



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA MAESTRÍA Y DOCTORADO EN
LINGÜÍSTICA

**Estudios de léxico disponible: diccionarios de español e inglés
de alumnos de inglés del CELE de la UNAM**

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRA EN LINGÜÍSTICA APLICADA

PRESENTA:
ANAHÍ RAMÍREZ ALFARO

TUTORA PRINCIPAL:
DRA. Marisela Colín Rodea
CENTRO DE ENSEÑANZA DE LENGUAS EXTRANJERAS

MÉXICO, D.F. Diciembre de 2015



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Comienzo contigo Gerardo, porque vuelves a enamorarme con tus palabras y tu amor incondicional. Gracias por estar a mi lado.

A Alonso, por llenar mis días de amor, alegría y sonrisas.

A mis padres, por su apoyo y amor en cada momento de mi vida.

A nuestra querida Aurelia, por todas las mañanas y tardes que cuidaste a mi niño con tanta paciencia.

A Martina, por escucharme siempre que necesitaba ánimos para continuar con este trabajo. ¡Mil gracias!

A Susana, por enseñarme a entenderme mejor y quererme más.

A Isabel y Aurora porque siguen siendo prueba de que la verdadera amistad es incondicional.

Agradezco principalmente a mi tutora, la Dra. Marisela Colín Rodea, por invitarme a participar en este importante proyecto en colaboración con los miembros del seminario de Léxico Disponible.

A ustedes compañeros, Leonardo, Paty, Barbara y Eréndira.

Finalmente agradezco a mis sinodales, Natalia Ignatieva, Silvia López, Patricia Cánovas y Leonardo Herrera por sus valiosos comentarios que enriquecieron esta tesis.

Índice

Índice de imágenes	v
Resumen	vi
Introducción	1
Capítulo 1 - Marco Teórico.....	11
1.1 ¿Qué es una palabra?.....	11
1.2 ¿Qué es un vocablo?.....	14
1.3 ¿Qué es el léxico?	15
1.4 Tipos de léxico	16
1.4.1 Léxico disponible	16
1.5 Modelo Gravitacional de Daniel Gile	18
1.6 Grado de disponibilidad y Centros de interés.....	20
Capítulo 2 - Enseñanza y Aprendizaje de Vocabulario	24
2.1 Tratamiento del léxico en los enfoques de enseñanza de lenguas.....	24
2.2 El francés elemental: unidades didácticas.....	35
2.3 Aplicaciones pedagógicas para la enseñanza del léxico	38
2.4 Enseñanza de campos léxicos como propuesta pedagógica.....	39
Capítulo 3 - Metodología.....	43
3.1 Objetivos específicos.....	43
3.2 Preguntas de investigación.....	43
3.3 Informantes	44
3.4 Encuestas de léxico disponible	44
3.5 Corpus.....	46
3.6 Edición de encuestas	49
3.7 Tiempo de llenado de encuestas	53
3.8 LEDICAL	53
3.8.1 De la idea al software	55
3.8.2 Salida.....	56
3.8.3 Arquitectura.....	57
3.8.3.1 Página principal: Lenguas materna y extranjeras.....	58
3.8.3.2 Ingreso y Edición.....	60
3.8.3.3 Maestro	60
3.8.3.4 Grupo	60
3.8.3.5 Encuestado	61
3.8.3.6 Encuesta	62
3.9 Información de la encuesta (Lista de vocabulario):.....	62
3.9.1 Página Principal: Índice computado (Compute Index)	63
Capítulo 4 - Evaluación de LEDICAL	65
4.1 Aplicación y selección de las encuestas	66
4.2 Edición de las encuestas previo ingreso a LEDICAL	67
4.3 Funcionamiento de LEDICAL.....	70
4.3.1 Ingreso del maestro a LEDICAL.....	70
4.3.2 Ingreso del grupo a LEDICAL	72
4.3.3 Ingreso del encuestado a LEDICAL	72

4.3.4 Ingreso de la encuesta a LEDICAL	73
4.3.5 Ingreso de los datos de la encuesta a LEDICAL	74
4.4 Valoración global de LEDICAL.....	74
Capítulo 5 - Resultados Cuantitativos	76
Capítulo 6 - Discusión y Conclusiones.....	83
Referencias bibliográficas.....	91
Apéndice 1.....	95
Apéndice 2.....	97

Índice de imágenes

Imagen 1. Ilustración del Modelo Gravitacional de Daniel Gile Copyright de John Benjamins B.V. (Gile, 2009:279).	19
Imagen 2. Dibujo de una bicicleta parecida a la utilizada por Ilari en su texto sobre campos léxicos.	41
Imagen 3. Páginas uno y dos de la encuesta de léxico disponible.	45
Imagen 4. Página principal de LEDICAL.	58
Imagen 5. Página del menú principal.....	59

Resumen

Esta investigación se inscribe en la línea de léxico disponible de lenguas extranjeras que contempla el almacenamiento de las encuestas y de los diccionarios en la base de datos LEDICAL. La tesis es parte de un proyecto mayor desarrollado por María Luisa Frey Pereira (2004)¹ en su investigación doctoral “El léxico disponible en los escritos de alumnos de español como lengua extranjera. Estudio comparativo de dos corpora” y posteriormente fue propuesta en su investigación postdoctoral “Corpus multilingüe del léxico disponible (CMLD) de los alumnos del Centro de Enseñanza de Lenguas Extranjeras (CELE), Universidad Nacional Autónoma de México (UNAM)”.

El objetivo de mi investigación es estudiar el léxico disponible de alumnos universitarios mexicanos de español e inglés. El primer tipo fue el léxico disponible del español como lengua materna, y el segundo de la lengua extranjera que esos mismos alumnos estudian, en este caso el inglés. Las encuestas fueron los datos para alimentar y evaluar el desarrollo tecnológico LEDICAL. Se trata de una base de datos diseñada específicamente para este proyecto por Jim Frey y Santiago Cortés en 2010. La información se procesó mediante programas computacionales a fin de obtener los diccionarios de léxico disponible de cada uno de los idiomas recogidos.

Abstract

This research is developed within foreign languages lexical availability studies that consider entering both polls and dictionaries into LEDICAL database. This thesis is part of a major project developed by Maria Luisa Frey Pereira (2004) in her doctorate research called “Available Lexicon in the Writings of Students of Spanish as Foreign Language. A Comparative Study of Two Corpora”. It was later proposed as her postdoctoral research Multilingual Corpora of Available Lexicon (CMLD in Spanish) of Students of the Centre of Foreign Language Teaching (CELE) at the National Autonomous University of Mexico (UNAM).

The aim of this research is to study both English and Spanish available lexicon of Mexican university students. The first type of lexicon was in Spanish as a mother language, and the second one, in the foreign language that these same students are studying —English.

Data was taken from the polls and it was entered and used to evaluate the technological development LEDICAL. Jim Frey and Santiago Cortés exclusively designed this database for this project in 2010. The information was processed with computer programs in order to obtain available lexicon dictionaries of each one of the collected languages.

¹ Después del fallecimiento de Helen Frey en mayo de 2009, se despertó el interés por estos estudios entre varios docentes que se integraron al trabajo de conformación del CMLD. Actualmente el Seminario de Léxico Disponible ha apoyado la parte esencial de ese proyecto CMLD y ha impulsado otras investigaciones de léxico disponible. De esta manera, el Proyecto Estudios de Léxico Disponible (CBHC091127) busca dar seguimiento a estas investigaciones, colaborar en la formación de sus participantes en el área y promover el trabajo colaborativo e interdisciplinario entre sus participantes mediante reuniones, seminarios y participaciones en espacios de internet.

Introducción

La presente investigación se inscribe en el área de aprendizaje de léxico de segundas lenguas, específicamente se aboca al estudio de léxico disponible en una población estudiantil universitaria que estudia inglés como lengua extranjera, cuya lengua materna es el español. El léxico disponible se define como aquel tipo de léxico que viene a la mente cuando se presenta un elemento disparador que activa esa área del léxico en el hablante.

El objetivo del estudio es estudiar el léxico disponible en español e inglés de alumnos universitarios mexicanos. El primer tipo de léxico disponible fue del español como lengua materna, y el segundo de la lengua extranjera que esos mismos alumnos estudian, en este caso el inglés. Las encuestas fueron los datos para alimentar y evaluar el desarrollo tecnológico LEDICAL, el cual fue diseñado específicamente para este proyecto por Jim Frey y Santiago Cortés en 2010. La información se procesó mediante programas computacionales a fin de obtener los diccionarios de léxico disponible¹ de cada uno de los idiomas recogidos, así como los índices de disponibilidad y frecuencia de sólo un par de grupos encuestados. El cómputo de dichas encuestas fue llevado a cabo a fin de evaluar el desarrollo tecnológico LEDICAL como se verá posteriormente.

Planteamiento del problema

Al ser el léxico un área de gran importancia para el aprendizaje de una lengua extranjera existe un gran interés en la lingüística aplicada por saber de qué manera se aprende el léxico y de qué manera permanece en nuestra mente. En esta investigación pretendo obtener los diccionarios de español e inglés de los alumnos de inglés avanzado del Centro de Enseñanza de Lenguas

¹ Cabe aclarar que los diccionarios de léxico disponible son recuentos léxicos de palabras disponibles recolectadas por medio de encuestas diseñadas para dicho fin.

Extranjeras (CELE) con el fin de conocer el léxico en primera y segunda lenguas del que dispone esta comunidad de habla en particular. Los resultados obtenidos en esta investigación son de vital importancia en el desarrollo y creación de material didáctico del léxico en segundas lenguas. Además, se obtendrá un corpus en español de estudiantes universitarios mexicanos el cual podrá utilizarse para futuras investigaciones en lingüística aplicada y sus muchas líneas de investigación.

Los estudios de léxico disponible surgen en la década de 1950 en Francia con las investigaciones realizadas por los franceses Georges Gougenheim y René Michéa. Ellos se dieron cuenta de que las listas de frecuencia que existían del vocabulario frecuente no coincidían con el léxico utilizado comúnmente por cualquier francófono. Con ello surge la intención de crear listas de vocabulario frecuente y disponible, a partir de los cuales diseñar material didáctico dirigido a extranjeros aprendientes de francés. En esos momentos, Michéa se refería al léxico disponible como palabras temáticas y las definió como el léxico que conoce el hablante y se usa sólo cuando se producen asociaciones específicas. Es así que las investigaciones en este terreno se dan en un contexto de didáctica de lenguas y en específico en el diseño de materiales enfocados a la enseñanza del léxico. En el Marco Teórico de este trabajo me referiré en detalle a la importancia de estas investigaciones para conocer mejor el estudio de léxico disponible desde sus orígenes.

En el mundo hispano el lingüista asturiano Alberto Carcedo González alrededor de 1970 comienza el estudio de léxico disponible de alumnos finlandeses preuniversitarios que estudian español como lengua extranjera. Sus investigaciones estudiaron el léxico disponible entre comunidades dialectales distintas, tales como Puerto Rico y Cádiz. A lo largo de este tiempo se han realizado importantes investigaciones valiéndose del análisis estadístico sobre la disponibilidad y frecuencia léxicas, incluyendo siempre un análisis cualitativo de los resultados arrojados. En México, el lingüista Juan López Chávez (2003) y otros investigadores han

trabajando en la creación de diccionarios de léxico disponible en español como lengua materna a nivel básico (primaria), medio (secundaria) y medio-superior (preparatoria).

El área de lenguas extranjeras es en realidad el punto de inicio del estudio del léxico disponible. A través de él se buscó acercarse al léxico cotidiano conocido por los hablantes nativos, el cual sería de gran relevancia para el aprendizaje de la lengua por alumnos extranjeros. El proyecto iniciado por Helen Frey Pereyra en 2008, “Corpus multilingüe del léxico disponible de los alumnos del CELE, UNAM”, planteado como un estudio de Posdoctorado y posteriormente transformado en un proyecto de investigación titulado “Estudios de Léxico Disponible” busca estudiar el léxico disponible dentro del campo de la enseñanza de lenguas extranjeras.

La presente investigación derivó del proyecto anterior y es la primera en guardar registro del léxico disponible de los alumnos de inglés avanzado del CELE, así como de los datos obtenidos en un primer cómputo; como un primer acercamiento a los índices de frecuencia y disponibilidad léxica del mismo grupo. La obtención de datos estadísticos de los dos grupos computados en este trabajo fue realizada bajo un criterio evaluador de LEDICAL como prototipo computacional.

Justificación

Existe el interés por parte de un grupo de investigadores mexicanos trabajando en la creación de diccionarios de léxico disponible en español como lengua materna a nivel básico (primaria) medio (secundaria) y medio-superior (preparatoria), como son los casos antes mencionados. La presente investigación es la primera en guardar registro de los diccionarios en español e inglés de léxico disponible de los alumnos de inglés avanzado del CELE. Se presenta por primera vez el desarrollo tecnológico LEDICAL, el cual fue creado exclusivamente para la realización de este

proyecto y de la eventual recolección del diccionario multilingüe de los alumnos del CELE de la UNAM, así como de los índices de cohesión y disponibilidad léxica de los mismos alumnos.

Las implicaciones didácticas de nuestro estudio son de gran importancia para este centro de enseñanza de lenguas extranjeras, encargado de formar alumnos autosuficientes y calificados para enfrentarse a un mundo intercultural donde el aprendizaje de lenguas se vuelva una necesidad cada vez mayor en el desarrollo académico y profesional de los aprendientes.

Objetivos del estudio

Para el desarrollo de esta investigación se planteó un objetivo general y tres específicos.

Objetivo general

Estudiar el léxico disponible en inglés y español de alumnos universitarios mexicanos del Centro de Enseñanza de Lenguas Extranjeras, UNAM.

Objetivos específicos

Los objetivos específicos fueron los siguientes:

- Obtener los diccionarios de español e inglés de léxico disponible de los alumnos de inglés avanzado del CELE.
- Describir el proceso de edición de las encuestas para alimentar la base de datos LEDICAL.
- Identificar el léxico disponible de las encuestas de acuerdo al modelo gravitacional de Daniel Gile.
- Valorar el funcionamiento del desarrollo tecnológico

Preguntas de investigación

- ¿Cuál es el léxico disponible de los alumnos de inglés y español del CELE de la UNAM?
- ¿En qué medida el proceso de edición es una primera etapa del estudio del léxico disponible de los alumnos?
- ¿Cómo se valora el funcionamiento del programa computacional LEDICAL?

Antecedentes del tema

L'élaboration du français fondamental (1er degré) de 1967 fue la culminación de un proyecto iniciado por G. Gougenheim, R. Michéa, P. Rivenc y A. Sauvegeot, cuyo origen se relaciona con una iniciativa de la UNESCO por difundir la enseñanza de ciertas lenguas en la educación primaria de algunos países y que impulsó la enseñanza del francés en las ex-colonias africanas. La difusión de la lengua francesa comenzó a darse sobre todo en estos países colonizados por Francia en un intento por enseñar la lengua francesa como herramienta de comunicación. Esto marcó los fundamentos de los estudios de léxico disponible a nivel mundial.

Anterior a *L'élaboration du français fondamental (1er degré)* está el libro intitulado *Le Français Élémentaire* de 1954, cuya reedición en 1959 recibió el nombre de *Français Fondamental (1er degré)*. Esta obra fue reconocida como instrumento de difusión de la lengua francesa con una lista definitiva de 1,475 palabras, cuyas 1,222 fueron reconocidas como palabras léxicas y 253 como gramaticales. *Le Français Élémentaire* es reconocido como el texto fundador de esta tradición para la enseñanza del francés en la federación de territorios conocidos como *Fédération Française*; posteriormente se reorientó como un apoyo para mantener la

relación con el francés como lengua oficial o extranjera. A lo largo de estos años no sólo creció el número de estudios, también se ampliaron sus líneas de investigación y el tipo de aplicaciones.

En los diversos estudios de léxico disponible a nivel mundial se han tomado en cuenta múltiples variables, como lengua materna, periodo de inmersión o referentes sociales y culturales de los sujetos de estudio. En la investigación realizada por Marta Samper (2001) de la Universidad de las Palmas de Gran Canaria se trabajó con la variable lengua materna. En sus diversos estudios, Alberto Carcedo (1998), por su parte, ha considerado para el cuestionario sociolingüístico el conocimiento de otras lenguas, además de las variables sexo y lengua materna. En el 2000 publica un artículo intitulado “La lengua como manifestación de otredad cultural” donde muestra cómo se da la dimensión cultural en el vocabulario de estudiantes finlandeses de español. A partir del corpus obtenido el investigador ha realizado también análisis de errores en la interlengua de dichos alumnos.

Frey Pereyra (2008) estudió también la interlengua de estudiantes de español como lengua extranjera, así como la relación que tiene la inmersión en su aprendizaje y en la producción escrita de los mismos alumnos. Posteriormente inició el proyecto postdoctoral: “Corpus multilingüe del léxico disponible (CMLD) de los alumnos del Centro de Enseñanza de Lenguas Extranjeras, UNAM”.

Humberto López Morales (1999) en *El léxico disponible de Puerto Rico*, tomó en cuenta el nivel sociocultural, sexo y geografía, entre otros, para describir la disponibilidad léxica. El investigador cubano dirige el Proyecto Panhispánico de Léxico Disponible en el mundo hispano, el cual ha impulsado un sin fin de investigaciones cuyo criterios buscan realizar comparaciones lingüísticas, etnolingüísticas, etnográficas y culturales, así como de la enseñanza del español como lengua extranjera.

La sociolingüística ha ofrecido diversas perspectivas en los innumerables estudios de léxico disponible, debido a que el léxico de una lengua no sólo contiene características morfológicas, fonéticas y semánticas, sino también culturales y sociales. Además, las encuestas de vocabulario que se aplican para la obtención del corpus de una lengua no pueden dissociarse de la primera lengua de los hablantes y de la que estudian. De tal manera, se obtienen listas del léxico del que disponen en su propia lengua y de la que estudian.

Los estudios de léxico disponible han podido conjuntar los estudios lingüísticos con diferentes enfoques que muestran el lexicón de una población determinada como testimonios de las influencias socioculturales adquiridas durante el aprendizaje del léxico en lengua extranjera y aquel en la materna. Desde los estudios realizados por Gougenheim y Michéa se observó que la frecuencia de las palabras concretas es poco estable porque su estabilidad está relacionada con la experiencia de los hablantes. Ambos investigadores aseguraron que los índices de distribución dan muestra de dicha inestabilidad, la cual proviene precisamente de que las palabras surgen en circunstancias específicas y en determinados temas de conversación. Si uno encontrara un amigo saliendo del dentista, él hablará de sus dientes; lo mismo ocurriría si hubiera habido un accidente en el metro. (Gougenheim y Michéa, 1967:139)

Los franceses continúan afirmando que en los estudios que realizaron sobre el léxico fundamental la frecuencia de la palabra *rojo*, por ejemplo, cambió de un campo semántico a otro; su frecuencia varía dependiendo la situación en cuestión como podría ser la naturaleza de los textos o conversaciones donde se encuentran, el país o incluso su régimen político. Ellos descubrieron que una palabra concreta no tiene una frecuencia estable porque ésta depende de la circunstancia o el pronunciamiento efectivo de una palabra. Los ejemplos expuestos los llevaron a cuestionar la pertinencia de la terminología que se estaba usando en esos días. Hubo que definir eficazmente la diferencia entre las *palabras frecuentes (mots fréquents)*, las *palabras usuales*

(*mots usuels*) y las *palabras comunes* (*mots courants*) por su falta de precisión conceptual. Fue así que surgieron dos conceptos más —*palabras temáticas* y *palabras atemáticas*— y de ellos, la distinción entre el *léxico frecuente* y el *léxico disponible*:

La continuidad de nuestras reflexiones nos llevó a oponer las palabras frecuentes de las disponibles. Tal es el nombre que daremos a esas palabras de una frecuencia débil y poco estable, pero que son palabras comunes y útiles. Les llamamos así porque aunque no son habladas o escritas a menudo (excepto en el caso en que la escritura y la conversación traten de un tema específico) están a nuestra disposición, aunque escapen o incluso desaparezcan momentáneamente bajo el efecto de la enfermedad.²

A partir de la definición, identificación y caracterización de los tipos de vocabulario es que los lingüistas franceses reconocieron el vocabulario necesario como aquel constituido por el vocabulario frecuente y el disponible. La disponibilidad, por otro lado, logró reconocerse con más precisión y entonces se plantearon un nuevo reto relacionado con la investigación de un método que no sólo ofreciera la posibilidad de trabajar con las palabras disponibles, sino determinar su grado de disponibilidad. Dicha noción, afirmaron, corresponde más o menos a la presencia de las palabras que están dentro de nuestra memoria. Con ello se despertó interés lingüístico, psicológico y pedagógico, lo cual resulta ser de gran importancia para el establecimiento de una metodología léxica de base y no sólo a la creación de diccionarios de frecuencia como se había estado haciendo. Las investigaciones realizadas por Gougenheim y Michéa ofrecieron la posibilidad de relacionar los estudios lexicológicos como base en la creación de métodos y material didáctico exclusivo de la enseñanza del léxico.

² La suite de nos réflexions communes nous a amenés à opposer aux mots fréquents les mots disponibles. Tel est le nom que nous donnerons à ces mots d'une fréquence faible et peu stable, qui sont cependant des mots usuels et utiles. Nous les appelons ainsi parce que, quoiqu'ils ne soient pas souvent prononcés ou écrits effectivement (hormis le cas où l'écrit ou la conversation traitent d'un sujet déterminé), ils sont à notre disposition, quitte à nous échapper momentanément, ou même à disparaître sous l'effet de la maladie. (Gougenheim y Michéa, 1967:145)

Limitaciones del estudio

Las limitaciones tienen que ver con la cantidad de encuestas necesaria para llevar a cabo el análisis estadístico del léxico de los alumnos. Recolectar, editar y almacenar el léxico disponible de los universitarios para ser tratado por LEDICAL implica tener un número mínimo de encuestas. En el caso de inglés avanzado, la población estudiantil es lo suficientemente grande para llevar a cabo esta tarea, pero no sucede lo mismo en los niveles avanzados de portugués o ruso, por ejemplo. El total de alumnos avanzados varía dependiendo la lengua de estudio, por lo que es necesario plantear estudios de léxico disponible cuyas listas sean almacenadas en etapas para cubrir el número mínimo de encuestas para su análisis estadístico. Asimismo, la edición y captura de las encuestas es una tarea que requiere mucho tiempo. Sin embargo, una vez almacenados los datos, LEDICAL presenta un gran potencial para la investigación.

Estructura de la tesis

En la introducción de esta tesis planteo los objetivos general y específicos que buscamos comprobar en la realización de esta investigación, así como de los antecedentes del tema y limitaciones del estudio. La justificación y planteamiento del problema en la introducción reafirman la importancia de este trabajo en los estudios de léxico disponible a nivel superior en México. A continuación mencionaré brevemente los capítulos que conforman el presente trabajo:

Capítulo 1. Marco teórico. En este capítulo defino los conceptos y nociones relevantes de esta tesis con el fin de darle un sustento teórico sólido que la determine dentro de los estudios de léxico disponible específicamente y de las investigaciones lexicológicas.

Capítulo 2. Estudio y enseñanza de léxico. En este capítulo presento las principales corrientes de investigación y métodos de enseñanza del léxico en primeras lenguas y lenguas extranjeras, así como algunas propuestas de unidades didácticas enfocadas al léxico disponible.

Capítulo 3. Metodología. En este capítulo expongo la metodología seguida para el desarrollo de nuestra investigación, la cual –como ya hemos mencionado– retoma la metodología general propuesta para los estudios de léxico disponible, consistente en la aplicación de la encuesta diseñada para este propósito. De forma específica, seguimos la propuesta de Helen Frey Pereyra para el estudio del léxico disponible y la adquisición de lenguas extranjeras. La novedad radica en el diseño, instalación y llenado de la base de datos LEDICAL.

Capítulo 4. Evaluación de LEDICAL. En este capítulo evalúo el proceso de recolección, edición e ingreso de las encuestas de léxico disponible al programa computacional y base de datos LEDICAL. También reflexiono en el proceso de construcción de LEDICAL (como base de datos), así como en las facetas que se llevaron a cabo en este proyecto.

Capítulo 5. Resultados cuantitativos. En este capítulo muestro un primer acercamiento a los índices de disponibilidad y frecuencias relativas de dos grupos de inglés R-003 y S-004 de lectura extensiva y producción oral; en LEDICAL ambos corresponden a las encuestas 13 y 22 en español y 20 y 24 en inglés. Los resultados cuantitativos expuestos en esta parte de la tesis son una aportación evaluadora del funcionamiento de LEDICAL.

Capítulo 6. Discusión y Conclusiones. En este capítulo defino el tipo de léxico disponible recabado en las encuestas y presento las posibilidades de investigación futuras en esta área de estudio surgidas de esta investigación. Los resultados y conclusiones obtenidas reafirman la importancia de esta tesis en la recolección de los diccionarios de inglés y español por medio de LEDICAL y su uso exclusivo del CELE de la UNAM.

Capítulo 1 - Marco Teórico

En este capítulo defino los conceptos y nociones relevantes de esta tesis con el fin de darle un sustento teórico sólido que la determine dentro de los estudios de léxico disponible específicamente y de las investigaciones lexicológicas.

En esta investigación pretendo obtener y conocer el diccionario de inglés-español de los alumnos de inglés avanzado del CELE con el fin de conocer el léxico en la lengua materna y en una lengua extranjera que conoce esta comunidad de habla en particular. El léxico es el grupo de palabras del que dispone un hablante en su lengua y al ser éste parte esencial del presente estudio primero defino lo es una palabra de que acuerdo a Luis Fernando Lara en su *Curso de lexicología* (2006). Después refiero el léxico y los tipos de léxico que existen para indagar específicamente en el disponible de los alumnos del CELE refiriéndome al Modelo Gravitacional de Daniel Gile y la diferenciación que hace del léxico en las zonas activa y pasiva. Al final, muestro las primeras nociones que se tuvieron sobre la disponibilidad léxica, grados de frecuencia y centros de interés a raíz del trabajo publicado por Gougenheim y Michéa en Francia en la década de 1950.

1.1 ¿Qué es una palabra?

Es relevante considerar lo que Lara entiende por la palabra *palabra* al ser una mejor manera de comprender el área en la que se inscribe esta investigación y por ende de los estudios de léxico en general —lexicología—. Elegí su definición por ser una aportación valiosa a las discusiones que existen en las teorías lexicológicas sobre esta. Tradicionalmente consideran que una palabra es aquella que gráficamente se encuentra entre dos espacios blancos, definición que se circunscribe

al lenguaje escrito únicamente y que carece de un sustento teórico que la explique en sus diversas formas de expresión.

Lara dice que los hablantes tienen una noción de palabra, ellos son capaces de decir palabras, en el caso de que se les pregunte ¿qué es una palabra? Lara señala que en primer lugar una palabra es:

Toda unidad de denominación cuya forma de la expresión consta de un número entero de sílabas y cuyos límites pueden estar definidos por la función demarcativa de algunos de sus fonemas, y cuya forma del contenido consta de un morfema léxico, ligado en cohesión estrecha con otro(s) morfema(s) léxicos para construir núcleos morfemáticos, y con morfemas pertenecientes a paradigmas cerrados de carácter flexional, conjugacional o derivativo. (Lara, 2006:138)

También puede definirse como:

Todo signo deslindado posteriormente entre las unidades de denominación, mediante una cohesión débil o inexistente, perteneciente a paradigmas cerrados de carácter sintáctico. (Lara, *op. cit.*)

Ambas definiciones son producto de una indagación lingüística minuciosa que se mueve en diferentes niveles de análisis y que cumplen con reglas y condiciones específicas en los planos fonológico, semántico, morfológico y sintáctico.

En el *Curso de Lexicología*, pareciera que existen diferentes tipos de palabra dependiendo el nivel de análisis que se lleva a cabo; una fonológica, otra morfológica y otra semántica. El primer nivel de análisis del que habla Lara surge casi naturalmente cuando los hablantes escuchamos por primera vez una lengua desconocida y realizamos una segmentación de continuos sonoros con el fin de identificar elementos que, al repetirse, son considerados compositivos de dicha lengua. (Lara, 2006: 21). Este nivel se conoce como el de la *inspección fonética* del cual "tomamos en cuenta las peculiaridades acústicas y perceptuales del sonido de las lenguas, [para poder] hacer una detallada clasificación de los sonidos y sus combinaciones en la

lengua estudiada, aun sin saber, todavía, ni cuáles son, como unidades, ni lo que ayudan para averiguar lo que se quiere decir". (Lara, 2006:22)

El *fonológico* es el segundo nivel de análisis que se lleva a cabo por medio de la identificación de fonemas que existen en una lengua y cuyas combinaciones dan origen a un cierto número de palabras en dicha lengua. Las relaciones estructurales que se establecen entre los fonemas son una primera condición para determinar una palabra en cada lengua. (Lara, 2006:28) La definición que Lara hace de la palabra fonológica es “una sucesión de fonemas de la lengua en cuestión, organizados en sílabas, según las restricciones fonéticas (articulatorias y acústicas) que determina su estructura fonológica”. (Lara, 2006:34) La función demarcativa de la primera definición de la palabra *palabra* se da en este nivel de análisis y se explica como la combinación de fonemas que igualmente delimitan las palabras en las lenguas.

Tanto el primero como el segundo nivel de análisis cuentan con condiciones específicas para determinar lo que es una palabra, pero como dice Lara, no es suficiente. Lo mismo sucede con el *acto de denominación*, el cual pertenece al tercer nivel de análisis lingüístico, ya que según el lexicólogo la palabra no es una simple secuencia de caracteres escritos separada por espacios delimitados y fijos. De hecho, la segmentación y cohesión morfológica entre morfemas constitutivos en el latín dependía del conocimiento lingüístico que se tenía de una lengua particular. Los copistas latinos comenzaron a dividir las palabras haciendo un esfuerzo por reconocer unidades de denominación. Lara apunta que “la combinación del análisis gramatical, que ayudaba a reconocer, sobre todo, paradigmas de morfemas de inventario cerrado, junto con el reconocimiento de unidades de denominación, habría de conducir a una determinación final de la unidad palabra en la escritura”. (Lara, 2006:130)

Esta exploración de la segmentación escrita de la palabra y la cohesión con otros morfemas para construir núcleos morfológicos se encuentra en el cuarto nivel de análisis: el

morfológico y que está ligado con el quinto nivel, el de la *lexicología*. Desde el cuarto nivel, Lara señala que ya es posible definir una unidad de denominación de acuerdo a los morfemas que la constituyen y que aparecen ordenados en una secuencia determinada. Al respecto, el lingüista dice:

Para encontrar el esquema morfotáctico de una unidad de denominación y decidir si se trata de una unidad palabra hace falta: a) segmentar la secuencia entre morfemas [...]; b) probar la cohesión que hay entre ellos y c) determinar el orden en que aparecen. (Lara, 2006:66)

Una vez más, Lara expone condiciones que deben cumplirse en la identificación de una “palabra”. El quinto nivel de análisis es el de la *lexicología* y después está la *sintaxis*. Cabe aclarar que no ahondo en el nivel sintáctico debido a que el presente estudio se enfoca en la palabra únicamente.

1.2 ¿Qué es un vocablo?

Otro de los conceptos esenciales en la investigación lexicológica es la palabra *vocablo*, el cual es entendido como:

Unidad de cita ³ que representa todo el paradigma de flexiones, derivaciones o conjugaciones que se forman a partir de una raíz o de un vocablo morfemático y que se reconocen como palabras (Lara, 2006:138)

El vocablo forma parte del método en una investigación porque logra definir los paradigmas que existen en una lengua. Es un “término de descripción y primera teorización que sirve como instrumento abstracto de la investigación lingüística y de la descripción lexicográfica” (Lara, 2006:139). Dicha edición transforma en una medida el léxico que conocen los estudiantes

³ Una unidad de cita es la representante del paradigma que forma la raíz o el núcleo morfemático de la palabra y los morfemas ligados en cohesión estrecha con ellos, sobre la base de la unidad de denominación. (Lara: 2006, 137)

de inglés del CELE. Por tal motivo, una de las preguntas de investigación de este trabajo tiene que ver con el proceso de edición, considerándolo como una primera etapa de estudio del léxico disponible de dichos alumnos.

En esta investigación se ha elegido editar las palabras de las encuestas de léxico disponible al ser lo que tradicionalmente se hace en los estudios de este campo a nivel mundial. Aunque existen otras propuestas, como la de Frey Pereyra,⁴ para diferenciar lo que es una palabra y un vocablo, nosotros elegimos la definición de Lara porque sustenta los criterios de edición elegidos en la realización de este estudio. Tomamos en cuenta las unidades de cita sin flexiones o derivaciones, es decir y con base en la definición de Lara, únicamente los vocablos. De tal manera que si la palabra *manos* se encuentra en el centro de interés *Partes del cuerpo* solamente se considera el vocablo *mano*. Dicha noción —la de *vocablo*— también facilitó la creación de los diccionarios de léxico disponible que recolectamos por medio de las encuestas para ser posteriormente ingresadas a LEDICAL.

1.3 ¿Qué es el léxico?

En el Diccionario de la Real Academia Española el léxico es definido como “el conjunto de las palabras de un idioma, o de las que pertenecen al uso de una región, a una actividad determinada, a un campo semántico dado, etc”. De acuerdo a esta definición, las palabras que conocen los hablantes de una lengua comprenden el léxico de una lengua, el cual es adquirido de muchas formas e implica cuestiones sociales, que tienen que ver con la experiencia de los hablantes al momento de la adquisición. En el capítulo cuarto, “El significado de la palabra”, del Curso de

⁴ Para ella, las palabras son todas las que “aparecen en un contexto determinado sin importar si son iguales o no entre sí. Por ejemplo, el grupo: “casa, casa, edificio” está conformado por tres palabras. Vocablos en cambio nos servirá para designar palabras distintas entre sí; de manera que en el grupo: “casa, casa, edificio” contamos sólo con dos vocablos”. (Frey Pereyra, 2008:81)

lexicología, se menciona que en los estratos de formación de significados (prototipo, estereotipo, cultura y término) se involucran objetivaciones lingüísticas, experiencias y percepciones, acciones verbales de una comunidad, universales y al mismo tiempo una cultura verbal. Por lo tanto, existen diversas formas de investigación léxica que resultan vastísimas y, por ende, dadoras de información lingüística, psicolingüística, social, cultural y neurolingüística.

Los alumnos que completaron las encuestas de léxico disponible evocaron respuestas determinadas a partir de su propia experiencia. Por lo tanto, las encuestas de léxico pueden considerarse no sólo evidencias lexicológicas de una comunidad de habla en particular sino también muestras escritas que van más allá de lo meramente lingüístico y que pueden ser analizadas desde diferentes perspectivas (etnográficas, psicolingüísticas y de enseñanza-aprendizaje de segundas lenguas, entre otras). Yo me concentro en los estudios de léxico disponible porque son una forma de conocer el léxico de una comunidad particular. El conjunto de palabras disponibles obtenido con las encuestas forman los diccionarios de léxico disponible tan relevantes para esta investigación. A continuación me gustaría aclarar lo que se entiende por léxico disponible y entonces abordar ciertos términos básicos que Lara hace sobre los tipos de léxico que existe. Esto con la finalidad de relacionarlo con las zonas activa y pasiva que menciona Daniel Gile en su Modelo Gravitacional.

1.4 Tipos de léxico

1.4.1 Léxico disponible

El léxico disponible consiste en las palabras que vienen a la mente del hablante en el momento de la enunciación y bajo un contexto comunicativo muy específico. El léxico disponible son las “unidades que, sin ser particularmente frecuentes, son evocadas y empleadas en ciertas

situaciones comunicativas para hablar de determinados asuntos. Una palabra es disponible si acude de manera fácil y rápida a la mente del hablante cuando se trata un tema específico” (Samper, 2002), como lo veremos más adelante cuando abordemos el grado de disponibilidad y los centros de interés.

1.4.2 Léxico pasivo y léxico activo

Éstos están en constante transformación dependiendo las necesidades del hablante y los momentos de comunicación en que participa. Luis Fernando Lara dice que el vocabulario activo es el “que domina el hablante, que maneja espontáneamente y con el que se desempeña en su vida cotidiana.” (Lara, 2006:149) Este vocabulario comprenderá miles de palabras y es aquel que se ha ido adquiriendo desde la infancia. Hay variabilidad en la cantidad de vocablos, pero gira alrededor de unas dos mil a ocho o nueve mil palabras. Dicha cantidad también varía con la profesión de un hablante. Un médico o un escritor no manejará la misma cantidad de vocablos y de hecho el léxico del que se valen para ejercer sendas carreras es diferente.

El vocabulario pasivo, por otro lado, comprende ciertos vocablos que “quedan grabados en su memoria y le permiten comprender nuevas conversaciones, nuevos textos, con poco esfuerzo”. (Lara, 2006:150) En otras palabras, es el vocabulario que no se utiliza de manera activa y que surge espontáneamente dependiendo, una vez más, de las necesidades comunicativas del hablante.

Lara también dice que el vocabulario activo y pasivo de un sujeto puede variar porque depende de una gran cantidad de factores:

Es un fenómeno de la memoria, que puede estar influido por muchos factores, como la edad, el sexo, la educación, el conocimiento recibido de la tradición verbal, la especialidad de cada quien y hasta, quizá, por flujos represivos del inconsciente, pero manifiesta la disponibilidad de vocablos en la memoria individual. (Lara, 2006:151)

En la siguiente sección describo el Modelo Gravitacional de Daniel Gile, el cual confirma las nociones de Lara respecto al vocabulario activo y pasivo que conocen los hablantes de una lengua.

1.5 Modelo Gravitacional de Daniel Gile

Existen diversas perspectivas desde las que se describen las características y situaciones comunicativas en que se usan los diferentes tipos de léxico y que buscan comprender el lexicon mental de los estudiantes de lenguas. En el Modelo Gravitacional de Daniel Gile, publicado en su libro *Basic Concepts and Models for Interpreter and Translator Training*, se proponen algunas teorías sobre el almacenamiento, adquisición y recuperación del léxico. Asimismo, se explica el acceso al lexicon mental y conocimiento de estructuras gramaticales de traductores e intérpretes, cuyo proceso de adquisición de la lengua meta es presumiblemente semejante al de los alumnos estudiantes de una lengua extranjera.

Para Gile (2009) su modelo representa el estatus de dominio oral y escrito que un individuo tiene de la lengua en un momento dado y en circunstancias particulares, describiendo la disponibilidad⁵ relativa de las unidades léxicas y reglas lingüísticas. El modelo contempla una parte variable y otra invariable; esta última se refiere a los elementos de la lengua cuya disponibilidad se asume constante o que varía muy lentamente. Esto aplica a las reglas gramaticales más básicas, como son las conjugaciones verbales o a la formación de plurales, así como a un pequeño número de la mayoría de las palabras de la lengua que son más frecuentemente utilizadas. La parte variable es más grande que la anterior porque hay varios niveles de magnitud, ya que incluye al menos decenas de reglas y cientos de palabras y frases

⁵ Gile usa la palabra “*availability*” en inglés, la cual es traducida como disponibilidad en español. Sin embargo, él no se refiere al término disponibilidad léxica propia de los estudios de léxico disponible sino a un significado amplio de la palabra.

idiomáticas. La parte variable está constituida por palabras y reglas girando en órbitas alrededor del núcleo. Las orbitas del modelo contemplan la disponibilidad léxica de los alumnos dependiendo la lejanía o cercanía de las palabras con respecto al centro. Dichas órbitas representan los grados de disponibilidad de las palabras. Mientras más lejana está una órbita del centro, mayor capacidad de procesamiento y de tiempo se requiere para acceder a las palabras de la órbita. (Gile, 2009:279)

Lo anterior explica que los grados de disponibilidad determinan el carácter activo o pasivo que tiene el léxico de un hablante. En la imagen del Modelo Gravitacional se pueden identificar dos zonas: la activa y la pasiva, que explico a continuación:

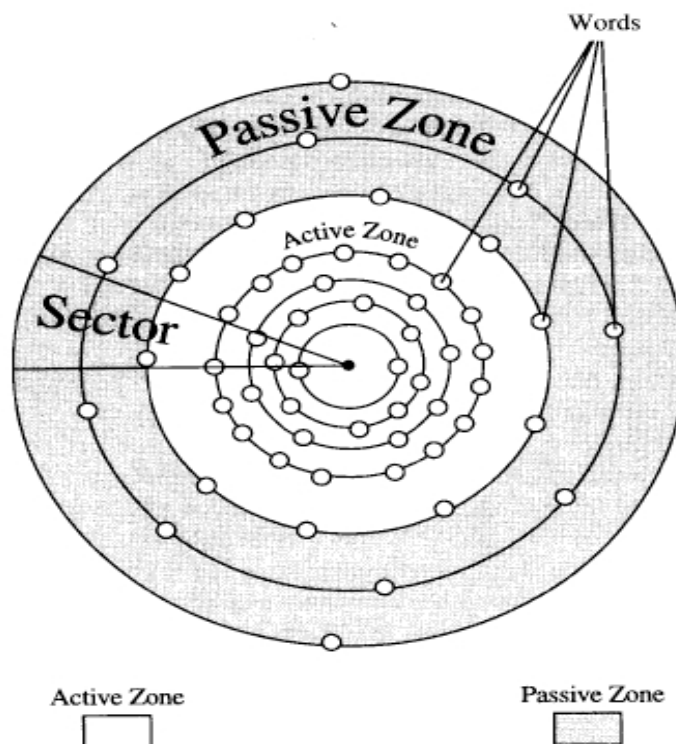


Imagen 1. Ilustración del Modelo Gravitacional de Daniel Gile Copyright de John Benjamins B.V. (Gile, 2009:279).

El léxico que se aloja en la zona activa del modelo tiene un conocimiento “de procedimiento”. Este vocabulario es conocido por el alumno tanto en su forma y significado como en el uso que le da. El léxico que se encuentra en la zona pasiva, que es aquel que no es común en el uso cotidiano, pero que es conocido por el hablante y que puede comprenderse y repetirse, es de tipo “declarativo”; porque no puede usarse para expresarse.

Tanto Lara como Gile coinciden en que la variabilidad del contexto es una constante en el conocimiento del vocabulario de los alumnos. Este carácter del léxico fue estudiado por Gougenheim y Michéa, quienes buscaron explicar cómo dicha variabilidad registra cuantitativamente el grado de disponibilidad que se obtiene del léxico de los alumnos. De ello surgieron los conceptos: grado de disponibilidad y centros de interés que a continuación describo brevemente.

1.6 Grado de disponibilidad y Centros de interés

El concepto de grado de disponibilidad surgió a partir de un estudio realizado por Michéa, del cual pudo reconocerse que el interés y la atención son la base de la memorización de términos particulares. Los participantes recibieron la instrucción de hacer una lista de veinte palabras que hayan pensado tras la lectura de un pasaje de *Tonnelier de Nuremberg* de *E. T. A. Hoffmann*. En el análisis de los resultados Michéa se dio cuenta de que las palabras concretas son las que tienen un mayor grado de estabilidad, seguidas de los verbos y al final de las gramaticales, como son las preposiciones o las conjunciones, por ser poco mencionadas y, por lo tanto, menos estables.

La comparación de la dispersión de las palabras con base a su clasificación, es decir, conceptuales o gramaticales, afirmaron lo que Michéa ya había concluido en la investigación realizada antes en Périgueux. Decía que:

En presencia de una situación determinada, las palabras que primero vienen a la mente son aquellas que están totalmente vinculadas a una situación y la caracterizan, es decir, los nombres. Las otras partes de la oración vienen a la mente en la medida en que buscan expresar la relación entre las cosas. La memoria, basada en la asociación de ideas, es selectiva: no obedece una ley probabilística simple. (Gougenheim y Michéa, *op. cit*)⁶

Las encuestas mostraron que la asociación de ideas en torno a un centro de interés confirma la estabilidad de los sustantivos concretos y la inestabilidad de los verbos. Dicho ejercicio asociativo es realizable haciendo uso de los centro de interés antes mencionados, sobre todo en la recolección de palabras concretas. El léxico disponible es precisamente reconocido como aquel utilizado en situaciones comunicativas específicas y su grado de disponibilidad es obtenido por medio de ejercicios de dispersión matemática realizados actualmente por programas computacionales diseñados exclusivamente para ese fin (en este caso, LEDICAL) y cuyo resultado es estable de acuerdo a una situación comunicativa determinada, o bien, a un centro de interés particular.

Ahora bien, la fórmula que calcula el índice de disponibilidad de cada palabra fue inventada por los lingüistas López Chávez y Strassburger Frías (1992), por medio de la cual se computa un valor disponible basada en las posiciones que la palabra ocupa en una lista determinada, el número de participantes que escriben las palabras en esas posiciones y la última posición que la palabra ocupa en cualquiera de las listas. En otras palabras, la fórmula asigna un valor más alto a las palabras que fueron escritas por un mayor número de participantes y que aparecen primero en las listas; y exponencialmente un valor menor a las palabras escritas por menos participantes y que aparecen al final de las listas. (Hernández-Muñoz, *et al*, 2006:736)

⁶ En présence d'une situation donnée, les mots qui viennent les premiers à l'esprit sont ceux qui sont liés tout spécialement à cette situation et la caractérisent, c'est-à-dire les noms. Les autres parties du discours ne se présentent à l'esprit que dans la mesure où il s'agit d'exprimer les rapports entre les choses. La mémoire, basée sur l'association des idées, est sélective : elle n'obéit pas à une loi probabilitaire simple. (Gougenheim y Michéa, 1967:151)

Los noción de centros de interés

La invención de los centros de interés se entiende a partir de lo encontrado por los lingüistas franceses en el experimento de Périgueux, cuya necesidad es de primer orden en los estudios de léxico disponible y que fungen como una manera de dar al sujeto de estudio las situaciones comunicativas específicas con el fin de obtener el léxico disponible en particular. Cabe aclarar que no existe una definición específica que defina este concepto, lo importante es comprender que dichos centros de interés fueron creados para evocar en la mente del hablante situaciones comunicativas diversas con el fin de que hiciera uso de un tipo de léxico particular.

Los términos de activo y pasivo de Lara y de Gile tiene la misma orientación en tanto que comprenden el léxico activo como aquel que viene de manera fácil a la mente del hablante y es de uso cotidiano. Este se diferencia del pasivo porque no surge espontáneamente en el habla del alumno y aunque lo conoce no lo usa para expresarse y comunicarse cotidianamente. La distinción entre los dos tipos de vocabulario me hizo pensar en los estudios de léxico disponible y su herramienta de recolección de datos: las encuestas de léxico disponible. En ellas existe una gran cantidad de léxico activo y pasivo que se mezcla en un mismo centro de interés.

En el de *vocabulario especializado*, por ejemplo, se espera que el léxico sea predominantemente pasivo ya que los alumnos pocas veces usan vocablos especializados en la vida cotidiana, como *DNA, pelvis, assets, debts, narrator*, entre otras. Sin embargo, existen otros como *cabeza, mano, bus, dog* o *metro* que pueden considerarse activos porque son comunes en el

uso cotidiano⁷ que los estudiantes hacen de su lengua materna o bien, de la lengua que están estudiando, y al mismo tiempo, pasivos porque vienen a la mente en contextos determinados.

Otro aspecto importante en el cual coinciden los dos lingüistas es el de la variabilidad del contexto porque influye en el dinamismo del tipo de léxico que se recaba y esto se observa en las encuestas de léxico disponible de los alumnos avanzados de inglés del centro. Con base en lo anterior, podría afirmar que el léxico disponible es primordialmente pasivo precisamente porque las situaciones comunicativas en que fue recabado (por medio de los centros de interés) no es una constante en la cotidianidad del alumno. Esto sucedería tanto en la lengua materna como en la meta porque en el caso de la primera el dominio de su conocimiento está por encima del uso y simple reconocimiento de palabras por campos semánticos, como lo hace un niño en las primeras etapas de la adquisición de una lengua. En el caso de la lengua meta pasa lo mismo porque en los planes de estudio y en los libros de texto de niveles avanzados el léxico no se estudia por campos semánticos sino por tareas comunicativas generales y poco especializadas.⁸

En esta investigación estudio el tipo de léxico diferenciando las palabras disponibles que se encuentran en la zona activa del modelo y las que están en la zona pasiva. Sería aventurado afirmar que el léxico disponible de estos sujetos de estudio se encuentra en un cien por ciento en la zona pasiva del modelo; no así, decir que las palabras con una disponibilidad mayor son las ubicadas en la zona activa y aquellas con índices menores y en el último orden de aparición en las listas son las encontradas en la zona pasiva de acuerdo a Lara y Gile.

⁷ Hablar del uso cotidiano que los universitarios hacen de su lengua materna es demasiado amplio. Sin embargo, se espera que todos compartan usos lingüísticos determinados que dependen de su edad, nivel académico y ciudad donde viven.

⁸ Cabe aclarar que espero que no sea éste el caso del centro de interés *vocabulario especializado* precisamente porque su objetivo es ése; el de aprender palabras especializadas en situaciones comunicativas determinadas, tales como en el trabajo, la lectura de un artículo o durante una junta entre colegas.

Capítulo 2 - Enseñanza y Aprendizaje de Vocabulario

En este capítulo presento las principales corrientes de investigación y métodos de enseñanza del léxico en primeras lenguas y lenguas extranjeras, así como algunas propuestas de unidades didácticas enfocadas al léxico disponible.

Actualmente en el área de adquisición de lenguas se reconoce el papel central del léxico en el aprendizaje de una lengua. Sin embargo, trabajos como el de Boyd Zimmerman (1997) muestran que la tendencia histórica en la enseñanza de vocabulario en segundas lenguas advierte que el léxico ha sido poco valorado y hasta hace poco no había sido objeto de investigación. Zimmerman considera que es fundamental para la lengua y de vital importancia para el estudiante de lenguas. El lingüista continúa afirmando que la enseñanza y aprendizaje del vocabulario ha sido menospreciado en el área de adquisición de segundas lenguas a lo largo de sus etapas y hasta el día de hoy. Los investigadores y maestros han privilegiado la sintaxis y la fonología para teorizar, considerándolos centrales en la teoría lingüística y más importantes en la pedagogía. Boyd Zimmerman (1997:5) atribuye este vacío a las tendencias seguidas por la investigación lingüística, más centrada en el estudio de la sintaxis, la gramática y la fonología. La propia lingüística aplicada olvidó la enseñanza del vocabulario a favor de la sintaxis.

2.1 Tratamiento del léxico en los enfoques de enseñanza de lenguas

Cuando se revisan las principales metodologías seguidas en la enseñanza de lenguas se puede comprender la razón del por qué el estudio del léxico fue ignorado durante tanto tiempo.

En el método de Traducción gramatical, primer enfoque de enseñanza de lenguas en escuelas públicas en Prusia durante el siglo XVIII y hasta inicios del siglo XX, estableció como

objetivos primarios la preparación de los estudiantes para leer y escribir textos clásicos y para pasar exámenes estandarizados. Los estudiantes recibían explicaciones gramaticales en su lengua nativa, paradigmas para memorizar y listas bilingües de palabras para aprender. Este conocimiento los preparaba para hacer la traducción de largos pasajes de los clásicos. El énfasis se ponía en el análisis de la estructura sintáctica y la conjugación de verbos. El vocabulario ilustraba reglas gramaticales y para cualquier problema se recurría a las raíces griegas y latinas, dando especial atención a los étimos y derivativos porque se creía que de esa manera se evitaba la degeneración de la lengua. La enseñanza del vocabulario se basaba en la definición y en la etimología y en listas bilingües de palabras organizadas por campos semánticos. Los diccionarios bilingües también eran herramientas de uso frecuente que se usaban como recursos didácticos. Y conforme se avanzaba en el conocimiento de las familias de lenguas y en el cambio del proceso natural de las lenguas se daba atención gradualmente al uso de cognados. Se inicia un movimiento de reforma que daba atención al vocabulario de acuerdo a su frecuencia. El trabajo de los lingüistas como T. Prendergast (1864), quien elaboró una lista de palabras frecuentes, fue una innovación importante que trajo léxico de uso cotidiano a la enseñanza de lenguas.

En el Movimiento Reformista en Inglaterra se encuentran lingüistas como Henry Sweet (1899) a la cabeza, quien insistía que el fracaso del método anterior se debió a que fue basado en un conocimiento deficiente de la ciencia del lenguaje. Por lo tanto, los reformistas enfatizaron la importancia del lenguaje hablado y el entrenamiento fonético. Se redefinió el concepto de fluidez como la habilidad de pronunciar con precisión fragmentos y de mantener asociaciones entre discursos y referentes del mundo. El método propuesto por Sweet fue representativo de su tiempo. En éste, los alumnos estudiaban fonética y transcripción primero y continuaban con la etapa gramatical, en la cual estudiaban gramática y un vocabulario muy básico para después seguir con la tercera etapa, la idiomática. Aquí estudiaban el vocabulario a profundidad y el

estudio filológico era realizado posteriormente en las etapas cuatro y cinco (Literaria y Arcaica), pero estaba destinado exclusivamente al trabajo universitario.

Henry Sweet también dejó de lado la enseñanza del léxico porque las clases era basada en un lenguaje hablado muy controlado, en el cual se evitaba enseñar listas de palabras separadas y oraciones aisladas. Solamente después de que un texto era estudiado en detalle es que se debían aislar los puntos gramaticales y los elementos léxicos con fines didácticos. Él creía que desde un punto de vista práctico y también científico la oración es la unidad de lenguaje y no la palabra porque aunque el lenguaje está formado por palabras, nosotros no hablamos en palabras sino en oraciones.

No obstante las ideas que Sweet tenía sobre la enseñanza del léxico se puede ver que los reformistas realizaron un cambio importante con respecto a los métodos anteriores. Las palabras fueron asociadas con la realidad, no así con otras palabras o ciertos patrones sintácticos. El vocabulario era elegido de acuerdo a su simplicidad y utilidad. De hecho, Sweet comenzó a discutir sobre las posibilidades de desarrollar listas de vocabulario basándose en medidas estadísticas (Kelly, 1969).

Hacia finales del siglo XIX se introdujo el Método Directo, uno de los métodos “naturales” más conocidos, que se benefició de los debates propuestos por el movimiento reformista. Su nombre surge de la prioridad de relacionar el significado directamente con la lengua meta sin pasar por la traducción. Fue desarrollado por Sauveur (1874) en Estados Unidos y se hizo famoso por Berlitz. En este método se usaba la lengua meta como lengua de instrucción en grupos pequeños y clases intensivas, las cuales consistían en una cuidadosa progresión gradual de preguntas y respuestas, así como el uso de vocabulario y oraciones cotidianas. En el Método Directo se creía que la interacción era importantísima en la adquisición natural de una lengua. Se

critica este método por la sobre simplificación de las similitudes entre la L1 y la L2 y su falta de consideración en la organización de las clases públicas.

En su manual para maestros de 1874, intitulado *An Introduction to the Teaching of Living Languages without Grammar or Dictionary*, Sauveur propuso dos principios en la enseñanza de lenguas que contemplan la selección de vocabulario y la instrucción que recibían los alumnos de los profesores. El primer principio consistía en solicitar a los maestros que realizaran preguntas auténticas con el fin de que los alumnos ofrecieran respuestas genuinas a sus cuestionamientos, y en segundo lugar que las preguntas estuvieran relacionadas entre sí con el fin de que las ideas fueran hilándose una tras otra promoviendo el aprendizaje en contexto.

Cuando los cursos de Berlitz comenzaron a impartirse se incluían palabras simples y familiares, tales como objetos del salón de clases, ropa o las partes del cuerpo, seguidos del verbo *to be* y los adjetivos comunes. El vocabulario concreto era explicado con fotos y demostraciones mientras que el vocabulario abstracto era enseñado por medio de la asociación de ideas. Se hacía uso de cuadros, dibujos y objetos y surgieron los términos *realia* y *realien* adoptados en esos años. Este método fue muy criticado y nunca fue adoptado en escuelas norteamericanas ni europeas; sin embargo, ganó un gran número de seguidores en las escuelas Berlitz.

Alrededor de las décadas de 1920 y 1930 surgieron en Estados Unidos y Gran Bretaña respectivamente, los métodos de Lectura y el de Enseñanza Situacional de lenguas (Zimmerman, 1997:9). El objetivo principal del método de Lectura era el desarrollo de las habilidades lectoras de los estudiantes norteamericanos, quienes, según un reporte de Coleman en 1929, tenían deficiencias serias en la comprensión de lectura. En Gran Bretaña, Michael West (1930) consideró necesario subrayar la necesidad de facilitar las habilidades lectoras mejorando las habilidades léxicas. Él decía que la adquisición del vocabulario en el aprendizaje de una lengua era esencial y que los seguidores del método directo habían recalcado la importancia del discurso

—*speech*, en inglés— sin ofrecer los lineamientos adecuados al seleccionar los contenidos de los libros de texto. La adquisición del vocabulario comenzó a reconocerse como el elemento principal en el aprendizaje de una lengua extranjera; sin embargo, no existía un concepto que explicara lo que era el vocabulario.

Los alumnos que llevaban más de tres años estudiando una lengua extranjera, decía West, no cuentan ni siquiera con un ciento de palabras básicas. Él fue uno de los primeros en recomendar el uso de la frecuencia léxica para seleccionar las palabras y ordenar el vocabulario en los materiales estudiantiles. En 1930 recomendó el uso de la lista de frecuencia léxica de Thorndike y en 1953 publicó *A General Service List of English Words*. Aunque no ha sido actualizada desde 1936 sigue siendo considerada como la lista de frecuencia léxica más utilizada en lengua inglesa.

En Estados Unidos y Gran Bretaña, H. E. Palmer y A.S. Hornby (1916), considerados líderes del movimiento de Enseñanza Situacional de Lenguas, creían que la lengua debía enseñarse practicando estructuras básicas en actividades significativas situacionales. Consideraban que el discurso era la base y la estructura que hacía posible el discurso (Zimmerman, 1997:10). Su principal propósito era desarrollar bases científicamente más sólidas para los métodos orales ya dados a conocer por los metodologistas directos. Por primera vez el vocabulario fue considerado uno de los aspectos más importantes en la enseñanza de segundas lenguas. También se le dio prioridad al desarrollo de una base científica y racional en la selección de vocabulario de los cursos de lenguas. Las investigaciones de Palmer y Michael West (1930) dieron inicio al desarrollo de principios de enseñanza en el control del vocabulario. Ellos fueron los primeros en establecer principios específicos en el diseño curricular de la enseñanza de lenguas.

El Método Audiolingual fue desarrollado por lingüistas estructuralistas estadounidenses durante la Segunda Guerra Mundial cuando los apoyos gubernamental e institucional existían en la enseñanza de lenguas extranjeras. Su fundador Charles Fries (1945) lo consideraba un enfoque nuevo en la pedagogía de la gramática y fue utilizado por primera vez en la Universidad de Michigan (Zimmerman, 1997:10). Él decía que los principales problemas a los que se enfrentaban los estudiantes de lenguas extranjeras se relacionaban con las diferencias que existen en los sistemas estructurales. Por lo tanto, en el método creado por Fries los alumnos aprendían los puntos gramaticales a través de ejemplos y ejercicios (*drills*) de ciertos patrones básicos de la oración enfatizando la pronunciación y los ejercicios intensivos. Su curso de tres meses proponía el estudio intensivo de estructuras esenciales de la lengua inglesa.

El vocabulario era seleccionado de acuerdo a la simplicidad y familiaridad, debido a que el objeto principal de estudio en la enseñanza de lenguas era la adquisición de patrones estructurales. Las palabras nuevas eran incluidas en los ejercicios pero sólo las suficientes para hacer factibles los ejercicios. Se asumía que las estructuras podían complementarse posteriormente cuando los estudiantes fueran mucho más conscientes de sus necesidades léxicas en situaciones particulares. Las creencias que tenía Fries sobre el aprendizaje del léxico fueron expuestas en 1945 en un texto que escribió sobre el tema. Decía que existía una sobre simplificación del vocabulario atribuida a tres falsas suposiciones que dictaban que las palabras tienen equivalencias exactas en diferentes lenguas, que una palabra tiene significado único cuando una palabra puede tener varias acepciones, y que una palabra solo cuenta con un significado “real” y “único” desacreditando los otros significados como figurativos o ilegítimos. Fries sostenía que las palabras son símbolos que derivan la totalidad de su contenido y sus limitaciones de significado a partir de las situaciones en que son usadas (Fries, 1945).

La enseñanza y aprendizaje de vocabulario habían sido ignorados durante mucho tiempo y en este momento comienza a adquirir un mayor peso en la planeación de cursos y selección de materiales. Sin embargo, también se creía que el temprano aprendizaje de una gran cantidad de vocabulario en el proceso de adquisición de una lengua daba a los estudiantes una falsa seguridad. Al respecto, Wilga Rivers comenta en un texto publicado en 1968, intitulado *Teaching Foreign Language Skills* que cuando los alumnos aprendían un exceso de vocabulario en las etapas tempranas del aprendizaje se creaban una impresión falsa de que lo más importante en el aprendizaje de una lengua tiene que ver con la acumulación de nuevas palabras como equivalentes a conceptos, los cuales ya expresan en su primera lengua. Con frecuencia no se dan cuenta de que el significado es expresado en grupos de palabras y combinaciones de segmentos de la lengua, y que el significado de una sola palabra es comúnmente difícil de determinar cuando es separada de contexto de otras palabras o frases. (Rivers, 1981:254)

Ella recomendaba que el vocabulario nuevo fuera presentado en actividades orales y que las palabras fueran reutilizadas con mucha frecuencia para asegurar una retención a largo plazo. Los alumnos debían practicar variaciones morfológicas y estructuras sintácticas con palabras conocidas para que de esa manera no se distrajeran de las estructuras meta. Rivers cambió de opinión posteriormente pero no fue el caso de Freeman Twaddell, colega de Fries. En 1980, él aseguraba que la adquisición léxica fue sobrevalorada por los alumnos, los maestros y teóricos de la lengua quienes reaccionaron enfatizando el rol de la gramática por encima del léxico. Esta noción podía observarse en el currículo y los materiales didácticos que presentaban el léxico como el medio para ilustrar temas gramaticales en lugar de valorarlo como un elemento comunicativo. Las ideas de Twaddell respecto al léxico dieron cabida a la enseñanza de habilidades compensatorias, tales como adivinar el significado de las palabras y tolerar la ambigüedad.

La Enseñanza Comunicativa de las Lenguas, propuesto por Dell Hymes (1972), surge como una reacción en contra de la noción chomskyana de competencia lingüística autónoma. La publicación del libro Estructuras Sintácticas de Noam Chomsky en 1957 planteó por primera vez la idea de que la lengua está representada en la gramática mental del hablante por medio de un grupo de reglas que se reflejan en una intuición inconsciente que posee el hablante sobre la lengua y que no refleja las creencias que tiene sobre su uso. Sin rechazar el modelo de Chomsky, Hymes creó el concepto de Competencia Comunicativa, en el cual enfatizó los factores sociolingüísticos y pragmáticos que rigen el uso efectivo de la lengua. La competencia comunicativa es entendida como el conocimiento interiorizado de una apropiación situacional de la lengua (Hymes, 1972), y en ella se incorpora la competencia lingüística con un sentido de creatividad lingüística. El resultado fue, una vez más, un cambio total en la enseñanza de lenguas extranjeras, cuyo eje central se relacionó más con el desarrollo de una competencia comunicativa en lugar de un dominio de estructuras gramaticales.

En este método tampoco se le dio importancia al vocabulario sino a la apropiación de categorías comunicativas y al discurso lingüístico. En 1972, el lingüista británico David Wilkins desarrolló las categorías nocionales y funcionales con el análisis de significados comunicativos y que más tarde fueron utilizados para desarrollar los programas del Consejo Europeo. Las categorías nocionales y funcionales de Wilkins fueron basadas en estándares temáticos y situacionales, por lo que se ha sugerido que se determinaron más por la semántica que por la sintaxis. Estos programas marcaron un parte aguas en la enseñanza de lenguas extranjeras y los objetivos de la misma, sin dar mayor importancia al léxico.

Posteriormente se comienza a estudiar el sistema léxico por su complejidad, tal como lo propuso el lingüista norteamericano Edward M. Anthony (1973). Él creía que la complejidad del léxico debía trabajarse en su relación con un contexto cultural, evitando la sobre simplificación y

desconfiando de la traducción como factor relevante en los enfoques, metodologías didácticas o técnicas de enseñanza. En el método comunicativo se ven reflejadas estas ideas basando los contenidos de los cursos en actividades contextualizadas, centrándose en un nivel discursivo más que en uno oracional, dando a los alumnos oportunidades para desarrollar estrategias de interpretación y usando la lengua como lo hacen los nativo-hablantes.

En los materiales comunicativos se han desplazado las frecuencias léxicas debido a los problemas que se han detectado en ellas y que se relacionan con el orden en que aparecen algunas palabras. Se considera que las primeras mil o dos mil palabras incluidas en las listas de frecuencia léxica son las que forman el vocabulario de un principiante. Sin embargo, esto no siempre es así debido a que las primeras palabras que aparecen en dichas listas no siempre son consideradas como las más importantes, o bien, el orden de las listas no refleja un orden ideal de enseñanza del léxico, ya que algunos pronombres posesivos no aparecen en los primeros lugares de frecuencia. Esto supone que las listas de frecuencia no son la mejor guía en la enseñanza de léxico “útil” a los alumnos. En el método comunicativo se cree que el desarrollo del vocabulario ocurre de manera natural en la L1 a través de un lenguaje contextualizado y naturalmente secuenciado, que se desarrollará de manera natural con una exposición comunicativa en la L2.

El Enfoque Natural fue diseñado por Tracy Terrell y Stephen Krashen (1983) a partir de las experiencias personales de ambos dentro del salón de clases en conjunto con la teoría propuesta por Krashen sobre la adquisición de segundas lenguas. Su idea fue principalmente hacer posible que el estudiante principiante alcance niveles aceptables de una habilidad comunicativa dentro del salón de clases. El vocabulario en este enfoque tiene una gran importancia en el proceso de adquisición como dador de significado. El input debe ser siempre comprensible y este carácter depende directamente de la habilidad de reconocer el significado de

elementos claves en una producción. Por lo tanto, la adquisición no ocurrirá sin la comprensión del vocabulario. (Krashen y Terrell, 1983)

En el caso de los estudiantes que están más arriba del nivel básico, Krashen sugiere que la lectura es el medio más eficaz por el cual puede adquirirse nuevo vocabulario. Cree que la lectura voluntaria es la forma óptima de lograr un input comprensible, ya que la lectura provee mensajes comprensibles presentados en una situación de bajo estrés (Krashen, 1989).

La investigación lexicográfica iniciada en la década de 1980 se dio cuenta de la necesidad que había de alcanzar una descripción lingüística más precisa y marcó un parteaguas en el diseño de programas comunicativos y enseñanza de lenguas. El proyecto de la Universidad de Birmingham, conocido como *The Collins-Birmingham University International Language Database* (COBUILD), es un análisis computacional detallado basado en un corpus de 20 millones de palabras, diseñado para dar cuenta del uso actual de la lengua. El corpus está formado principalmente por información escrita y el proyecto ha generado una variedad de diccionarios y materiales didácticos.

Antes de la publicación del diccionario COBUILD, el editor en jefe John Sinclair anticipó el papel que los procesadores computacionales de textos jugarían en la reorientación de la descripción de la lengua y que requeriría de cambios mayores en la instrucción lingüística. Dicha reorientación ha provocado que muchos resignifiquen la naturaleza de la lengua y el rol del vocabulario. El trabajo realizado por medio del análisis de corpus y la lingüística computacional ha despertado gran interés en la importancia de largas “frases” (*chunks*, en inglés) de lengua, conocidos como elementos léxicos, frases léxicas y unidades prefabricadas. En su libro *Lexical Phrases and Language Teaching*, James R. Nattinger y Jeanette S. De Carrico (1992) han analizado grandes muestras de lenguaje para demostrar el papel central de los segmentos léxicos. Ellos usan unidades léxico gramáticas, conocidas como frases léxicas, como la base de su análisis

y afirman que la competencia pragmática se determina por la habilidad que tiene el estudiante de acceder y adaptar frases prefabricadas de la lengua. Michael Lewis pensaba que los elementos léxicos eran centrales para el uso de la lengua y que deberían serlo en la enseñanza de lenguas. Decía que la lengua consiste en el léxico gramaticalizado y no en la gramática lexicalizada. (Lewis 1993) Asimismo, desafió la dicotomía gramático-léxica y, por el contrario, demostró que la lengua consiste en “frases multipalabra”⁹ (*multiword chunks* en inglés). Su sugerencia pedagógica incluía la integración de un enfoque comunicativo centrándose en un léxico surgido naturalmente.

Los trabajos de Sinclair, Nattinger, De Carrico y Lewis (1992) representaron un cambio significativo en la teoría y pedagogía por varios motivos, ya que subrayaron la importancia de una descripción lingüística exacta, además de que cambiaron el punto de vista tradicional de los límites léxicos, enfatizando la necesidad de los aprendientes de lenguas por percibir y usar patrones léxicos y colocaciones como parte de su proceso de aprendizaje. Lo más importante es que existe una aseveración subyacente que afirma que la producción lingüística no es un proceso regido por reglas sino la recuperación de grandes unidades fraseológicas de la memoria. Dicha afirmación transforma las ideas chomskianas sobre la sintaxis, como la base del conocimiento lingüístico internalizado por el hablante. Las implicaciones de este cambio son muchas tanto en la investigación lingüística como en la pedagogía de lenguas extranjeras del futuro.

El repaso histórico realizado por Boyd Zimmerman en un capítulo del libro *Second Language Vocabulary Acquisition. A Rationale for Pedagogy* da cuenta de la poca importancia que se le ha dado a la enseñanza del léxico en las metodologías y corrientes pedagógicas de enseñanza de segundas lenguas. Sin embargo, el surgimiento de listas de frecuencia léxica, en las

⁹ Aunque la palabra *chunks* no puede ser traducida como *frase*, utilizo el término *frases multipalabras* porque así se encuentra en la literatura sobre el tema.

décadas de 1920 y 1930, es un ejemplo de la inquietud que había por conocer el léxico obtenido por los estudiantes de lenguas extranjeras, cuya tradición sigue vigente en la lexicografía actual por medio de proyectos léxicos que recaban listas de vocabulario de uso frecuente entre hablantes de todo el mundo. De hecho, las listas de frecuencia y disponibilidad del léxico recabado en Francia por Gougenheim y Michéa dan origen al diseño de unidades didácticas enfocadas en el léxico. Decidimos presentar los enfoques y métodos de enseñanza de segundas lenguas porque puede verse cómo la enseñanza del léxico se da como un paso evolutivo mediante el cual los maestros y los lingüistas reconocen la utilidad de enseñar vocabulario como objetivo primario en el desarrollo de la interlengua de los alumnos. Trabajos como este, de recolección y descripción de un tipo de vocabulario, ofrecen material léxico rico para el diseño de materiales didácticos y métodos pedagógicos enfocados a la enseñanza del léxico.

A continuación, regreso a los orígenes de los estudios de léxico disponible y a continuación presento dos ejercicios pedagógicos surgidos del *Francés fundamental* de 1956 como una forma de ejemplificar la práctica a partir de la teoría.

2.2 El francés elemental: unidades didácticas

Le français élémentaire contiene lecciones enfocadas al léxico diseñadas para alumnos principiantes de escuelas africanas. En la introducción al libro del maestro, intitulado *El francés elemental*, nuevo curso de lenguas para las clases principiantes en escuelas africanas de A. Davesne (1955), el autor aclara que tuvo el cuidado de adaptar el vocabulario del francés fundamental (1er grado) a las necesidades particulares de su público. Él dejó de lado lo que definió como vocabulario “cívico y social” por ser inaccesible a un público muy joven. La

determinación del vocabulario concreto fue decidido por los modos de vida de las poblaciones africanas a las cuales se dirigía la lección. (Davesne, 1955).

El primer ejemplo de unidad didáctica se llama *Los oficios (Les Métiers)* e inicia con una nota donde Davesne (1955) solicita al profesor que adapte las lecciones del libro al contexto africano (fuera urbano o rural) donde se imparten los cursos de francés. El objetivo de la lección es que los alumnos vayan construyendo oraciones sencillas conjugando el verbo *trabajar* y conociendo el nombre de los oficios en lengua francesa. El maestro inicia con las oraciones: *Yo trabajo, Yo realizo un trabajo*; después se presenta la segunda persona del singular y sigue con la tercera: “*Fatou barre la casa : ella trabaja. Mi madre prepara la comida : ella trabaja. El panadero trabaja en la panadería. El carnicero trabaja en la carnicería. El tendero trabaja en la tienda. [...], etc*” (Gougenheim y Michéa, 1995:257). Posteriormente se formulan preguntas sobre las actividades que hacen el carnicero y el panadero, por ejemplo. La pregunta es: ¿qué hace el panadero?; la respuesta es: Él hace el pan. ¿Qué hace el carnicero? y responden: prepara la carne y la vende. A estas preguntas le siguen otras como: ¿Cuál es el oficio de X? : Él es un carnicero. ¿Cuál es el oficio de Y? : Ella es una cocinera.

Después de las preguntas siguen las oraciones afirmativas para conocer un nuevo verbo: *descansar*. Entonces el maestro dice: El que trabaja es un trabajador y el que no trabaja es un perezoso. Al final de la unidad se solicita al alumno que conjugue los verbos *ser* o *estar*, *trabajar* y *descansar* en presente, pasado compuesto y futuro. Antes de finalizar se hace una revisión general de la actividad. El objetivo de esta unidad es que el alumno utilice el vocabulario referido a los oficios y que al mismo tiempo conozca por primera vez algunos verbos y su conjugación.

El segundo ejercicio que presento puede encontrarse en el libro de G. Mauger y G. Gougenheim (1956) como método de enseñanza del francés hablado. Los autores aclaran que la obra está conformada estrictamente de las conclusiones de las comisiones del Francés elemental

tanto en el plano del vocabulario como en el de la gramática. Se propone facilitar el acceso del francés vivo a los jóvenes extranjeros o a los adultos interesados en adquirir una práctica acelerada de la lengua francesa. En la obra se incluye una sección de expresiones comunes bajo la rúbrica “frase francesa” destinadas al aprendizaje mecánico de las estructuras de la lengua hablada y de las innovaciones del francés. Anexo al libro existe una enciclopedia sonora grabada en dos discos con los cuales el alumno puede conocer la entonación de las frases. Estas frases son presentadas como ejercicios de sintaxis y fonética y se encuentran al final de la unidad. Ejemplo: “*Yo soy el que construyó la casa, Fuiste tú el que construyó el teatro y Fue él el que construyó la escuela.*”¹⁰

La unidad didáctica se llama *La ciudad (La ville, en francés)* y el alumno lee un fragmento descriptivo de una ciudad con ciertos sustantivos (plaza, tribunal, teatro) y adjetivos (bello, viejo, nuevo) marcados en negro. El objetivo es presentar vocabulario referido a una ciudad con el cual los alumnos aprenderán los géneros masculino y femenino de los sustantivos y adjetivos. Las frases que enseñan lo anterior son: un jardín *público* (un jardin *public*) y una plaza *pública* (une place *publique*) para ilustrar las inflexiones de género masculino y femenino respectivamente en los ejemplos anteriores.

Asimismo surgen diversos diccionarios como el *Primer diccionario en imágenes* de P. Fourré (1963), el cual incluye las principales acepciones de 1,500 palabras que constituyen el vocabulario del *Francés fundamental, 1er grado*; y el *Diccionario fundamental de la lengua francesa* de Gougenheim de 1958. Este diccionario comprende 3,000 palabras más 1,700 aproximadamente que son las que más se usan en el Francés Fundamental. El autor dice que los añadidos fueron hechos con el mismo espíritu con el que se hizo el *Francés fundamental, 1er*

¹⁰ C’est moi qui ai construit la maison, C’est toi qui as construit le théâtre y C’est lui qui a construit l’école.

grado, y partiendo de la encuesta estadística sobre el francés hablado y las encuestas por centro de interés. El desarrollo de los textos fue hecho por G.E. Vander Beke. (Gougenheim y Michéa, 1955)

Las unidades didácticas anteriores ilustran el uso que los franceses le dieron a los índices de frecuencia y disponibilidad obtenidos en el proyecto que iniciaron en la década de 1950. Ellos demostraron que de los resultados estadísticos es posible llevar a cabo una selección de vocabulario para enseñar léxico a estudiantes de lenguas extranjeras, tal como en el siguiente apartado donde continúo exponiendo muestras en la enseñanza de léxico pero en tiempos más recientes.

2.3 Aplicaciones pedagógicas para la enseñanza del léxico

Los estudios lexicológicos han buscado determinar los procesos de adquisición de palabras nuevas en aprendientes básicos y avanzados, así como probar métodos distintos en la enseñanza del vocabulario dentro del método comunicativo. Existen diferentes propuestas pedagógicas en la enseñanza del léxico como lo planteado por Jean Aitchison en 1987 sobre la enseñanza-aprendizaje de las palabras en términos de sus conexiones semánticas. Si bien esta idea podría relacionarse con lo definido como léxico disponible pocas propuestas han estado enfocadas principalmente a la enseñanza de este tipo de léxico. A continuación, daré un ejemplo breve sobre los ejercicios propuestos por Ilari en su *Introdução ao estudo do Léxico brincando com as palavras* (2012) y la noción de campos léxicos que presenta, ya que, éstos tienen una cierta relación con lo que entendemos por léxico disponible.

2.4 Enseñanza de campos léxicos como propuesta pedagógica

De acuerdo a Rodolfo Ilari (2012) los campos léxicos están constituidos por palabras que nombran un conjunto de experiencias que de alguna manera tienen un sentido similar. Existen dos recursos para organizar un campo léxico. El primero se refiere a un análisis componencial y el segundo a un análisis de prototipos. El componencial o por componentes parte de la base de que el significado de una palabra puede “romperse” en unidades menores, a las cuales generalmente se le llaman “componentes” o “características semánticas”, y que las unidades encontradas en el análisis de una palabra determinada reaparecerá en otras palabras. Es así que puede encontrarse lo que tienen en común dos o más palabras.

En el análisis por prototipos, se identifican las personas que mejor representan toda una categoría y, dice Ilari, se tratan de entender los otros elementos de la categoría a partir de nuestra experiencia (no será traducido a rasgos) de esos individuos. Es a lo que se refería Ilari cuando dice que los hablantes pueden reconocer que una perdiz es un ave sin que necesariamente sepan el tipo de ave que es porque tiene características específicas que la reconocen como tal. Este autor diseñó algunas actividades pedagógicas de enseñanza de vocabulario que mencionaremos a continuación por su creatividad e innovación y porque se asemejan a la idea de ofrecer al alumno una situación comunicativa específica en la cual haga uso de un tipo de vocabulario determinado equivalente al léxico disponible en este trabajo.

En el primer ejercicio se solicita a los alumnos que busquen la mejor manera de definir un objeto desconocido a una persona y se les pide que traten de dar a un salvaje (sic) un buen ejemplo de un coche, un bolígrafo, un sándwich, un funcionario público, un teléfono celular, una zapatilla, un yogurt o una cerveza. Se pide a los alumnos que piensen en el ejemplo que usarían

procurando ser muy específicos.¹¹ Lo esperado es que los alumnos hagan uso de las características de cada objeto para describirlo. Esta forma de clasificar objetos dependiendo de sus características se relaciona con la prototipicidad, es decir, lo que los hablantes de cualquier lengua pueden reconocer como característica válida para describir un objeto animado, por ejemplo, de otro que no es animado. Se trata de la capacidad clasificatoria de la que todo ser vivo cuenta.

La segunda actividad solicita a los alumnos que de un cuadro con imágenes varias se elijan las imágenes que no “cabén” en cada uno de los cuadros. Por ejemplo: el primer cuadro presenta una casa con un camino, una playa, un desierto, una isla y un gato. Dependiendo de las características que comparten las imágenes, el alumno elegirá aquella que no tiene una relación semántica con las otras. En este caso, el gato. En la segunda imagen, hay una cámara de video, una grabadora, un radio, un cinematógrafo y un pájaro. La elección de la imagen debe ser sustentada y justificada por el alumno.

Ilari comenta que los lingüistas utilizan a veces el nombre de campo léxico para indicar el conjunto de palabras que designan las partes de un objeto o conjunto de dos términos que se refieren a una misma técnica, y a un mismo ramo de actividades. (Ilari, 2012:44) Para los que no son especialistas, dice, es difícil enumerar todas las palabras de un campo léxico; en general, nombran algunas, comprenden otras cuando las escuchan y desconocen las restantes.

En la tercera actividad se solicita a los alumnos que recuerden y escriban todas las partes que componen una bicicleta. Después, deben comparar su lista con otra que se les da y que tiene el dibujo de una bicicleta que muestro a continuación:

¹¹ Suponha que você quisesse fornecer a um selvagem um bom exemplo do que é um carro, uma caneta esferográfica, um sanduíche, um funcionário público, etc., um telefone celular, um chinelo, um iogurte, uma cerveja. Que exemplos você usaria? Procure ser bastante específico. (Ilari, 2012: 43)

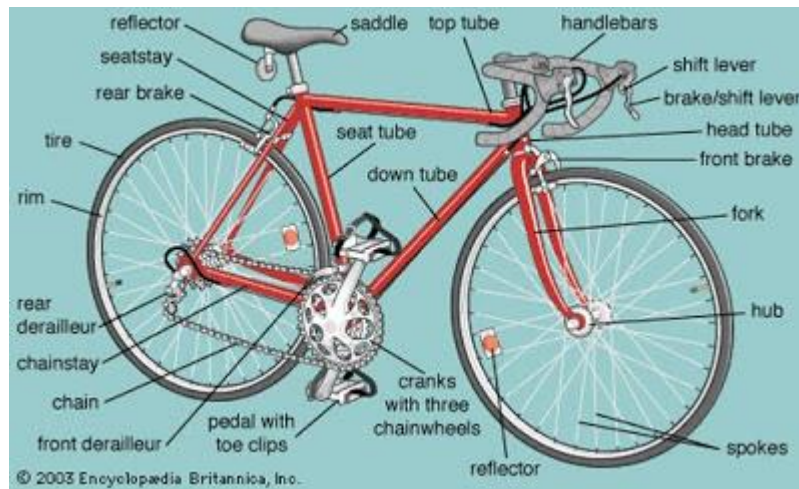


Imagen 2. Dibujo de una bicicleta parecida a la utilizada por Ilari en su texto sobre campos léxicos.

La actividad finaliza cuando los alumnos dividen las palabras en tres listas: a) vocabulario activo; b) vocabulario pasivo¹² y c) vocabulario desconocido. Suponemos que la división de las palabras no será la misma si se repitiera el ejercicio en diferentes momentos de un curso de enseñanza de una lengua, dado el dinamismo del cual habla Gile en su Modelo Gravitacional. Un alumno que acaba de comprar una bicicleta recordaría y clasificaría las palabras más fácilmente que otro que no ha andado en bicicleta desde la infancia. La experiencia, en esta caso, determinaría la disponibilidad y por ende la frecuencia del vocabulario descriptivo de una bicicleta. Algunas palabras pueden ser clasificadas como activas (pedal, por ejemplo) y otras pasivas (horquilla, por ejemplo).

¹² Cabe aclarar que la distinción de vocabulario pasivo y activo propuesto en esta actividad es aquella que se refiere a las palabras que conoce el alumno y que puede o no producir y utilizar. Este uso comúnmente utilizado en la lingüística aplicada se diferencia del léxico pasivo y activo de Gile.

La última actividad de Ilari pretende obtener el léxico disponible de los alumnos a través de un objeto común y corriente mas no necesariamente cotidiano en la vida estudiantil del aprendiente. En ese aspecto recae precisamente el diseño de unidades pedagógicas enfocadas a la enseñanza del léxico disponible. Los alumnos del CELE pueden conocer este tipo de léxico por medio de actividades cuyo objetivo no sólo sea la enseñanza de nuevo vocabulario, sino el reconocimiento y uso de palabras que vienen a su mente en situaciones comunicativas específicas. El conocimiento léxico en una lengua, por ende, se reconoce como una adquisición que engloba niveles de conocimiento varios dependiendo de la tarea solicitada por los docentes.

En conclusión me gustaría mencionar que, como parte de las unidades didácticas expuestas en este trabajo, propongo una unidad didáctica, en el apéndice 1, diseñada a partir del uso de las encuestas de léxico disponible en la formación de colocaciones en inglés.

Capítulo 3 - Metodología

En este capítulo se expone la metodología seguida para el desarrollo de nuestra investigación, la cual —como ya hemos mencionado— retoma la metodología general propuesta por los estudios de léxico disponible, consistente en la aplicación de una encuesta diseñada para este propósito. La novedad radica en el diseño, instalación y llenado de la base de datos LEDICAL, así como de una primera obtención de los índices de disponibilidad y frecuencias de las encuestas obtenidas en español e inglés de dos grupos participantes en el estudio. Con el breve análisis de datos estadísticos que realizo, posteriormente evalué la eficacia del prototipo de dicho programa computacional.

3.1 Objetivos específicos

- Obtener los diccionarios de español e inglés de léxico disponible de los alumnos de inglés avanzado del CELE.
- Describir el proceso de edición de las encuestas para alimentar la base de datos LEDICAL.
- Identificar el léxico disponible de las encuestas de acuerdo al Modelo Gravitacional de Daniel Gile.
- Valorar el funcionamiento del desarrollo tecnológico.

3.2 Preguntas de investigación

- ¿Cuál es el léxico disponible de los alumnos de inglés del CELE de la UNAM?

- ¿En qué medida el proceso de edición es una primera etapa del estudio del léxico disponible de los alumnos?
- ¿Cómo se valora el funcionamiento del programa computacional LEDICAL?

3.3 Informantes

Los informantes de mi investigación son alumnos universitarios que estudian inglés como lengua extranjera y/o alguno de los cursos de especialización que se ofrecen en el centro, tales como gramática, conversación, comprensión auditiva y escrita, lectura extensiva, fonética y preparación para el TOEFL. Estos alumnos cuentan con un nivel B1 de acuerdo al Marco Común de Referencia Europeo o han cursado los seis niveles de inglés en el CELE. De esta manera, se infiere que el nivel de interlengua que tienen es muy similar y por ende, es factible extraer características comunes entre los aprendientes.

3.4 Encuestas de léxico disponible

Para la obtención del léxico disponible de español e inglés de los alumnos del CELE utilicé las encuestas aplicadas por miembros del Seminario de Léxico Disponible del CELE, dirigido por Marisela Colín Rodea. En la primera página obtengo las características sociales generales de los alumnos, tales como nacionalidad, lengua materna, sexo, y área de especialización.

Cabe aclarar que dicha información fue recabada con el fin de obtener las variables para el análisis de resultados de acuerdo al centro de interés, ya que se pueden llevar a cabo análisis comparativos entre grupos tomando en cuenta la edad o sexo de los encuestados. Sin embargo, en este trabajo sólo elegí algunas palabras o frases léxicas simples con el fin de ilustrar evidencias de la interlengua de los aprendientes de acuerdo a Larry Selinker (1972), así como un par de

errores ortográficos o sobregeneralizaciones. Los primeros datos arrojados por LEDICAL son analizados cuantitativamente como evidencia del funcionamiento del programa y base de datos LEDICAL. Las encuestas de léxico disponible en papel pueden conservarse con el fin de corroborar la información léxica que se ha ingresado al programa y aquella que no está en él (las listas de *vocabulario especializado*, por ejemplo).

A continuación presento la primera y segunda páginas de la encuesta de léxico disponible que se utilizó en la recolección de datos:

CMLD
CELE, UNAM

SS Code

--	--	--	--

--	--	--

Teacher Group

Name Surname

Tel. (optional) e-mail (optional)

Date of birth Nationality

Major Semester

Mother tongue Female Male

How long have you been studying English?

How long have you been studying at CELE?

Do you speak other languages?.....

How long have you studied each of them?

How many books do you read for pleasure per year?(aprox).

How often do you go to the cinema?

Thanks for your cooperation!

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
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

Imagen 3. Páginas uno y dos de la encuesta de léxico disponible.

La página dos es igual a las subsecuentes que contienen las columnas donde los alumnos del CELE llenaron los 16 centros de interés. El único centro de interés que no es tomado en

cuenta es el de *Vocabulario especializado* porque los alumnos son de carreras diversas y no es posible establecer relaciones estadísticas entre dichas palabras. A partir de este momento me refiero sólo a los 15 centros de interés ingresados en LEDICAL reiterando que el de *vocabulario especializado* se encuentra sólo en las encuestas en papel. Una vez obtenida la información léxica de los alumnos se usa el programa de computación y base de datos LEDICAL para obtener los diccionarios de español e inglés. De igual manera se evalúa su funcionamiento y arquitectura al haber sido utilizado por primera vez para esta investigación.

3.5 Corpus

El corpus está conformado por las encuestas de disponibilidad léxica en español e inglés, respondidas por estudiantes universitarios que estudian inglés como lengua extranjera en el CELE, UNAM.

Los 16 centros de interés de las encuestas son los siguientes:

1. Las partes del cuerpo
2. La ropa
3. Alimentos y bebidas
4. La escuela
5. El campo
6. La ciudad
7. Medios de transporte
8. Los animales
9. Juegos y distracciones
10. Profesiones y oficios
11. Colores
12. Sensaciones y sentimientos
13. Adjetivos
14. Medios de comunicación
15. Verbos
16. Vocabulario especializado

Estos centros de interés¹³ son frecuentes en los estudios de léxico disponible en diferentes partes del mundo y este proyecto no es la excepción. Sin embargo, en LEDICAL existen cerca de 50 centros de interés disponibles que pueden dar pie a la realización de otras investigaciones en el área. A continuación enlistaré los 16 centros de interés en inglés porque, aunque son los mismos que en español, los alumnos tuvieron que haber traído a la mente palabras diferentes en la lengua meta a las de su lengua materna. No sólo porque el cambio de lengua supone una entrada diferente al léxico mental, sino por un cambio en la situación comunicativa o centro de interés evocados. Es decir, *las partes del cuerpo y el cuerpo humano* (*The Human Body*, en inglés) no son equivalentes; *la escuela y los muebles y materiales de la escuela* (*School furniture and materials*, en inglés) tampoco lo son.

- 1 The human body
- 2 Clothes
- 3 Food and drink
- 4 School furniture and materials
- 5 The countryside
- 6 The city
- 7 Means of transport
- 8 Animals
- 9 Games and amusement
- 10 Professions and jobs
- 11 Colors
- 12 Feelings and sensations
- 13 Adjectives
- 14 Means of communication, travel
- 15 Verbs
- 16 Specialized vocabulary

En ambas lenguas se espera que los alumnos recurran a un vocabulario de tipo nominal para llenar las encuestas, salvo en el caso de los centros de interés de *verbos* y *adjetivos*. En

¹³ Para una explicación más concreta de esta noción ver la página 26.

algunas ocasiones podrían presentarse frases léxicas complejas debido a que los alumnos tienen total libertad de llenar las encuestas con lo primero que les viene a la mente.

Colín *et al* (2013) observan que las encuestas ofrecen asociaciones libres, en las que un individuo establece relaciones semánticas con la palabra propuesta. Las autoras citan a Lara (2006:183) quien dice que estas palabras recogidas por asociación libre nos dan un campo asociativo, el cual está formado por:

- a) vocablos relacionados por el tema o los objetos de la experiencia a los que refieren (incluso términos de una especialidad, conocidos por los informantes)
- b) vocablos con la misma etimología
- c) vocablos agrupados por su categoría gramatical
- d) vocablos organizados por sus categorías morfológicas y fonológicas
- e) vocablos de diferentes dialectos o variedades regionales o sociales de la misma lengua.

Por ejemplo, si a un hispanohablante se le propone la palabra *alimentos* como parte inicial de la tarea de asociación libre, obtenemos una constelación de vocablos como la siguiente: pan, chocolate, plátano, nuez, miel; si el informante ha pensado en el desayuno como referencia. Cuando hemos obtenido una constelación, es decir, significados alrededor de un núcleo o centro de interés, podemos organizarla siguiendo diferentes criterios, tales como: tipos de alimentos, procedencia de los alimentos, frutas, verduras, lácteos, postres, condimentos e ingredientes. Así, una vez organizados, entonces se puede dar paso al análisis.

Las autoras, enfocan la manera en que nos acercamos a las asociaciones retomando el trabajo de Samper (Ibid.:279) quien señala los siguientes acercamientos: a) el campo léxico en el que aparece la palabra: cabeza, ojo, brazo, mano y nariz; b) referencias a una categoría semántica en las respuestas c) información sobre el hablante; d) la localización de una palabra que se

encuentra entre otras que hacen ciertas referencias y e) la existencia de otras referencias del mismo tipo.

Retomando nuevamente los señalamientos de Lara (Ibíd.:185), en cuanto al ejercicio asociativo que realizan los alumnos cuando llenan las encuestas, se dice que el análisis semántico del léxico es un método interpretativo cuya meta es garantizar que los resultados sean intersubjetivamente verificables y den lugar a una verosimilitud ampliamente aceptable. Se trata de un fenómeno de la mente humana cuyo análisis plantea un problema de objetividad, los signos pueden usarse de maneras ilimitadamente diferentes, ampliando o reduciendo su significado o modificándolo metafóricamente, los significados obtenidos varían de acuerdo al contexto, a los conocimientos y sensibilidad del lingüista que realiza el análisis.

3.6 Edición de encuestas

Los estudios de léxico disponible siguen dando frutos en cuanto a la enseñanza–aprendizaje de léxico en lenguas extranjeras. Cada día surgen más propuestas en el análisis de los índices de disponibilidad y cohesión obtenidos de las encuestas de léxico disponible por centros de interés. Por ejemplo Alberto Carcedo, *Disponibilidad Léxica en español como lengua extranjera: el caso finlandés* (2000); Humberto López Morales, *Léxico disponible de Puerto Rico* (1999) o más recientemente Rosa María Jiménez (ed.) *Lexical Availability in English and Spanish as a Second Language* (2014).¹⁴ Por lo tanto, la edición de dichas encuestas es parte fundamental en esta investigación como una forma de conocer el léxico del que dispone esta comunidad en particular.

Existe unificación de criterios de edición de los datos para lograr la mayor confiabilidad y validez de los mismos. Este trabajo fue realizado en Bilbao en 1998 por especialistas de léxico

¹⁴ Ver bibliografía completa al final de este trabajo.

disponible del mundo hispano, quienes al darse cuenta de las diferencias dialectales que surgían al obtener encuestas de diversos países hispanohablantes, como España, República Dominicana y Puerto Rico, se dieron a la tarea de unificar los criterios de edición. Éstos fueron rescatados de un trabajo realizado por José Antonio Samper en 1998, y que a continuación detallo:

1. En la edición ortográfica, Samper propone la “corrección gráfica” para el aspecto de edición “que aborda disfunciones relacionadas con deficiencias en el aprendizaje. Es decir, no se trata de faltas de ortografía concretas, que además resultan homogéneas en muchos casos, como la situación de “b” por “v” o de “g” por “j” [en español], sino que nuestros informantes en la mayoría de los casos no han llegado a asimilar del todo esas palabras en las que cometen los errores, al menos en lo que se refiere a la nómina activa”. (Samper, 1998). Este tipo de errores, dado la falta de conocimiento de la interlengua de los aprendientes, puede ser de dos tipos. Un tipo es donde están aquellas palabras que carecían de correlato en español y el segundo tipo, contempla las palabras que se parecían a la lengua meta pero no podían encontrarse equivalentes seguros. En algunas ocasiones, Samper optó por ignorar estas palabras a falta de comprobación directa con el aprendiente.
2. Cuando existen vocablos con variación en la grafía como *mahonesa* o “mayonesa” no se realiza la corrección y se toman en cuenta ambas opciones.
3. Los extranjerismos y los vocablos adaptados al español como *basketbol* y *baloncesto* se conservan. En esta investigación opto por conservar ambas variantes aunque siempre cuidando la ortografía españolizada y la variante del español mexicano; de tal manera, *basketbol* fue editado en la encuesta en papel e ingresado a LEDICAL como *basquetbol*.

4. Cuando un vocablo es repetido se toma en cuenta sólo el primero y se eliminan el segundo o los subsecuentes.

5. Todos los vocablos son neutralizados cuando existen flexiones de género y número. Los sustantivos son editados a masculino y singular, a menos que existan casos en los que al ser editados haya un cambio de significado, como en el caso de *empanizados* refiriéndose a la comida empanizada que carecería de sentido si lo editamos a singular. Tampoco hay edición, por ejemplo, en los vocablos como *vaqueros* donde el centro de interés es *la ropa*.

6. Es común encontrar extranjerismos en las encuestas de español e inglés que obedecen a la necesidad de los aprendientes de “compensar” la falta de conocimiento por medio de la creación de algunas palabras. Por lo tanto, Samper, divide los extranjerismos en tres tipos:

a. Neologismos especializados: palabras como *jeans, barman, piercing*, ya que están presentes en diferentes culturas y han sido adoptados del inglés. En nuestras encuestas de léxico disponible existe una gran variedad de neologismos como *leggings* o *bóxers* en el centro de interés *la ropa*.

b. Palabras de la lengua materna de los hablantes: Es difícil determinar cuando se trata de este tipo de extranjerismos o si se trata de una muestra de la interlengua que van desarrollando los aprendientes.

c. Extranjerismos adoptados formalmente. Éstos “constituyen mecanismos de creación léxica en lengua extranjera: los estudiantes no conocen exactamente cuál es la palabra castellana correspondiente a un determinado concepto, y optan entonces por recurrir a su lengua materna para crearla.” (Samper, 2001:282) Ejemplo: *forastero*, se toma de la raíz *forest (bosque)* en inglés y se agrega el

sufijo –eso para indicar la profesión. En otras ocasiones, el extranjerismo remite a una traducción literal de la lengua materna, tal como lo encontramos en sintagmas nominales como tablón negro (*blackboard*) o espacio de fuego (*fireplace*). (Samper, 2001:282)

7. Creación de neologismos partiendo de la lógica. Los alumnos crean palabras haciendo uso de las terminaciones o raíces propias de la lengua meta. En el análisis cito algunos ejemplos de los alumnos que sobregeneralizan reglas para crear palabras en inglés.

Me parece relevante señalar que aunque los criterios de edición fueron consensuados dentro del mundo hispánico, las encuestas en la lengua extranjera (en este caso el inglés) también presentan este tipo de casos y con ello la necesidad de acordar criterios de edición en las listas de vocabulario en inglés. Un aspecto importante en la edición fue la homogeneización de los vocablos para evitar posibles errores estadísticos una vez que se realiza el cómputo de las listas. En el caso de las palabras en español donde había que decidir por una variante u ortografía, consulté el Diccionario del Español de México del COLMEX, dirigido por Luis Fernando Lara.

A continuación enuncié un ejemplo en las encuestas de los alumnos del CELE que muestra la forma de crear neologismos partiendo de la lógica. Se inventan palabras en la lengua meta basándose en los conocimientos de su propia lengua. Es decir, un alumno que estudia inglés como lengua extranjera buscará crear una palabra en la lengua meta haciendo una traducción léxica literal del español. Si en español existe el concepto “áreas verdes” para referirse a zonas con pasto localizadas dentro de las ciudades, el mismo concepto puede existir en inglés. Entonces el alumno crea *green areas* para referirse a las áreas verdes que conoce en su lengua meta y que ahora traslada a la lengua que estudia.

Este tipo de casos no sólo puede abordarse desde el aspecto de la adquisición de lenguas extranjeras o segundas lenguas, sino también desde el cultural y del significado. ¿Cuál es la relación que existe entre la palabra y la realidad del alumno en un entorno cultural determinado? Desde el punto de vista del alumno es similar y a partir de esta relación es que establece conexiones léxicas entre culturas y lenguas, las cuales se reflejan en la creación de palabras en la lengua meta.

3.7 Tiempo de llenado de encuestas

El tiempo destinado a los alumnos para llenar las encuestas también obedeció a un criterio de homogenización en la metodología que se ha seguido en los estudios de léxico disponible desde hace unos años. El tiempo total fue de dos minutos para llenar cada una de las encuestas con los 16 centros de interés antes mencionados.

3.8 LEDICAL

Es un software gratuito que computa la disponibilidad y frecuencias relativas de las palabras por centro de interés dentro de un grupo definido de encuestados. LEDICAL fue diseñado por el ingeniero Jim Frey en coordinación con Santiago Cortés, quienes se dieron a la tarea de crear un programa computacional estadístico sin las carencias de los otros software utilizados tradicionalmente en los estudios de léxico disponible: Lexidisp y Dispolex.

En su investigación posdoctoral, Frey Pereyra había detectado algunas fallas en ambos softwares que se relacionaban con la falta de accesibilidad y uso para aquellos interesados en los estudios de léxico disponible. Además, era necesario contar con características computacionales muy específicas para realizar la instalación de los programas en los equipos de cómputo.

LEDICAL, en cambio, fue creado con la idea de solucionar estos problemas y tener mayor alcance de estudio. El programa puede calcular los índices de disponibilidad y frecuencias considerando diversas variantes y haciendo comparaciones entre encuestas, encuestados y lenguas extranjeras, así como entre instituciones educativas dedicadas a la enseñanza de lenguas extranjeras.

En 2010, Frey y Cortés terminaron LEDICAL y casi inmediatamente comencé a ingresar las listas de vocabulario de los alumnos encuestados por los miembros del Seminario de Léxico Disponible, en el 2009. Un año más tarde, LEDICAL fue instalado en el servidor del CELE de la UNAM con acceso abierto a todo maestro interesado en ingresar el léxico obtenido de sus alumnos tanto en español como en la lengua extranjera que estudian, tales como: inglés, francés, portugués y ruso. La idea es que LEDICAL sea una base de datos masiva en la cual puedan ingresarse tantas listas de vocabulario por centro de interés deseen los maestros de lenguas extranjeras. La dirección electrónica del programa es: www.cele.unam.mx/ledical.

El uso de LEDICAL en investigaciones sobre el acceso al léxico mental es de vital importancia por ser una base de datos multilingüe que contiene datos para realizar trabajos en análisis de errores y la interlengua de los alumnos, o bien en estudios sociolingüísticos haciendo uso de los datos y del léxico especializado. Al ser LEDICAL una base de datos multilingüe pueden realizarse estudios en decenas de lenguas. En la presente investigación se trabajó solamente con encuestas en español e inglés y 16 centros de interés anteriormente referidos; sin embargo, es necesario recordar que las opciones que ofrece LEDICAL van desde el francés, portugués y ruso hasta swahili, y chino mandarín, incluso existe la posibilidad de ingresar encuestas en maya y zapoteco. Asimismo existen 50 centros de interés en total que han sido utilizados en diferentes investigaciones de léxico disponible desde sus inicios, por lo que pueden realizarse estudios diversos entre dos mismos centros de interés en varias lenguas. Por ejemplo:

comparar los centros de interés “educación” o “problemas globales” en español e inglés, o en cualquiera de las lenguas que se estudian en el CELE.

LEDICAL ofrecería llevar un control detallado sobre el número de sujetos de estudio, así como de datos etnográficos y lingüísticos. Los centros de interés obtienen sustantivos principalmente, por lo que podría evaluarse el conocimiento de los sufijos que tienen los alumnos en la formación de sustantivos o adjetivos (si se trata del centro de interés *adjetivos* o *sensaciones y sentimientos*). Una vez que se realice el cálculo de los índices de disponibilidad y frecuencia puede observarse el uso de los sufijos para obtener los sustantivos o adjetivos, según sea el caso.

Si se propone algún estudio relacionado con el análisis de errores, tal como Frey Pereyra esperaba realizar dentro de esta misma investigación, las posibilidades se multiplican, ya que pueden realizarse estudios por niveles de interlengua e incluso hacer una tipificación sobre los errores que se cometen dependiendo el nivel que estudian los alumnos de la lengua extranjera en cuestión. El ingeniero Jim Frey comentaba sobre las posibilidades de estudio que existen entre instituciones nacionales e internacionales. Es posible que LEDICAL funcione no sólo para llevar a cabo estudios dentro del CELE de la UNAM, sino en colaboración con otras universidades o centros de lenguas extranjeras en el mundo.

3.8.1 De la idea al software

Desde sus inicios y como ya se mencionó antes, LEDICAL fue creado con la idea de ser una base de datos cuyas listas de vocabulario sean de utilidad para llevar a cabo estudios de diversa índole. Las posibilidades que tiene el programa permitirán realizar investigaciones de carácter lingüístico, sociolingüístico, cultural y cognitivo. En este momento es el único software que

utilizo para obtener los diccionarios español e inglés de los alumnos, así como un primer cálculo de los índices de cohesión y disponibilidad de las encuestas antes mencionadas para mi tesis.

A continuación detallo su funcionamiento y arquitectura y me refiero a las posibilidades de almacenamiento de información que tiene el programa. Es decir, ¿por qué LEDICAL guarda información de acuerdo a *maestro*, *grupo*, *encuesta* y *encuestado*? ¿Cuáles son las múltiples posibilidades que existen para efectuar un análisis cuantitativo de las listas de vocabulario? En el análisis del presente trabajo incluyo todos los aspectos relacionados con el ingreso de las encuestas al programa y la evaluación del mismo.

3.8.2 Salida

LEDICAL computa la disponibilidad léxica y frecuencias relativas de las palabras por centro de interés dentro de un grupo definido de encuestados. Asimismo mantiene una base de datos con información correspondiente a *Maestro (Teacher)*, *Grupo (Group)*, *Encuestado (Respondent)*, *Encuesta (Poll)* e *información de la encuesta (Poll Data)*.

La disponibilidad léxica es calculada con la fórmula de López Chávez y Strassburger Frías.

$$D(P_j) = \sum_{i=1}^n e^{-2.30 \frac{i-1}{n}} \cdot \frac{f_{ji}}{I_1}$$

En ésta la *i* equivale a la posición (de la palabra) en la lista de vocabulario del encuestado; la *n*, es la posición máxima alcanzada en la encuesta (extensión o longitud de la lista más larga) y la *e*, es la constante matemática (conocido como número de Euler, el cual equivale a 2.71828 18284 59045... .)

$D(P_j)$ = disponibilidad léxica de la palabra j

i = posición en la lista de vocabulario

n = la posición máxima alcanzada en la encuesta (longitud de la lista más larga)

e = constante matemática (número de Euler) = 2.71828 18284 59045 ...

$-2.303 \frac{\log \frac{f_{ji}}{n}}{e}$ = factor de dispersión

f_{ji} = Frecuencia absoluta de la palabra j en un nivel i

I^1 = Número de encuestados que participaron en la encuesta

Las frecuencias relativas son:

Freq1 = frecuencia relativa por número total de palabras

Freq2 = frecuencia relativa por el número total de encuestados

3.8.3 Arquitectura

LEDICAL es una aplicación en línea desarrollada con PHP, el cual es una base de datos Mysql y, el buscador Firefox. Actualmente está restringido a un usuario por sesión en una computadora con Linux o Mac operando como servidor y cliente. Puede ser extendida para convertirse en un programa de varios usuarios que funcione en una intranet o Internet. LEDICAL requiere usuarios registrados con una *clave de acceso (password)* y diferentes niveles de acceso, tales como *administrador*, *mantenedor (maintainer en inglés)* y usuario. El programa funciona con los buscadores Firefox, Internet Explorer y Google Chrome.

3.8.3.1 Página principal: Lenguas materna y extranjeras

La página principal solicita el nombre de usuario y la contraseña (imagen 4). Una vez que se ingresa, se despliega la página del menú en inglés y español. No importa cuál de los dos botones elija el usuario, hasta este momento la información siguiente aparece únicamente en inglés (imagen 5).

LEDICAL

IN MEMORIAM: MARÍA LUISA HELEN FREY PEREYRA de CORTÉS 1978 - 2009

Software for Linguists
Programación para Linguistas

Sign in to Lederal:

User Login:

Password:

Imagen 4. Página principal de LEDICAL.



Imagen 5. Página del menú principal.

El menú de la imagen 4 lleva al usuario a dos menús diferentes; en el primero las opciones que se despliegan son *Ingreso (Input)*, *Edición (Edit)*, *Cómputo (Compute)*, *Contraseñas (Passwords)* y *Administrador (Admin, abreviado)*. En el segundo se puede elegir entre *Maestro (Teacher)*, *Grupo (Group)*, *Encuestado (Respondent)*, *Encuesta (Poll)* e *Información de la encuesta (Poll Data)*. El usuario puede ingresar un *maestro* eligiendo las opciones ingreso del primer menú y *maestro* del segundo. Hay información obligatoria como *país, exposición, lengua, nivel de lengua y nacionalidad*. Las dos últimas tienen más de 200 entradas por cada una.

Después de ingresar los datos del *maestro* continuamos con el *grupo, el encuestado, la encuesta* y al final la *información de la encuesta*. En el menú de los centros de interés que se despliega en Información de la encuesta se encuentran los utilizados en este trabajo más otros centros seleccionados de Internet hasta sumar un total de 50.

A continuación doy un breve repaso de las opciones que ofrece el programa en el ingreso de la información requerida para llevar a cabo el presente estudio.

3.8.3.2 Ingreso y Edición

La información marcada con un asterisco es considerada obligatoria. Existe la opción de llenar el cuadro con algún otro carácter y editarlo posteriormente. Sin embargo, en ocasiones el programa no permite editar información de grupos cuyas listas de vocabulario han sido ingresadas en su totalidad en el sistema. Algunos iconos contienen listas con opciones más concretas, como es el caso de *maestro* y *lengua meta*. En todas las páginas existe la opción de ingresar notas lo cual resulta de gran utilidad para el maestro o el lingüista interesados en concentrarse en un aspecto de las listas y buscar hacer anotaciones respecto a ello. La única opción que no contiene notas es *información de la encuesta*.

3.8.3.3 Maestro

La información solicitada al maestro tiene que ver con nombre, apellidos, fecha de nacimiento (la cual es opcional), sexo, cursos o diplomas en enseñanza de alguna lengua extranjera, carrera, nacionalidad, país, educación, lengua materna y notas.

3.8.3.4 Grupo

El grupo es asignado forzosamente a un maestro. Es posible cambiar un grupo a otro maestro usando la opción de edición. En *Encuestado*, los *encuestados* son asignados a un grupo como se muestra posteriormente. Palabra clave es utilizado para identificar al grupo por medio de una

clave o un nombre más breve que pueda ayudar a reconocer el grupo de manera sencilla. Lengua meta (*target language*) es la lengua de estudio del grupo que se está analizando. Puede tratarse de la lengua materna o de la lengua extranjera que estudian los alumnos y es posible hacer una comparación entre las lenguas usando la opción Comparación (*Comparison language*), cuyo icono despliega una lista muy larga de opciones. Ahora bien, si el usuario no requiere llenar la opción de Lengua meta puede elegir la última opción de la lista que dice: 'Lang_not_used'.

3.8.3.5 Encuestado

En esta sección existen campos y operaciones que necesitan ser explicados a detalle:

1. Lugar de origen: Este campo está disponible y es opcional dado que las respuestas de un encuestado podrían variar dependiendo su nacionalidad e incluso región lingüística. Es decir, un mexicano del sur no hablará igual que uno del norte y por ende su vocabulario también cambia dependiendo de su lugar de origen.
2. Debajo de la opción *Grupo* existe un bloque de entradas específicas para el CELE. En el título se muestra la lengua meta del grupo y una casilla es utilizada para seleccionar un nivel de lengua básico o avanzado. Si se selecciona avanzado, los cursos avanzados del CELE deben ser activados y asignados en orden.
3. También es posible activar más de tres lenguas estudiadas por el encuestado, así como el nivel correspondiente y años de uso.
4. Para editar *encuestados* es necesario elegir primero *maestro* y *grupo* porque de esa manera se evita tener que recorrer todos los encuestados hasta llegar al buscado.

3.8.3.6 Encuesta

Al seleccionar un grupo encuestado, la lengua meta aparecerá a la derecha de la casilla seleccionada. Si el usuario selecciona Comparación entonces podrá seleccionar de una lista larga una segunda lengua (de comparación). Es importante aclarar que no es posible hacer cambios en una encuesta cuando ya se ha realizado un cálculo de la información ingresada.

3.9 Información de la encuesta (Lista de vocabulario):

- Ingreso – Primero se deben seleccionar las opciones *Maestro*, *Grupo* y *Encuesta*, y después existen tres opciones para mostrar la información que ya fue ingresada por grupo por medio del icono *Mostrar centros (Show Centers)*:
 - *Nuevo (New)*: Muestra todos los centros de interés que no han sido capturados. Los nombres de los alumnos de este grupo son las siguientes opciones que se despliegan.
 - *Parcial (Partial)*: Muestra opciones desplegables de los centros de interés que contienen información de algunos encuestados del grupo. En esta opción se puede ingresar alguna lista de vocabulario faltante y se puede elegir grupo y encuestado para ingresar la información nueva. Es de gran utilidad porque si por algún motivo el usuario no pudo terminar de ingresar la lista de un alumno puede hacerlo posteriormente. Los centros de interés que ya tienen información ingresada no se muestran como opción, sólo se muestran los que están incompletos.
 - *Completo (Complete)*: Muestra las opciones de los centros que tienen información de todos los encuestados en un grupo. También existe una lista con todos los nombres de los encuestados. En esta opción no se despliega la caja de palabras porque no existe nada nuevo para ingresar. Se trata de la opción “completo”.

- *Edición* - Se selecciona *Encuesta*, *Centro de interés* y *Encuestado*. Si la encuesta no tiene información aparece un mensaje que lo indica. Si existe información ingresada aparecen las opciones *centro* y *encuestado* y se muestran sólo aquellos que tienen información. Entonces se despliega la caja de texto con las palabras ingresadas o en blanco (dependiendo el caso) para que el usuario ingrese la información deseada.

3.9.1 Página Principal: Índice computado (Compute Index)

Esta casilla es utilizada para mostrar un menú desplegable para seleccionar *maestro*, *grupo*, *encuesta* y *centro de interés*, los cuales serán utilizados para hacer el cálculo. La casilla *GO* es utilizada para calcular el *Output (Salida)*, y después se despliega una tabla con la siguiente información:

Encabezado (Header): *Maestro, encuesta (día, hora y tipo), Centro de Interés, día y hora del cálculo, Grupo (palabra clave y lengua meta), encuestados (número de participantes/total por grupo) y lugar de la encuesta.*

Cuerpo (Body): Cinco columnas que muestran una línea (o fila) de números, palabra, índice, frecuencia 1, frecuencia 2. Arriba de la tabla están las casillas para exportar la información en formato CSV o impresa. Si se elije la opción imprimir es posible obtener la tabla en formato PDF, a partir del cual se puede hacer otra conversión en diferentes formatos: JPEG, JPEG-2000, Open EXR, PDF, PNG y TIFF.

En esta última sección presenté los detalles técnicos de LEDICAL como base de datos y programa computacional que calcula los índices de frecuencia y disponibilidad de las palabras recolectadas. Su presentación en este trabajo es detallada por las posibilidades que representa en los estudios lexicológicos en México a nivel superior, así como por ser la primera vez que el

funcionamiento de este programa es puesto a prueba. En el siguiente capítulo señalo el proceso en sus diferentes etapas (recolección, edición e ingreso de las encuestas) como una forma de evaluar el programa y el estudio de léxico disponible de los alumnos avanzados del CELE.

Capítulo 4 - Evaluación de LEDICAL

En la Metodología describí todo el proceso de almacenamiento del léxico disponible de los alumnos del CELE de la UNAM en el programa LEDICAL. En este capítulo evalué el funcionamiento del desarrollo tecnológico detallando los pasos seguidos durante la recolección, edición e ingreso de las encuestas utilizadas para dicho fin. Su creación, como ya se mencionó anteriormente, parte de la idea de usar LEDICAL como una base de datos por medio de la cual se calculan los índices de disponibilidad y frecuencia del léxico disponible en diversas lenguas extranjeras. Mi tesis presenta dichos resultados estadísticos en sólo dos lenguas: español e inglés como lengua extranjera, además de los diccionarios de léxico disponible.

En el capítulo 5 me refiero a los datos cuantitativos obtenidos, y en éste reflexiono sobre el proceso de construcción de LEDICAL como base de datos en las dos lenguas mencionadas. Asimismo hago mención de las facetas que se llevaron a cabo en vías de obtener en los próximos años los diccionarios de léxico disponible (entendidos como compendios de palabras disponibles) no sólo en lengua inglesa sino en cada una de las lenguas que se estudian en el CELE.

Describo brevemente algunos puntos sobre la aplicación de las encuestas al ser el primer paso que se llevó a cabo en la recolección de las listas de vocabulario de los alumnos. De tal manera, el lector puede darse una idea del ejercicio que se realizó en el salón de clases durante la recolección de los datos y que fue objeto de la selección del léxico que se ingresó al programa posteriormente. En el siguiente punto presento algunas ventajas y desventajas que existen en el proceso de edición, selección e ingreso de las encuestas a LEDICAL para después ahondar sobre las posibilidades del programa en la obtención de los diccionarios en español e inglés de los sujetos que participaron en el estudio.

4.1 Aplicación y selección de las encuestas

La aplicación de las encuestas estuvo dividida en dos fases. La primera se realizó en mayo de 2009 en el Auditorio Rosario Castellanos del CELE de la UNAM y la segunda en octubre de 2012 en las aulas del mismo centro por algunos miembros del seminario de Léxico Disponible.¹⁵ En 2009 se aplicaron las encuestas a un total de siete grupos de inglés avanzado, previa autorización de los maestros. Las instrucciones se dieron en español con la idea de evitar confusiones u omisiones en el llenado de las encuestas. No obstante, algunos alumnos omitieron información completa en la primera página (como nombre del profesor o la clave del grupo) y en ocasiones, centros de interés completos. El tiempo que se da a los alumnos una vez que comienzan a llenar las listas de vocabulario son dos minutos por cada uno de los dieciséis centros de interés.

En la primera fase de aplicación de las encuestas fue común encontrar que los alumnos no llenaron los dieciséis centros de interés en su totalidad, como es el caso del centro de interés *verbos* en español perteneciente a un par de encuestas de un grupo de TOEFL. Un caso similar fue el de un grupo de comprensión auditiva (identificado como L-001) cuyas encuestas carecen de las listas de *partes del cuerpo, alimentos y bebidas y la ropa*. Estas fueron descartadas pues carecen de elementos necesarios para llevar a cabo el análisis cuantitativo por centro de interés. De un total de siete grupos que participaron en la primera fase sólo se seleccionaron cuatro (TOEFL-1, L-001, G-001 y S-001) para realizar la edición en papel previo ingreso a LEDICAL. Con base en la recolección de encuestas en la primera fase es que se decidió realizar una segunda aplicación para contar con el número necesario de listas de vocabulario.

¹⁵ El motivo por el cual las encuestas fueron aplicadas después de tanto tiempo se debió a que el ingreso de las listas comenzó mucho después de la primera fase de aplicación y hasta ese momento detectamos errores en algunas encuestas que tuvieron que ser eliminadas.

En la segunda fase participaron sólo tres grupos de inglés; uno de lectura extensiva (R-003), el segundo y el tercero de producción oral (S-004 y S-003 respectivamente). Se decidió llevar a cabo el análisis cuantitativo de los primeros dos grupos (R-003 y S-004) porque ambos contaban con el mismo número de alumnos (doce), lo cual facilitaba la comparación entre grupos por centro de interés.

La primera fase de aplicación de las encuestas fue fructífera porque una gran cantidad de encuestas fue recolectada y se inició con la edición de las mismas de acuerdo a lo expuesto por Samper. De lo observado podemos concluir que se reduce el margen de error si se realiza con un menor número de alumnos, lo cual impacta positivamente en la selección de las listas que serán ingresadas a LEDICAL. Los alumnos cometieron errores similares durante el llenado de las encuestas, lo cual evidencia fallas en la instrucción que recibieron en ese momento. A partir de lo aprendido en la primera fase es que se lograron reducir los errores u omisiones considerablemente en la segunda.

La aplicación de las encuestas es considerado el primer paso en la obtención de los diccionarios de léxico disponible en español e inglés que presentamos en este trabajo. Recolectamos un total de siete grupos con un total de 77 alumnos; cada uno lleno los dieciséis centros de interés en español e inglés.

4.2 Edición de las encuestas previo ingreso a LEDICAL

Los criterios de edición de las encuestas de léxico disponible fueron detallados en el capítulo de Metodología del presente trabajo (en las páginas 54, 55 y 56). El más importante mencionado en dicho apartado se refiere a los vocablos neutralizados cuando existen flexiones de género y número, el cual coincide con lo planteado por Lara cuando dice que los vocablos son las unidades

de cita sin flexiones o derivaciones que se reconocen como palabras (Lara, 2006). Si en la encuesta en papel aparecen las palabras *árboles*, *cielo* y *nubes* en el centro de interés *el campo* se ingresan a LEDICAL como *árbol*, *cielo* y *nube*. Todos los sustantivos deben editarse a masculino y singular salvo cuando se dé un cambio de significado después de la neutralización. La eliminación de las flexiones es vital para efectuar el análisis cuantitativo de todas las listas de léxico disponible, así como en la adecuada recolección de los diccionarios. Se presentaron excepciones en palabras como buscar ejemplos porque eliminar las marcas de género y número implicarían un cambio en el significado. En este caso se ingresan al programa tal cual fueron llenadas por los alumnos.

Los criterios de edición utilizados en esta investigación, como sugiere Samper, también contemplan la creación de neologismos partiendo de la lógica de los alumnos, quienes inventan palabras en la lengua meta basándose en los conocimientos de su propia lengua. Cuando el autor divide los extranjerismos en tres tipos no se realiza ningún cambio en la morfología de la palabra. Su relevancia recae en la identificación de evidencias en la interlengua de los alumnos, como puede verse en la frase léxica *bad traffic* del centro de interés *la ciudad* localizado en una encuesta del grupo G-001. En este ejemplo se demuestra el proceso de transferencia léxica que el alumno realizó para unir dos palabras que dan la idea de que en la ciudad hay *mal tráfico*. Si en español existe el concepto *mal tráfico* para referirse a la pesada circulación de vehículos en una ciudad, el mismo concepto puede existir en inglés. Con esto no quiero decir que un nativo hablante de inglés no entienda a lo que se refiere alguien que dice *bad traffic*, pero no es el término que se usa en inglés. *Heavy traffic* es una unidad fraseológica, conocida en la lexicología como colocación, la cual ha sido utilizada en la enseñanza de lenguas extranjeras desde la década de 1990 cuando Nattinger y De Carrico exaltaron la importancia de los segmentos léxicos. Ellos

dicen que la competencia pragmática de los alumnos se determina por la habilidad que tienen de acceder y adaptar frases prefabricadas de la lengua que estudian.¹⁶

Otro ejemplo que se relaciona con la creación de neologismos es el de una alumna que escribió la palabra *cooker* en el centro de interés de Profesiones porque su conocimiento de la lengua meta le indica que el sufijo *-er* en la palabra *cook* se utiliza para hablar de la persona que realiza una acción, en este caso cocinar. José Antonio Samper (1997) afirmaría que ella aplicó un mecanismo de creación léxica en la lengua extranjera y catalogaría este ejemplo como extranjerismo adoptado formalmente, ya que los estudiantes se valen de su lengua materna y de la meta (como es este caso) para crear palabras correspondientes a un determinado concepto.

En la edición ortográfica de Samper no se consideran los errores ortográficos como tal sino una “corrección gráfica” que se explica por la falta de conocimiento que los estudiantes tienen de la lengua extranjera que estudian. Este evento se considera como una falta de asimilación de dicha lengua y por ende no pueden ser errores ortográficos. Es éste el caso de la omisión de alguna letra como en “*dificult*” con una *f*, o bien de “*stuborn*” sin doble “*b*”, como debe escribirse. En ambas palabras las omisiones de las dobles letras (*ff* para *difficult* y *bb* para *stubborn*) no son significativas y lo que se hace es hacer la corrección ortográfica y continuar con el ingreso de las palabras a la base de datos.

En Profesiones ocurre que existen palabras en femenino por ser las mujeres quienes han desempeñado tradicionalmente trabajos específicos. Se piensa en una ama de casa, una niñera y una cocinera (en femenino) y en un jardinero, un mecánico y un mesero (en masculino). Una vez más, puede aseverarse que la designación de las palabras surge de la experiencia del hablante y que los criterios de edición propuestos hasta ahora deberían de contemplar este tipo de fenómenos que apelan a la experiencia y conocimiento del mundo que posee el estudiante. Si

¹⁶ Ver Capítulo 2, páginas 38 y 39 del presente trabajo.

bien es cierto que la edición elegida en este trabajo se apega a la tradición de los estudios de léxico disponible, sería interesante realizar un estudio donde el léxico de los alumnos se considere tal como “sale” de sus mentes. Estadísticamente sería un reto casi imposible de realizar, pero en términos cualitativos se antoja como material invaluable.

4.3 Funcionamiento de LEDICAL

A continuación detallo los pasos seguidos en el funcionamiento de LEDICAL como un programa computacional diseñado exclusivamente para la realización de estudios de léxico disponible en el CELE de la UNAM, a partir del proyecto multilingüe posdoctoral propuesto por Frey Pereyra en el 2008.

4.3.1 Ingreso del maestro a LEDICAL

Una vez que las encuestas fueron aplicadas y seleccionadas para su ingreso al programa como se describió anteriormente, se da de alta al profesor. Es el primer paso a seguir en el ingreso de las encuestas al programa. Al profesor se le piden datos generales como nombre, apellido, teléfono, e-mail, género y fecha de nacimiento, asimismo se requiere conocer la carrera que tiene, o certificado de enseñanza, nacionalidad y lengua materna. Algunos campos de la sección *maestro* son opcionales (*fecha de nacimiento y teléfono*) porque la información no es relevante para realizar investigaciones de léxico disponible. Los aplicadores pueden tener antecedentes académicos y laborales varios y participar de igual manera en dichos estudios.

Es importante mencionar que aunque lo ideal es que cada maestro ingrese sus propias listas de vocabulario, un sólo maestro no podría contar con el mínimo número de alumnos requerido para ingresar los datos a LEDICAL y obtener los índices de disponibilidad y

frecuencia. Por tal motivo, se solicita el apoyo de otros maestros para aplicar las encuestas en sus grupos e incrementar el número de alumnos encuestados. Eso fue lo que sucedió con las encuestas en español e inglés que se utilizaron para realizar esta investigación. La tarea que realicé fue la de ingresar las listas de vocabulario de todos estos grupos aún y cuando yo no estuve presente en la aplicación de dichas encuestas. Por tal motivo el único maestro dado de alta en el programa que cuenta con grupos y listas de vocabulario ingresadas hasta este momento soy yo. El resto de los maestros forman parte del Seminario de Léxico Disponible pero no han ingresado información al programa.

Ya dijimos que los maestros-aplicadores no siempre aplican las encuestas con sus mismos grupos. Las razones son varias pero la más importante es que para que el programa pueda calcular los índices de disponibilidad y frecuencia léxicas es necesario ingresar un mínimo de 50 encuestas por lengua con sus 16 centros de interés, es decir, 50 en español y 50 en inglés, lo que equivale a 100 encuestas en ambos idiomas o 32 centros de interés en inglés y español por alumno. Esto nos da un total de 1,600 listas de vocabulario en español e inglés.

En el caso de las lenguas extranjeras que se estudian en el CELE y que tienen un menor número de alumnos (como portugués, ruso y sueco) existe un problema para alcanzar el mínimo requerido y realizar los cálculos estadísticos deseados. De hecho, los compañeros del seminario que trabajaron con otras lenguas, como portugués y ruso realizaron las aplicaciones y analizaron los datos obtenidos sólo cualitativamente.

Con base en lo anterior, se establece una relación entre profesor, número de encuestas, lenguas analizadas e índices de disponibilidad. Es necesario encontrar la manera de llevar a cabo los análisis cuantitativos del léxico disponible de los alumnos que estudian lenguas menos demandadas. Un buen inicio para alcanzar dicho objetivo es que se comience a llenar LEDICAL con lenguas como portugués y ruso para que simultáneamente se trabaje con el ingeniero Jim

Frey en una alternativa estadística que pueda arrojar datos cuantitativos sobre el léxico de estas lenguas. Si bien, los datos carecerían de una validez más sólida sería un buen ejercicio para aproximarnos al análisis de datos de dichas lenguas extranjeras y que tienen la misma importancia que la lengua inglesa.

4.3.2 Ingreso del grupo a LEDICAL

En esta sección se llenan los datos del grupo encuestado pero sólo aquellos relevantes para la investigación, como es la lengua meta y la clave del grupo que corresponde con la asignada por el mismo centro. De esa manera, se evitan confusiones al momento de identificarlos. La lengua meta es aquella que estudian los alumnos (en este caso el inglés), misma que será comparada con su lengua materna (en este caso el español) al momento de realizar los análisis cuantitativos.

En el campo de descripción y notas se espera encontrar información general del grupo, como puede ser número de alumnos, horario y lugar de clases, o bien, si existen alumnos extranjeros cuya lengua materna no sea el español. En el campo de maestro es importante aclarar que no se refiere al maestro del grupo, sino al que ingresa los datos al programa, ya que éste es el que realizará la investigación. En este trabajo, ingresé los datos léxicos en español e inglés de todos los alumnos encuestados a la fecha por lo que hasta este momento no existe una opción diferente a mi nombre —Anahí Ramírez Alfaro— de la lista de ocho maestros dados de alta en LEDICAL.

4.3.3 Ingreso del encuestado a LEDICAL

Los datos ingresados en la sección encuestado son los mismos que se obtuvieron en la primera página de las encuestas escritas. Existen campos obligatorios y otros optativos (correo electrónico y lugar de origen) además del dominio de lengua que tienen los alumnos basado en los niveles

propuestos por el Marco Común de Referencia Europeo (A1, A2, B1, B2, C1, C2). En esta investigación los alumnos encuestados cursaban los niveles avanzados del CELE por lo que se decidió estandarizarlos a B2. De acuerdo a los planes de estudio del Departamento de Inglés del CELE de la UNAM los alumnos deben de contar con dicho nivel al término de los seis que se ofrecen.

Lugares de exposición y número de veces que usa internet a la semana son los dos campos que no fueron considerados en LEDICAL porque no se les solicitó a los alumnos dicha información en las encuestas escritas. En estos casos y en aquellos en los cuales el alumno olvidó llenar la información se decidió dejar los espacios en blanco o elegir la opción *none*.

4.3.4 Ingreso de la encuesta a LEDICAL

Cuando se ingresa la encuesta al programa en la sección *Encuesta* debe elegirse una vez más el *maestro* para que se desplieguen los *grupos* encuestados ingresados previamente por dicho maestro. De esa manera se lleva un control de los grupos con los que está trabajando un maestro en particular y se contará con el registro de la lengua. Asimismo la lengua meta ya está definida porque se entiende que el maestro ha ido trabajando con los mismos grupos en la misma lengua meta.

En este estudio la lengua meta fue el inglés. Por lo tanto cuando las encuestas en inglés fueron ingresadas a LEDICAL se eligió en tipo de encuesta la opción *Meta*, mientras que cuando fueron ingresadas en español se eligió *Comparación*, y luego la *lengua de comparación*: Español. El resto de la información solicitada en esta sección tiene que ver con el lugar y la fecha de encuesta, así como la duración de la encuesta cuyos rangos van de 30 segundos a 10 minutos. En todas las encuestas ingresadas se eligieron dos minutos.

4.3.5 Ingreso de los datos de la encuesta a LEDICAL

En la sección *Información de la encuesta* se elige *maestro*, *grupo* y *encuesta* y cuando se comienza a ingresar las listas de un grupo nuevo se elige al *alumno* y en el campo *Mostrar centros*, la opción *Nuevo*. De esa manera se despliegan los 53 centros de interés de entre los cuales se eligieron los 15 que fueron ingresados a LEDICAL. No es necesario ingresar todos los centros de interés en una misma sesión en LEDICAL, ya que existe la opción *Parcial* que te da la posibilidad de ingresar otro día al programa y completar el ingreso de todas las listas. Las listas de vocabulario son ingresadas a manera de párrafo dividiendo cada palabra con una coma. Al término se elige *Enviar (Submit)* y puedes continuar ingresando otras listas, o bien cerrar el programa y terminar la tarea posteriormente.

El resto de las funciones del programa se relacionan con la edición de la información ingresada, la cual es muy sencilla de realizar. Se elige la opción *Editar* y después las opciones que te ofrece el programa para editar la información que necesita corregirse o completar. En ocasiones tuve dificultades al momento de leer las listas de los alumnos y no pude realizar el ingreso de las palabras sino hasta que lo consulté con otros profesores. En estos casos, volví a entrar al programa al campo *Editar* para hacer los cambios necesarios. Es importante aclarar que aunque el usuario del programa pueda elegir entre inglés y español para navegar por la página, todas sus instrucciones y campos están en inglés.

4.4 Valoración global de LEDICAL

A grandes rasgos, LEDICAL es un programa computacional de uso práctico y accesible para todo público y además ofrece una gran variedad de lenguas (meta y de comparación) con las cuales se puede trabajar en futuros estudios de léxico disponible. El único inconveniente que presenta el

programa es que en ocasiones no permite editar la información de los grupos cuyas listas de vocabulario han sido ingresadas en su totalidad en el sistema.

El funcionamiento de LEDICAL fue evaluado desde el proceso de edición de las encuestas y su posterior ingreso al programa, debido a que durante el proceso se realizaron cambios y ajustes sustanciales que culminaron en su exitosa ejecución. Este trabajo se presenta como el primero en hacer uso del programa. Esperamos que en los próximos años los diccionarios de léxico disponible de cada una de las lenguas que se estudian en el centro sean recabados y con ello el posterior diseño de material didáctico enfocado a la enseñanza de léxico disponible.

El siguiente capítulo presenta los índices de disponibilidad y frecuencia de dos de los grupos que participaron en el estudio.

Capítulo 5 - Resultados Cuantitativos

En este capítulo muestro los índices de disponibilidad y frecuencias relativas de algunas encuestas con la idea de comprobar la eficacia del programa. Los resultados cuantitativos expuestos en esta parte de la tesis son una aportación evaluadora del funcionamiento de LEDICAL. El análisis de resultados cuantitativos comprende los obtenidos con el cómputo de dos centros de interés de *la ropa y alimentos y bebidas* pertenecientes a los grupos R-003 y S-004 en inglés y español. La obtención de los índices de disponibilidad y frecuencia de los dos grupos elegidos dieron un total de 60 listas que equivalen a 15 centros de interés en inglés y español de los 16 recolectados en total en las encuestas en papel.

LEDICAL se ha presentado en esta tesis como una primera versión o prototipo de un programa computacional perfectible para su uso posterior en el CELE de la UNAM. Por lo tanto, consideramos pertinente mostrar las siguientes encuestas a manera de ejemplo de lo que puede realizarse en términos de análisis cuantitativo, así como los formatos que ofrece LEDICAL para trabajar con los índices de disponibilidad y frecuencia obtenidos en las listas. El índice de disponibilidad, como ya se mencionó anteriormente, se define como aquel que mide la cantidad de veces que una palabra viene a la mente del alumno cuando piensa en un centro de interés o situación comunicativa específica. Los dos índices de frecuencia miden la frecuencia relativa por número de palabras —frecuencia uno o Freq1—, y la frecuencia relativa por número de encuestados —frecuencia dos o Freq2—. El total de las listas analizadas que se obtuvieron de los grupos R-003 y S-004 se encuentran al final de la tesis a manera de apéndice.

Los resultados cuantitativos obtenidos se presentan comparando los datos por lengua (primero en inglés y después en español) entre los dos grupos ya mencionados. La primera encuesta en la lengua meta (inglés) que muestro se tomó del grupo R-003. El centro de interés es

Clothes en el cual existe un total de 52 vocablos. En su mayoría se trata de sustantivos concretos relacionados con prendas de vestir y sólo un par de palabras relacionadas con la compra de ropa, como *price* y *store*. Los vocablos que presentaron mayor índice de disponibilidad son *skirt* con 0.77326 y *T-shirt* con 0.68365. *Swimsuit* fue la palabra con menor índice de disponibilidad: 0.01023, lo cual supone que esta prenda de vestir no viene de manera inmediata a la mente de los encuestados. Ambas frecuencias son las mismas en las primeras cuatro palabras de la lista: *skirt*, *T-shirt*, *pants* y *sock*. La Freq1 tiene un índice de 5.88 y la Freq2 de 90.91 mostrando que estas palabras se presentaron 10 de 11 veces en las listas de vocabulario por número de palabras y de encuestados. Mientras que *swimsuit* sólo se registró una de once veces por palabra y encuestado con una Freq1 de 0.59 y una Freq2 de 9.09. De las 52 palabras de la lista 23 registraron las mismas frecuencias que *swimsuit*; entre éstas enlisto: *uniform*, *panties*, *tight*, *store* y *boat*.

La encuesta anterior fue comparada con la del grupo S-004, cuyos resultados también pertenecen al centro de interés *Clothes* con un total de 51 vocablos. El total de palabras recolectadas en ambas encuestas es casi el mismo (52 en la primera y 51 en la segunda) y puede verse que las palabras con mayor índice de disponibilidad fueron las mismas que en el grupo anterior (R-003). En S-004 *T-shirt* ocupa el primer lugar de la lista con una disponibilidad de 0.74055 y *pants*, en el segundo lugar, con 0.67765. Las Freq1 y Freq2 de la primera palabra (*T-shirt*) son de 6.02 y 90.91 respectivamente, mientras que las dos frecuencias de *pants* son de 5.42 (Freq1) y 81.82 (Freq2). Puede verse que con respecto a la encuesta del grupo R-003, en ésta, *T-shirt* tiene una disponibilidad un poco más alta, lo cual indica que es un vocablo que viene de manera inmediata a la mente de los hablantes cuando piensan en el centro de interés *Clothes*. De hecho, su Freq2 es idéntica en ambos grupos (90.91) comprobando que 10 de 11 alumnos (en cada grupo) pensaron en *T-shirt*. La tercera palabra que presenta mayor disponibilidad del grupo S-004 es *skirt* con un índice de 0.55395 y las Freq1 de 4.22 y Freq2 de 63.64.

En los grupos R-003 y S-004 se encuentran *T-shirt, pants* y *skirt* ocupando los primeros tres lugares de las encuestas en ambos grupos, lo cual indica que dichas prendas de vestir pertenecen al centro de interés *La ropa* en la lengua meta y son referentes de gran importancia para los alumnos de los dos grupos.

La siguiente encuesta en inglés muestra el centro de interés *food and drink* del grupo R-003, en la cual se encuentran en primero y segundo lugar los vocablos *water* y *soda* con sendas disponibilidad de 0.59403 y 0.46883. Los dos índices de frecuencia de *water* fueron de 3.67 en la Freq1 y 72.73 en la Freq2, y presentaron una diferencia considerable con respecto a *soda* (la segunda palabra de la lista). Ésta presenta una Freq1 de 2.75 y una Freq2 de 54.55. Estos ejemplos muestran que las palabras más frecuentes no son las que necesariamente tienen la disponibilidad más alta. El índice de disponibilidad mide la frecuencia de las palabras por número de palabras o encuestados en un grupo y la conecta con una situación comunicativa particular. De tal manera, la palabra *milk* (tercera palabra de la lista) presenta una frecuencia 1 más alta que la de *soda*, la cual ocupa el segundo lugar en la lista. *Milk* presenta una Freq1 de 2.75 y un índice de disponibilidad menor: 0.32670.

El total de palabras de esta lista fue de 105, lo cual demuestra la diversidad léxica que poseen los alumnos en este centro de interés aun en la lengua meta. Las dos frecuencias son constantes en más de la mitad de las palabras. Las palabras *sauce* y *condiment* fueron las que presentaron una menor disponibilidad con índices de 0.1298 y 0.00996, respectivamente. Sus frecuencias relativas muestran que aparecieron una sola vez en toda la lista tanto por número de palabras como de encuestados.

Los siguientes resultados son del mismo centro de interés que el anterior, *food and drink*, del grupo S-004 en la lengua meta. Esta lista contiene un total de 104 vocablos. Existen algunas colocaciones como: *orange juice, chicken breast, french fries, mashed potatoes, pork chop, broil*

vegetables, junk food y malted milk con índices de disponibilidad muy variables. Esta lista es la que contiene un mayor número de frases léxicas simples con respecto a la anterior, que sólo recolectó cuatro: *fried potato, pork chop, fast food, energetic drink y scrambled egg*.

En ambos grupos las frases léxicas no presentan índices de disponibilidad altos. Sin embargo, pueden considerarse importantes como muestras de la interlengua de los alumnos que estudian los cursos avanzados del CELE. Todas las colocaciones evidencian un uso exclusivo y un nivel avanzado de la lengua meta, las cuales muy posiblemente serían desconocidas para un alumno en los niveles básicos.

En el grupo S-004 los dos vocablos que registraron los índices de disponibilidad más altos fueron *soda* (0.49743) y *beer* (0.41300). Sus Frecuencias relativas son de 3.57 (Freq1) y 72.73 (Freq2) para *soda* y 2.68 (Freq1) y 54.55 (Freq2) para *beer*. *Melon* (en inglés) es la palabra con menor índice de disponibilidad (0.01052) y frecuencias (0.45 y 9.09). Los valores registrados en esta palabra son iguales en casi la mitad de las palabras. En ambos grupos la palabra *soda* ocupó los dos primeros lugares de las encuestas con índices de disponibilidad muy parecidos.

Los centros de interés *clothes y food and drink* en inglés presentaron las mismas dos o tres palabras ocupando los primeros lugares de las listas, lo cual supone un conocimiento compartido entre los alumnos encuestados aun y cuando estudian cursos de especialización diferentes.

La primera encuesta en español de grupo R-003 muestra los 62 vocablos recolectados en el centro de interés *La ropa*. *Pantalón, zapato y camisa* fueron las tres primeras palabras de la lista con sendos índices de disponibilidad de 0.72063, 0.65952 y 0.60714. Sus frecuencias relativas son las mismas en los tres vocablos: la Freq1 es de 4.41 y la Freq2, de 90.91. En esta lista también se observa la presencia de frases léxicas simples como *ropa interior, traje de baño, pantalón corto y pantalón de vestir*, lo cual demuestra una diversidad léxica mayor de los alumnos en la lengua materna que en la meta, así como una especificidad más alta en el uso de su

propia lengua. El índice de disponibilidad de estas palabras es mayor que palabras como *pañuelo*, *blazer* y *disfraz* que es la última palabra de la lista. Tiene un índice de disponibilidad de 0.01088 y frecuencias uno y dos de 0.44 y 0.09, respectivamente. Ambas frecuencias fueron las mismas que las de otras palabras como *chal*, *rebozo*, *chaqueta*, *blazer*, *botón* y *bóxer*, entre otras.

El centro de interés *La ropa* de la encuesta en español perteneciente al grupo S-004 incluye un número mayor que la del grupo anterior con un total de 79 vocablos. Las primeras dos palabras de esta lista son *pantalón* y *camisa* con índices de disponibilidad de 0.65673 y 0.64217 respectivamente. Las frecuencias uno y dos de *pantalón* son de 3.73 y 90.91, mientras que *camisa* presentó una frecuencia uno de 4.10 y una frecuencia dos de 100 por ciento. Como ya lo mencioné anteriormente la frecuencia dos registra la frecuencia relativa de una palabra entre el número de encuestados. Es interesante observar que la palabra *camisa* en este grupo registró un 100 en la frecuencia dos, lo cual comprueba que un cien por ciento de los alumnos (hombres y mujeres) pensó en esta palabra al referirse a este centro de interés en particular. En el grupo R-003 la palabra que registró una frecuencia dos del 100 por ciento fue: *falda*.

Las primeras cuatro palabras recolectadas en ambos grupos fueron: *pantalón zapato*, *camisa* y *falda*. Si bien es cierto que estas no fueron idénticas sí presentaron índices de disponibilidad y frecuencia muy cercanos, lo cual indica que vinieron a la mente de los alumnos de manera inmediata cuando pensaron en el centro de interés *la ropa* en su lengua materna.

En cuanto a *alimentos y bebidas* en español se refiere, el grupo R-003 hubo un total de 142 vocablos. Las primeras dos palabras —*agua* y *refresco*— comparten las mismas frecuencias relativas (Freq1 3.23 y Freq2 81.82). El índice de disponibilidad de la palabra *agua* fue de 0.61673 y de *refresco*, 0.55551. A partir del vocablo 31 aproximadamente las frecuencias uno y dos son muy similares, las cuales varían entre un 0.72 y 0.36 (Freq1) y un 18.18 y 9.09 (Freq2). La mayoría de las palabras de esta lista son sustantivos concretos y llama la atención la presencia

de cuatro adjetivos: *hambriento*, *sediento*, *cocido* y *quemado*. Sus índices de disponibilidad son bajos con respecto a los de otras palabras como *taco* (0.10625), *fresa* (0.6853) o *betabel* (0.05662). Lo último indica que los adjetivos son palabras que surgen en la mente del estudiante con menos rapidez que los sustantivos concretos en el centro de interés *alimentos y bebidas*.

En el lugar 123 aparece la palabra *emparedado* con un índice de disponibilidad de 0.02052 y las frecuencias relativas uno y dos de 0.36 y 9.09 respectivamente. El hecho de que *emparedado* pertenezca a una variante dialectal distinta a la del 99 por ciento de los encuestados no modifica los índices obtenidos con respecto a otras palabras como *aguacate* y *horchata*, si entendemos que ambos artículos son más cercanos a la cotidianidad de los alumnos mexicanos que participaron en el estudio. Cabe mencionar que aunque se asume que la palabra *sándwich* corresponde a un uso más común entre dicha población se observó que esta palabra no apareció en el vocabulario recolectado de este grupo. En el S-004 sí encontramos la palabra *sándwich* ocupando el lugar 13 de la lista con un índice de disponibilidad de 0.22535, una freq1 de 2.28 y una freq2 de 45.45. Dicha frecuencia indica que de un total de 11 encuestados cinco pensaron en la palabra cuando llenaron su encuesta.

Los últimos resultados que presento corresponden al centro de interés *alimentos y bebidas* del grupo S-004. Esta encuesta tiene un total de 97 vocablos en español a diferencia de los 142 del anterior. En cuanto a la presentación de los resultados cuantitativos obtenidos en esta encuesta, podemos mencionar las primeras dos palabras en la lista: *refresco* y *carne*. Ambas presentan índices de disponibilidad de 0.47997 y 0.40864 respectivamente. Sus frecuencias relativas son las mismas; la Freq1 presenta un valor de 3.65 y la Freq2 de 72.73. A partir de la palabra 39 aproximadamente la frecuencia uno es de 0.46 y la frecuencia dos, de 9.09 aunque en ambos índices existe una que otra variante. Las últimas cinco palabras de la lista son frutas: *piña*, *cereza*, *plátano*, *papaya* y *guayaba*. Esta última registró un índice de disponibilidad de 0.00987 y

las mismas frecuencias relativas antes mencionadas. El resto de las encuestas de ambos grupos se presenta al final de esta tesis en los apéndices.

Los resultados cuantitativos expuestos en este capítulo son meramente descriptivos, ya que nos acercan al tipo de léxico disponible que poseen los alumnos en niveles avanzados en el CELE de la UNAM, tanto en la lengua meta como en la propia. Asimismo validan el funcionamiento de LEDICAL como base de datos y programa computacional diseñado exclusivamente para la eventual recolección del diccionario multilingüe de lenguas extranjeras del CELE de la UNAM. La relevancia de los datos obtenidos se relaciona con el reconocimiento del léxico que los alumnos poseen de acuerdo al nivel de lengua que tienen y el conocimiento léxico que poseen en su lengua materna, el cual será planteado en el siguiente capítulo del presente trabajo.

A continuación presentaré y discutiré algunos puntos importantes que surgieron a partir de la obtención de los índices de disponibilidad y frecuencia de las encuestas ya mencionadas.

Capítulo 6 - Discusión y Conclusiones

A continuación presento algunas nociones surgidas a partir de los resultados cuantitativos presentados en el capítulo anterior, así como del proceso de recolección, edición y almacenamiento del léxico disponible en español e inglés de los alumnos universitarios mexicanos a fin de evaluar la eficacia del desarrollo tecnológico LEDICAL. Por un lado, la descripción del proceso de edición e ingreso de las encuestas y su posterior cómputo comprobaron el buen funcionamiento del programa. Por otro, se obtuvieron los índices de disponibilidad y frecuencia de las listas de vocabulario de dos grupos participantes en el estudio. Ambas vertientes se reconocen como maneras diferentes e igualmente válidas de estudio del léxico disponible de estos alumnos.

Se logró recolectar alrededor de 200 encuestas en papel con sendos 16 centros de interés en inglés y español. Sin embargo, las encuestas ingresadas a LEDICAL fueron sólo 154 que equivalen a 77 encuestados de 7 grupos diversos de los cursos de especialización. Los diccionarios en ambas lenguas son los primeros que existen en México a nivel universitario y están disponibles en LEDICAL. El ingreso de las encuestas al programa representa un gran avance para conocer el léxico disponible de los alumnos del CELE. Los docentes de lenguas extranjeras e interesados en los estudios de léxico podrán acceder a la base de datos y con base en ello contribuir a la elaboración de ejercicios didácticos que estimulen la adquisición de vocabulario en lenguas extranjeras.

A continuación presento una relación del total de grupos y alumnos encuestados, cuyas listas de vocabulario ya fueron ingresadas a LEDICAL para su eventual cómputo.

Grupos dados de alta en LEDICAL

	Grupo	Alumnos encuestados	Encuestas en inglés y español
	TOEFL-1	16	32
	L-001	14	28
	G-001	9	18
	S-001	8	16
	S-003	8	16
	R-003	11	22
	S-004	11	22
	Total	77	154

La segunda etapa de edición fue la más difícil en cuanto al cuidado que se requiere para editar las listas de vocabulario. Los criterios de edición son muy específicos en los estudios de léxico disponible, por lo que se requiere de un entrenamiento exigente y laborioso por parte del encuestador. Éste debe aprender los criterios de edición antes mencionados cuidando la homogenización de ciertos extranjerismos o neologismos, así como influencias culturales y sociales que se encuentran comúnmente en las listas de vocabulario de léxico disponible. Antes de la edición, las encuestas en inglés y español contenían sustantivos principalmente con flexiones y derivaciones propias de la lengua. La edición toma en cuenta las palabras o “unidades de cita” sin flexiones ni derivaciones, es decir, los vocablos. De acuerdo a Luis Fernando Lara, los vocablos son muy relevantes en la investigación lexicográfica porque logran definir los paradigmas que existen en una lengua. Por lo tanto, es importante aclarar que aunque con el proceso de edición hay información que se pierde al momento de eliminar flexiones de género y número para su posterior ingreso a LEDICAL, también es cierto que al trabajar con los vocablos se puede lograr una “descripción y primera teorización que sirve como instrumento abstracto de la investigación lingüística y de la descripción lexicográfica”. (Lara, 2006)

Al momento de realizar la edición de las listas de vocabulario, el encuestador se acerca por primera vez al léxico bilingüe de los alumnos participantes cuyo análisis, y de acuerdo a Luis Fernando Lara, plantea un problema de objetividad porque se trata de un fenómeno de la mente humana. El tipo de léxico que se encontró en dichas listas demuestra que efectivamente los signos pueden usarse de maneras ilimitadamente diferentes, ampliando o reduciendo su significado o modificándolo metafóricamente. El análisis que se realizó sobre los significados obtenidos efectivamente varían de acuerdo al contexto (i.e. los centros de interés), a los conocimientos y a la sensibilidad del lingüista.

Sin embargo, se encontraron características comunes en todas las encuestas como las flexiones de género y número de acuerdo a los centro de interés que llenaban, tales como *alimentos y bebidas, el campo, la ciudad y profesiones y oficios*. Los alumnos se han creado un concepto de la realidad y con ello una categorización semántica. Es precisamente por eso que el llenado de las encuestas implica un ejercicio de autocensura por parte de los alumnos porque el campo semántico les indicó lo que podían o no escribir basándose en su experiencia y conocimientos lingüísticos. Con base en lo anterior, podemos preguntarnos lo siguiente: ¿Qué diferencias se establecen entre el léxico disponible en español y el léxico disponible en inglés? Las diferencias del léxico en español e inglés de los alumnos del CELE se deben a que la variabilidad del contexto influye en el dinamismo del tipo de léxico que se recaba.

La actividad que realizan los alumnos es la de “recordar” ciertas palabras dependiendo una situación comunicativa, por ejemplo, en el centro de interés *la ropa* hubo una presencia mayor de sustantivos mas no de adjetivos. O en el centro de interés *alimentos y bebidas* en inglés casi no hay presencia de palabras de origen hispano, como *tacos, papaya o tortilla* que ya forman parte del léxico anglófono de Estados Unidos o Inglaterra.

Las diferencias en los índices de disponibilidad de las primeras palabras de las encuestas comparadas fue muy baja, por lo que asumimos una similitud mayor de la esperada entre grupos y lenguas de estudio. Esto puede deberse a que los alumnos de estos dos grupos comparten características comunes, tales como el nivel de interlengua, que se reflejan en el tipo de léxico que se recabó. Ahora bien, las posibles razones que pudieron haber influido en el tipo de vocabulario que se recolectó son especulativas porque para cerciorarnos sobre el proceso cognitivo por el que pasaron los alumnos y la selección en el vocabulario es necesario solicitar a los alumnos que hagan una reflexión semántica posterior al llenado de las encuestas para conocer las decisiones que tomaron al momento de completar un centro de interés.

El objetivo de esta investigación es el estudio del léxico disponible de los alumnos y su definición a partir de la zona —activa o pasiva— que ocupan de acuerdo al modelo Gravitacional de Daniel Gile. Asimismo, los resultados obtenidos se presentan como indicios del nivel de interlengua que tienen los alumnos de inglés avanzado del CELE por medio de la identificación de omisiones ortográficas o creación de extranjerismos adoptados formalmente. Estos últimos no tienen una presencia estadística elevada en los resultados cuantitativos obtenidos y sus índices de disponibilidad y frecuencia no están muy alejados a los de otras palabras de uso más “frecuente”. Recordemos la encuesta en español del grupo R-003, cuyo centro de interés *la ropa* recolectó palabras de origen anglófono como *blazer* y *bóxer* con índices de disponibilidad muy parecidos a las de otras de origen hispano como *chal* y *rebozo*, y frecuencias relativas idénticas (Freq1 de 0.44 y Freq2 de 9.09).

En su *Curso de Lexicología*, Luis Fernando Lara habla del léxico pasivo como aquel que el hablante usa solamente en situaciones comunicativas específicas, el cual se diferencia del activo en que es menos frecuente mas no así menos conocido. Tanto Lara como Gile en su Modelo Gravitacional identifican este tipo de léxico (el pasivo) como dinámico y dependiente de

la experiencia del hablante y de la situación comunicativa en la que usa una u otra palabra; es decir, de acuerdo al centro de interés. Sin embargo, el léxico disponible de nuestros sujetos de estudio está definido a partir del lugar que ocupa una palabra en la lista y, por ende, de su grado de disponibilidad. En un mismo centro de interés existen vocablos activos y pasivos.

De lo observado en los datos estadísticos es posible afirmar que el léxico disponible que conocen los alumnos en su lengua materna es mayor al que tienen en la lengua meta —el inglés— como muestran las listas: el número de vocablos en español fue mayor. En siete de las ocho encuestas que fueron comparadas por centro de interés y lengua se recolectó un mayor número de vocablos en la lengua materna que en la meta. La excepción fue la encuesta de *Food and drink* del grupo S-004 en inglés que contiene 104 vocablos a diferencia de los 97 encontrados en la encuesta del grupo S-004 en la lengua materna. Las diferencias cuantitativas que existen entre las encuestas en inglés y español no fueron considerable. Esto podría refutar la idea de que las palabras en español son mayores en cantidad con índices de disponibilidad más altos por ser la lengua materna de los alumnos.

La “corrección gráfica” y creación de extranjerismos adoptados formalmente fueron los indicios hallados al momento de la edición de las encuestas sobre el nivel de interlengua que tienen los alumnos de inglés avanzado del CELE. Las correcciones gráficas presentes en estas encuestas son omisiones de letras como en las palabras *difficult* y *stubborn* antes mencionadas. La influencia de la lengua materna —español— en el inglés fue casi nula por lo que puede asegurarse que el nivel de dominio de inglés de los alumnos es el esperado: B1 de acuerdo al Marco Común de Referencia Europeo. En las encuestas en español, en cambio, hubo muy pocas “correcciones gráficas”. Por otro lado, resulta complicado medir los índices de disponibilidad y frecuencia de los neologismos partiendo de la lógica contemplados en los criterios de edición de las encuestas propuesto por Samper (1998) porque en algunos casos desaparecieron con el

proceso de edición de las listas. Los datos cuantitativos arrojados de los extranjerismos escritos por los alumnos—*leggings*, *bóxer*, *pants* o *shorts*— no presentaron índices altos y fueron similares a los de otras palabras de origen hispano.

En un futuro sería interesante comparar la presencia de extranjerismos y neologismos de los alumnos del CELE que cursan los primeros niveles de inglés con los de esta investigación partiendo de la suposición de que en el caso de los primeros se presentaría una mayor cantidad de estos. A partir de esta investigación se pueden diseñar estudios de léxico disponible realizando comparaciones lingüísticas por edad, sexo, lengua y curso de especialización que estudian los alumnos, así como establecer diferencias etnográficas, etnolingüísticas y culturales. La investigación que dio origen al *Français Fondamental* de Gougenheim y Michéa fue precedida por una gran cantidad de comparaciones estadísticas por edad, sexo y región de los sujetos de estudio. Nuestros diccionarios dan cabida a este mismo tipo de estudios.

La metodología de este trabajo contempla la elección de los centros de interés dentro de otras 50 opciones aproximadamente, así como el tiempo de llenado de las encuestas de 2 minutos, la edición y el uso de LEDICAL como base de datos. La elección de los centros de interés se realizó en consenso con los miembros del Seminario de Léxico Disponible. Ninguno de los centros representó un problema para encuestadores y alumnos al momento del llenado de las encuestas y su ingreso al programa.

Las dos fases en las que se aplicaron las encuestas; la primera en 2009 y la segunda en 2012 no afectaron en lo absoluto el proceso de edición e ingreso de las listas a LEDICAL. Sin embargo, se recomienda que se revisen las listas de vocabulario en papel previo a su ingreso al programa, ya que en ocasiones falta información en las encuestas que se aplican y resulta más complicado volver a trabajar con dichos grupos. De la aplicación de las encuestas al ingreso

puede tomar algunos días o semanas por lo que los maestros o investigadores podrían encontrar omisiones importantes al momento de realizar los cálculos o análisis pertinentes.

En la tercera etapa se realizó el almacenamiento de la listas a LEDICAL. Dicho proceso tomó varios meses de trabajo en constante comunicación con Jim Frey porque nunca antes se había puesto a prueba LEDICAL. Su funcionamiento es sencillo y accesible para cualquier persona que tenga conocimientos básicos de computación. LEDICAL cuenta con la usabilidad de la que se habla en informática y tecnología para referirse a “la capacidad de un software de ser comprendido, aprendido, usado y ser atractivo para el usuario, en condiciones específicas de uso”. Esta definición de la Organización Internacional para la Estandarización (ISO) explica sencillamente los atributos con los que debe contar cualquier programa computacional de uso cotidiano para el usuario. LEDICAL fue creado exclusivamente para realizar el proyecto multilingüe de léxico disponible en el CELE de la UNAM y durante estos tres años de prueba no ha registrado ninguna falla considerable. El siguiente paso a seguir es continuar trabajando en la obtención de los índices de disponibilidad y frecuencia del resto de las encuestas que ya se encuentran en LEDICAL y de esa manera realizar los cambios necesarios junto con el ingeniero Frey.

La relevancia de este proyecto es la obtención de los diccionarios de español e inglés de siete grupos de inglés avanzado del CELE de la UNAM; con un total de 15,635 vocablos recolectados en 154 encuestas en inglés y español e ingresados manualmente a la base de datos por la autora de este trabajo. Los alumnos encuestados suman un total de 77 alumnos LEDICAL es una herramienta única en su tipo, pues fue creado exclusivamente para la obtención del diccionario multilingüe del centro. Con este proyecto se ha contribuido a la consolidación de una base de datos léxica de uso exclusivo para el CELE de la UNAM y el único a nivel universitario realizado en México hasta la fecha. La obtención de los diccionarios en español e inglés ha

iniciado con este proyecto, el cual resulta de la mayor importancia para el desarrollo de las investigaciones lexicográficas y de la enseñanza de lenguas extranjeras a nivel didáctico y en la recolección de los diccionarios multilingües del CELE.

Referencias bibliográficas

- Aitchison, J. (1987). *Words in the mind: an introduction to the mental lexicon*. Oxford: Blackwell.
- Anthony, E. M. (1973) *Towards a theory of lexical meaning: An essay*. USA: University of Pittsburgh
- Carcedo, Alberto. (2000). *Disponibilidad léxica en español como LE: el caso finlandés*. Turku: Turun Yliopisto.
- _____ . (1999). “Análisis de errores léxicos del español en la interlengua de los finlandeses”. En *Español como lengua extranjera: enfoque comunicativo y gramática*. Tomás Jiménez Juliá (Ed). Universidad Santiago de Compostela. ASELE.
- _____ . (1999). Estudio comparativo del vocabulario español (LE) disponible de estudiantes finlandeses y el de la sintopía madrileña: propuestas didácticas. *DEA*, 1, 73-87.
- _____ . (1997-1998). “Desarrollo de la competencia léxica en español LE: Análisis de cuatro fases de disponibilidad”. *Pragmalingüística*, 5-6, 75-94.
- Cifuentes Honrubia, J. L. (1992). “Teoría de prototipos y funcionalidad semántica”. En *E.L.U.A.*, 8, 133-177.
- Coady, J. & Huckin, T. (1997). *Second Language Vocabulary Acquisition. A Rationale for Pedagogy*. Australia: Cambridge University Press.
- Colín, M. et al. (2013.) *Diccionarios de léxico disponible en español en comparación con los diccionarios de lengua extranjera (inglés, portugués) de alumnos universitarios hispanohablantes*. Memoria del 15 Encuentro Nacional de Profesores de Lenguas Extranjeras (ENPLE).
- Colín, M. et al (2009). *Estudios de Léxico Disponible (CBHC091127)* Departamento de Lingüística Aplicada. CELE: UNAM.
- Davesne, A. (1955). *Le français élémentaire, nouveau cours de langage pour les classes de débutants dans les écoles africaines*. Paris: Istra.
- Diccionario del Español de México (DEM) <http://dem.colmex.mx>, El Colegio de México, A.C., [8 de Septiembre de 2015].
- Diccionario de la lengua española (RAE) <http://www.rae.es/rae.html>, Real Academia Española. [8 de Septiembre de 2015].

- Ellis, R. (2003). *Task-based Language Learning and Teaching*. China: Oxford University Press.
- Fourré, P. (1956). *Premier dictionnaire en images*. Paris: Didier.
- Frey Pereyra, L. H. (2008). *El léxico disponible en los textos de alumnos de ELE*. Tesis doctoral no publicada. Universidad de Alcalá, España.
- Frey Pereyra L. H. (2009). “Corpus multilingüe de léxico disponible de los estudiantes del CELE.” (Proyecto Posdoctoral, Posgrado en Lingüística, UNAM, CONACYT.)
- Fries, C. C. (1945). *Teaching and learning English as a foreign language*. Ann Arbor: University of Michigan Press.
- Gile, D. (2009). *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam: John Benjamins Publishing Company.
- Gougenheim, G. & R. Michéa. (et al.) (1967). *L'Élaboration du Français Fondamental (1er degré)*. Paris: Didier.
- Hymes, D. (1972). “On communicative competence”. In *Sociolinguistics*, eds. J.B. Pride & J. Holmes, 269-293. New York: Penguin.
- Ilari, R. (2012). “Campos Lexicais”. En *Introdução ao estudo do Léxico brincando com as palavras*. Brasil: Editura Catexto.
- Jiménez Catalán, R. M. (ed.) (2013). *Lexical Availability in English and Spanish as a Second Language*. Netherlands: Springer.
- Kelly, L.G. (1969). *Centuries for Language teaching*. Rowley, MA: Newbury House.
- Krashen, S. (1989). “We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis.” *Modern Languages Journal*, 73(4), 440-464.
- Krashen, S. D. & Tracy. D. Terrel. (1996). *The Natural Approach: Language Acquisition in the Classroom*. USA: Bloodaxe Books Ltd.
- López Chávez, J. (2003). *¿Qué te viene a la memoria? La disponibilidad léxica: teoría, métodos y aplicaciones*. Zacatecas: Universidad Autónoma de Zacatecas.
- López Morales, H. (1999). *Léxico disponible de Puerto Rico*. Madrid: Arco/Libros.
- Mahlberg, M. (2006). “Lexical cohesion. Corpus linguistic theory and its application in English language teaching”. *International Journal of Corpus Linguistics*, 11:3, 363-283.

- Mauger, G. & G. Gougenheim. (1956). *Le Français élémentaire, méthode progressive du français parlé*. Paris: Hachette.
- Nattinger, J.R. & De Carrico, J.S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Palmer, H. E. (1968). *The scientific study and teaching of foreign languages*. England: Oxford University Press.
- Prendergast, T. (1864). *The mastery of languages or, the art of speaking foreign tongues idiomatically*. London: R. Bentley.
- Rastier, F. (1991b). “Catégorization, typicalité et lexicologie. Préliminaires critiques”. Ed. D. Dubois. 259-277
- _____ (1991c). *Sémantique et recherches cognitives*. Paris: PUF.
- Rivers, W.M. (1981). *Teaching Foreign-Language Skills*. Chicago: University of Chicago Press.
- Samper, Marta. (2001-2002). “La variable ‘lengua materna’ en la disponibilidad léxica de estudiantes de ELE”. DEA, 3-4, 147-172.
- _____. (2001) “Dificultades de los estudios de disponibilidad léxica en ELE: los criterios de edición de materiales”. En Nuevas aportaciones al estudio de la lengua española. Investigaciones Filológicas. J.A. Bartol, Hernández. et al (Eds.). España. Luso-Española de Ediciones.
- Samper, José Antonio. (1997). “Criterios de edición de léxico disponible: sugerencias. Lingüística”, 10, 311-333.
- Sauver, L. (1874). *An introduction to the teaching of living languages without grammar or dictionary*. Boston: Schoenhof & Moeller.
- Selinker, Larry. (1972). “Interlanguage.” *International Review of Applied Linguistics in Language Teaching* 10(1-4): 209-232.
- Sweet, H. (1964). *The practical study of languages: A guide for teachers and learners*. London: Oxford University Press.
- Twadell, F. (1980). “Vocabulary expansion in the TESOL classroom”. En *Readings of English as a second language*, ed. K. Croft, 439-457. Cambridge: Winthrop.
- Vander Beke, G.E. (1935). *French Word Book*. New York: The MacMillan Company.

- West, M. (1930). Speaking-vocabulary in a foreign language. *Modern Language Journal*, 14, 509-521.
- Wilkins, D. (1974). "Linguistics in language teaching". En *Second-language learning and teaching*. London: Edward Arnold.
- Zimmerman, C. B. (1997). "Historical trends in second language vocabulary instruction". En Coady J. & Huckin T. (Eds.) *Second Language Vocabulary Acquisition*. (pp. 6-19). UK: Cambridge University Press.

Apéndice 1

COLOCACIONES EN INGLÉS

Las colocaciones son tipos de unidades fraseológicas que constan de grupos de palabras que se combinan de manera específica.

En inglés, por ejemplo, hay *tall trees* y no *high trees*. El adjetivo *high* podría quedar para referirse a la *temperatura alta*: *high temperature*.

La combinación pertinente de dos o más palabras puede ser de distintos tipos: verbo + sustantivo (fly a kite); adjetivo + sustantivo (Strong coffee); adverbio + verbo (to boldly go); adverbio + adjetivo (totally different); adjetivo + preposición (similar to) y sustantivo y sustantivo (a collocation dictionary).¹⁷

ACTIVIDADES PROPUESTAS A PARTIR DE LAS ENCUESTAS DE LÉXICO DISPONIBLE DE LOS ALUMNOS DEL CELE

A. Del centro de interés *los medios de transporte* de tu encuesta (de léxico disponible) escoge un verbo para crear colocaciones de tipo verbo + sustantivo. Haz tantas combinaciones como puedas y escríbelas sobre las líneas.

- | | | |
|-----------------------------|----------|-------|
| 1. bus | a. ride | _____ |
| 2. car | b. take | _____ |
| 3. airplane | c. fly | _____ |
| 4. helicopter | d. drive | _____ |
| 5. subway | e. run | _____ |
| 6. ship | | _____ |
| 7. boat | | _____ |
| 8. bike | | _____ |
| 9. motorcycle ¹⁸ | | _____ |

¹⁷ Los tipos de colocaciones fueron tomados de la página de Internet: <http://stewardess.inhatc.ac.kr/philoint/usage/collocation-4.htm>, en su versión del día 22 de marzo de 2011.

¹⁸ Estos ejemplos los obtuve de una encuesta de léxico disponible de los alumnos de inglés VI del CELE de la UNAM. Es el léxico que ya conocen porque fue utilizado en el llenado de las encuestas.

Apéndice 2

Encuestas de léxico disponible de los grupos R-003 y S-004

11/9/15, 1:48 PM

LEXICAL AVAILABILITY				
Ramirez Alfaro, Anahi		Date: 09-11-2015 1:42:14 PM		
Poll: 26-10-2012 10:00:00		Group: R-003 - English		
Type: Target		Respondents: 11/11		
Center: Clothes		Place: Centro de Enseñanza de Lenguas E		
#	Word	Index	Freq1	Freq2
1	skirt	0.77326	5.88	90.91
2	T-shirt	0.68365	5.88	90.91
3	pants	0.67028	5.88	90.91
4	sock	0.54801	5.88	90.91
5	shoe	0.41799	4.12	63.64
6	sweater	0.39807	5.29	81.82
7	jacket	0.35619	5.29	81.82
8	jeans	0.34885	4.12	63.64
9	dress	0.31844	4.12	63.64
10	shirt	0.28620	3.53	54.55
11	suit	0.23870	4.12	63.64
12	scarf	0.20741	2.35	36.36
13	hat	0.20526	4.12	63.64
14	shorts	0.15058	2.94	45.45
15	tennis	0.13528	1.76	27.27
16	coat	0.11588	2.35	36.36
17	trousers	0.10846	1.18	18.18
18	sneaker	0.09180	1.18	18.18
19	blouse	0.09121	2.35	36.36
20	underwear	0.09072	2.35	36.36
21	sandal	0.08387	1.76	27.27
22	cap	0.08345	1.18	18.18
23	leggings	0.08305	1.18	18.18
24	boat	0.08103	0.59	9.09
25	glove	0.07633	1.76	27.27
26	cotton	0.06943	1.18	18.18
27	high heels	0.06438	0.59	9.09
28	nylon	0.06189	1.18	18.18
29	uniform	0.05739	0.59	9.09
30	tie	0.05684	1.18	18.18
31	tennis shoe	0.05662	1.18	18.18
32	costum	0.04064	0.59	9.09
33	panties	0.04064	0.59	9.09
34	tight	0.04064	0.59	9.09

35	store	0.03623	0.59	9.09
36	wonderbra	0.03623	0.59	9.09
37	color	0.03229	0.59	9.09
38	price	0.03229	0.59	9.09
39	top	0.03229	0.59	9.09
40	belt	0.03061	1.18	18.18
41	leather jacket	0.02879	0.59	9.09
42	stocking	0.02879	0.59	9.09
43	pyjamas	0.02287	0.59	9.09
44	brassiere	0.02039	0.59	9.09
45	sleepers	0.02039	0.59	9.09
46	purse	0.01817	0.59	9.09
47	swimming suit	0.01817	0.59	9.09
48	backpack	0.01444	0.59	9.09
49	boot	0.01444	0.59	9.09
50	thong	0.01287	0.59	9.09
51	bikini	0.01147	0.59	9.09
52	swimsuit	0.01023	0.59	9.09

LEXICAL AVAILABILITY				
Ramirez Alfaro, Anahi			Date: 10-11-2015 9:56:49 AM	
Poll: 26-10-2012 12:00:00			Group: R-003 - English	
Type: Comparison: Spanish			Respondents: 11/11	
Center: Clothes			Place: Centro de Enseñanza de Lenguas E	
#	Word	Index	Freq1	Freq2
1	pantalón	0.72063	4.41	90.91
2	zapato	0.65952	4.41	90.91
3	camisa	0.60714	4.41	90.91
4	falda	0.54151	4.85	100
5	playera	0.45499	3.96	81.82
6	chamarrá	0.43630	4.85	100
7	suéter	0.39952	3.08	63.64
8	vestido	0.38706	3.52	72.73
9	tenis	0.32914	3.52	72.73
10	blusa	0.28089	2.64	54.55
11	short	0.27830	3.08	63.64
12	calcetín	0.27127	3.08	63.64
13	media	0.26570	3.08	63.64
14	bufanda	0.23751	3.08	63.64
15	abrigo	0.23638	3.08	63.64
16	brasier	0.23637	2.20	45.45
17	calceta	0.23236	1.76	36.36
18	saco	0.22362	2.64	54.55
19	camiseta	0.20970	1.32	27.27
20	gorra	0.18215	2.20	45.45
21	tacón	0.17356	1.32	27.27
22	sombrero	0.17079	2.20	45.45
23	jeans	0.14163	0.88	18.18
24	calzón	0.13272	1.76	36.36
25	guante	0.12700	2.64	54.55
26	traje	0.11627	1.32	27.27
27	chaleco	0.11550	1.76	36.36
28	corbata	0.10102	1.76	36.36
29	sudadera	0.09605	1.76	36.36
30	pantimedia	0.09384	0.88	18.18
31	pantaleta	0.09374	0.88	18.18
32	cinturón	0.08316	1.76	36.36
33	ropa interior	0.07915	0.88	18.18
34	sandalia	0.07470	0.88	18.18

35	tin	0.06972	0.44	9.09
36	bota	0.06587	1.32	27.27
37	huarache	0.06382	0.44	9.09
38	pijama	0.06382	0.44	9.09
39	trusa	0.05347	0.44	9.09
40	uniforme	0.05347	0.44	9.09
41	gorro	0.04894	0.44	9.09
42	botón	0.04480	0.44	9.09
43	bóxer	0.04480	0.44	9.09
44	maquillaje	0.03753	0.44	9.09
45	traje de baño	0.03631	0.88	18.18
46	sombra	0.03436	0.44	9.09
47	bermuda	0.02879	0.44	9.09
48	malla	0.02879	0.44	9.09
49	pantalón de vestir	0.02879	0.44	9.09
50	chaqueta	0.02635	0.44	9.09
51	rompevientos	0.02635	0.44	9.09
52	pantalón corto	0.02412	0.44	9.09
53	mascada	0.02377	0.88	18.18
54	zapatilla	0.01850	0.44	9.09
55	blazer	0.01550	0.44	9.09
56	pañuelo	0.01550	0.44	9.09
57	pulsera	0.01550	0.44	9.09
58	chal	0.01418	0.44	9.09
59	collar	0.01418	0.44	9.09
60	arete	0.01298	0.44	9.09
61	rebozo	0.01298	0.44	9.09
62	disfraz	0.01088	0.44	9.09

LEXICAL AVAILABILITY

Ramirez Alfaro, Anahi
 Poll: 26-10-2012 10:00:00
 Type: Target
 Center: Food and drink

Date: 09-11-2015 1:46:29 PM
 Group: R-003 - English
 Respondents: 11/11
 Place: Centro de Enseñanza de Lenguas E

#	Word	Index	Freq1	Freq2
1	water	0.59403	3.67	72.73
2	soda	0.46883	2.75	54.55
3	milk	0.32670	3.67	72.73
4	spaghetti	0.32157	2.29	45.45
5	juice	0.30668	3.21	63.64
6	hot dog	0.29530	2.29	45.45
7	coffee	0.24926	2.75	54.55
8	meat	0.22248	1.83	36.36
9	chicken	0.21501	3.21	63.64
10	rice	0.19908	1.83	36.36
11	cheese	0.19875	1.38	27.27
12	apple	0.18453	1.83	36.36
13	sandwich	0.17338	1.38	27.27
14	lettuce	0.17104	1.38	27.27
15	hamburger	0.16082	1.38	27.27
16	vegetable	0.15932	1.38	27.27
17	strawberry	0.14703	0.92	18.18
18	lemon	0.13877	1.83	36.36
19	tea	0.13344	1.83	36.36
20	cake	0.13140	1.83	36.36
21	soup	0.13105	1.38	27.27
22	carrot	0.13042	1.38	27.27
23	bread	0.12428	1.38	27.27
24	tomato	0.12131	1.38	27.27
25	beer	0.11969	0.92	18.18
26	bean	0.11072	0.92	18.18
27	salad	0.10956	0.92	18.18
28	chocolate	0.10933	1.38	27.27
29	cookie	0.10674	1.38	27.27
30	jam	0.10509	0.92	18.18
31	alcohol	0.10252	0.92	18.18
32	coke	0.09091	0.46	9.09
33	fried potato	0.09091	0.46	9.09
34	meal	0.09091	0.46	9.09

35	pineapple	0.08960	0.92	18.18
36	ice-cream	0.08657	1.38	27.27
37	Mexican	0.08321	0.46	9.09
38	pork chop	0.08330	0.92	18.18
39	onion	0.08039	0.92	18.18
40	sugar	0.08039	0.92	18.18
41	salt	0.07625	0.92	18.18
42	blueberry	0.07617	0.46	9.09
43	cosmopolitan	0.07617	0.46	9.09
44	orange	0.07617	0.46	9.09
45	candy	0.07198	1.38	27.27
46	fish	0.07049	1.38	27.27
47	banana	0.06972	0.46	9.09
48	dessert	0.06970	1.83	36.36
49	dish	0.06382	0.46	9.09
50	fast food	0.06382	0.46	9.09
51	hot cake	0.06382	0.46	9.09
52	pizza	0.06382	0.46	9.09
53	squash	0.06382	0.46	9.09
54	pie	0.06314	0.92	18.18
55	steak	0.06308	0.92	18.18
56	cereal	0.05961	0.92	18.18
57	blackberry	0.05841	0.46	9.09
58	donut	0.05841	0.46	9.09
59	garbage	0.05841	0.46	9.09
60	energetic drink	0.05347	0.46	9.09
61	melon	0.05347	0.46	9.09
62	fruit	0.04894	0.46	9.09
63	milkshake	0.04894	0.46	9.09
64	red bull	0.04894	0.46	9.09
65	potato	0.04728	0.92	18.18
66	pork	0.04619	0.92	18.18
67	gum	0.04480	0.46	9.09
68	honey	0.04480	0.46	9.09
69	meatball	0.04480	0.46	9.09
70	scotch	0.04480	0.46	9.09
71	egg	0.04428	0.92	18.18
72	cream	0.04261	0.92	18.18
73	lemonade	0.04101	0.46	9.09
74	pasta	0.04101	0.46	9.09

75	spicy	0.04101	0.46	9.09
76	tangerine	0.04101	0.46	9.09
77	sausage	0.03870	0.92	18.18
78	cocoa	0.03753	0.46	9.09
79	grilled	0.03753	0.46	9.09
80	mutton	0.03753	0.46	9.09
81	peach	0.03753	0.46	9.09
82	pumpkin	0.03753	0.46	9.09
83	scrambled egg	0.03753	0.46	9.09
84	watermelon	0.03753	0.46	9.09
85	bacon	0.03714	0.92	18.18
86	shot	0.03436	0.46	9.09
87	wine	0.03436	0.46	9.09
88	hungry	0.03145	0.46	9.09
89	avocado	0.02879	0.46	9.09
90	starving	0.02879	0.46	9.09
91	spinach	0.02635	0.46	9.09
92	thirsty	0.02635	0.46	9.09
93	Coca-Cola	0.02412	0.46	9.09
94	barbecue	0.02412	0.46	9.09
95	bisteck	0.02412	0.46	9.09
96	pancake	0.02412	0.46	9.09
97	restaurant	0.02208	0.46	9.09
98	cocktail	0.02021	0.46	9.09
99	kitchen	0.02021	0.46	9.09
100	ham	0.01850	0.46	9.09
101	oil	0.01850	0.46	9.09
102	celery	0.01693	0.46	9.09
103	hot	0.01693	0.46	9.09
104	sauce	0.01298	0.46	9.09
105	condiment	0.00996	0.46	9.09

LEXICAL AVAILABILITY				
Ramirez Alfaro, Anahi		Date: 10-11-2015 9:58:03 AM		
Poll: 26-10-2012 12:00:00		Group: R-003 - English		
Type: Comparison: Spanish		Respondents: 11/11		
Center: Food and drink		Place: Centro de Enseñanza de Lenguas E		
#	Word	Index	Freq1	Freq2
1	agua	0.61673	3.23	81.82
2	refresco	0.55551	3.23	81.82
3	pollo	0.55099	3.23	81.82
4	carne	0.47522	2.51	63.64
5	leche	0.42577	2.51	63.64
6	jugo	0.34874	2.15	54.55
7	café	0.34713	2.15	54.55
8	pescado	0.33608	2.51	63.64
9	pan	0.25938	2.15	54.55
10	huevo	0.25689	2.15	54.55
11	arroz	0.21318	1.08	27.27
12	tortilla	0.21026	2.51	63.64
13	zanahoria	0.20787	1.43	36.36
14	cerveza	0.19111	1.43	36.36
15	ensalada	0.18837	1.43	36.36
16	alcohol	0.18087	1.08	27.27
17	fruta	0.17737	1.08	27.27
18	té	0.16414	1.08	27.27
19	dulce	0.14644	1.43	36.36
20	chocolate	0.14415	1.08	27.27
21	pastel	0.13810	1.43	36.36
22	pizza	0.13760	1.08	27.27
23	cereal	0.13654	1.43	36.36
24	galleta	0.13578	1.43	36.36
25	vino	0.13442	0.72	18.18
26	postre	0.13101	1.08	27.27
27	hamburguesa	0.13008	1.43	36.36
28	malteada	0.12994	0.72	18.18
29	verdura	0.12540	0.72	18.18
30	queso	0.12350	1.43	36.36
31	sopa	0.12144	0.72	18.18
32	jamón	0.12022	0.72	18.18
33	salchicha	0.11714	0.72	18.18
34	papa	0.11441	0.72	18.18

35	tomate	0.11104	0.72	18.18
36	taco	0.10625	1.08	27.27
37	enchilada	0.10462	0.72	18.18
38	pavo	0.09286	0.72	18.18
39	brocheta	0.09091	0.36	9.09
40	mermelada	0.09091	0.36	9.09
41	platillo	0.09091	0.36	9.09
42	plato	0.09091	0.36	9.09
43	barbecue	0.08496	0.36	9.09
44	traste	0.08496	0.36	9.09
45	vaso	0.08496	0.36	9.09
46	torta	0.08151	1.08	27.27
47	helado	0.07982	0.72	18.18
48	albóndiga	0.07941	0.36	9.09
49	chuleta	0.07941	0.36	9.09
50	cocina	0.07941	0.36	9.09
51	naranja	0.07636	0.72	18.18
52	restaurante	0.07421	0.36	9.09
53	lechuga	0.07295	0.72	18.18
54	bebida	0.06936	0.36	9.09
55	cerdo	0.06936	0.36	9.09
56	fresa	0.06853	0.72	18.18
57	pasta	0.06727	0.72	18.18
58	nopal	0.06409	0.72	18.18
59	atole	0.06115	1.08	27.27
60	betabel	0.05662	0.36	9.09
61	milanesa	0.05662	0.36	9.09
62	papaya	0.05662	0.36	9.09
63	hot dog	0.05405	0.72	18.18
64	champiñón	0.05291	0.36	9.09
65	limón	0.05291	0.36	9.09
66	tienda	0.05291	0.36	9.09
67	sushi	0.05236	0.72	18.18
68	salsa	0.05040	0.72	18.18
69	crema	0.04945	0.36	9.09
70	salud	0.04945	0.36	9.09
71	manzana	0.04864	0.72	18.18
72	envase	0.04622	0.36	9.09
73	licor	0.04622	0.36	9.09
74	panqué	0.04622	0.36	9.09

75	sal	0.04432	0.72	18.18
76	alcachofa	0.04320	0.36	9.09
77	empaque	0.04320	0.36	9.09
78	licuado	0.04320	0.36	9.09
79	pan dulce	0.04320	0.36	9.09
80	lugar	0.04037	0.36	9.09
81	comer	0.03773	0.36	9.09
82	leguminosa	0.03773	0.36	9.09
83	nachos	0.03773	0.36	9.09
84	tequila	0.03773	0.36	9.09
85	avena	0.03526	0.36	9.09
86	hambriento	0.03526	0.36	9.09
87	palomita	0.03526	0.36	9.09
88	sandía	0.03526	0.36	9.09
89	enfrijolada	0.03296	0.36	9.09
90	malteado	0.03296	0.36	9.09
91	marisco	0.03296	0.36	9.09
92	sediento	0.03296	0.36	9.09
93	tocino	0.03296	0.36	9.09
94	vodka	0.03296	0.36	9.09
95	espagueti	0.03080	0.36	9.09
96	esquimo	0.03080	0.36	9.09
97	flor de calabaza	0.03080	0.36	9.09
98	quesadilla	0.02954	0.72	18.18
99	calabaza	0.02879	0.36	9.09
100	guacamole	0.02879	0.36	9.09
101	mora	0.02879	0.36	9.09
102	tamal	0.02879	0.36	9.09
103	chilaquil	0.02690	0.36	9.09
104	frijol	0.02690	0.36	9.09
105	mayonesa	0.02690	0.36	9.09
106	mostaza	0.02514	0.36	9.09
107	ternera	0.02514	0.36	9.09
108	trigo	0.02514	0.36	9.09
109	condimento	0.02350	0.36	9.09
110	cosmopolitan	0.02350	0.36	9.09
111	Coca-Cola	0.02196	0.36	9.09
112	aderezo	0.02196	0.36	9.09
113	asado	0.02196	0.36	9.09

114	gelatina	0.02196	0.36	9.09
115	horchata	0.02196	0.36	9.09
116	maíz	0.02196	0.36	9.09
117	miel	0.02196	0.36	9.09
118	aguacate	0.02052	0.36	9.09
119	azúcar	0.02052	0.36	9.09
120	burrito	0.02052	0.36	9.09
121	caviar	0.02052	0.36	9.09
122	empanada	0.02052	0.36	9.09
123	emparedado	0.02052	0.36	9.09
124	espinaca	0.02052	0.36	9.09
125	jamaica	0.02052	0.36	9.09
126	tostado	0.02052	0.36	9.09
127	alfalfa	0.01918	0.36	9.09
128	arrachera	0.01918	0.36	9.09
129	hongo	0.01918	0.36	9.09
130	parrilla	0.01918	0.36	9.09
131	bistec	0.01793	0.36	9.09
132	cocido	0.01793	0.36	9.09
133	harina	0.01793	0.36	9.09
134	verdolaga	0.01793	0.36	9.09
135	quemado	0.01675	0.36	9.09
136	camarón	0.01566	0.36	9.09
137	lasaña	0.01566	0.36	9.09
138	mesero	0.01566	0.36	9.09
139	dinero	0.01463	0.36	9.09
140	nata	0.01195	0.36	9.09
141	elote	0.01043	0.36	9.09
142	esquites	0.00975	0.36	9.09

LEXICAL AVAILABILITY

Ramirez Alfaro, Anahi
 Poll: 01-10-2012 10:00:00
 Type: Target
 Center: Clothes

Date: 10-11-2015 12:30:27 PM
 Group: S-004 - English
 Respondents: 11/11
 Place: Centro de Enseñanza de Lenguas E

#	Word	Index	Freq1	Freq2
1	T-shirt	0.74055	6.02	90.91
2	pants	0.67765	5.42	81.82
3	skirt	0.55395	4.22	63.64
4	jacket	0.54959	5.42	81.82
5	shoe	0.54931	6.02	90.91
6	dress	0.33679	4.22	63.64
7	sweater	0.30089	3.61	54.55
8	coat	0.29866	4.22	63.64
9	underwear	0.29729	3.01	45.45
10	sock	0.29651	4.82	72.73
11	shorts	0.29010	3.01	45.45
12	jeans	0.28184	3.01	45.45
13	shirt	0.27434	3.01	45.45
14	trousers	0.24725	2.41	36.36
15	blouse	0.23874	3.01	45.45
16	tie	0.22397	3.01	45.45
17	hat	0.20417	3.01	45.45
18	glove	0.19944	3.01	45.45
19	belt	0.17876	2.41	36.36
20	boot	0.17842	2.41	36.36
21	tennis	0.13085	1.81	27.27
22	scarf	0.11553	1.81	27.27
23	sandal	0.11172	1.81	27.27
24	leggings	0.09091	0.60	9.09
25	tuxedo	0.08539	1.20	18.18
26	suit	0.07468	1.81	27.27
27	bag	0.06292	0.60	9.09
28	watch	0.06292	0.60	9.09
29	high heels	0.05709	1.20	18.18
30	cut	0.04774	0.60	9.09
31	glasses	0.04774	0.60	9.09
32	overall	0.04774	0.60	9.09
33	raincoat	0.03972	0.60	9.09
34	thong	0.03972	0.60	9.09

35	parka	0.03623	0.60	9.09
36	smoking	0.03623	0.60	9.09
37	cap	0.03014	0.60	9.09
38	jersey	0.03014	0.60	9.09
39	pyjama	0.03014	0.60	9.09
40	stocking	0.03014	0.60	9.09
41	belly	0.02749	0.60	9.09
42	undershirt	0.02749	0.60	9.09
43	short	0.02507	0.60	9.09
44	brassiere	0.02287	0.60	9.09
45	slip	0.02287	0.60	9.09
46	bikini	0.02086	0.60	9.09
47	suitcase	0.02086	0.60	9.09
48	sneaker	0.01444	0.60	9.09
49	cardigan	0.01317	0.60	9.09
50	pullover	0.01201	0.60	9.09
51	polo shirt	0.00999	0.60	9.09

LEXICAL AVAILABILITY				
Ramirez Alfaro, Anahi			Date: 10-11-2015 12:33:10 PM	
Poll: 01-10-2012 12:00:00			Group: S-004 - English	
Type: Comparison: Spanish			Respondents: 11/11	
Center: Clothes			Place: Centro de Enseñanza de Lenguas	
E				
#	Word	Index	Freq1	Freq2
1	pantalón	0.65673	3.73	90.91
2	camisa	0.64217	4.10	100
3	falda	0.60779	3.36	81.82
4	zapato	0.56709	3.36	81.82
5	blusa	0.51192	3.36	81.82
6	chamarra	0.42042	4.10	100
7	playera	0.41561	2.99	72.73
8	tenis	0.41052	2.99	72.73
9	calcetín	0.36653	3.36	81.82
10	vestido	0.36203	2.61	63.64
11	corbata	0.35180	3.36	81.82
12	media	0.27711	2.61	63.64
13	sudadera	0.26363	2.61	63.64
14	suéter	0.23675	2.24	54.55
15	abrigo	0.22543	2.61	63.64
16	bufanda	0.22499	3.36	81.82
17	calceta	0.22245	1.87	45.45
18	traje	0.20297	1.49	36.36
19	sandalia	0.20224	1.49	36.36

20	bota	0.19524	1.87	45.45
21	ropa interior	0.18159	1.12	27.27
22	zapatilla	0.17672	1.12	27.27
23	camiseta	0.17378	1.12	27.27
24	saco	0.16552	2.61	63.64
25	gorra	0.14886	1.49	36.36
26	calzón	0.14400	1.49	36.36
27	pants	0.14297	1.12	27.27
28	sombrero	0.14248	1.49	36.36
29	short	0.14192	2.24	54.55
30	trusa	0.13487	1.49	36.36
31	jean	0.12870	1.49	36.36
32	brasier	0.12200	1.12	27.27
33	faja	0.09091	0.37	9.09
34	bóxer	0.09066	1.12	27.27
35	correa	0.08891	0.75	18.18
36	moño	0.08398	0.37	9.09
37	guante	0.08284	1.49	36.36
38	reloj	0.07774	0.75	18.18
39	cinturón	0.07653	1.12	27.27
40	bikini	0.07470	0.75	18.18
41	pantimedia	0.07166	0.37	9.09
42	mall		0.75	18.18

		0.06172		
43	pulóver	0.05649	0.37	9.09
44	gabardina	0.05524	0.75	18.18
45	tacón	0.05266	0.75	18.18
46	camisón	0.05218	0.37	9.09
47	impermeable	0.05218	0.37	9.09
48	licra	0.04722	0.75	18.18
49	tirante	0.04628	0.75	18.18
50	chaleco	0.04570	0.75	18.18
51	jumper	0.04453	0.37	9.09
52	smoking	0.04453	0.37	9.09
53	anillo	0.04375	0.75	18.18
54	jersey	0.04362	0.75	18.18
55	chaqueta	0.04113	0.37	9.09
56	tanga	0.04113	0.37	9.09
57	arete	0.04042	0.75	18.18
58	overol	0.03799	0.37	9.09
59	balerina	0.03242	0.37	9.09
60	camisa interior	0.03242	0.37	9.09
61	chancla	0.03242	0.37	9.09
62	pinza para el cabello	0.02995	0.37	9.09
63	jeans	0.02767	0.37	9.09
64	bermuda	0.02556	0.37	9.09

65	mascada	0.02556	0.37	9.09
66	pijama	0.02556	0.37	9.09
67	crinolina	0.02181	0.37	9.09
68	leggings	0.02181	0.37	9.09
69	diadema	0.02015	0.37	9.09
70	blusón	0.01861	0.37	9.09
71	huarache	0.01861	0.37	9.09
72	vaqueros	0.01861	0.37	9.09
73	corbatín	0.01588	0.37	9.09
74	chal	0.01355	0.37	9.09
75	manga	0.01252	0.37	9.09
76	traje de baño	0.01252	0.37	9.09
77	tutú	0.01156	0.37	9.09
78	pantalónera	0.00987	0.37	9.09
79	poncho	0.00987	0.37	9.09

LEXICAL AVAILABILITY				
Ramirez Alfaro, Anahi			Date: 10-11-2015 12:31:06 PM	
Poll: 01-10-2012 10:00:00			Group: S-004 - English	
Type: Target			Respondents: 11/11	
Center: Food and drink			Place: Centro de Enseñanza de Lenguas E	
#	Word	Index	Freq1	Freq2
1	soda	0.49743	3.57	72.73
2	beer	0.41300	2.68	54.55
3	chicken	0.36562	3.57	72.73
4	hamburger	0.31532	2.68	54.55
5	apple	0.31271	2.23	45.45
6	water	0.28981	2.23	45.45
7	meat	0.28544	2.23	45.45
8	salad	0.26991	1.79	36.36
9	rice	0.26737	2.23	45.45
10	soup	0.26058	2.23	45.45
11	grape	0.23846	1.79	36.36
12	sandwich	0.23449	1.79	36.36
13	wine	0.22999	1.79	36.36
14	bread	0.20808	2.23	45.45
15	watermelon	0.19797	2.23	45.45
16	juice	0.19477	1.79	36.36
17	milk	0.19393	1.79	36.36
18	potato	0.19334	1.79	36.36
19	orange	0.19330	1.79	36.36
20	tomato	0.18114	2.23	45.45
21	fish	0.17911	1.79	36.36
22	orange juice	0.16965	0.89	18.18
23	beef	0.16463	1.34	27.27
24	ham	0.16419	0.89	18.18
25	sugar	0.15186	1.34	27.27
26	tea	0.15131	1.34	27.27
27	onion	0.13362	1.34	27.27
28	jam	0.12928	0.89	18.18
29	spaghetti	0.12316	0.89	18.18
30	pasta	0.12273	1.79	36.36
31	fruit	0.12117	1.34	27.27
32	pizza	0.12032	0.89	18.18
33	coffee	0.11909	1.34	27.27
34	sausage	0.11784	0.89	18.18

35	steak	0.11462	0.89	18.18
36	artichoke	0.10777	0.89	18.18
37	lettuce	0.10588	1.34	27.27
38	dessert	0.10367	0.89	18.18
39	tequila	0.09337	0.89	18.18
40	egg	0.09313	0.89	18.18
41	plum	0.09277	0.89	18.18
42	pork	0.08765	1.34	27.27
43	cheese	0.08460	0.45	9.09
44	duck	0.08460	0.45	9.09
45	cognac	0.07874	0.45	9.09
46	coke	0.07874	0.45	9.09
47	lemonade	0.07874	0.45	9.09
48	meet	0.07874	0.45	9.09
49	ice-cream	0.07440	0.89	18.18
50	breakfast	0.07328	0.45	9.09
51	chip	0.07328	0.45	9.09
52	tuna	0.07328	0.45	9.09
53	strawberry	0.07237	0.89	18.18
54	vegetable	0.06930	0.89	18.18
55	cookie	0.06920	0.89	18.18
56	chicken breast	0.06819	0.45	9.09
57	lunch	0.06819	0.45	9.09
58	scotch	0.06819	0.45	9.09
59	cucumber	0.06735	0.89	18.18
60	butter	0.06346	0.45	9.09
61	dinner	0.06346	0.45	9.09
62	eggplant	0.05906	0.45	9.09
63	pie	0.05906	0.45	9.09
64	french fries	0.05497	0.45	9.09
65	cake	0.05115	0.45	9.09
66	mashed potato	0.05115	0.45	9.09
67	oats	0.05115	0.45	9.09
68	chilli	0.04761	0.45	9.09
69	margarita	0.04431	0.45	9.09
70	carrot	0.04386	0.89	18.18
71	T-bone	0.04123	0.45	9.09
72	liver	0.04123	0.45	9.09
73	whiskey	0.04123	0.45	9.09
74	pork chop	0.03837	0.45	9.09

75	hot dog	0.03571	0.45	9.09
76	lemon	0.03571	0.45	9.09
77	mushroom	0.03571	0.45	9.09
78	sauce	0.03571	0.45	9.09
79	veal	0.03571	0.45	9.09
80	bean	0.03324	0.45	9.09
81	candy	0.03324	0.45	9.09
82	broil vegetable	0.03093	0.45	9.09
83	yogurt	0.03093	0.45	9.09
84	junk food	0.02879	0.45	9.09
85	alcohol	0.02679	0.45	9.09
86	snickers	0.02679	0.45	9.09
87	malted milk	0.02493	0.45	9.09
88	muffin	0.02493	0.45	9.09
89	bacon	0.02320	0.45	9.09
90	donut	0.02320	0.45	9.09
91	kiwi	0.02320	0.45	9.09
92	salt	0.02320	0.45	9.09
93	flour	0.02159	0.45	9.09
94	meal	0.02159	0.45	9.09
95	vodka	0.02159	0.45	9.09
96	baguette	0.02010	0.45	9.09
97	coriander	0.02010	0.45	9.09
98	peanut	0.02010	0.45	9.09
99	pear	0.01870	0.45	9.09
100	pineapple	0.01740	0.45	9.09
101	banana	0.01306	0.45	9.09
102	papaya	0.01215	0.45	9.09
103	guava	0.01131	0.45	9.09
104	melon	0.01052	0.45	9.09

LEXICAL AVAILABILITY				
Ramirez Alfaro, Anahi			Date: 10-11-2015 12:34:07 PM	
Poll: 01-10-2012 12:00:00			Group: S-004 - English	
Type: Comparison: Spanish			Respondents: 11/11	
Center: Food and drink			Place: Centro de Enseñanza de Lenguas E	
#	Word	Index	Freq1	Freq2
1	refresco	0.47997	3.65	72.73
2	carne	0.40864	3.65	72.73
3	cerveza	0.40362	2.74	54.55
4	pollo	0.34712	3.65	72.73
5	manzana	0.30602	2.28	45.45
6	hamburguesa	0.29970	2.74	54.55
7	jamón	0.28858	1.83	36.36
8	agua	0.27847	2.28	45.45
9	ensalada	0.26598	1.83	36.36
10	arroz	0.26032	2.28	45.45
11	sopa	0.24775	2.28	45.45
12	uva	0.23750	1.83	36.36
13	sándwich	0.22535	1.83	36.36
14	vino	0.22095	1.83	36.36
15	leche	0.20446	2.28	45.45
16	pan	0.19955	2.28	45.45
17	melón	0.19256	2.28	45.45
18	naranja	0.18998	1.83	36.36
19	papa	0.18795	1.83	36.36
20	jugo	0.18342	1.83	36.36
21	jitomate	0.17233	2.28	45.45
22	jugo de naranja	0.16848	0.91	18.18
23	pescado	0.16763	1.83	36.36
24	azúcar	0.14785	1.37	27.27
25	te	0.14530	1.37	27.27
26	cebolla	0.12733	1.37	27.27
27	carne de res	0.12384	0.91	18.18
28	espagueti	0.11838	0.91	18.18
29	pizza	0.11640	0.91	18.18
30	pasta	0.11421	1.83	36.36
31	salchicha	0.11393	0.91	18.18
32	café	0.11199	1.37	27.27
33	fruta	0.11175	1.37	27.27
34	alcachofa	0.10228	0.91	18.18

35	postre	0.10118	0.91	18.18
36	lechuga	0.09977	1.37	27.27
37	tequila	0.09027	0.91	18.18
38	huevo	0.08800	0.91	18.18
39	pato	0.08398	0.46	9.09
40	queso	0.08398	0.46	9.09
41	cerdo	0.08340	1.37	27.27
42	Coca-Cola	0.07757	0.46	9.09
43	ciruela	0.07757	0.46	9.09
44	coñac	0.07757	0.46	9.09
45	limonada	0.07757	0.46	9.09
46	atún	0.07166	0.46	9.09
47	desayuno	0.07166	0.46	9.09
48	papas fritas	0.07166	0.46	9.09
49	fresa	0.06937	0.91	18.18
50	helado	0.06813	0.91	18.18
51	pepino	0.06806	0.91	18.18
52	lunch	0.06620	0.46	9.09
53	pechuga de pollo	0.06620	0.46	9.09
54	galleta	0.06467	0.91	18.18
55	verdura	0.06276	0.91	18.18
56	cena	0.06115	0.46	9.09
57	mantequilla	0.06115	0.46	9.09
58	berenjena	0.05649	0.46	9.09
59	pay	0.05649	0.46	9.09
60	papa a la francesa	0.05218	0.46	9.09
61	avena	0.04820	0.46	9.09
62	bistec	0.04820	0.46	9.09
63	pastel	0.04820	0.46	9.09
64	puré de papa	0.04820	0.46	9.09
65	chile	0.04453	0.46	9.09
66	margarita	0.04453	0.46	9.09
67	whisky	0.04113	0.46	9.09
68	zanahoria	0.04023	0.91	18.18
69	T-bone	0.03799	0.46	9.09
70	hígado	0.03799	0.46	9.09
71	limón	0.03510	0.46	9.09
72	venado	0.03510	0.46	9.09
73	champiñón	0.03242	0.46	9.09
74	hot dog	0.03242	0.46	9.09

75	salsa	0.03242	0.46	9.09
76	dulce	0.02995	0.46	9.09
77	frijol	0.02995	0.46	9.09
78	verdura hervida	0.02767	0.46	9.09
79	yogurt	0.02767	0.46	9.09
80	comida chatarra	0.02556	0.46	9.09
81	alcohol	0.02361	0.46	9.09
82	panqué	0.02361	0.46	9.09
83	sal	0.02361	0.46	9.09
84	dona	0.02181	0.46	9.09
85	harina	0.02181	0.46	9.09
86	kiwi	0.02181	0.46	9.09
87	cacahuate	0.02015	0.46	9.09
88	tocino	0.02015	0.46	9.09
89	vodka	0.02015	0.46	9.09
90	cilantro	0.01861	0.46	9.09
91	pera	0.01861	0.46	9.09
92	baguet	0.01719	0.46	9.09
93	piña	0.01588	0.46	9.09
94	cereza	0.01355	0.46	9.09
95	plátano	0.01252	0.46	9.09
96	papaya	0.01156	0.46	9.09
97	guayaba	0.00987	0.46	9.09