosotros consideramos de gran interés a la hora de conocer más detalles sobre la competencia léxica de estudiantes de ELE, pero también para detectar posibles deficiencias en relación con sintopías o comunidades de habla nativas, controlar el aducto léxico al que están expuestos y establecer, en consecuencia, mejoras en la selección del vocabulario de los distintos materiales didácticos empleados.

Dentro de este ámbito, el de conocer el vocabulario de los estudiantes de lenguas extranjeras para así seleccionar el léxico que se necesita enseñar, nosotros nos hemos valido de los resultados de dos investigaciones en disponibilidad léxica con informantes chinos para llevar a cabo un cotejo entre ambas muestras que nos sirva para: 1) determinar la evolución diacrónica del léxico disponible de estudiantes sinohablantes de ELE; 2) localizar las posibles diferencias y/o carencias; y 3) tenerlo en cuenta para la selección y mejora del léxico contenido en manuales de ELE —entre otros materiales— muy consolidados, sobre todo, en la China, o para la creación de nuevos materiales didácticos, pues es evidente que el mercado chino supone un gran atractivo para editoriales y creadores.

## 1. LA DISPONIBILIDAD LÉXICA

La disponibilidad léxica es un campo de investigación que se desarrolló en Francia en la década de 1950 de la mano de investigadores como René Michéa, George Gougenheim, Aurélien Sauvageot y Paul Rivenc con la elaboración del *Français Élémentaire* (Gougenheim *et al.*, 1954) —que años más tarde pasó a ser *Français Fondamental* (Gougenheim *et al.*, 1964) —, una obra que supuso la base y guía para la aplicación de esta línea de investigación. Esta surgió como respuesta a las limitaciones derivadas de la selección léxica basada solamente en la *frecuencia* (p. ej. la excesiva cantidad de palabras gramaticales —mucho más fácilmente catalogables y recopilables— o la exclusión de palabras comúnmente usadas y útiles en determinadas situaciones o temas, entre otras). Bartol Hernández (2010: 91) se refiere a la disponibilidad léxica en los siguientes términos:

[...] supone un nuevo enfoque sobre el léxico. No se trata de las palabras utilizadas al producir determinados textos; sino de las palabras que, organizadas, están en nuestro lexicón mental y están «disponibles» para ser utilizadas cuando las circunstancias comunicativas lo requieran; cuando un estímulo active el nodo cerebral adecuado.

El principal propósito de estos lingüistas era el de mejorar la enseñanza del francés como lengua extranjera a través de la localización en hablantes nativos de francés de las palabras más disponibles en relación con temas que ellos consideraron universales —es decir, las que primero acuden a la mente ante un estímulo temático dado—. Estos temas, a los que denominaron *centros de interés* (CI), quedaron fijados en dieciséis<sup>1</sup> y se les atribuía ser universales y representativos del conocimiento humano. Sin embargo, más tarde, la universalidad y representatividad de estos centros se han visto cuestionadas, y muchos estudios los han adaptado o incluso sustituido en función de sus propósitos. De esta manera, tenemos que la disponibilidad léxica, además de en la didáctica de lenguas, se ha aplicado en campos como la dialectología, la sociolingüística, la etnolingüística, la psicolingüística o la neurolingüística.

Años más tarde, López Morales (1973), consciente del potencial de este campo de la lingüística, introdujo la disponibilidad léxica en el ámbito hispánico y creó el «Proyecto Panhispánico de Disponibilidad Léxica» (PPHDL), que pretende elaborar un diccionario de léxico disponible de las distintas regiones de habla hispana que permita establecer comparaciones. Desde entonces se han sucedido estudios de diferente índole y con diferentes objetivos, tanto en muestras de informantes nativos, principalmente con finalidades dialectales y sociolingüísticas, como en estudiantes de ELE, con objetivos fundamentalmente didácticos.

Es, de hecho, en lengua española donde mayor recorrido cuenta esta disciplina, además de haber contribuido al perfeccionamiento de la fórmula matemática que calcula este tipo de léxico (López Chávez y Strassburger,

---

<sup>1</sup> 01. Partes del cuerpo, 02. La ropa, 03. Partes de la casa (sin los muebles), 04. Los muebles de la casa, 05. Alimentos y bebidas, 06. Objetos colocados en la mesa para las comidas, 07. La cocina y sus utensilios, 08. La escuela: muebles y materiales, 09. Calefacción e iluminación, 10. La ciudad, 11. El campo, 12. Medios de transporte, 13. Trabajos del campo y del jardín, 14. Los animales, 15. Juegos y distracciones y 16. Profesiones y oficios.

1987)<sup>2</sup>. Tras realizarse estudios en casi todas las provincias españolas y en buena parte de Hispanoamérica, los estudios en disponibilidad léxica se han acrecentado sobre todo en el campo del español como lengua extranjera, cuyo primer exponente es Carcedo González (2000), y donde encontramos una amplia variedad de aplicaciones y usos. De este modo, y de forma general, contamos con trabajos que describen la competencia léxica de los estudiantes, ya sea en su país de origen o en inmersión; y con trabajos que tratan de mejorar la selección del vocabulario que se va a enseñar y que servirá para la elaboración y ajuste de programas curriculares, planificaciones lingüísticas y materiales didácticos.

También apreciamos una especificación de las aplicaciones, tales como cotejar con los ideales expresados por comunidades nativas de español o con los listados de léxico disponible de estudiantes de ELE de otras nacionalidades y lenguas; analizar la influencia de diferentes factores sociolingüísticos; contrastar con materiales y métodos para la enseñanza de español como lengua extranjera con objeto de evaluarlos y mejorar su contenido léxico; determinar necesidades específicas de estudiantes; evaluar el proceso de aprendizaje en sí mismo, ya sea a través de la variable «nivel de español» o bien mediante una evaluación diacrónica del léxico disponible de la misma muestra o muestras parecidas; adquirir información sobre la interlengua de los aprendientes; analizar las relaciones semánticas que se producen entre las palabras más disponibles; investigar el componente cultural presente en la competencia léxica; adaptar metodologías, enfoques o técnicas de enseñanza y realizar propuestas didácticas basadas en la disponibilidad léxica; o usar la disponibilidad léxica no solo para la selección del léxico, sino también como metodología y material real que puede llevarse a nuestras aulas de ELE.

---

<sup>2</sup> Para recopilar el léxico disponible se emplean unas pruebas específicas en las que los informantes deben escribir en forma de columna todas las palabras que asocien con el centro de interés dado en un límite de dos minutos. Tras esto, los datos son editados y tratados mediante una aplicación informática que, a través de una fórmula matemática, considera la frecuencia y el orden de aparición de las respuestas para elaborar un listado o diccionario de léxico disponible.

En este marco de aplicaciones, es la evaluación diacrónica del proceso de aprendizaje la que centra nuestro interés y la que expondremos en los apartados siguientes.

## 2. METODOLOGÍA

Para llevar a cabo nuestro cotejo debemos tener en cuenta las características de las muestras y de las dos investigaciones que tomamos como referencia, además de los centros de interés que se incluyen, con las correspondientes designaciones.

### 2.1. Las muestras

Las muestras que comparamos aquí pertenecen a dos investigaciones diferentes. La primera muestra procede del estudio de Lin (2006 y 2012) y está compuesta por 263 estudiantes universitarios de ELE que cursan estudios en nueve universidades chinas:

- La Universidad de Estudios Extranjeros de Pekín.
- La Universidad de Pekín.
- La Universidad de Estudios Internacionales de Pekín.
- La Universidad de Comunicación de China.
- La Universidad de Estudios Internacionales de Shanghái.
- La Universidad de Estudios Extranjeros de Cantón.
- La Universidad de Estudios Internacionales de Sian.
- La Universidad de Estudios Extranjeros de Sichuan.
- La Universidad de Estudios Extranjeros de Luoyang.

La segunda muestra proviene de una investigación propia<sup>3</sup> realizada en el año 2016 sobre 400 estudiantes de ELE de seis universidades chinas:

- La Universidad de Estudios Extranjeros de Pekín.
- La Universidad de Pekín.
- La Universidad de Jilin.
- La Universidad de Estudios Extranjeros de Cantón.
- La Universidad de Estudios Internacionales de Sian.
- La Universidad de Estudios Extranjeros de Sichuan.

Los dos conjuntos muestrales han sido seleccionados siguiendo un proceso de muestreo aleatorio con la condición de que fueran estudiantes a partir del segundo curso del Grado en Lengua Española.

Los factores sociolingüísticos que estratifican la muestra en el trabajo de Lin (2006) son «sexo», «edad», «nivel de español» y «zona geográfica» (en función de la universidad de estudio), mientras que nuestras variables han sido «sexo», «curso académico», «nivel de español», «prestigio de la universidad», «prestigio del Departamento de Español», «lengua materna», «conocimiento de otras lenguas extranjeras (aparte del español)» y «estancia en un país hispanohablante».

Al no tratarse este de un trabajo de corte sociolingüístico, nuestro principal propósito ha sido que los informantes presentaran un alto grado de equivalencia en relación con las universidades de procedencia, el tipo de formación y el nivel de español. De este modo, apreciamos que existen grandes coincidencias en cuanto a los centros de estudio y al tipo de formación, pues coinciden cinco de las universidades y todos los encuestados se encontraban en la misma franja académica (a partir del segundo curso). Con respecto al nivel de español —que supone la variable más influyente a la

---

<sup>3</sup> Los datos para la elaboración de este artículo forman parte de una investigación doctoral pendiente de defensa en el presente curso académico 2016/2017.

hora de perfilar el tipo de informante—, tanto nuestro trabajo como el de Lin (2006) establece tres niveles: nivel A (para A1 y A2), nivel B (para B1 y B2) y nivel C (para C1 y C2)<sup>4</sup>, en nuestro caso, y nivel elemental, intermedio y avanzado, en el caso de la investigadora china. Esta taxonomía implica que las muestras representan distintos niveles de aprendizaje, pero equivalentes entre sí.

Debemos mencionar aquí que el período de recopilación de los datos es similar en ambas investigaciones, pues en las dos la administración de las encuestas tuvo lugar durante el primer semestre del curso académico.

Con todo esto, las características mencionadas revelan un alto grado de correspondencia entre los informantes. Sin embargo, no ocurre lo mismo en cuanto al tamaño muestral, por lo que habrá que tener en cuenta esta diferencia a la hora de realizar las comparaciones.

## 2.2. Los centros de interés

Los centros de interés son los núcleos o campos nocionales que desencadenan las respuestas de los informantes y que sirven para elaborar, mediante una fórmula matemática, los diccionarios de léxico disponible del total de cada muestra.

El estudio de Lin (2006) se basa —con ciertas modificaciones— en los dieciséis centros de interés clásicos, es decir, los originados en las investigaciones francesas y más tarde incorporados al PPHDL: 01. Partes del cuerpo humano, 02. La ropa, 03. La casa (sin muebles), 04. Los muebles de la casa, 05. Alimentos, 06. Objetos situados en la mesa para las comidas, 07. La cocina y sus utensilios, 08. La universidad: muebles y material escolar, 09. Iluminación y aire acondicionado, 10. La ciudad, 11. El campo, 12. Medios de

---

<sup>4</sup> Según los niveles del *Marco común europeo de referencia para las lenguas* (Consejo de Europa, 2002).

transporte, 13. Trabajos del campo y del jardín, 14. Animales, 15. Juegos y distracciones y 16. Profesiones y oficios.

En nuestra investigación, siguiendo criterios didácticos y basándonos en el trabajo de Sánchez-Saus Laserna (2011), hemos incluido centros de interés nuevos y hemos alterado la etiqueta designativa de algunos. De este modo, contamos con los siguientes dieciocho centros: 01. El cuerpo humano, 02. La ropa y complementos, 03. La casa, 04. Comidas y bebidas, 05. Escuela y universidad, 06. La ciudad, 07. El medio ambiente, 08. Medios de transporte, 09. Los animales, 10. Ocio y tiempo libre, 11. Profesiones y trabajos, 12. Tiempo meteorológico y clima, 13. Acciones y actividades habituales, 14. Aspecto físico y carácter, 15. La familia, 16. Viajes y vacaciones, 17. El dinero y 18. Tecnología e Internet.

Esta diversidad de los campos nocionales nos obliga a restringir el número de ellos que podemos comparar. De este modo, solo contemplaremos aquellos centros de interés que coinciden en ambas investigaciones —pese a las posibles diferencias nominativas—, por lo que la nómina final para el cotejo son: CI01. El cuerpo humano [CUE], CI02. La ropa y complementos [ROP], CI03. La casa [CAS], CI04. Comidas y bebidas [COM], CI05. Escuela y Universidad [ESC], CI06. La ciudad [CIU], CI07. El medio ambiente [AMB], CI08. Medios de transporte [TRAN], CI09. Los animales [ANI], CI10. Ocio y tiempo libre [OCI] y CI11. Profesiones y oficios [PROF]<sup>5</sup>.

## 2.3. Obtención de los datos

Para la recopilación de los datos en cada una de las investigaciones se ha seguido la misma metodología. Se administró una encuesta a los informantes en la que se incluían dos partes diferenciadas: un cuestionario sociológico, para recoger información acerca de las variables sociolingüísticas, y una

---

<sup>5</sup> Las designaciones que aquí aparecen se corresponden con las que hemos empleado en nuestra propia investigación.

prueba de léxico disponible (véase Anexo)<sup>6</sup>, en la que los estudiantes debían escribir en forma de columna todas aquellas palabras y expresiones que asociasen con el centro de interés en cuestión. Para aportar sus respuestas — por escrito— disponían de 2 minutos para cada centro, por lo que la duración total de la encuesta dependerá de los centros de interés incorporados en el estudio. Una vez agotado ese tiempo, se cambiaba de centro de interés y los alumnos no podían, bajo ningún concepto, volver a los centros anteriores. También se advirtió a los informantes de que no se preocuparan en el caso de tener dudas sobre la ortografía o la pertenencia de sus respuestas al campo nocional.

Tras esta fase, se procede a la digitalización y edición de los datos. Para el primero de estos pasos se sigue un formato determinado —que varía según el programa informático que vaya a procesar los datos—, y para el segundo se aplica una serie de criterios comunes en las investigaciones de léxico disponible en nativos (Samper Padilla, 1997: 311 y ss.) y otros propios de los estudios con informantes extranjeros (Sánchez-Saus Laserna, 2011: 235-244).

Seguidamente, todos estos datos son procesados informáticamente a través de un programa específico que calcula la disponibilidad léxica de las respuestas de los informantes. Lin (2006) se sirve de *Lexidisp* (Moreno Fernández, 1995), mientras que nosotros nos hemos valido de *Dispogen* (Echeverría *et al.*, 2005). No obstante, ambos programas se basan en la misma fórmula desarrollada por López Chávez y Strassburger (1987) que, *grosso modo*, calcula el léxico disponible en función de la *frecuencia de aparición* de los vocablos, es decir, de cuántos informantes actualizan dicha lexía, y de la *posición* en la que aparece (de ahí que los estudiantes escriban sus respuestas en forma de columna). Cuantas más veces se repita un término y en posiciones más altas, mayor será su índice de disponibilidad.

---

<sup>6</sup> En el anexo solo aportamos una muestra de la prueba de léxico disponible, ya que los cuestionarios para recopilar la información sociolingüística difieren en ambas investigaciones al divergir los factores contemplados.

Finalmente, los resultados de léxico disponible se presentan de forma cuantitativa, mediante datos que dan cuenta del total de palabras, del total de vocablos, del promedio de vocablos y del índice de cohesión (que determina la medida en que convergen las respuestas de los informantes, pero que se ve afectado por el número de encuestados); y de forma cualitativa, a través de un diccionario o listado en el que se incluyen los vocablos actualizados por los informantes ordenados por centro de interés en función de su índice de disponibilidad.

## 2.4. Vocablos de referencia para el cotejo

Otra delimitación aplicada a nuestro contraste, además de la relativa a los centros de interés, está relacionada con el número de vocablos que compararemos. Para el cotejo cuantitativo emplearemos los resultados totales de cada una de las investigaciones en cuanto al número total de palabras, al promedio de palabras y al número total de vocablos. Sin embargo, para el cotejo cualitativo, por cuestiones de rentabilidad y espacio y con intención de disminuir la influencia de los tamaños muestrales, solo utilizaremos una parte teniendo en cuenta su índice de disponibilidad.

Tradicionalmente, en los estudios contrastivos en disponibilidad léxica se han empleado distintas medidas de corte: los 20 o 50 primeros vocablos, el *índice de disponibilidad*, la *frecuencia de aparición*, la *frecuencia acumulada* o, más recientemente, el *índice de compatibilidad* (Ávila Muñoz, 2016). Todos estos criterios de selección cuantitativa, pese a ser relevantes por sí mismos, responden a planteamientos subjetivos, puesto que, como afirman Ávila Muñoz y Villena Ponsoda (2010: 60), «no poseemos ningún valor de referencia objetivo que marque la mayor o menor proximidad de los vocablos al núcleo prototípico [el centro de interés]». En nuestro caso, para la comparación cualitativa nos serviremos del tradicional *índice de disponibilidad* (ID) superior o igual a 0,1, un índice matemático que ordena los vocablos de 1 a 0, representando 1 el máximo grado de disponibilidad y 0 el mínimo. Por lo



tanto, solo compararemos aquellos vocablos de ambos estudios que cumplan con este índice.

## 2.5. Procedimiento del cotejo

Nuestro cotejo intermuestral se compone, como ya hemos anticipado, de dos tipos de contrastes: uno cuantitativo y otro cualitativo. En el primero compararemos los datos procedentes de los resultados generales de nuestro estudio con los de Lin (2006). Estos datos, que nos informan sobre la productividad total, es decir, el número total de palabras por centro de interés, y sobre la productividad individual, o sea, la media de palabras por informante y centro de interés; sobre la riqueza léxica, o lo que es lo mismo, el número total de vocablos; y sobre los vocablos más disponibles, nos ayudarán a determinar el rendimiento léxico de las muestras comparadas.

En el segundo contraste localizaremos las similitudes y diferencias entre los vocablos más disponibles, evaluando para ello los distintos vocablos activados con un índice de disponibilidad igual o superior a 0,1, como ya apuntamos en el apartado anterior. En este último caso, tendremos en cuenta el porcentaje de permanencia léxica, es decir, el número de vocablos disponibles que coincide entre la investigación de 2006 y la de 2016.

## 3. COTEJO CUANTITATIVO

En este apartado emplearemos los valores totales procedentes de los resultados generales de las dos investigaciones comparadas. Nos limitaremos a los once centros de interés ya mencionados (véase 2.2.) y abordaremos el número total de palabras (cada una de las respuestas), el número total de vocablos (cada una de las entradas o palabras diferentes), el promedio de palabras por informante y centro de interés y el número de vocablos más disponibles.

## 3.1. Total de palabras y de vocablos

A través del total de palabras y del total de vocablos podemos extraer información acerca de la *productividad* y de la *riqueza léxica* de las muestras para los distintos centros de interés. A continuación podemos ver representados estos datos:

Tabla 1. – Total de palabras

CI	PALABRAS	
	China (2006)	China (2016)
CI01. CUE	3.969	7.536
CI02. ROP	2.575	<b>5.229</b>
CI03. CAS	2.496	7.229
CI04. COM	4.021	<b>7.972</b>
CI05. ESC	<b>2.098</b>	7.382
CI06. CIU	3.058	7.327
CI07. AMB	2.362	6.574
CI08. TRAN	2.234	5.723
CI09. ANI	<b>4.304</b>	6.703
CI10. OCI	2.392	7.731
CI11. PROF	4.051	7.207
<b>TOTAL</b>	33.560	76.613

El total de palabras para el conjunto de las muestras revela claramente las diferencias y la influencia de los tamaños muestrales, con una supremacía evidente por parte de los informantes del estudio de 2016. Por este motivo, no consideraremos las cifras absolutas, sino que nos centraremos en el rango u orden de productividad que presentan los distintos centros de interés. De este modo, constatamos que no existe correspondencia en los extremos de productividad. Los centros más productivos son el 09. «Animales» en el estudio de Lin (2006) y el 04. «Comidas y bebidas» en nuestro trabajo,

mientras que los que menor rendimiento arrojan son el 08. «Medios de transporte» en el primero, y el 02. «La ropa y complementos» en el segundo.

En el resto de centros de interés, encontramos correspondencia en las posiciones quinta, novena y décima, ocupadas por los centros 06. «La ciudad», 07. «El medio ambiente» —que en el estudio de 2006 recibe el nombre de «El campo»— y 08. «Medios de transporte», respectivamente en ambas investigaciones.

Si ponemos el foco en el número de vocablos, es decir, en la riqueza léxica, también comprobamos estas diferencias muestrales.

Tabla 2. – Total de vocablos

CI	VOCABLOS	
	China (2006)	China (2016)
CI01. CUE	119	451
CI02. ROP	125	<b>388</b>
CI03. CAS	123	674
CI04. COM	237	713
CI05. ESC	132	996
CI06. CIU	295	<b>1.171</b>
CI07. AMB	<b>315</b>	1.091
CI08. TRAN	<b>99</b>	718
CI09. ANI	137	488
CI10. OCI	271	885
CI11. PROF	272	868
<b>TOTAL</b>	2125	8.443

En cuanto a los campos nocionales más ricos, observamos que la primera posición la ocupan el centro 07. «El campo», en la investigación más antigua, y el centro 06. «La ciudad», en la más reciente. No obstante, existe cierta proximidad entre ambos estudios con respecto a las posiciones más altas, ya que los citados centros de interés ocupan el segundo puesto en la

investigación contraria. En la última posición, en cambio, encontramos el centro de interés 08. «Medios de transporte», para los informantes analizados en 2006, y el centro 02. «La ropa y complementos», para los evaluados en 2016. En este último caso notamos que el centro menos rico y el menos productivo coinciden.

En relación con los demás centros de interés, hallamos dos coincidencias en la cuarta y undécima posición: el centro 10. «Ocio y tiempo libre» —«Juegos y distracciones» en Lin (2006)— y el centro 01. «El cuerpo humano» —«Partes del cuerpo humano» en el trabajo de autoría china—.

### 3.2. Promedio de palabras

Hasta aquí solo hemos podido aportar información relativa a los centros de interés, ya que la disparidad de los tamaños muestrales que hemos observado dificulta conocer de forma más precisa el comportamiento léxico de los informantes. Por este motivo, consideramos que a través de un cotejo de los promedios de palabras —por informante y centro de interés— podemos, en cierta medida, atenuar la influencia del número de informantes sobre los resultados y determinar así el rendimiento léxico de las dos muestras sinohablantes.

Tabla 3. – Promedio de palabras por informante y CI

CI	PROMEDIO DE PALABRAS	
	China (2006)	China (2016)
CI01. CUE	15,09	<b>17,12</b>
CI02. ROP	9,79	<b>11,88</b>
CI03. CAS	9,49	<b>16,43</b>
CI04. COM	15,28	<b>18,12</b>
CI05. ESC	7,97	<b>16,78</b>
CI06. CIU	11,62	<b>16,65</b>
CI07. AMB	8,98	<b>14,94</b>
CI08. TRAN	8,49	<b>13,01</b>



<b>CI09. ANI</b>	16,36	<b>15,23</b>
<b>CI10. OCI</b>	9,06	<b>17,57</b>
<b>CI11. PROF</b>	15,4	<b>16,38</b>
<b>MEDIA TOTAL</b>	11,59	<b>15,83</b>

A partir de la tabla anterior podemos deducir que los promedios son mayores en nuestra investigación para todos los centros de interés cotejados, algo que refleja también la media total (15,83 palabras por informante y centro de interés frente a las 11,59). No obstante, debemos recordar aquí que entre ambas investigaciones existe una variación en la designación de los campos nocionales empleados, lo que sin duda habrá repercutido en una mayor capacidad asociativa de los mismos y, por ende, en una mayor producción de respuestas. Pese a ello, también constatamos que en aquellos centros de interés donde la designación no cambia o donde la modificación es mínima los promedios son igualmente mayores que los proporcionados por los informantes del años 2006. Por lo tanto, esta situación podría interpretarse como una mejora en la productividad léxica de los informantes de 2016.

### 3.3. Vocablos más disponibles

Una comparación, en términos cuantitativos, de los vocablos más disponibles, es decir, de aquellos que poseen un índice de disponibilidad superior o igual a 0,1 —recordamos que esta es una de las medidas de corte que se emplean tradicionalmente para determinar los vocablos más disponibles—, nos permitirá formarnos una idea del nivel de cohesión o similitud que existe entre las respuestas de los informantes: una mayor semejanza de las respuestas supondrá una cantidad mayor de vocablos en las primeras posiciones.

Tabla 4. – Vocablos con un ID  $\geq$  0,1

CI	PROMEDIO DE PALABRAS	
	China (2006)	China (2016)
<b>CI01. CUE</b>	24	<b>25</b>
<b>CI02. ROP</b>	15	<b>20</b>
<b>CI03. CAS</b>	16	<b>24</b>
<b>CI04. COM</b>	28	<b>30</b>
<b>CI05. ESC</b>	21	<b>25</b>
<b>CI06. CIU</b>	<b>20</b>	17
<b>CI07. AMB</b>	17	17
<b>CI08. TRAN</b>	12	<b>13</b>
<b>CI09. ANI</b>	<b>30</b>	26
<b>CI10. OCI</b>	12	<b>25</b>
<b>CI11. PROF</b>	<b>30</b>	28
<b>MEDIA TOTAL</b>	225	<b>250</b>

Comprobamos que nuestros informantes presentan mayor número de vocablos más disponibles en todos los centros de interés salvo en el CI06. «La ciudad», el CI07. «El medio ambiente» —donde la cantidad es la misma en las dos investigaciones—, el CI09. «Los animales» y el CI11. «Profesiones y oficios». Esto significa que sus respuestas guardan mayor parecido en cuanto a frecuencia y orden de aparición.

## 4. COTEJO CUALITATIVO

En este apartado expondremos aquellos vocablos que coinciden y divergen entre ambas muestras, pero solo de aquellos con un ID  $\geq$  0,1, es decir, empleamos los vocablos más disponibles para así mitigar la influencia de la diferencia de tamaños de las muestras, de modo que comparemos hechos de norma y evitemos la interferencia de respuestas que representan fenómenos particulares de un grupo reducido de informantes o incluso de un

único encuestado. Tomaremos como referencia la investigación de Lin (2006), y estableceremos el porcentaje de vocablos más disponibles que se han mantenido tras los años. A tal efecto, a continuación aportaremos los vocablos que coinciden en ambas listas y los que son exclusivos de cada una de ellas, indicando entre paréntesis la cantidad —que nos permitirá calcular el porcentaje de permanencia—.

#### 4.1. CI01. El cuerpo humano

**Coincidencias:** *ojo, mano, nariz, cabeza, pie, boca, pierna, oreja, brazo, pelo, dedo, pecho, cara, espalda, cuello, diente, estómago, lengua, piel, corazón, uña, ceja y labio* (23)

**Divergencias:**

China (2006): *lomo* (1)

China (2016): *hombro(s) y cuerpo* (2)

La separación temporal apenas demuestra diferencias entre las dos muestras para este centro de interés, ya que 23 de las 24 unidades más disponibles de la muestra de Lin (2006), es decir, un 92 %, permanecen como tales en nuestra investigación. Esto es lógico si tenemos en cuenta que se trata de un campo léxico muy estable donde los referentes extralingüísticos no suelen variar.

#### 4.2. CI02. La ropa y complementos

**Coincidencias:** *pantalón, camisa, chaqueta, blusa, calcetín, abrigo, zapato, falda, jersey, camiseta, traje, vestido, sombrero, bufanda y gorra* (15)

**Divergencias:**

China (2006): (0)

China (2016): *gafas, anillo, pendiente(s), ropa y gorro* (5)

El 100 % de los vocablos más disponibles del estudio de 2006 subsiste en el presente estudio. De hecho, las nuevas unidades añadidas por nuestra muestra se deben a la ampliación de la designación del centro de interés («los complementos»).

#### 4.3. CI03. La casa

**Coincidencias:** *cuarto de baño, cocina, habitación, sala (de estar), dormitorio, ventana, puerta, balcón, salón, techo, piso y jardín* (12)

**Divergencias:**

China (2006): *pared, lavabo, suelo y comedor* (4)

China (2016): *mesa, silla, sofá, tele(visión), lámpara, armario, cuarto, televisor, edificio, mueble y estudio* (11)

Ambos estudios, pese a la evidente variación en el nombre del centro de interés («La casa» frente a «Partes de la casa (sin muebles)»), comparten 12 unidades, lo que se traduce en que un 75 % del léxico más disponible de la primera investigación en sinohablantes está presente también en la nuestra. Hemos considerado equivalentes dos términos que aparecen en cada una de las investigaciones: *sala (de estar)*, del primer estudio, y *sala*, del segundo, pese a que siguen criterios de edición algo diferentes.

#### 4.4. CI04. Comidas y bebidas

**Coincidencias:** *arroz, pan, carne, pescado, verdura, huevo, leche, fruta, pollo, manzana, sopa, tomate, café, agua, jamón, patata, chocolate, paella, naranja, ensalada, vino y Coca-Cola* (22)

**Divergencias:**

China (2006): *trigo, queso, postre, helado, maíz y cerdo* (6)

China (2016): *cerveza, zumo, té, tortilla, tapa(s), (vino) tinto, alcohol y sangría* (8)

De las 28 unidades más disponibles de la muestra de 2006, 22 aparecen también en la muestra de 2016. Esto significa que el 78,57 % de léxico más disponible no ha variado en el transcurso de los años. Además, comprobamos que buena parte de las diferencias se debe al segundo elemento nocional del centro de interés que usamos en nuestra investigación: «bebidas».

#### 4.5. CI05. Escuela y universidad

**Coincidencias:** *mesa, libro, biblioteca, comedor, aula, edificio y dormitorio* (7)

**Divergencias:**

China (2006): *silla, pizarra, ordenador, televisor, pupitre, lámpara, lápiz, pluma, campo de deportes, sala de clase, papel y tele(visión)* (12)

China (2016): *profe(sor), estudiante, alumno, clase, facultad, (escuela) secundaria, examen, compañero, (escuela) primaria, cole(gio), maestro, bachiller(ato), escuela y universidad* (14)

La alteración de la etiqueta designativa del centro de interés en nuestra investigación —en Lin (2006) se usa «La universidad: muebles y material escolar»—, con el consiguiente aumento en el número y tipología de respuestas, explica que en este caso solo se mantengan 7 de los vocablos más disponibles (un 28 %) entre las dos muestras.

#### 4.6. CI06. La ciudad

**Coincidencias:** *edificio, calle, coche, tienda, parque, supermercado, autobús, hospital, escuela, restaurante, gente y tráfico* (12)

**Divergencias:**

China (2006): *plaza, universidad, jardín, avenida, mercado, cine, hotel y casa* (8)

China (2016): *Madrid, metro(politano/a), Barcelona, contaminación y ciudadano* (5)

Este centro de interés es el primero, hasta ahora, en el que la muestra más antigua supera a la muestra de nuestra investigación. Esto explica el hecho de que existan varias unidades que no coinciden entre ambos estudios. Algunas de estas diferencias representan nuevas asociaciones activadas: los topónimos *Madrid* y *Barcelona* de nuestra muestra. Aun así, la convergencia arroja un valor de 12 unidades o, lo que es lo mismo, un 60 % de los vocablos más disponibles de los informantes evaluados en 2006 están presentes también en los resultados referentes al 2016.

#### 4.7. CI07. El medio ambiente

**Coincidencias:** *río, árbol, flor, tierra, montaña y bosque* (6)

**Divergencias:**

China (2006): *campesino, huerto, casa, vaca, campo, perro, trigo, arroz, cerdo, hierba y trigal* (11)

China (2016): *aire, contaminación, agua, mar, cielo, animal, sol, lluvia, planta, lago y viento* (11)

Aquí tenemos otro caso de la incidencia del nombre del centro de interés sobre la naturaleza de las respuestas. Mientras que la primera investigación —que usa la designación «El campo»— se centra en elementos paisajísticos y animales, fundamentalmente, la segunda añade fenómenos naturales y problemas ambientales. Con todo esto, se aprecia una estabilidad léxica

relativamente baja del 35,29 %, es decir, ambos estudios comparten solo 6 vocablos.

#### 4.8. CI08. Medios de transporte

**Coincidencias:** *(auto)bús, bici(cleta), coche, avión, tren, metro, barco, moto(cicleta), taxi, carro y a pie* (11)

**Divergencias:**

China (2006): *camión* (1)

China (2016): *aeropuerto y andar* (2)

Una gran estabilidad define este centro de interés, pues 11 de las 12 lexías más disponibles de la investigación de Lin (2006) se mantienen en la nuestra, esto es, un 91,66 %. De hecho, las diferencias por parte de nuestra muestra están causadas por vocablos que no son medios de transporte propiamente dichos.

#### 4.9. CI09. Los animales

**Coincidencias:** *perro, gato, cerdo, tigre, oso, león, caballo, mono, zorro, vaca, lobo, pato, pez, gallina, serpiente, pájaro, conejo, toro, oveja, gallo, ratón, elefante, pollo, oso panda y ave* (25)

**Divergencias:**

China (2006): *cuervo, papagayo, asno, buitre y buey* (5)

China (2016): *jirafa* (1)

Hasta un 83,33 % del léxico más disponible (25 unidades) de la muestra de 2006 permanece invariable con el transcurso de los años. Es más, la pérdida de riqueza léxica en nuestra investigación (26 unidades frente a las 30

del estudio de 2006) es la causa principal de las divergencias existentes entre los dos estudios.

#### 4.10. CI10. Ocio y tiempo libre

**Coincidencias:** *fútbol, baloncesto, cantar, nadar y correr* (5)

**Divergencias:**

China (2006): *bailar, cartas, tenis, discoteca, voleibol, ping-pong y jugar a las cartas* (7)

China (2016): *leer, música, dormir, libro, deporte, pelí(cula), jugar, cine, tele(visión), viajar, comer, Internet, charlar, comprar, pasear, ordenador, bailar, amigo, estudiar y compras* (20)

Las nuevas asociaciones activadas por la denominación del centro de interés en nuestra investigación han hecho que solo 5 de las unidades más disponibles del estudio de 2006 —que emplea la etiqueta «Juegos y distracciones»— permanezcan, lo que supone únicamente un 21,74 %.

Debemos hacer referencia a un caso en el que no estamos de acuerdo con los criterios de edición utilizados por Lin (2006). Se trata de las opciones *cartas y juegos de cartas*, pues pensamos que según parámetros didácticos deberían haber sido agrupadas bajo una misma entrada.

#### 4.11. CI11. Profesiones y trabajos

**Coincidencias:** *profesor, médico, cantante, estudiante, cocinero, funcionario, maestro, obrero, doctor, enfermera, abogado, policía, camarero, campesino, director, actor, empleado, ingeniero, escritor y presidente* (20)

**Divergencias:**

China (2006): *enfermero, peluquero, taxista, gerente, chófer, intérprete, mecánico, empresario, secretario y guía* (10)

China (2016): *jefe, ama de casa, trabajador, comerciante y vendedor* (5)

Por último, este centro de interés arroja un total de 20 unidades compartidas por ambas muestras, es decir, un 66,67 % del léxico más disponible no se ha alterado a pesar de los diez años transcurridos entre las investigaciones.

Nuevamente no estamos de acuerdo con la inclusión de dos entradas separadas para las formas *enfermera* y *enfermero*, ya que no pensamos que el género sea una justificación suficiente para tomar esta decisión.

## 5. CONCLUSIONES

Esta comparación diacrónica, pese a las diferencias de tamaño de las muestras y del nombre de algunos centros de interés, nos ha revelado una alta consistencia o estabilidad léxica entre ambos conjuntos muestrales, ya que apreciamos una permanencia del 70,22 % del vocabulario más disponible de los informantes sinohablantes de 2006 (158 unidades de un total de 225). Hemos contemplado que los valores más altos de coincidencia se han producido en centros de interés bastante compactos, es decir, con posibilidades más limitadas o controladas en la variedad de respuestas. Estos centros, pese a las ligeras variaciones nominativas, han sido: 02. «La Ropa y complementos» (100 %), 01. «El cuerpo humano» (92 %) y 07. «Medios de transporte» (91,66 %). Por el contrario, las mayores diferencias se han observado en tres centros que se han visto afectados por el cambio designativo y que se caracterizan por una gran capacidad asociativa: 10. «Ocio y tiempo libre» (21,74 %), 05. «Escuela y universidad» (28 %) y 07. «Medio ambiente» (35,29 %). Todo esto refleja la importancia de los centros de interés que se incorporan y los nombres de los mismos.

Asimismo, hemos detectado una mejoría en términos cuantitativos de los resultados de nuestra investigación en relación con la de Lin (2006), con un mayor rendimiento en siete de los once centros de interés contrastados: 01. «El cuerpo humano», 02. «La ropa y complementos», 03. «La casa», 04. «Comidas y bebidas», 05. «Escuela y universidad», 08. «Medios de transporte» y 10. «Ocio y tiempo libre».

Estos rendimientos seguramente se hayan visto propiciados por los cambios nominativos de los centros de interés, lo que corrobora su influencia y nos indica la necesidad de llegar a un consenso en esta materia para próximas investigaciones, sobre todo en estudiantes de ELE. Asimismo, quizá también son síntoma de un mayor asentamiento didáctico del español, de materiales y metodologías más variados y de la diferencia formativa de los docentes. No obstante, no debemos ignorar la influencia de las eventuales variaciones en el perfil del alumnado, aunque debido al carácter diacrónico de este estudio, no disponemos de datos para concretar dichas variaciones. Una solución a esta limitación podría ser la realización de análisis más restringidos en función, por ejemplo, del centro educativo, ya que esto nos permitiría recopilar información más precisa sobre el contexto de enseñanza (materiales, metodología, alumnado, profesorado...). Esto también nos conduce a reflexionar sobre la posible influencia del tamaño de las muestras, aunque consideramos que el empleo de los vocablos más disponibles ha atemperado esta influencia, ya que en las primeras posiciones estarán aquellas unidades que presentan más cohesión entre los informantes.

Otra limitación que apuntamos es la de no haber considerado las listas de léxico disponible en su totalidad y habernos delimitado a los vocablos más disponibles. Creemos que su utilización en estudios con muestras de igual tamaño y características puede aportarnos una visión más precisa de la evolución del vocabulario y, también, de las relaciones y variaciones semánticas que se activan. Derivado de esta homogeneidad a la que hacemos referencia, otro posible ámbito de actuación sería el estudio diacrónico de una misma muestra, es decir, realizar un seguimiento del proceso de aprendizaje

de lengua española de un mismo grupo de informantes a lo largo de los distintos cursos académicos.

Con todo esto, este tipo de contraste, centrándonos sobre todo en las diferencias, nos permitirá localizar aquellas unidades léxicas cuya enseñanza deberá someterse a escrutinio y, en su caso, reforzarse, sobre todo teniendo en cuenta la —necesaria— actualización de los materiales didácticos. Por lo tanto, sería interesante utilizar la disponibilidad léxica y sus resultados para analizar el componente léxico de manuales y de profesores en aras de tener una idea más completa de las necesidades del alumnado y cómo estos las cubren.

Por último, nos hemos percatado de la diferencia de criterios de edición empleados (p. ej. el tratamiento por separado de dos unidades que representan la misma realidad extralingüística, pero que no responde a motivos de sinonimia, sino morfológicos), una diferencia que puede entrañar efectos en los índices de disponibilidad de los vocablos. Por lo tanto, desde aquí reivindicamos una revisión y fijación de los criterios de edición empleados en trabajos con estudiantes de ELE o en los que tienen como muestra informantes nativos y cuyo objetivo es establecer un corpus léxico para la mejora de la selección del vocabulario que debe enseñarse. El campo del ELE debe apropiarse de su parcela dentro de la disponibilidad léxica para llevar a cabo estudios encaminados a la enseñanza del léxico, lo que conllevará una modificación en dichos criterios de edición y una revisión y sistematización de los centros de interés y de sus respectivas denominaciones.

## BIBLIOGRAFÍA

- Ávila Muñoz, Antonio Manuel. «El léxico disponible y la enseñanza del español. Propuesta de selección léxica basada en la teoría de los conjuntos difusos». *Journal of Spanish Language Teaching*, 3(1), 2016, 31-43. Web. 23 de diciembre de 2016

- Ávila Muñoz, Antonio Manuel y Villena Ponsoda, Juan Andrés (eds.). *Variación social del léxico disponible en la ciudad de Málaga: diccionario y análisis*. Editorial Sarriá, 2010.
- Bartol Hernández, José Antonio. «Disponibilidad léxica y selección del vocabulario». *De moneda nunca usada. Estudios filológicos dedicados a José M. Enguita Utrilla*, editado por Rosa M.ª Castañer Martín y Vicente Lagüéns Gracia, Institución «Fernando el Católico», 2010, 85-107.
- Carcedo González, Alberto. *Disponibilidad léxica en español como lengua extranjera: el caso finlandés (estudio del nivel preuniversitario y cotejo con tres fases de adquisición)*. Turun Yliopisto, 2000.
- Consejo de Europa. *Marco común europeo de referencia para las lenguas: Aprendizaje, Enseñanza, Evaluación*. Traducción del Instituto Cervantes. Madrid: Ministerio de Educación, Cultura y Deporte, 2002. Web. 24 de enero de 2017.
- Echeverría, Max S. et al. *Dispogen II. Programa computacional para el análisis de la disponibilidad léxica*. Concepción, Chile: Universidad de Concepción, 2005.
- Gougenheim, George et al. *L'élaboration du Français Élémentaire (1er degré). Étude sur l'élaboration d'un vocabulaire et d'une grammaire de base*. Didier, 1954.
- ———. *L'élaboration du Français Fondamental (1er degré). Étude sur l'élaboration d'un vocabulaire et d'une grammaire de base*. Didier, 1964.
- Hoey, Michael. *Lexical priming: a new theory of words and language*. Routledge, 2005.
- Lewis, Michael. *The Lexical Approach. The State of ELT and a way forward*. Language Teaching Publications, 1993.
- Lin, Jing. «El estudio de disponibilidad léxica de los estudiantes chinos de español como lengua extranjera». Memoria de máster, Universidad de Alcalá de Henares, 2006.



- ——— «El estudio de disponibilidad léxica de los estudiantes chinos de español como lengua extranjera». *MarcoELE*, 14, 2012. Web. 25 de enero de 2017.
- López Chávez, Juan y Strassburger Frías, Carlos. «Otro cálculo del índice de disponibilidad léxica». *Presente y perspectivas de la investigación computacional en México. Actas del IV Simposio de la Asociación Mexicana de Lingüística Aplicada*. UNAM, 1987.
- López Morales, Humberto. *Disponibilidad léxica en escolares de San Juan*. Manuscrito, 1973.
- López Morales, Humberto. *Enseñanza de la lengua materna. Lingüística para maestros de español*. Madrid: Playor, 1984.
- Moreno Fernández, Francisco *et al.* «Cálculo de la disponibilidad léxica. El programa Lexidisp». *Lingüística*, 7, 1995, 243-249.
- Samper Padilla, José Antonio. «Criterios de edición del léxico disponible: sugerencias». *Lingüística*, 10, 1997, 311-333.
- Sánchez-Saus Laserna, Marta. «Bases semánticas para el estudio de los centros de interés del léxico disponible. Disponibilidad léxica de informantes extranjeros en las universidades andaluzas». Tesis doctoral, Universidad de Cádiz, 2011.

## ANEXO

### Ejemplo (adaptado) de la prueba de léxico disponible

Centro 1: <b>EL CUERPO HUMANO</b>	Centro 2: <b>LA ROPA Y COMPLEMENTOS</b>	Centro 3: <b>LA CASA</b>
1.	1.	1.
2.	2.	2.
3.	3.	3.
4.	4.	4.
5.	5.	5.
6.	6.	6.
7.	7.	7.
8.	8.	8.
9.	9.	9.
10.	10.	10.
11.	11.	11.
12.	12.	12.
13.	13.	13.
14.	14.	14.
15.	15.	15.
16.	16.	16.
17.	17.	17.
18.	18.	18.
19.	19.	19.
20.	20.	20.
21.	21.	21.
22.	22.	22.
23.	23.	23.
24.	24.	24.
25.	25.	25.
26.	26.	26.
27.	27.	27.
28.	28.	28.
29.	29.	29.

30.	30.	30.
31.	31.	31.
32.	32.	32.
33.	33.	33.
34.	34.	34.
35.	35.	35.
36.	36.	36.
37.	37.	37.
38.	38.	38.
39.	39.	39.
40.	40.	40.