

**Universidad Autónoma de Querétaro**

**Facultad de Lenguas y Letras**

**Doctorado en Lingüística**

Propuesta Metodológica para la elaboración de pruebas  
de decisión léxica de vocabulario  
académico especializado  
Que como parte de los requisitos para obtener el grado de

Doctor en Lingüística

**Presenta:**

Adriana Medellín Gómez

**Dirigido por:**

Dr. Ignacio Rodríguez Sánchez

Mayo, 2012

México



Universidad Autónoma de Querétaro  
Facultad de Lenguas y Letras  
Doctorado en Lingüística

Propuesta Metodológica para la elaboración de pruebas  
de decisión léxica de vocabulario  
académico especializado

Que como parte de los requisitos para obtener el grado de

Doctor en Lingüística

**Presenta:**

Adriana Medellín Gómez

**Dirigido por:**

Dr. Ignacio Rodríguez Sánchez

SINODALES

Dr. Ignacio Rodríguez Sánchez  
Presidente


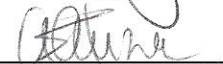

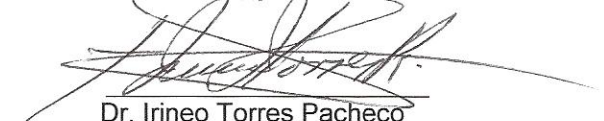
Dra. Donna Terry Jackson Lembark  
Secretario

Dra. Alejandra Auza Benavides  
Vocal

Dra. Luisa Josefina Alarcón Nieve  
Suplente

Dr. Eduardo Patricio Velázquez Patiño  
Suplente

Mtra. Ma. Eugenia Castillejos Solís  
Directora de la Facultad de Lenguas y Letras

  
Firma  
Firma  
Firma  
Firma  
Firma  
Dr. Irineo Torres Pacheco  
Director de Investigación y  
Posgrado

Centro Universitario  
Querétaro, Qro.  
Mayo 2012  
México

## RESUMEN

Esta investigación presenta la metodología, el desarrollo, la aplicación y los resultados de una propuesta metodológica para la elaboración de pruebas de decisión léxica capaces de medir con precisión el reconocimiento de vocabulario académico especializado. Para los objetivos específicos de este trabajo, se construyeron dos corpus académicos especializados, uno para el área de Nutrición y el segundo para el área de Odontología. Para cada una de estas áreas se llevó a cabo una primera selección de términos especializados utilizando diversas herramientas de procesamiento de corpus. Posteriormente se seleccionó una serie de prefijos de alta frecuencia dentro de cada uno de los corpus, los cuales fueron combinados con bases también de alta frecuencia para generar una serie de no palabras. Un grupo de expertos en cada área fue consultado para confirmar la selección final. Finalmente se elaboraron dos pruebas de decisión léxica, una para el área de Nutrición y otra para Odontología. Cada prueba consistía en un total de 200 palabras de las cuales 160 eran términos académicos especializados y 40 eran no palabras. En el estudio participó un grupo control de 40 estudiantes de bachillerato, un grupo de 157 estudiantes de la Licenciatura en Nutrición y un grupo de 93 estudiantes de la Licenciatura en Odontología de dos universidades: una pública y una privada de la ciudad de Querétaro, México. Los resultados indicaron que el número de palabras que los participantes conocían y reconocían se incrementó en relación al semestre del programa de estudios que cursaban y era significativamente diferente entre semestres. Estos resultados demuestran la viabilidad y efectividad de utilizar este tipo de pruebas de decisión léxica como medio para evaluar la evolución del conocimiento de vocabulario académico especializado.

**(PALABRAS CLAVE:** Vocabulario Académico, Evaluación, Lingüística de Corpus)

## SUMMARY

This research work presents the methodology, development, application and results of a methodological proposal for the elaboration of a checklist assessment procedure that accurately measures specialized academic vocabulary identification. Two corpora were made for the specific purposes of this research, one on Food Science and the other one on Odontology. For each one of these areas a first items selection was made by using several corpus analysis tools. Then, a series of highly frequent prefixes of each area was combined with also highly frequent word roots to create non-words. Next, a group of experts in each field was asked to confirm the final selection of items. Finally, two checklists, one for Food Science and one for Odontology were made. The final checklists included 200 words divided in 160 real words and 40 non-words each. The participants were a group of 40 high school students, a group of 157 undergraduates of Food Science and a final group of 93 students of Odontology at two different universities, one public and one private in Querétaro, México. The results revealed that the number of words the participants knew increased according to the number of semesters they had studied in their study programs and were significantly different among semesters. These findings demonstrate the viability and effectiveness of using this type of checklist as a means of assessing the development of specialized academic vocabulary knowledge.

**(KEY WORDS:** Academic Vocabulary, Assessment, Corpus Linguistics)

A Griselda Gómez P.

Porque otra vez estuviste en cada momento.

Y porque no importa cuanto crezca, nunca voy a ser tan grande como tú.

Agradezco infinitamente la invaluable colaboración y ayuda de:

- El Dr. Ignacio Rodríguez Sánchez, por su paciencia, su apoyo, su solidaridad; y porque sin él todo hubiera sido muy diferente, y ahora es mejor.
- Al Dr. Gilberto Herrera Ruíz, por acogerme e impulsarme, y al Dr. Aurelio Domínguez por ese último empujoncito.
- Los profesores del Doctorado en Lingüística: Dra. Alejandra Auza, Dra. Donna Jackson, Dra. Luisa Josefina Alarcón y Dr. Eduardo Velázquez por cumplimiento mucho más allá del deber.
- Los profesores y alumnos de las Licenciaturas en Nutrición y en Odontología de la UAQ y de la UVM, que tan gentilmente me regalaron su tiempo.
- Griselda y Cristina, porque me dieron una casa, un hogar y una familia.
- Los miembros del elenco de “la casa tomada”, que siempre estarán ahí, y a los que amo cada día más.

## INDICE

1. INTRODUCCIÓN.....	12
1.1 Diseño de la investigación.....	16
1.2 Estructura del estudio.....	16
2. OBJETIVOS.....	18
2.1 Objetivo General: .....	18
2.2 Objetivo Particular:.....	18
3. HIPÓTESIS.....	19
4. PREGUNTAS DE INVESTIGACIÓN .....	20
5. MARCO TEÓRICO .....	23
5.1 ¿Qué es “vocabulario”?.....	23
5.1.1 El Vocabulario Académico.....	26
5.2 Evaluación de Vocabulario.....	28
5.2.1 Diseño y validez de las Pruebas de Evaluación de Vocabulario .....	34
5.2.1.1 Pruebas de Decisión Léxica con formato Checklist .....	41
5.2.1.1.1 Estudios previos realizados con el formato de <i>Checklist</i> .....	42
5.3 Lingüística de corpus .....	50
5.3.1 Corpus Lingüísticos .....	51
5.3.2 Consideraciones generales para la elaboración de corpus .....	57

5.3.3	El uso de Corpus en la Lingüística .....	59
6.	ANTECEDENTES.....	62
6.1	Listas de Vocabulario especializado .....	62
6.2	Frecuencia .....	67
6.3	Terminología .....	70
6.3.1	La Terminología y la lexicología .....	72
6.3.2	La terminología y la lingüística .....	73
7.	METODOLOGÍA.....	75
7.1	Criterios generales para la construcción de los corpus.....	75
7.2	Construcción del corpus de Nutrición.....	76
7.3	Construcción del corpus de Odontología .....	77
7.4	Procesamiento de los corpus .....	77
7.4.1	Etiquetado .....	77
7.4.2	Herramientas para el análisis de los corpus .....	81
7.4.3	Selección de los términos académicos especializados .....	83
7.4.4	Generación de no palabras .....	83
7.5	Validación de términos y no palabras con expertos del área de Nutrición y Odontología .....	87
7.6	Participantes .....	87
7.6.1	Grupo Control .....	88



7.6.2	Grupo Experimental para el área de Nutrición.....	88
7.6.2.1	Grupo Experimental para el área de Nutrición- Maestría .....	89
7.6.3	Grupo Experimental para el área de Odontología .....	89
7.7	Elaboración y aplicación de las pruebas .....	90
8.	RESULTADOS Y DISCUSIÓN .....	92
8.1	Resultados de la Prueba de Nutrición.....	93
8.1.1	Licenciatura en Nutrición de la UAQ.....	93
8.1.2	Licenciatura en Nutrición UVM .....	95
8.1.3	Maestría en Nutrición UAQ.....	97
8.1.4	Análisis por frecuencia de los términos y no palabras de la Prueba de Nutrición.....	99
8.2	Resultados de la Prueba de Odontología.....	101
9.	CONCLUSIONES .....	106
10.	APÉNDICES.....	134
10.1	Prueba de Nutrición .....	134
10.2	Prueba de Odontología.....	137
10.3	Selección de Palabras Reales del Corpus de Nutrición por Frecuencia de Palabra.....	140
10.4	No Palabras elaboradas a partir de prefijos y bases del Corpus de Nutrición.....	157

10.5	Selección de Palabras Reales del Corpus de Odontología por Frecuencia de Palabra .....	159
10.6	No Palabras elaboradas a partir de prefijos y bases del Corpus de Odontología .....	163
10.7	Prefijos de alta frecuencia en el corpus de Nutrición .....	165
10.8	Prefijos de alta frecuencia en el corpus de Odontología .....	167

Tabla 1 Componentes de conocimiento de vocabulario (Nation, 1990) .....	29
Tabla 2. Dimensiones de evaluación del vocabulario propuestas por Read (2000) .....	30
Tabla 3. Posibles respuestas a la Prueba de formato Checklist (Huibregtse y Admiraal, 2002) .....	45
Tabla 4 Criterios para la construcción de los Corpus, basados en Sinclair (2005)	75
Tabla 5 Número de participantes por institución educativa y por semestre para la Prueba de Nutrición.....	89
Tabla 6 Número de participantes por semestre para la prueba de Odontología ...	90

Figura 1. Etiquetado del corpus. Pantalla de ingreso de datos del Programa TreeTagger.....	78
Figura 2 Etiquetado del corpus. Pantalla de resultado de etiquetado del Programa TreeTagger.....	79
Figura 3 Etiquetas y términos procesados para el análisis del Corpus .....	80
Figura 4 Herramientas utilizadas en el procesamiento de los corpus, Wordlist ....	82
Figura 5 Herramientas utilizadas en el procesamiento de los corpus, Concordance .....	82
Figura 6 Resultados de la Prueba de Nutrición aplicada a estudiantes de la UAQ + Grupo Control.....	94
Figura 7 Resultados de la Prueba de Nutrición aplicada a estudiantes de la UVM96	
Figura 8 Resultados de la Prueba de Nutrición aplicada a todos los Grupos UAQ y UVM .....	98
Figura 9 Análisis por frecuencia de los términos y no palabras.....	100
Figura 10 Resultados de la Prueba de Odontología aplicada a estudiantes de la UVM .....	102
Figura 11 Comparativo de los resultados de la Prueba de Nutrición y de Odontología en la UVM .....	103

## 1. INTRODUCCIÓN

El conocimiento del vocabulario académico en cualquier campo de estudio diferencia a los estudiantes universitarios bien preparados de aquellos con carencias académicas (Kuehn, 1996). Esto ha generado un creciente interés por el estudio del vocabulario académico básico que los estudiantes deben conocer para poder desempeñarse exitosamente en el ámbito universitario, dando como resultado una serie de listas de vocabulario académico general con propósitos pedagógicos. Por citar algunos ejemplos, encontramos la *Academic Vocabulary List* (Campion y Elley, 1971), la *Academic University List* (Praninskas, 1972), la *University Word List* (Xue y Nation, 1984) y la *Academic Word List* (Coxhead, 2000).

En años más recientes, otros investigadores se han dedicado al estudio del vocabulario académico por disciplinas específicas, sosteniendo que existen rasgos individuales que identifican cada área o sub-área de estudio. Es a partir de estas investigaciones que se han obtenido listas de vocabulario relacionadas con las Ciencias Computacionales (Lam, 2001), la Ingeniería (Mudraya, 2006), la Medicina (Baker, 1988; Chen y Ge, 2007; Wang et al, 2008) y la Agricultura (Martínez et al, 2009) entre otras.

Este trabajo se enfoca únicamente en el estudio del vocabulario académico especializado, es decir, vocabulario que pertenece a un área ó campo de conocimiento específico. El objetivo de esta investigación es el de presentar una propuesta metodológica para evaluar de una manera práctica, eficiente y confiable el desarrollo del incremento del vocabulario académico especializado en primera lengua de alumnos universitarios, en este caso particular el conocimiento de vocabulario especializado de los estudiantes de la Licenciatura en Nutrición y de la Licenciatura en Odontología, presentando los resultados obtenidos en dos instituciones educativas, una pública y otra privada, de la ciudad de Querétaro en México.

La investigación que se presenta a continuación se inserta en los principios, bases y metodologías de la lingüística aplicada. Ésta provee los fundamentos teóricos y descriptivos para la investigación y resolución de problemas relacionados con la lengua (DeKeyser, 2007) y aunque originalmente sólo era reconocida para la investigación de enseñanza de lenguas, actualmente es concebida como una ciencia interdisciplinaria, la cual involucra un amplio espectro de temas como son las primeras y segundas lenguas, las lenguas extranjeras, su enseñanza y aprendizaje, bilingüismo y multilingüismo, análisis del discurso, traducción e interpretación, metodología de la investigación, pedagogía, evaluación del lenguaje, literatura, retórica, terapia del lenguaje, lingüística forense, terminología, lexicografía, alfabetismo, entre muchas otras que podemos mencionar. En general, podemos concluir que hoy en día la lingüística aplicada toma el lenguaje y las teorías del lenguaje como una base para proponer planteamientos sobre cómo funciona la comunicación real, con el propósito de identificar los problemas que pueden presentarse en muy diversos contextos, y analizarlos para poder proponer soluciones prácticas y útiles para las personas reales del mundo real (Richads et al., 1985). Esta es una de estas propuestas, ya que busca proponer una metodología para evaluar el conocimiento del vocabulario de una forma práctica y objetiva.

Para los propósitos de esta investigación estableceremos que el vocabulario está compuesto por *palabras*, que son las unidades básicas de significado que componen estructuras más largas como son oraciones, párrafos y textos completos (Read, 2005). Es decir, la metodología propuesta evalúa unidades léxicas (palabras) y no unidades léxicas múltiples (frases, colocaciones, expresiones idiomáticas). Además, es importante definir que el vocabulario que compete a este estudio es únicamente el vocabulario académico especializado. Por vocabulario académico entenderemos todo aquel vocabulario “formal, independiente del contexto, constituido por palabras de muy alta frecuencia o con un alto rango de ocurrencia a lo largo de diversas disciplinas científicas” (Farrell, 1990, p.11). Por vocabulario académico especializado entenderemos todo aquel vocabulario académico que pertenece a un área o disciplina científica en particular, como es la Medicina, la Ingeniería o la Arquitectura. Algunos autores

denominan a este tipo de vocabulario como “vocabulario técnico” o simplemente términos ó términos especializados (Cabré, 1998). De hecho, esta última definición correspondiente a la terminología (Ver 6.3) es la que más se acerca a la conceptualización de vocabulario académico especializado correspondiente a esta investigación.

El estudio propone una metodología para la evaluación del vocabulario académico especializado. Tradicionalmente, la evaluación del vocabulario ha sido dividida en dos grandes ramas: la evaluación del tamaño del vocabulario y la evaluación de la profundidad del conocimiento de vocabulario. El tamaño de vocabulario se refiere a la cantidad de palabras que el hablante posee, mientras que la profundidad de vocabulario implica la calidad del conocimiento de esas palabras, es decir los diversos significados de la palabra, su pronunciación, su categoría gramatical, las colocaciones más comunes, su registro, etc. También podemos encontrar una distinción entre la evaluación del vocabulario receptivo, y la evaluación del vocabulario productivo. Mientras que el vocabulario receptivo se refiere a la comprensión que los hablantes tienen acerca del vocabulario, es decir, a las palabras que los hablantes son capaces de entender, el vocabulario productivo implica la capacidad de los hablantes para generar palabras (Wesche y Paribakht, 1993). La metodología que aquí se presenta utiliza una prueba de tamaño de vocabulario receptivo, es decir, presenta el vocabulario meta y mide el número de palabras que los participantes reconocen.

Existen muy diversas conceptualizaciones y métodos para evaluar el vocabulario de un hablante (Ver 5.2). Esta investigación reporta los resultados de un estudio experimental realizado a partir de la aplicación de una prueba de decisión léxica con formato de *checklist*. Este formato consiste en una lista de palabras las cuales deben ser marcadas sólo en caso de que el hablante conozca su significado. Si el hablante no reconoce la palabra, debe dejarla en blanco, de ahí el nombre de “decisión léxica”. En sus orígenes, esta prueba fue usada en el campo de adquisición de la primera lengua (Sims, 1929; Tilley, 1936; Zimmerman et al, 1977); posteriormente el formato se modificó para incluir no palabras (Zimmerman, Broder, Shaughnessy y Underwood, 1977; Anderson y Freebody,

1983) y finalmente se aplicó a los procesos de adquisición de segunda lengua (Meara y Buxton, 1987; y Mochida y Harrington, 2006).

La prueba diseñada y aplicada en esta investigación consiste en una lista de ítems de los cuales 100 son palabras de baja frecuencia, 60 son de alta frecuencia y 40 son no palabras. Las no palabras son “aparentes palabras” que comparten características fonéticas, morfológicas y sintácticas con las palabras reales de una lengua, pero que en realidad no existen. Los rangos de frecuencia fueron establecidos de acuerdo a número total de ocurrencias de cada palabra dentro de dos corpus especializados, los cuales fueron especialmente desarrollados para los propósitos de este estudio, uno para el área de Nutrición y otro para el área de Odontología. Para la generación de las no palabras también se siguieron criterios de frecuencia, utilizando herramientas de procesamiento de corpus para detectar los prefijos y las bases más frecuentes dentro de los corpus y unirlos en combinaciones no existentes.

Los resultados obtenidos muestran que el número de palabras que los participantes manifestaron reconocer se incrementa en proporción directa al número de semestres que los estudiantes han cursado dentro de su programa de estudios y es significativamente diferente entre semestres. Estos resultados demuestran la viabilidad y efectividad de utilizar este tipo de pruebas de decisión léxica como medio para evaluar la evolución del conocimiento de vocabulario académico especializado.

Es así como este estudio que tiene sus bases en la lingüística aplicada, toma conceptos teóricos y metodológicos de áreas especializadas como son la terminología, el procesamiento de corpus y la evaluación de primera lengua para diseñar un método de evaluación de vocabulario académico especializado, que nos permita medir de forma rápida y objetiva el nivel de conocimiento de vocabulario en español de los estudiantes universitarios de dos disciplinas científicas, Nutrición y Odontología.



## 1.1 Diseño de la investigación

El enfoque que se utilizó para realizar esta investigación fue de carácter cuantitativo, ya que utiliza la recolección y el análisis de datos para contestar las preguntas de investigación e hipótesis establecidas (Hernández et al., 2003), aplicando el análisis estadístico para medir el nivel de reconocimiento de vocabulario especializado en diversos grupos de los semestres del programa de estudios universitarios de dos disciplinas, Nutrición y Odontología.

El diseño general de la investigación es de tipo experimental, ya que aplica un instrumento de decisión léxica a una serie de participantes con el objetivo de analizar la relación entre el número de semestres que han estudiado en su respectivo plan de estudios (variable independiente) y su nivel de reconocimiento de vocabulario especializado (variable dependiente).

Como dicho instrumento de decisión léxica fue aplicado en una sola ocasión a cada uno de los participantes, la investigación es de tipo transversal, ya que evalúa el nivel de reconocimiento de vocabulario académico especializado de los participantes en un momento específico.

Finalmente mencionaremos que el estudio realizado es de carácter descriptivo, ya que se reportan los resultados obtenidos por cada uno de los grupos experimentales los cuales se analizan aplicando métodos estadísticos.

## 1.2 Estructura del estudio

El presente estudio se organiza de la siguiente manera:

En las siguientes secciones, dedicadas al Marco Teórico y a los Antecedentes, se describen brevemente los conceptos básicos así como las investigaciones previas relacionadas con este estudio.

Posteriormente, en la sección correspondiente a la Metodología, se describe el procedimiento para la construcción y procesamiento de los corpus especializados elaborados específicamente para esta investigación.

Dentro de esta misma sección se detalla la metodología propuesta para la elaboración de los instrumentos de evaluación utilizados en este estudio y se describe brevemente a los diferentes grupos de estudiantes que participaron en el mismo.

En la sección de resultados se presentan los datos obtenidos para cada grupo de participantes así como el análisis estadístico de los mismos.

Finalmente, se presenta la discusión de los resultados así como las conclusiones generales del estudio.

## 2. OBJETIVOS

### 2.1 Objetivo General:

Esta investigación tiene como objetivo general proponer una metodología para la elaboración de pruebas de evaluación de vocabulario académico especializado en español para estudiantes universitarios nativo hablantes del español.

### 2.2 Objetivo Particular:

El objetivo particular de la investigación es el de reportar los resultados obtenidos en la aplicación de dos pruebas de evaluación de vocabulario académico especializado en español elaboradas según los parámetros de la metodología propuesta, una para el área de Nutrición y la segunda para el área de Odontología. Los resultados se analizarán y se reportarán en función de las siguientes variables:

- a) Disciplina de estudio, Nutrición y Odontología, respectivamente.
- b) Tiempo de exposición al vocabulario especializado, el cual se reportará de acuerdo al número de semestres cursados dentro del programa de estudio de la Licenciatura en Nutrición y la Licenciatura en Odontología.
- c) Institución, que corresponde a las dos instituciones académicas a las que pertenecían los participantes de este estudio: la Universidad Autónoma de Querétaro y la Universidad del Valle de México, Campus Querétaro.

### 3. HIPÓTESIS

A continuación se presentan las hipótesis que guían la realización de este estudio:

H1. A través de la lingüística de corpus y de diversas herramientas para el procesamiento de corpus es posible seleccionar una serie de palabras clave para una disciplina científica específica.

H2a. El conocimiento del vocabulario académico especializado aumenta según el tiempo de exposición al mismo.

H3b. Los estudiantes que tienen mayor tiempo de exposición a un vocabulario académico especializado obtienen mejor puntuación en una prueba de reconocimiento léxico.

H4. Una prueba de decisión léxica es capaz de determinar cuantitativamente el tiempo de exposición a un vocabulario académico especializado.

H5. El nivel de conocimiento de vocabulario académico especializado puede ser comparado aún entre distintas disciplinas de estudio.

H6. No existen diferencias significativas entre el nivel de reconocimiento de vocabulario académico especializado de los estudiantes de una misma disciplina que estudian en distintas instituciones educativas.

H7 Los estudiantes con un mayor tiempo de exposición a un vocabulario académico especializado son capaces de distinguir entre el léxico más frecuente de su disciplina y el léxico que no pertenece a su área de conocimiento.

#### 4. PREGUNTAS DE INVESTIGACIÓN

A continuación se presentan las preguntas de investigación planteadas para la realización de este trabajo:

1. ¿Es posible diseñar un instrumento que de forma práctica, rápida y objetiva sea capaz de diferenciar a través del reconocimiento de léxico entre los estudiantes que han estado expuestos al vocabulario académico de una disciplina científica y aquellos estudiantes que nunca han sido expuestos al mismo?

El diseño de un instrumento práctico, rápido y objetivo necesariamente comprende los siguientes aspectos: 1) La elaboración del instrumento: debe de ser simple, los ítems seleccionados deben de seguir criterios estrictos pero sencillos, las instrucciones deben de ser claras y el formato debe ser familiar para los participantes a los que va dirigido. 2) La resolución del instrumento: los participantes deben ser capaces de resolver rápidamente la prueba, y deben de comprender fácilmente el funcionamiento de la misma, 3) la obtención de los resultados: el instrumento debe de poder arrojar resultados rápidamente, es decir, poder ser calificada en un lapso corto de tiempo, además esos resultados deben ser medibles cuantitativamente y poder ser analizados estadísticamente con el propósito de facilitar la toma de decisiones.

2. ¿Existe una diferencia medible de forma cuantitativa que nos permita distinguir a través del reconocimiento de léxico el tiempo al que han estado expuestos al vocabulario especializado los estudiantes de una disciplina científica?

2b. ¿Es posible detectar y medir de forma cuantitativa el tiempo de exposición a un vocabulario académico especializado al que los estudiantes de una disciplina científica han sido expuestos?

Por sentido común se podría responder que entre más tiempo esté expuesto un alumno o un hablante a un input específico más conocimiento tendrá del mismo. Sin embargo, esperamos poder medir este conocimiento de una forma cuantitativa, obteniendo niveles o rangos que se relacionen directamente con el tiempo de exposición al vocabulario meta. Para el caso particular de esta investigación, el tiempo de exposición se medirá en años escolares. Por año escolar entenderemos dos semestres del programa oficial de la licenciatura ya sea en Nutrición o en Odontología. Así pues, aplicaremos y analizaremos los resultados de los estudiantes del 2do, 4to, 6to y 8vo. semestres de cada licenciatura.

3. ¿Los estudiantes de posgrado continúan ampliando su nivel de conocimiento de léxico especializado aún después de haber concluido sus estudios de licenciatura?

Si el tiempo de exposición es uno de los factores que influyen en la adquisición del vocabulario especializado, entre más tiempo de exposición, mayor debe ser el conocimiento del mismo. Así pues, aún cuando los participantes ya hayan concluido sus estudios de licenciatura, al ingresar a un programa de posgrado ¿deben continuar ampliando su vocabulario? ¿O acaso los conocimientos adquiridos durante la licenciatura son los únicos necesarios para tener un completo dominio del vocabulario de una disciplina?

4. ¿Serían comparables los niveles de reconocimiento de vocabulario de dos grupos de estudiantes especializados en distintas disciplinas científicas?

Ésta tal vez esta sea la pregunta más ambiciosa planteada en esta investigación, ya que necesariamente implicaría que la prueba de decisión léxica propuesta por esta metodología sea suficientemente sensible al tiempo de exposición y totalmente independiente a elementos paralelos como son las diferencias en el programa de estudios, las prácticas profesionales, los libros de texto, las diferencias en el discurso académico

de cada una de las disciplinas, etc. Los resultados de la prueba tendrían que arrojar “rangos” esperados para cada semestre, y estos rangos tendrían que ser los mismos tanto en diferentes universidades como en diferentes programas de licenciatura.

5. ¿Es posible realizar un comparativo de dos programas de estudio de la misma disciplina, pero en instituciones educativas diferentes, a través de la medición del reconocimiento de vocabulario especializado?

Es decir, a pesar de las diferencias en los programas de estudio de cada institución, y a pesar de los diferentes antecedentes académicos y las diferencias en los requisitos de ingreso y egreso de un programa de licenciatura, ¿se puede comparar a través de reconocimiento de léxico, el nivel de conocimiento de un profesionalista? Y si estos resultados son iguales, entonces ¿hay mejores universidades que otras? ¿Hay mejores programas que otros? ¿Realmente la universidad cumple la función social de igualar el conocimiento? Y si estos resultados son diferentes ¿qué hace la diferencia? ¿El número de alumnos por grupo? ¿Sus antecedentes personales? ¿Los profesores que imparten cátedra en sus universidades?

6. Un instrumento simple de reconocimiento léxico ¿puede ser una medida confiable de conocimiento de vocabulario especializado?

A pesar de que existen múltiples investigaciones que han aplicado las pruebas de reconocimiento léxico para la evaluación de vocabulario, éstas han sido dedicadas en su gran mayoría a la adquisición de segunda lengua, principalmente el inglés, y no han sido utilizadas para la evaluación de primera lengua y mucho menos relacionadas con el vocabulario académico de una o más disciplinas. Esta investigación servirá para evaluar su comportamiento en nuevos ámbitos de la adquisición del lenguaje y como un antecedente para su posible aplicación en la evaluación de la educación media superior.

## 5. MARCO TEÓRICO

### 5.1 ¿Qué es “vocabulario”?

Tradicionalmente, el vocabulario se define como el inventario de palabras y los significados asociados a las mismas que un hablante posee (Read, 2005). Este vocabulario puede referirse tanto a unidades léxicas (palabras aisladas) como a unidades léxicas múltiples (frases, colocaciones, expresiones idiomáticas). Las unidades léxicas o palabras aisladas, pueden dividirse en palabras de contenido y palabras funcionales. Mientras que las palabras de contenido son unidades con significado propio: sustantivos, verbos, adjetivos y adverbios, las palabras funcionales únicamente cumplen funciones de carácter gramatical y poseen poco o ningún significado al encontrarse aisladas, como son artículos, pronombres, preposiciones y conjunciones.

Hoy en día, entre la mayoría de los investigadores de adquisición de lengua existe un consenso sobre la importancia de los procesos de adquisición de vocabulario, sobre como éste debe ser redefinido y cómo debe ser considerado como un constructo multidimensional, el cual debe comprender por lo menos amplitud (tamaño) y profundidad (calidad). La amplitud del vocabulario involucra el número de palabras que el aprendiz conoce incluso en un nivel superficial, mientras que la profundidad se refiere a qué tan bien conoce dicha palabra. De acuerdo con Qian (1999) esta última dimensión incluye factores como la pronunciación, escritura, significado, registro, frecuencia y colocación, así como propiedades morfológicas y sintácticas de la palabra.

Para Chapelle (1998) la definición del constructo del vocabulario debe incluir las siguientes dimensiones: 1) tamaño de vocabulario; 2) conocimiento de las características de la palabra; 3) organización del lexicón y 4) procesos de acceso al lexicón. Mientras que para Henriksen (1999) esta definición debe incluir: 1)



precisión del conocimiento; 2) profundidad del conocimiento y 3) conocimiento receptivo y productivo.

Quian (2002) por su parte menciona los siguientes elementos: 1) tamaño de vocabulario; 2) profundidad del conocimiento de vocabulario; 3) organización del léxico, referente al almacenamiento, conexión y representación de las palabras en la mente del hablante y 4) automatización del conocimiento receptivo-productivo, el cual implica los procesos mediante los cuales una palabra es localizada tanto para procesos receptivos como productivos.

Es así como encontramos que para la mayoría de los investigadores relacionados con los procesos de adquisición de vocabulario existe una clara diferencia entre el tamaño y la profundidad del mismo. Hasta aproximadamente el año 2000 las principales investigaciones se centraron en establecer el tamaño del vocabulario de los hablantes o aprendices de una lengua, principalmente a nivel receptivo, aunque también hubo algunas a nivel productivo. En general estas pruebas se desarrollaron con base en la frecuencia de la palabra. La frecuencia se refiere al grado de probabilidad de encontrar una palabra (o varias palabras) en el habla o la escritura. Los trabajos que arrojaron más luz sobre este aspecto –la amplitud de vocabulario- fueron las pruebas realizadas por Nation (*Vocabulary Levels Tests*; 1983) y las de Meara (*Yes/No Test*, Meara & Buxton 1987). Ambas permiten establecer de modo fiable, preciso y de manera rápida y sencilla, el tamaño del vocabulario de los hablantes.

En comparación con la cantidad de investigaciones para establecer y medir la amplitud de vocabulario, existen relativamente pocas investigaciones respecto a la profundidad de éste. La profundidad del conocimiento de vocabulario se refiere a todo el conocimiento que el hablante tiene sobre una palabra específica. Entre las pruebas más representativas se encuentran aquellas desarrolladas por Vives y Meara (1994) quienes realizaron una prueba de asociaciones léxicas, la prueba de asociaciones de palabra (WAT) realizada por Read (1993), la escala de conocimiento de vocabulario (VKS, *Vocabulary Knowledge Scale*) desarrollada

por Wesche y Paribakht (1996) y la prueba de asociaciones en segunda lengua publicada por Meara y Wolter (2004).

Sin embargo es importante mencionar que existen otros investigadores como Vermeer (2001), para los cuales esta relación –amplitud y profundidad- más que formar parte de dos polos del conocimiento de vocabulario forma parte de un continuo, ya que al desarrollar nuestro tamaño de vocabulario podemos conformar redes más densas dentro de nuestro lexicón, y por lo tanto, desarrollar más la profundidad del conocimiento que tenemos de las palabras que ya conocíamos. Existen diversos estudios empíricos que respaldan esta afirmación (Aitchison, 1987, Meara 1996; Vermeer 2001). En ese mismo sentido habría una fuerte relación entre los instrumentos de medición de ambas dimensiones. Por citar un ejemplo encontramos los estudios realizados por Vermeer (2001). En una primera experimentación se enfocó en una serie de tareas de descripción y asociaciones de vocabulario receptivo en un grupo de 50 estudiantes de holandés, en tanto en su segundo estudio examinó la relación entre la probabilidad de conocer una palabra y la frecuencia de entrada (*input*) de dicha palabra en 1,600 estudiantes de educación básica de entre 4 y 7 años de edad que aprendían el holandés como primera o segunda lengua. Los resultados para ambos estudios mostraron calificaciones más altas en los estudiantes cuya primera lengua era el holandés, aunque no hubo diferencias significativas entre el tamaño del vocabulario y el número y tipos de aspectos de significado entre aquellos cuya primera lengua era el holandés y aquellos que lo estudiaban como segunda lengua.

El argumento de Vermeer es que estas correlaciones apoyan la idea del desarrollo de una cierta estabilidad en el proceso de adquisición del vocabulario. Es decir, las palabras no se aprenden en un orden arbitrario sino que las más básicas se aprenden antes que aquellas que corresponden a lo que llama “supercategorías” y “subcategorías”. La profundidad del vocabulario se relacionaría con el conocimiento del hablante de dichas categorías. Vermeer explica que estas correlaciones altas son una consecuencia lógica a partir del hecho de que los elementos léxicos dentro del lexicón consisten en una serie de nodos interrelacionados en una red, la cual especifica el significado de un

elemento. Incluso afirma que es posible el hecho de que un niño que conozca más palabras puede conocer las palabras a mayor profundidad. De igual manera, el autor considera que la frecuencia juega un papel importante en el conocimiento de palabras al grado de ser ésta un factor que afecta tanto la amplitud como la profundidad del conocimiento del vocabulario.

Así pues, a manera de conclusión, se debe considerar estas observaciones y posteriormente generar investigaciones que estudien la correlación entre diversas pruebas de amplitud y profundidad de vocabulario para así generar instrumentos válidos y confiables que logren una valoración conjunta del conocimiento del hablante.

#### 5.1.1 El Vocabulario Académico

En años recientes, diversos investigadores se han dedicado al estudio del vocabulario académico para disciplinas específicas, sosteniendo que existen rasgos individuales que identifican cada área o sub-área de estudio.

Los resultados obtenidos por este tipo de investigaciones han demostrado que cada disciplina tiene su propia forma de argumentación (Hyland 2006, Hyland y Tse 2007), y que las palabras en cada disciplina adquieren un significado altamente especializado, el cual es significativo únicamente para los miembros de una comunidad académica específica. Muchas palabras presentan diferentes significados entre diferentes disciplinas o desarrollan colocaciones y significados asociados a campos de estudio específicos (Hyland y Tse, 2007).

Por ejemplo, el término “solución” en el área de las Matemáticas regularmente se interpreta como “resolver un problema matemático”, mientras que el mismo término en la química significa “líquido” (Mudraya, 2006). Este fenómeno no sólo

se presenta en las palabras aisladas, también se presenta en las colocaciones: “Medio de Cultivo” desde el punto de vista de la Biología es la “sustancia nutritiva para el crecimiento de microorganismos” mientras que para la Agronomía es el “instrumento para la siembra de plantas y vegetales” (Martínez et al., 2009).

Ahora bien, los textos científicos incluyen diferentes tipos de vocabulario. Entre las distinciones que se encuentran en la literatura podemos mencionar tres tipos: el vocabulario académico, el técnico y el sub-técnico. A continuación se describe brevemente cada uno:

#### A) Vocabulario Académico

De acuerdo con Coxhead y Nation (2001) el vocabulario académico es aquel que se encuentra en textos de tipo académico (libros de texto, artículos científicos, etc.) y que es común a diferentes campos de estudio. Los autores sostienen que este tipo de vocabulario es el que los estudiantes necesitan conocer si desean estudiar en una universidad de un país angloparlante. Las palabras que componen este tipo de vocabulario de hecho son consideradas de uso común y pueden ser muy útiles incluso para aquellos hablantes que no tienen metas académicas, ya que se encuentran frecuentemente en periódicos, revistas, novelas, y pueden ser escuchadas en la televisión, radio o películas. Algunos ejemplos de este tipo de palabras en inglés son: *area, energy, function, issue, similar, environment, capacity, vary, albeit, intrinsic, qualitative, hierarchy, equate*.

La importancia de este tipo de vocabulario radica en valor para la construcción de la experiencia científica, es decir, el discurso académico cuenta con una serie de características que lo diferencian del discurso de ficción y que son comunes e identificables para los miembros de una comunidad académica (Jones y Haywood, 2004).

## B) Vocabulario Técnico

El vocabulario técnico es aquel que también se incluye en los textos científicos, pero que se caracteriza por no contar con sinónimos exactos, por tener alta resistencia al cambio semántico y un rango muy limitado de uso (e.g. *vulcanizar, uretra*) (Mudraya, 2006). A diferencia del vocabulario académico, este depende enteramente de la disciplina científica a la que pertenece y rara vez se encuentra fuera de su contexto, es decir, no pertenecen al vocabulario general y cotidiano de una lengua, sino más bien al vocabulario especializado de una comunidad científica específica.

## C) Vocabulario Sub-Técnico

Algunos autores (Baker, 1988; Cowan, 1974; Flowerdew, 1993; Trimble, 1985) identifican una sub-categoría del vocabulario técnico a la cual denominan “Vocabulario Sub-técnico”. En esta categoría se encuentran palabras como *fuerza, tensión y corriente* las cuales cuentan con un sentido tanto técnico como de uso común, es decir que no son altamente técnicas ni específicas de un área de conocimiento, pero tampoco son palabras de uso informal que no aparecen en textos especializados (Baker, 1988). De hecho, se ha demostrado que las palabras sub-técnicas tienen muy alta distribución en textos académicos de diversos campos científicos (Yang, 1986).

## 5.2 Evaluación de Vocabulario

Para poder evaluar el vocabulario, antes es necesario considerar los elementos involucrados en el conocimiento de una palabra. Nation (1990) en su propuesta sobre los componentes del conocimiento de vocabulario plantea la diferencia entre el vocabulario receptivo, que es aquel que involucra la

comprensión de las palabras por parte del hablante y el vocabulario productivo, que radica en la capacidad del hablante para utilizar el vocabulario que conoce. La Tabla 1 muestra esta distinción junto con los elementos involucrados en el conocimiento del vocabulario:

<i>Forma</i>		
Forma oral	R	¿Cómo suena la palabra?
	P	¿Cómo se pronuncia la palabra?
Forma escrita	R	¿Cómo se ve la palabra?
	P	¿Cómo se escribe la palabra?
<i>Posición</i>		
Patrones gramaticales	R	¿En qué patrones ocurre la palabra?
	P	¿En qué patrones debemos usar la palabra?
Colocaciones	R	¿Qué palabras o tipos de palabras pueden ocurrir antes o después de la palabra?
	P	¿Qué palabras o tipos de palabras debemos usar con esta palabra?
<i>Función</i>		
Frecuencia	R	¿Qué tan común es la palabra?
	P	¿Con qué frecuencia debemos usar la palabra?
Propiedad de contexto	R	¿Dónde esperamos encontrar esta palabra?
	P	¿Dónde se puede usar esta palabra?
<i>Significado</i>		
Concepto	R	¿Qué significa la palabra?
	P	¿Qué palabra expresa este significado?
Asociaciones	R	¿En qué otras palabras nos hace pensar esta palabra?
	P	¿Qué otras palabras se pueden usar en lugar de ésta?

R= Vocabulario Receptivo; P= Vocabulario Productivo.

**Tabla 1 Componentes de conocimiento de vocabulario (Nation, 1990)**

Si observamos todos los elementos mencionados por Nation (1990) encontramos que el conocimiento del vocabulario no sólo implica el saber el significado que una palabra tiene ya sea aislada o dentro de un contexto específico. También involucra elementos como su frecuencia, su escritura y hasta la propiedad de su uso en situaciones específicas.

Esta clasificación no sólo plantea la complejidad que representa el conocer una palabra, sino que además obliga a pensar que la evaluación del conocimiento léxico no puede llevarse a cabo con un solo tipo de instrumento, por el contrario, es necesario recurrir a diversos formatos, metodologías y recursos que nos permitan evaluar tanto las habilidades productivas como la receptivas.

Pero ¿cómo debe evaluarse el vocabulario? Read (2005) analiza las posibilidades de la evaluación del vocabulario planteando una distinción entre considerar al vocabulario como un constructo independiente o dependiente; la manera de seleccionar los ítems para la evaluación (selectivo ó comprensivo) y el papel del contexto en que se encuentre o no inmerso el vocabulario a evaluar. La Tabla 2 muestra esta propuesta:

<p><b>Discreto</b></p> <p>Una medida del conocimiento o uso del vocabulario como constructo independiente</p>	<p><b>Insertado</b></p> <p>Una medida del vocabulario que forma parte de otro constructo mayor.</p>
<p><b>Selectivo</b></p> <p>Una medida en la que algunos ítems específicos de vocabulario son el centro de la evaluación.</p>	<p><b>Comprensivo</b></p> <p>Una medida que toma en consideración todo el contenido del vocabulario en tareas de comprensión lectora o auditiva, o la respuesta del sujeto en tareas orales o de escritura.</p>
<p><b>Independiente del contexto</b></p> <p>Una medida de vocabulario en la cual el estudiante puede producir la respuesta esperada sin referirse a algún contexto.</p>	<p><b>Dependiente del contexto</b></p> <p>Una medida del vocabulario que evalúa la habilidad del estudiante de considerar información del contexto para producir la respuesta esperada.</p>

**Tabla 2. Dimensiones de evaluación del vocabulario propuestas por Read (2000)**

En esta tabla, la primera dimensión –discreto o insertado- hace referencia al constructo de la prueba. Un constructo es, en evaluación de la lengua, un atributo mental o una habilidad que se pretende medir con el instrumento. Es decir, el tipo de prueba que se aplique para evaluar el vocabulario puede considerar al léxico como un constructo independiente o evaluar al vocabulario como parte del conocimiento general de una lengua dentro de un contexto más comunicativo. Una prueba discreta considera el conocimiento del vocabulario como un elemento independiente de otros elementos de la competencia lingüística, mientras que una prueba de vocabulario insertado contribuye a la evaluación de un constructo mayor.

La segunda dimensión plantea la distinción entre vocabulario selectivo y el comprensivo. Esto es el rango del vocabulario a evaluar. Mientras que una prueba selectiva incluye una serie de palabras meta previamente seleccionadas, una prueba comprensiva evalúa el uso general del vocabulario del hablante de acuerdo a ciertos criterios sin esperar la producción de una lista particular de palabras. Es decir, evalúa el vocabulario a partir de la producción libre del hablante.

La última dimensión de la evaluación de vocabulario propuesta por este autor hace referencia al contexto. Tradicionalmente, se ha considerado por contexto el hecho de que el vocabulario evaluado se presente de manera aislada (independiente del contexto) o dentro de una oración o texto comunicativo (dependiente del contexto).

Con base en estas distinciones, diversos investigadores han generado nuevos instrumentos para medir y evaluar el vocabulario. Al igual que la conceptualización general del vocabulario, estas medidas también se han separado tradicionalmente en tamaño y profundidad de conocimiento. Mientras que las medidas de tamaño implican un gran número de palabras evaluadas a las cuales los participantes deben otorgar una respuesta simple (*multiple choice, C-test, etc.*), las medidas de profundidad implican lo que Anderson y Freebody (1981) definen como “calidad de entendimiento”. Esta vaga definición ha



generado que los académicos sigan diferentes patrones para evaluar el conocimiento del vocabulario.

Nuevamente Read (2004) resume 3 líneas o conceptualizaciones utilizadas para la evaluación del vocabulario en diversas investigaciones:

1. Precisión de significado: La diferencia entre poseer una idea limitada del significado de una palabra y tener un conocimiento elaborado y específico sobre el mismo.
2. Conocimiento comprensivo: Conocimiento de una palabra que incluye no solamente sus rasgos semánticos, sino ortográficos, fonológicos, morfológicos, sintácticos, de colocación y pragmáticos.
3. Conocimiento de red: La incorporación de la palabra dentro de una red en el léxico, que permita relacionarla y diferenciarla del resto de las palabras que componen la misma.

Este mismo autor menciona varios puntos que deben tenerse en consideración en relación a estas conceptualizaciones:

- a) Los enfoques pueden traslaparse en diversos puntos.
- b) Del tipo de conceptualización adoptada, depende el tipo de instrumento desarrollado para evaluar el conocimiento del vocabulario.
- c) Las aproximaciones pueden conceptualizarse en términos del estado del conocimiento del aprendiz en un punto específico o como un proceso de desarrollo léxico.
- d) Estas aproximaciones típicamente corresponden a la evaluación del conocimiento declarativo – al cual se puede acceder y reportar en una prueba- y no al conocimiento de procesamiento.

A continuación se describen más detalladamente cada una de las tres aproximaciones, junto con algunos ejemplos de pruebas y estudios realizados a partir de cada una de ellas:

## I. Precisión de Significado:

Esta aproximación presenta la dificultad de que las palabras son ambiguas y polisemánticas dependiendo del contexto en que se encuentren. Sin embargo, este enfoque se ha optimizado a partir de tres líneas:

- a. Diseñar pruebas con ítems que requieran mayor precisión de conocimiento: Los ítems deben ser diseñados de forma tal que requieran de un entendimiento mayor de la palabra para poder ser contestados, como un ejercicio de *matching* o completar la oración en la cual se requiera el significado menos frecuente de una palabra polisémica.
- b. Escalas de autoreporte de conocimiento: Esta noción de escalas de grados de conocimiento de vago a preciso es altamente utilizada en los estudios de vocabulario. Estas escalas permiten que el alumno autoevalúe su conocimiento específico sobre una palabra meta.
- c. Solicitar definiciones: Este método requiere solicitar a los aprendices que expliquen el significado de una serie de palabras, para después evaluar la calidad de las definiciones de cada participante. Este método presenta la dificultad que existe entre conocer el significado de una palabra y ser capaz de brindar una definición del mismo. Además, una definición no siempre incluye los múltiples contextos en los que una palabra puede ocurrir.

## II. Conocimiento Comprensivo:

De acuerdo con Nation y Waring (1997) existen tres factores involucrados en el conocimiento de una palabra:

*Forma:* Pronunciación, escritura y morfología.

*Significado:* Relación forma-significado, concepto y referentes, asociaciones.

*Uso:* Funciones gramaticales, colocaciones, restricciones de uso, etc.

Cada una de estas categorías involucra tanto el conocimiento productivo como el receptivo, lo cual genera aún más complejidad en el momento de diseñar el instrumento de evaluación, ya que se deben determinar muy claramente cuáles son los aspectos específicos del conocimiento que se van a evaluar y si unos son más importantes que otros. Por otra parte, se presenta el problema de decidir sobre el número de palabras a evaluar. Enfocarse en un pequeño número de palabras puede conducir a importantes resultados en un estudio sobre adquisición de lengua, pero no nos dará información para la toma de decisiones dirigidas a mejorar los procesos de adquisición.

### III. Conocimiento de red:

Este conocimiento se refiere a la habilidad de los aprendices para distinguir palabras relacionadas semánticamente, así como a su conocimiento sobre palabras que de alguna manera están relacionadas entre sí, conformándolas en grupos dentro del léxico.

La técnica básica para evaluar este tipo de conocimiento ha sido por medio de las asociaciones, en busca de alguna de las tres relaciones básicas entre una palabra ítem y sus asociaciones: 1) paradigmática (superordinados, sinónimos) 2) sintagmática (colocaciones) y 3) analítica (palabras que representan un elemento clave para el significado de la palabra meta).

#### 5.2.1 Diseño y validez de las Pruebas de Evaluación de Vocabulario

En la actualidad, las pruebas dedicadas a la evaluación del vocabulario son utilizados con diversos propósitos tanto para la enseñanza como para la investigación de los procesos de adquisición de lengua, sin embargo aún no existe una base sólida para evaluar estos instrumentos o para elaborar nuevos; hoy en día, existen diversos estudios sobre el concepto y el acercamiento más adecuado a la evaluación del conocimiento del vocabulario, incluso difiriendo sobre si éste es o no un constructo evaluable independientemente.

Por una parte, encontramos a los investigadores que evalúan el vocabulario como un constructo independiente en términos de su tamaño y crecimiento, utilizando palabras de contenido independientes y descontextualizadas, y cuyos instrumentos se adecúan a la tradición psicométrica-estructuralista de la enseñanza del lenguaje, e.g: *Vocabulary Levels Test*.

Como una segunda aproximación, se encuentran aquellos investigadores que conciben la evaluación del vocabulario como una actividad más integral, en dónde las palabras meta deben ser evaluadas en un contexto. e.g. *C-test*. Y finalmente, encontramos a aquellos teóricos para los cuales la evaluación del vocabulario no constituye un constructo por sí misma, ya que el vocabulario forma parte del desempeño lingüístico en general.

Con el objeto de establecer una serie de guías para evaluar los distintos tipos de pruebas de vocabulario utilizados, Read y Chapelle (2001) proponen un marco de referencia que tome en consideración las implicaciones del propósito de la prueba para el diseño de la misma:

#### A) Definición del Constructo:

Es poca la atención que los investigadores y diseñadores de pruebas de vocabulario han prestado a la definición del constructo dentro del cual serán interpretados los resultados de las pruebas, y por lo tanto será evaluado el desempeño del aprendiz. En general encontramos tres tipos de definiciones de vocabulario como constructo:

1. Rasgo: El desempeño de la prueba depende de las características del aprendiz, por lo tanto las variables contextuales no son representativas y el contexto es eliminado en gran medida de los ítems de la prueba. e.g. *Vocabulary Levels Test*.

2. Conductista: El contexto tiene un rol central dentro de esta conceptualización. En este tipo de consideración se desea replicar las condiciones en las cuales el aprendiz sería capaz de utilizar el vocabulario evaluado, y el objetivo es evaluar su desempeño lingüístico general. Así pues, la definición del constructo depende de la definición del contexto. Por lo tanto estas pruebas son altamente comprensivas y dependientes del contexto, e.g. *TOEFL*, Sección de Lectura de Comprensión.
3. Interaccionalista: Este constructo es una combinación de los dos anteriores, ya que define el vocabulario a evaluar (rasgo) pero dentro de un contexto dado (conductista). e.g. *Lexical Frequency Profile*.

Tradicionalmente, es el constructo de rasgo el que ha predominado en pruebas discretas, selectivas e independientes del contexto, las cuales evalúan palabras aisladas, aunque de acuerdo con los autores es necesario adoptar un enfoque más interaccionista que permita evaluar el conocimiento de vocabulario en contextos significativos.

#### B) Objetivo de la prueba y su validez:

##### 1. Inferencias y validez del constructo:

Las inferencias son las conclusiones sobre la habilidad en la lengua meta del aprendiz a las cuales se puede llegar a través de la observación de su desempeño en la prueba, e.g. asignar una calificación o un nivel. Las inferencias deben estar soportadas por un constructo válido ya que de esta manera pueden ser interpretadas en varios niveles:

- a) Por rasgo.
- b) Como sub-prueba, cuando la evaluación del vocabulario se encuentra inmersa en un constructo mayor.

- c) Como prueba general de habilidad lingüística, cuando el objetivo de la prueba es evaluar la capacidad lingüística general del alumno, no sólo su conocimiento de vocabulario.

## 2. Aplicaciones, relevancia y utilidad:

Por aplicaciones entendemos los objetivos prácticos para los cuales son utilizadas las pruebas. Pueden dividirse en 3 categorías:

- a) Instrucción: La prueba se utiliza para la toma de decisiones en un contexto educativo. Es el uso más común de las pruebas, e.g. Colocación, logro, desempeño general.
- b) Investigación: La prueba es utilizada en investigaciones empíricas en el campo de la evaluación y adquisición de vocabulario. Los resultados no son utilizados para evaluar a los individuos, sino que son utilizados para responder hipótesis.
- c) Evaluación: La prueba se utiliza para tomar decisiones respecto a la calidad de los programas educativos y métodos.

## 3. Impacto y consecuencias:

El impacto de la prueba se refiere a los efectos que ésta tendrá sobre sus usuarios: los estudiantes, los maestros, las clases, los programas e incluso la sociedad en general. Estos efectos deben ser percibidos como parte integral del diseño de la prueba. Las consecuencias de implementar una prueba para cumplir con un propósito específico pueden ser evaluadas en relación a los efectos que tendrá la misma dentro desde el mismo proceso de diseño.

## C) Factores que intervienen entre el propósito y el diseño de la Prueba

## 1. Definición del Constructo

## 2. Reporte del desempeño del aprendiz

Este factor implica la forma en que los resultados de las pruebas serán presentados, así como la manera en que éstos serán evaluados. Pueden mencionarse tres niveles para establecer una calificación o reporte:

*Nivel 1.* Evalúa el desempeño lingüístico general del aprendiz. Regularmente la calificación se emite en un solo número o grado. Se utiliza para la toma de decisiones . E.g. TOEFL

*Nivel 2.* En este tipo de calificación, los estudiantes son continuamente evaluados durante un curso y los resultados generalmente son reportes que describen en términos funcionales las habilidades comunicativas del aprendiz.

*Nivel 3.* En este tipo de reporte, el profesor monitorea el desempeño del estudiante y le brinda retroalimentación informal constante sobre su desempeño.

## 3. Presentación de la Prueba

Este factor involucra tanto el formato de la prueba como la audiencia a la cual va dirigida la misma: alumnos, profesores, investigadores, etc.

### D) Validación de la Prueba:

La validación de la prueba debe consistir en una serie sólida de argumentos tanto teóricos como empíricos. La validación debe ser guiada por las necesidades de la prueba en un contexto específico en donde las elecciones

correspondientes a la definición del constructo, el reporte de calificaciones y la presentación de la prueba deben ser justificadas en un solo argumento.

Es justo aquí donde encontramos en la literatura que diversos investigadores relacionados con la adquisición y evaluación del vocabulario, continuamente se encuentran con la dificultad de garantizar la validez de las pruebas que utilizan para evaluar el conocimiento del vocabulario. Para poder llegar a una solución satisfactoria, Nation (2007) plantea los siguientes elementos a considerar:

#### 1) La actitud del aprendiz de lengua y la variabilidad individual

Es importante considerar que no todos los estudiantes responden a las pruebas con seriedad, y contestan rápidamente con el único objetivo de terminar la tarea asignada. Otro factor que puede afectar su desempeño es una mala experiencia al responder pruebas anteriores, ya que esto puede deteriorar su nivel de confianza respecto a su propia habilidad y conocimiento. De igual manera, es necesario considerar que no todos los estudiantes conocen las estrategias adecuadas para responder a una prueba.

Por otro lado, también se deben considerar las limitaciones de la prueba; el formato de la misma puede afectar su nivel de comprensión (falta de contexto, homonimia, etc.) ó los ítems incluidos pueden no ser los más adecuados para el nivel y características de los aprendices.

#### 2) Listas de Frecuencia

Las pruebas de vocabulario generalmente se basan en listas de frecuencia de vocabulario, las cuales últimamente se han visto complementadas con el trabajo desarrollado por la lingüística de corpus. Sin embargo, es necesario considerar que la naturaleza del corpus en el cual se basan las listas de frecuencia influye en el contenido del mismo, es decir, un corpus desarrollado para representar el lenguaje académico va a arrojar listas de frecuencia relacionadas con el lenguaje académico, y esta lista no podrá utilizarse para evaluar el conocimiento de vocabulario de niños o adolescentes, ya que ésta no es representativa del tipo de lenguaje que se pretende evaluar.

#### 3) Unidades de conteo



Es importante que las unidades de conteo que se utilicen para medir el vocabulario de los aprendices coincidan con las unidades de los datos que se utilizan en la prueba, ya que una unidad inapropiada puede conducir a una sobrestimación o subestimación del conocimiento del vocabulario de los sujetos. De acuerdo con Vermeer (2001) para la evaluación del uso productivo del lenguaje, la mejor unidad de conteo es la de lemas, ya que involucra diferentes colocaciones y construcciones gramaticales. Sin embargo, para la evaluación del lenguaje receptivo, las familias de palabra resultan más convenientes (Bertram et al., 2000).

#### 4) Medidas múltiples

Utilizar dos o tres pruebas para medir el conocimiento del vocabulario dentro de un mismo estudio permite observar la validez de los resultados obtenidos, a la vez que garantiza la obtención de datos utilizables independientemente de que una de las pruebas resulte muy difícil o muy fácil para los participantes, además permite evaluar las diferencias entre distintos métodos de aprendizaje y evaluación.

#### 5) Formatos de la Prueba

En años recientes, algunos formatos de prueba han demostrado obtener mejores resultados que otros. Es importante utilizar un formato que sea familiar y claro para los estudiantes cuyo vocabulario va a ser evaluado. Un formato demasiado novedoso o complejo puede no reflejar el verdadero conocimiento de los participantes.

#### 6) Evaluar el vocabulario en uso

Las medidas de vocabulario en uso tienen la característica de evaluar el vocabulario como parte de un constructo mayor, como son la lectura o la escritura, y consideran el vocabulario en su contexto comunicativo. Estas medidas han demostrado tener correlación con pruebas de carácter discreto, como son las pruebas de opción múltiple o de traducción. Sin embargo, no pueden determinar el tamaño del vocabulario de los participantes, ya que sólo están diseñadas para

evaluar el uso del vocabulario del aprendiz en una tarea comunicativa determinada.

La medida de vocabulario en uso más comúnmente utilizada son las escalas, en donde el vocabulario es una sub-escala de una escala mayor. Regularmente estas escalas constan de 4 o 5 niveles que catalogan el nivel del conocimiento del vocabulario de muy pobre hasta excelente y sofisticado.

#### 5.2.1.1 Pruebas de Decisión Léxica con formato *Checklist*

Las pruebas de decisión léxica con formato de *checklist* miden el conocimiento receptivo del vocabulario a través de una tarea de reconocimiento (Anderson y Freebody, 1983). Tradicionalmente, el formato presenta una lista extensa de vocabulario meta donde únicamente se debe responder a las palabras conocidas. Es una prueba independiente del contexto y únicamente evalúa el tamaño del vocabulario y no la profundidad de conocimiento.

Las palabras meta incluidas en el formato se seleccionan bajo el criterio de frecuencia dentro de listas o corpus especializados. Por su naturaleza este formato permite evaluar rápidamente un gran número de ítems de vocabulario.

En algunas ocasiones, estas tareas de decisión léxica pueden incluir no palabras. Una no palabra es una unidad del discurso o texto la cual aparenta ser una palabra real en un idioma, aunque en realidad no forme parte del léxico de una lengua específica. Estas son incluidas en algunas pruebas ya que reducen la posibilidad de que los participantes sobrestimen su conocimiento; es decir, si el participante indica que conoce todas las palabras reales mientras que no marca las no palabras, es muy probable que en realidad conozca las palabras reales que dice conocer. Si por el contrario, el participante manifiesta conocer tanto las palabras reales como las no palabras es posible que esté sobreestimando su vocabulario.

Las respuestas afirmativas a las palabras reales (*hits*) reflejan el conocimiento del vocabulario de estudiante, mientras que las respuestas

afirmativas a las no palabras (falsas alarmas) miden la tendencia individual a seleccionar arbitrariamente las respuestas. Tanto los *hits* como las respuestas negativas a las no palabras son consideradas respuestas correctas, y el número de *hits* modificado a partir del índice de falsas alarmas es el resultado obtenido en la prueba por el hablante evaluado.

Debido a las características de su formato estas pruebas son muy fáciles de elaborar y de calificar; además de que nos permiten evaluar un gran número de palabras en poco tiempo. Sin embargo, han sido fuertemente criticadas por diversos investigadores (Read, 2005) debido a que no evalúan el vocabulario dentro de contextos comunicativos, involucran únicamente el ámbito receptivo pero no en el productivo y porque, al ser una autoevaluación, dan lugar a que los estudiantes sobreestimen su propio conocimiento.

En sus orígenes, esta prueba fue usada en el campo de adquisición de primeras lenguas (Sims, 1929; Tilley, 1936); posteriormente, se modificó, incluyendo no-palabras (Zimmerman et al 1977; Anderson y Freebody, 1983) y después se probó su funcionamiento en los procesos de adquisición de vocabulario en segundas lenguas (Meara y Buxton, 1987; y Mochida y Harrington, 2006). Algunos de estos estudios se detallan en la siguiente sección.

#### 5.2.1.1.1 Estudios previos realizados con el formato de *Checklist*

En esta sección se comentan brevemente los estudios más representativos que se han llevado a cabo previamente utilizando la prueba de decisión léxica con formato de *checklist*. También se comentan las conclusiones de cada estudio así como la forma de validación del mismo al comparar dicho formato con otros formatos también utilizados para la evaluación del vocabulario.

En Anderson y Freebody (1983). compararon los resultados arrojados por dos pruebas de evaluación del vocabulario en primera lengua (L1). La primera prueba aplicada fue de opción múltiple mientras que la otra era una *checklist*. En su estudio participaron 120 estudiantes norteamericanos de educación primaria. La prueba de opción múltiple incluía 195 palabras en inglés; mientras que la

*checklist* incluía las mismas 195 palabras más 131 no palabras. Los resultados de ambas pruebas arrojaron una correlación de .84, lo cual significa que ambas pruebas se correlacionan significativamente.

Pocos años después Meara y Buxton (1987) realizaron un estudio similar pero aplicado a la adquisición de inglés como segunda lengua. Los autores compararon los resultados de la *Cambridge First Certificate Examination*, la cual es una prueba de opción múltiple, con los de una prueba diseñada por ellos mismos con formato de *checklist*. Este estudio se llevo a cabo 100 con estudiantes universitarios de *ESL (English as a Second Language)*. Nuevamente, se obtuvo una alta correlación entre los resultados de ambas pruebas (.829) validando nuevamente a las *checklists* como instrumentos para la evaluación del vocabulario.

Tras analizar sus resultados, estos investigadores plantearon la necesidad de confirmar la validez del formato de opción múltiple, ya que cuando el vocabulario a evaluar no es muy amplio, la tasa de muestreo es aceptable, pero a medida que el vocabulario aumenta la tasa de muestreo presenta malos resultados. En teoría, este problema se puede resolver al aumentar el número de ítems incluidos en la prueba, pero sostienen que esto no resultaría práctico, ya que la prueba resultaría demasiado compleja y extensa, lo cual presentaba más ventajas para el formato de *checklist*, debido a su fácil elaboración, su rápida aplicación y la posibilidad de incluir un mayor número de palabras a ser evaluadas.

En el 2001, Beckmans et al. realizaron un estudio en el cual compararon los resultados obtenidos por tres pruebas que incluían elementos de medición del vocabulario, una con formato de verdadero/falso, la segunda de opción múltiple y la tercera con formato de *checklist*. En el estudio participaron 352 estudiantes universitarios cuya primera lengua era el francés, pero aprendían el holandés como segunda lengua.

Para la prueba de verdadero/falso se elaboraron una serie de ítems de gramática conformados por múltiples oraciones representativas de las categorías gramaticales que los estudiantes de holandés como segunda lengua debían

conocer. Mientras que 50 de estas oraciones eran gramaticalmente correctas, 50 eran incorrectas. Para la prueba de opción múltiple se mantuvieron los mismos contenidos gramaticales, sólo que únicamente se elaboraron 78 ítems con 4 opciones de respuesta cada uno. Mientras que la prueba de formato de *checklist* tenía dos versiones, cada una con 100 ítems, de los cuales 60 eran palabras y 40 no palabras. Todas las palabras fueron tomadas de un corpus del holandés, el cual contiene 3700 palabras seleccionadas por su frecuencia. Las no palabras se crearon modificando la morfología de una palabra real, pero siguiendo las reglas fonotácticas y morfológicas de la lengua.

Los resultados obtenidos concluyen que los estudiantes omiten más respuestas en las pruebas de formato de verdadero/falso que en el de opción múltiple, lo cual repercute en la confiabilidad de la prueba de verdadero/falso. La confiabilidad de la prueba de formato de *checklist* se analizó con el *Alfa de Cronbach* (Ver 8) y la fórmula de *Spearman Brown*. El análisis arrojó en promedio correlaciones de 0.70, pero este índice aumenta cuando se consideran sólo las palabras reales. Sin embargo, los autores sostienen que el incremento en la confiabilidad debe interpretarse como una medida precisa del sesgo y no como garantía de una medida adecuada de conocimiento de vocabulario.

En sus conclusiones, Beckmans *et al* discuten sobre los factores (lingüísticos, socioculturales, metalingüísticos, etc.) que pueden afectar el tamaño del sesgo, así como la instrucción y el contenido de la prueba. Sugieren que hay que considerar estos factores para tener una elección más motivada de las palabras a evaluar y formular las no palabras con principios claros de alteración. Sostienen que la investigación más profunda en estos factores, apoyada con datos empíricos mostrará si la prueba de formato *checklist* es una medida válida de tamaño de vocabulario receptivo.

Para el año 2002, Huijbregtse y Admiraal realizaron un estudio cuyo objetivo era el de analizar tres métodos distintos de procesar los resultados obtenidos en una prueba de formato *checklist* para medir el tamaño de vocabulario en los procesos de adquisición de lengua. Los autores sostienen que ya que una *checklist* es una prueba de autoreporte de conocimiento donde se incluyen no palabras, es necesario considerar distintos métodos de estimación de

vocabulario y ajustar la posibilidad de adivinar la respuesta (*correction for guessing*) y el estilo de responder de los participantes (*correction for participants' response style*), ya que la resolución y por lo tanto los resultados obtenidos en la prueba, pueden verse influenciados por estos factores. Los autores sostienen que el formato tiene 4 posibles respuestas por parte de los participantes evaluados, de las cuales dos son correctas – responder afirmativamente a una palabra real y no responder a una no palabra- y dos son incorrectas –no responder a una palabra real y responder afirmativamente a una no palabra. Esta relación se ilustra en la siguiente tabla:

Alternativa de estímulo	Alternativa de respuesta	
	SÍ	NO
	<p>Sí/palabra Correcto (hit)</p>	<p>No/palabra Incorrecto (miss)</p>
<p>Sí/ no palabra Falsa alarma</p>	<p>No/no palabra Identificación correcta</p>	

**Tabla 3. Posibles respuestas a la Prueba de formato Checklist (Huibregtse y Admiraal, 2002)**

Estas variaciones en las repuestas tienen implicaciones al momento de evaluar el conocimiento de vocabulario de un hablante o aprendiz. Por ejemplo, puede ser que un estudiante piense que conoce una palabra pero no está seguro de su significado exacto. Si esto fuera cierto, el estudiante no estaría adivinando totalmente al azar, si no con cierto conocimiento previo. En cuanto al estilo de responder, encontramos que existen estudiantes con una alta tendencia a responder afirmativamente a todos los estímulos, mientras que otros estudiantes, más conservadores, prefieren no contestar afirmativamente a menos que estén totalmente seguros de la respuesta.

Considerando lo anterior, Huibregtse y Admiraal (2002) proponen que el cálculo de estimación de vocabulario de una prueba con formato de *checklist* debe considerar tres condiciones:

- Que existen dos tipos de respuestas correctas y dos tipos de respuestas incorrectas
- Que los participantes tienen la posibilidad de adivinar
- Que los participantes tienen diferentes estilos de respuesta
- 

Tomando en cuenta dichas condiciones, los autores analizan tres métodos para evaluar el conocimiento del vocabulario de los participantes que responden a una prueba de formato *checklist*:

- 1) El conteo neto del número de respuestas correctas menos las falsas alarmas

Este método consiste en restar las falsas alarmas (no palabras marcadas como reales) del número total de *hits* (las palabras reales marcadas afirmativamente). Es decir, penalizar a los estudiantes por la posibilidad de adivinar la respuesta o sobreestimar su vocabulario. La fórmula aplicada en este método es:

$h = \text{hits (palabras reales marcadas afirmativamente)}$   
 $f = \text{falsas alarmas (no palabras marcadas como reales)}$

2) Corrección por Adivinar (*Correction for guessing: cfg*)

El método está basado en el *Blind Guessing Model*. Este modelo asume que los estudiantes conocen completamente o no conocen en absoluto el ítem evaluado, sin considerar la posibilidad de que el participante adivine “sofisticadamente”, por lo que penaliza fuertemente todas las no palabras marcadas como conocidas. La fórmula aplicada en este método es la siguiente:

—

$h$  = hits (palabras reales marcadas afirmativamente)  
 $f$  = falsas alarmas (no palabras marcadas como reales)

3) Teoría de Detección de Señales (*Signal detection theory*)

Este modelo, como su nombre lo indica, está basado en la teoría de detección de señales y considera la posibilidad de adivinar sofisticadamente o con cierto nivel de conocimiento previo al asumir que cada alternativa de respuesta posee cierto grado de credibilidad. Sin embargo, subestima el conocimiento de los estudiantes con un estilo extremadamente conservador que responden afirmativamente sólo cuando están totalmente seguros de que conocen la palabra. La fórmula que se aplica en este método es:

— . .

$h$  = hits (palabras reales marcadas afirmativamente)  
 $f$  = falsas alarmas (no palabras marcadas como reales)

4) Nueva propuesta para la Teoría de Detección de Señales (*Signal detection theory: sdt*)



Finalmente, Huijbregtse y Admiraal (2002) proponen un nuevo método basado en la fórmula anterior, pero considerando el estilo de responder de los participantes, proponiendo la aplicación de la siguiente fórmula:

$$I_{SDT} = 1 - \frac{4h(1-f) - 2(h-f)(1+h-f)}{4h(1-f) - (h-f)(1+h-f)}$$

*h = hits (palabras reales marcadas afirmativamente)*  
*f = falsas alarmas (no palabras marcadas como reales)*

Por último, los autores discuten la necesidad de realizar más investigación relacionada con la forma de analizar los datos arrojados por las pruebas de formato *checklist*, ya que para ellos aún es necesario definir qué significa la calificación obtenida en este tipo de prueba.

En el año 2006 Mochida y Harrington, basados en los resultados obtenidos por Huijbregtse y Admiraal. (2002), realizan un estudio en el que comparan los resultados de una prueba de formato *checklist* con los resultados obtenidos por un grupo de estudiantes de *ESL* en la prueba de *Vocabulary Levels Test* propuesta por Nation en 1990.

Esta última es una prueba de opción múltiple ampliamente difundida y tiene el objetivo de evaluar el tamaño de vocabulario en inglés de los no nativos hablantes de esta lengua que estudian el inglés como L2. Esta versión de la prueba se basa en la *General Service List* (West, 1953) y la *University Word List* (Nation, 1990) (Ver 6.1), cuyas bases teóricas radican en la concepción del vocabulario del inglés, o de cualquier otra lengua, como una serie de niveles basados en la frecuencia de ocurrencia de las palabras (Nation, 1990). Así, los niveles pueden ser interpretados como grupos de 1000 palabras, dando como resultado una serie de listas que contienen en primer lugar las 1000 palabras más frecuentes de una lengua, seguidas por una lista de las 1000 palabras que aparecen en segundo orden de frecuencia, a la que le sigue la lista de las 1000 palabras en tercer lugar de frecuencia y así, consecutivamente.

Los niveles de tamaño de vocabulario que la prueba determina son los siguientes:

De 1000 a 1999 palabras	Nivel 1000 +
De 2000 a 2999 palabras	Nivel 2000 +
De 3000 a 3999 palabras	Nivel 3000 +
De 4000 a 5999 palabras	Nivel 5000 +
De 10 000 palabras en adelante	Nivel 10 000 +

La prueba está diseñada para medir el tamaño del vocabulario del alumno en relación al número de palabras que pueda utilizar productivamente en un contexto controlado. Esto es, que el alumno tenga la capacidad de completar una palabra meta faltante dentro de una oración determinada. El formato de la prueba es el siguiente:

Para cada palabra, se provee una oración significativa que contenga únicamente las primeras letras de la palabra meta. Esto evita que los estudiantes completen la oración con una palabra que, aunque sea semánticamente apropiada dentro del contexto de la oración, pertenezca a una lista de palabras que se encuentre en un nivel de frecuencia distinto al evaluado en ese momento.  
*E.g:*

The book covers a series of isolated epis\_\_\_\_\_ from history.

En el estudio realizado por Mochida y Harrington (2006) se encontró que los resultados obtenidos en esta última prueba eran buenos predictores de los resultados obtenidos por los estudiantes de *ESL* en una prueba de formato *checklist* que evaluaba el nivel de vocabulario presentando una serie de ítems seleccionados a través de la frecuencia, lo cual validaba a las pruebas de decisión léxica como medidas de vocabulario receptivo.

Todos estos estudios nos permiten suponer que la prueba de decisión léxica con formato de *checklist* cuenta con elementos valiosos que coinciden con

los objetivos de esta investigación: a) ha demostrado correlacionarse con pruebas más complejas de evaluación de vocabulario, b) diversos investigadores han encontrado que es un formato válido para la evaluación del vocabulario receptivo, c) al estar basado en frecuencia de ocurrencias responde a la necesidad de ser representativa y d) es un formato que permite diseñar, aplicar y calificar la prueba de forma rápida y práctica, todas características que se buscan para la propuesta metodológica que compete a este trabajo.

En esta sección se establecieron las bases teóricas y metodológicas relacionadas con la evaluación de vocabulario que se aplicaron en la elaboración de este estudio. A continuación se presentan los conceptos básicos relacionados con la lingüística de corpus que también fueron aplicados en esta investigación.

### 5.3 Lingüística de corpus

La lingüística de corpus puede definirse como el estudio y análisis de los fenómenos lingüísticos a través de muestras del habla del “mundo real”, las cuales ocurren en un contexto natural (Hunston, 2002). La lingüística de corpus se basa en una serie de procedimientos o métodos para el estudio del lenguaje. Estos procedimientos continúan desarrollándose y expandiéndose continuamente con el objeto de adaptarse a las necesidades de diversos estudios y áreas de la lingüística. Es en esta capacidad de adaptación donde radica su potencial para redefinir y reorientar las teorías del lenguaje, o por lo menos facilitar la exploración de nuevas teorías sobre el lenguaje. (McEnery y Hardie, 2011)

La base de la lingüística de corpus son los corpus (o *corpora* en latín) los cuales son grandes colecciones de textos que hoy en día son analizadas a través de herramientas automatizadas de procesamiento (Hunston, 2002). Si bien el uso de muestras reales del lenguaje para estudios lingüísticos no es nuevo, en las dos últimas décadas la lingüística de corpus se ha desarrollado considerablemente debido a las posibilidades que ofrece el procesamiento de lenguaje a través de computadoras. La disponibilidad tanto de computadoras como de textos digitalizados ha hecho posible disponer rápida y fácilmente de grandes cantidades

de datos listos para ser procesados. Regularmente, estos corpus son procesados y analizados utilizando herramientas de procesamiento que permiten a los usuarios realizar búsquedas rápidas y confiables. Estas herramientas permiten buscar palabras o series de palabras en contextos definidos- concordancias-, pueden brindar información sobre la frecuencia de los términos o colocaciones meta –datos de frecuencia-, e incluso pueden elaborar listas de frecuencia en las cuales se presenta cada palabra acompañada por el número de veces que aparece en un corpus. Estas concordancias y listas de frecuencia son sólo un ejemplo del tipo de análisis que se puede llevar a cabo utilizando estas herramientas, - análisis cualitativos o cuantitativos- y que son de igual importancia para la lingüística de corpus (McEnery y Hardie, 2011).

La lingüística de corpus tiene múltiples y muy diversas aplicaciones, que van desde estudios descriptivos de sintaxis o prosodia, hasta la enseñanza de primeras y segundas lenguas. La siguiente sección presenta más ejemplos de éstas aplicaciones, además describe brevemente diversos tipos de corpus, sus características y usos.

### 5.3.1 Corpus Lingüísticos

Como ya se mencionó anteriormente, un corpus lingüístico es un conjunto de textos almacenados y agrupados con el fin de estudiar una lengua o una determinada variedad lingüística. Su objetivo es constituirse como un elemento de referencia para el estudio de un cierto aspecto de la lengua, ya que se compone de frases que ocurren naturalmente dentro de la misma (Hudston, 2002).

Los corpus o colecciones de textos se empleaban ya en el siglo XIX e incluso con anterioridad para describir los cambios en una lengua, para justificar enunciados gramaticales, documentar la adquisición de una lengua, elaborar diccionarios o para hacer comparaciones entre diversas lenguas. Principalmente se elaboraban compilando diversos textos escritos del género literario. Debido a que el procesamiento tenía que realizarse de forma manual, era difícil efectuar

análisis cuantitativos. Además no había interés en la composición de los corpus y en consecuencia éstos no eran representativos. Los primeros corpus electrónicos aparecieron en los años 1960 en forma de tarjetas perforadas y los textos desde luego tenían que ser digitalizados. Actualmente existen corpus que cuentan con cientos de millones de palabras, las cuales aumentan día con día debido a la facilidad de recopilar el creciente número de textos electrónicos disponibles (Hudston, 2002).

Los corpus se definen según los objetivos para los que han sido creados y según su contenido. A continuación se presenta una breve descripción y ejemplos de algunos tipos de corpus utilizados actualmente en los estudios de lingüística de corpus (en Huntston, 2002):

#### A) Corpus de lengua general con fines generales

Su objetivo principal es constituir una fuente de información textual de una lengua para fines diversos. Puede incluir lenguaje oral o escrito, o ambos y utilizar textos producidos en uno a más países. Generalmente es muy robusto (incluye un gran número de palabras) y puede ser utilizado para objetivos de enseñanza, traducción o para ser comparado con corpus especializados. También se le conoce como corpus de referencia (McEnery y Hardie, 2011).

Un ejemplo de este tipo de corpus es el Corpus de Referencia del Español Actual (CREA). Desarrollado por el Instituto de Lexicografía de la Real Academia de la Lengua Española, contiene textos literarios, periodísticos, científicos y técnicos, así como transcripciones de grabaciones de la lengua oral y de medios de comunicación correspondientes a los últimos veinticinco años. Es considerado el corpus “oficial” de la lengua española y está disponible en línea de forma gratuita, codificado de forma tal que permite establecer criterios de búsqueda cronológica, geográfica, por autor, por tema y por medio.

Otro ejemplo de este tipo de corpus es el Bank of English (BoE) desarrollado en la Universidad de Birmingham. Este corpus fue inicialmente construido a

principios de la década de los ochentas (Hunston, 2002) y continua expandiéndose desde entonces. El *BoE* contiene más de 500 mil millones de palabras, organizadas por secciones de inglés general (450 mil millones de palabras) y una sección dedicada al lenguaje pedagógico (56 mil millones de palabras).

A pesar del enorme tamaño de los corpus mencionados, en años recientes ha surgido la idea de que todos contamos con un corpus aún más robusto y en constante crecimiento disponible en red: el internet o más claramente *la Web*. Este concepto de la *Web como Corpus* (Kilgarriff y Grefenstette, 2003) tiene como base la representatividad y disponibilidad de la Web en todo momento para llevar acabo estudios de carácter lingüístico, por ejemplo para estudiar los antónimos y sus usos regulares en inglés (Jones et al. 2007). Incluso, algunos investigadores han diseñado interfaces específicas para realizar búsquedas lingüísticas a través de motores de búsqueda comerciales como *Google* (*WebCorp* Renouf, 2003). Incluso, en el estudio que se presenta en este documento la Web o red fue utilizada para corroborar la existencia o no existencia de las no palabras generadas a través de la aplicación de la metodología propuesta (Ver 7.4.4). Sin embargo, utilizar la Web como corpus representa una serie de problemas muy específicos: mientras que los corpus son colecciones de textos desarrolladas muy cuidadosamente, el contenido de la web no está restringido por ningún tipo de lineamientos. Tampoco está dividido por géneros, o por años o por estilos, y el número inmenso de datos que presenta requiere de mucho tiempo y herramientas muy elaboradas de procesamiento. Además, muchos de los textos que actualmente se encuentran en red contienen diversos errores de ortografía, sintaxis, prosodia, etc. (McEnery y Hardie, 2011).

## B) Corpus de la lengua general con fines específicos

Este tipo de corpus pretende dar respuesta a un propósito concreto, como el estudio de determinados aspectos de la gramática o el léxico de la lengua, la extracción de datos estadísticos, el estudio del comportamiento lingüístico de una determinada población de hablantes, análisis comparativos de diversas

variedades lingüísticas, o el desarrollo y evaluación de sistemas de procesamiento del lenguaje. (McEnery y Hardie, 2011).

Entre estos corpus podemos encontrar al CorVerifSDGEE (Corpus de verificación del sistema de diccionarios y gramáticas electrónicos del español) el cual es un corpus textual directamente relacionado con el Sistema de Diccionarios y Gramáticas Electrónicas del Español.

### C) Corpus especializados

Estos corpus contienen textos de un tipo específico de lenguaje, estilo o registro como son periódicos, libros especializados, artículos científicos, etc. El objetivo de estos corpus es el de representar únicamente un tipo particular de texto con el propósito de investigar un tipo particular de lenguaje. En la investigación que se presenta en este documento, dos corpus especializados fueron desarrollados específicamente para los propósitos de este estudio (Ver 7.1). Se utilizaron artículos científicos del área de Nutrición y Odontología para construir dos corpus que nos permitieran seleccionar una serie de términos especializados de ambas disciplinas.

Otros ejemplos de este tipo de corpus son el CANCODE (*Cambridge and Nottingham Corpus of Discourse in English*) el cual incluye 5 millones de palabras de registros informales del inglés británico, o el MICASE (*Michigan Corpus of Academic Spoken English*) el cual consiste en registros de lengua oral del lenguaje académico utilizado en los Estados Unidos de Norteamérica.

### A) Corpus orales

Los corpus orales están constituidos por señales de voz y generalmente incluyen transcripciones de anotación fonética. Un corpus oral contiene grabaciones de llamadas telefónicas, entrevistas o programas de radio. Un ejemplo de este tipo de corpus es el Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC) el cual consiste en una base de datos

textual (corpus de lengua hablada): realizada a partir de la transcripción de textos grabados en cintas de audio del registro oral. 1.100.000 de palabras. Este corpus fue desarrollado por la Universidad Autónoma de Madrid a principios de la década de los 90's.

#### B) Corpus multimodales

Los corpus multimodales están constituidos por otros datos orales como prosodia, gestos, movimientos de la boca, ó grabaciones sonoras y fílmicas (noticias, documentales). Entre los pocos ejemplos disponibles que encontramos de este tipo de corpus están: el COSCODA desarrollado en Japón y el *TalkBank* que constituye el corpus multimodal más importante en la actualidad para la lengua inglesa.

#### C) Corpus sincrónicos vs. diacrónicos

Mientras que para la construcción de un corpus sincrónico se recopila material que se compone de la lengua contemporánea (por ejemplo a partir de 1960 a la fecha), para el corpus diacrónico se recogen textos de varias etapas históricas de la lengua con el propósito de poder observar los cambios en la lengua (por ejemplo, de los años 1200 hasta el 1900). Éstos últimos generalmente son utilizados para dar seguimiento al desarrollo de determinados aspectos de un lenguaje a lo largo del tiempo. Como ejemplo de corpus diacrónico podemos citar al *Helsinki Corpus*, el cual consiste en alrededor de 1700 textos del inglés abarcando un total de 1.5 millones de palabras.

#### D) Corpus monolingües vs. corpus multilingües

Los corpus monolingües contienen, en comparación con los corpus multilingües, textos únicamente en una lengua. Los corpus multilingües (incluidos los corpus bilingües) son muy escasos en comparación con los corpus monolingües porque los textos tienen que existir en versiones traducidas. Los



corpus bilingües son de gran utilidad para hacer comparaciones entre las lenguas, elaborar diccionarios o textos pedagógicos.

#### E) Bases de datos de árboles

Las bases de datos de árboles son textos etiquetados sintácticamente. En general, los análisis sintácticos tienen una estructura en forma de árbol, lo cual explica su nombre. Sin embargo, existen también bases de datos de árbol que tienen una estructura de gráfica con conexiones adicionales entre las palabras, en los que la construcción sintáctica no corresponde a un árbol simple. Estas bases de datos de árboles son un sistema de codificación y anotación sintáctica, y generalmente se utilizan para resolver problemas de ambigüedad y ambivalencia en un término, frase u oración.

Actualmente se encuentran disponibles diversas bases de este tipo como son: NEGRA/TIGER, PDT: *Prague Dependency Treebank*, el *Corpus Le Monde*, el TUT: *Turin University Treebank*, *Spanish Treebank* (UAM), ISST: *Italian Syntactic Semantic Treebank*.

#### G) Corpus paralelos

Este tipo de corpus consiste en dos o más corpus que pertenecen a diferentes idiomas o a diferentes variedades de lenguaje, pero que de alguna forma son equivalentes, por ejemplo una novela en inglés que ha sido traducida al español, o una novela en inglés británico que ha sido traducida al inglés canadiense. Su construcción debe de llevarse a cabo de forma cuidadosa, ya que debe de contener textos pertenecientes a diferentes formatos – periódicos, revistas, entrevistas, etc.- en una misma proporción. Este tipo de corpus generalmente es utilizado por traductores y estudiantes de lengua para detectar y analizar las diferencias y equivalencias entre una lengua y otra. El ICE (*International Corpus of English*) es un buen ejemplo de un corpus paralelo, ya que abarca 4 variedades diferentes del inglés incluyendo un millón de palabras por cada variedad.

## H) Corpus de estudiantes

Este tipo de corpus se caracteriza por ser una colección de textos producidos por los estudiantes de una lengua. Generalmente consiste en ensayos breves y se utiliza para identificar los aspectos del lenguaje en los cuales los aprendices pueden presentar dificultades o diferencias en relación a su lengua madre. Como ejemplo de este tipo de corpus podemos citar el *International Corpus of Learner English* (ICLE), el cual consisten en una colección de diferentes corpus, cada uno constituido por 20 mil palabras. Todos los textos en inglés fueron escritos por estudiantes nativo hablantes de otras lenguas, como por ejemplo el francés, el alemán, el checo, etc.

## I) Corpus pedagógicos

Este corpus está conformado por todo el input al cual un estudiante ha sido expuesto en un periodo determinado de tiempo. Debe incluir todos los libros de texto, reseñas, pláticas, etc. a los que los aprendices tuvieron acceso. Como puede suponerse, este tipo de corpus es muy difícil de compilar, ya que es muy complejo poder recolectar absolutamente todo el material al que el estudiante estuvo expuesto.

### 5.3.2 Consideraciones generales para la elaboración de corpus

De acuerdo con Hunston (2002) existen diversas consideraciones que deben tomarse en cuenta para la elaboración de corpus, ya sean generales o especializados. Estas consideraciones se describen brevemente a continuación:

#### A) Tamaño

Los notables avances producidos por las nuevas tecnologías computacionales permiten contar con corpus que día con día incrementan su tamaño. Aunque contar con más cantidad de datos y ejemplos puede

ser de gran utilidad para el investigador, es necesario considerar que la utilidad de corpus no radica en el número de palabras que contiene, sino en la velocidad, confiabilidad y sencillez de las herramientas desarrolladas para su procesamiento y análisis. En algunos casos es mejor tener un corpus más pequeño que permita obtener rápidamente los términos o colocaciones meta, a contar con millones de ejemplos que dificulten la búsqueda de investigador.

## B) Contenido

Un corpus no es bueno o malo, sino simplemente adecuado o no para los propósitos para los cuales fue construido. Las decisiones sobre que debe o no ser incluido en un corpus deben tomarse considerando no sólo el objetivo para el cual va a ser utilizado sino la información que esté disponible en el momento de elaboración del corpus. Es decir, el contenido del corpus debe de ser diseñado para cumplir con un propósito, por ejemplo, el corpus de Nutrición que fue elaborado para este estudio, cuyo propósito es el de detectar el lenguaje académico especializado de esa disciplina. Pero además debe contemplar los textos que se tienen disponibles, en este caso, artículos científicos arbitrados que estuviesen disponibles en revistas en línea.

## C) Balance y representatividad

Generalmente, los corpus pretenden ser representativos de un tipo particular de lenguaje. Si su objeto de estudio es la prosa académica, o el lenguaje periodístico debe de incluir textos relacionados con estos temas. Por otra parte, un corpus balanceado es aquel que consiste de un número de palabras igual para cada categoría del objeto de estudio, por ejemplo: en el caso del estudio del lenguaje periodístico el corpus elaborado debe contener noticias, editoriales, notas sociales, etc.

En el caso particular de la representatividad de corpus, es necesario considerar que los corpus deben ser actualizados continuamente, en caso contrario no resultan representativos del objeto de estudio, debido a que no contiene ejemplos de textos actuales y contemporáneos.

### 5.3.3 El uso de Corpus en la Lingüística

El análisis de corpus o lingüística de corpus puede aplicarse al estudio de diversos fenómenos lingüísticos en muy diversas áreas. A continuación se mencionan algunos ejemplos de las diversas aplicaciones de la lingüística de corpus (en Huntston, 2002):

La sociolingüística investiga en el corpus cómo se distinguen los dialectos/sociolectos entre sí y sus diferencias con la lengua estándar (en todos los niveles lingüísticos). Estudia, por ejemplo, cómo cambian los dialectos entre diferentes regiones geográficas o si las mujeres pertenecientes a una determinada comunidad tienen un léxico diferente al de los hombres de la misma comunidad.

La lingüística histórica analiza en un corpus los cambios ortográficos, fonéticos, sintácticos o semánticos que ha sufrido una palabra a lo largo de un periodo determinado. También puede realizar estudios de fraseología, colocaciones o modos de uso a lo largo de los años.

Los estudios psicolingüistas pueden analizar si la frecuencia de las palabras incluye en la velocidad de reacción de un hablante y qué papel desempeña la frecuencia de uso de una palabra dentro de los procesos de aprendizaje de una lengua.

Para crear un diccionario, el lexicógrafo analiza en el corpus en qué contexto aparece una palabra determinada, qué lecturas de una palabra existen, qué palabras aparecen con frecuencia en una combinación establecida (colocaciones), qué palabras no tienen uso dentro de un registro específico, qué

palabras se emplean en un lenguaje profesional o qué sustantivos tienen al mismo tiempo más de un género (femenino, masculino, neutro).

En la sintaxis se puede verificar la existencia de una estructura sintáctica específica, analizar el tipo de adjetivos que preceden o siguen a un verbo en particular, o en qué contexto un verbo se utiliza en modo subjuntivo y en cuál está en modo indicativo.

Los estudios de semántica pueden utilizar la lingüística de corpus para analizar los sentidos literales o metafóricos de una palabra específica

La lingüística aplicada a la fonología puede estudiar cómo se pronuncian los extranjerismos, si mediante la prosodia es posible distinguir las lecturas de una palabra, cómo se pueden clasificar los acentos, y por qué los aprendices de una lengua meta pronuncian de una manera determinada las palabras en esa lengua.

Los lingüistas computacionales utilizan los corpus como recurso para la elaboración automática de diccionarios, como recurso para la extracción de frases bilingües, para la extracción automática de colocaciones y fraseología.

Estas son sólo algunas de las aplicaciones que actualmente se le pueden dar a la lingüística de corpus. Existen además ejemplos de investigaciones llevadas a cabo con corpus generales y especializados las cuales tienen objetivos pedagógicos, que pueden involucrar procesos de adquisición de primera o segunda lengua, elaboración de textos de enseñanza de lengua a partir de frecuencias y colocaciones en corpus ó elaboración de pruebas de evaluación de lengua. Los traductores también usan corpus bilingües para comparar los usos de aparentes traducciones equivalentes en dos lenguas, o analizar como una expresión ha sido traducida anteriormente (McEnery y Hardie, 2011).

Para este estudio en particular, se utilizaron algunas técnicas de la lingüística de corpus para la elaboración y procesamiento de dos corpus especializados en lenguaje académico en español, uno para el área de Nutrición y otro para el área de Odontología. Ambos corpus sirvieron para la elaboración de

las pruebas de evaluación de lenguaje académico desarrolladas para sustentar la metodología propuesta por esta investigación.

## 6. ANTECEDENTES

En esta sección se mencionan y se explican brevemente las listas de vocabulario tanto generales como especializadas más comúnmente utilizadas en los estudios de enseñanza y evaluación de vocabulario. Debido a que todas estas listas tienen sus bases teóricas en la aplicación de criterios de frecuencia de ocurrencia, también se mencionan diversos estudios que respaldan la importancia de la frecuencia dentro de los diferentes tipos de procesamiento lingüístico.

### 6.1 Listas de Vocabulario especializado

La práctica de contar las palabras más frecuentes de una lengua tiene una larga historia que se remonta incluso hasta los tiempos de la cultura helénica (DeRocher, 1973). Varios de estos primeros conteos de frecuencia de palabra se mencionan en Fries y Traver (1960). Actualmente y para el caso específico de la lengua inglesa, existen varias listas de frecuencia de palabras que son ampliamente utilizadas por teóricos de adquisición de lengua así como por los diseñadores de textos y programas de estudio. Estas se utilizan tanto para la enseñanza de lengua, como para el diseño de material especializado para la enseñanza. A continuación se mencionan algunas de las listas de palabras más conocidas y utilizadas, citadas en Nation y Waring (1997):

- *The General Service List* o GSL (West, 1953) que contiene un total de 2000 familias de palabras frecuentes (e.g. *know*) y más de 11 mil palabras formas flexivas (e.g. *knows*) y derivadas (e.g. *unkown*) que se extrajeron de un corpus desarrollado durante los años cuarenta del siglo pasado.
- *The Teachers Word Book of 30,000 words* (Thorndike y Lorge, 1944) que contiene un total de 30 mil palabras divididas en 13 mil familias de palabras. Esta lista se desarrolló con base en un corpus de 18 millones de palabras escritas.

- *The American Heritage Word Frequency Book* (Carroll, et al., 1971) la cual se basa en los textos escritos utilizados en las escuelas de educación elemental y media de los Estados Unidos de América.
- *The Brown* (Francis y Kucera, 1982) basada en un corpus de un millón de palabras del inglés.

Además de estas listas de vocabulario general, se han llevado a cabo diversos estudios relacionados con la elaboración de listas de vocabulario académico a través del procesamiento de corpus especializados. Estos corpus especializados se desarrollan esto con el objetivo primordial de identificar la terminología más frecuente dentro de los textos científicos y facilitar así la adquisición de dicho vocabulario por parte de estudiantes e investigadores tanto en primeras como en segundas lenguas.

Muchos de estos estudios se enfocaron en el vocabulario académico utilizado de las disciplinas científicas en general. Por ejemplo, en 1971 Champion y Elley analizaron una serie de 301 800 palabras provenientes de libros de texto y artículos científicos provenientes de 19 disciplinas académicas diferentes. El resultado de este estudio fue una lista que contenía las 500 palabras más frecuentes del lenguaje académico en general y una lista de las 3 200 palabras que los estudiantes de diferentes áreas de estudio seguramente encontrarían en sus libros de texto.

En 1972 Praninskas elaboró una compilación llamada *The American University Word List*, la cual estaba basada en un corpus de 271 466 palabras provenientes de 10 libros de texto de 10 disciplinas académicas diferentes. Estudios posteriores (Lynn, 1973; Ghadessy, 1979) elaboraron sus listas de vocabulario académico a partir de las anotaciones que los estudiantes extranjeros realizando estudios de licenciatura escribían en sus cuadernos como palabras académicas difíciles de aprender.



Algunos años más tarde, Xue y Nation (1994) propusieron una lista que consistía en combinar las cuatro listas mencionadas anteriormente, generando así la *University Word List* (UWL), la cual incluía además 800 palabras que no estaban consideradas en la GSL (West, 1953) pero que eran de alta frecuencia dentro de los textos académicos universitarios. El propósito de esta lista era el de proporcionar a los estudiantes extranjeros universitarios una herramienta que facilitara sus estudios en los Estados Unidos de América.

En años más recientes Coxhead (2000) desarrolló la *Academic Word List* (AWL), utilizando un corpus de 3 millones y medio de palabras y diversas herramientas de procesamiento de corpus que le permitieron analizar tanto la frecuencia de ocurrencia como el rango de ocurrencia en diferentes disciplinas académicas. Los textos utilizados para la elaboración de este corpus fueron seleccionados de diferentes publicaciones científicas así como libros de texto provenientes de tres áreas de estudio universitarias: arte, comercio, leyes y ciencias naturales. Así pues, la AWL contiene un total de 570 familias de palabras seleccionadas a través de estrictos criterios de frecuencia y de rango y actualmente es una de las listas más citadas por los estudios de *ESL*.

A pesar de su importante rango de cobertura del vocabulario académico y la estricta metodología para su elaboración, en los últimos años la AWL ha sido fuertemente criticada por los diseñadores de textos y cursos para la enseñanza de *ESL* (Inglés como Segunda Lengua) y *ESP* (Inglés para Propósitos Específicos) debido a que diversos estudios han cuestionado y demostrado que existen muchas diferencias léxicas a lo largo de distintas disciplinas académicas, y que en algunos casos estas diferencias pueden ser más que las similitudes (Hyland, 2002) y que es necesario desarrollar el vocabulario de los estudiantes de acuerdo a las necesidades de su disciplina específica, ya que cada área de estudio tiene sus propias formas de argumentación (Bloor, 1998; Holmes 1997; Hyland, 1999, Samraj, 2002).

Incluso algunos autores sostienen que el utilizar la AWL para la enseñanza del vocabulario académico más que beneficiar a los estudiantes los perjudica, ya que

no sólo los expone a un mayor número de palabras del que realmente necesitan, sino que los distrae del vocabulario académico que realmente van a utilizar dentro de su área de estudio (Chen y Ge, 2007; Hyland y Tse, 2007; Paqout, 2007).

Es por esto que además de las listas interdisciplinarias, van surgiendo también en los últimos años estudios dedicados a la identificación de la terminología por áreas de estudio, considerando que cada disciplina cuenta con vocabulario especializado que difiere del vocabulario académico general. A continuación se describen brevemente algunos ejemplos representativos de estas listas:

En el 2001, Lam llevó a cabo un estudio empírico sobre el estudio del vocabulario académico relacionado con las Ciencias Computacionales con el objetivo de detectar los problemas de vocabulario que los estudiantes de esta área encontraban al realizar diversas lecturas de textos académicos. Sus conclusiones señalan que algunos términos académicos varían semánticamente si son considerados dentro del lenguaje académico general o si se encuentran inmersos en un contexto especializado en las Ciencias Computacionales, e.g: el término *switch*. En sus conclusiones la autora sugiere que este tipo de términos sean incluidos en un glosario académico especializado, que incluya información relacionada con la frecuencia de ocurrencia en corpus específicos.

Para el año 2006, Mudraya publica el *Student Engineering English Corpus* (SEEC), el cual incluye cerca de 2 millones de palabras seleccionadas de libros de texto de ingeniería de 13 disciplinas diferentes de especialización, e.g: mecánica de suelos, ingeniería civil, automatización, etc. El análisis de este corpus produjo una lista de 1 200 familias de palabras en inglés especializadas en el área de ingeniería. En sus conclusiones la autora resalta la importancia de que los profesores de *ESL* utilicen la lingüística de corpus con el propósito de proveer a sus alumnos con el lenguaje especializado que requieren para desarrollarse académica y profesionalmente.

En relación a la Medicina, Wang et al. (2008) elaboraron la *Medical Academic Word List* (MAWL), la cual es una lista de las palabras más frecuentes

encontradas en publicaciones científicas especializadas en el área médica. Esta lista se elaboró a partir de un corpus de 1 093 011 palabras compiladas de artículos científicos en línea. La MAWL contiene un total de 623 familias de palabras. Los autores sostienen que la alta frecuencia de las palabras seleccionadas así como el rango de representatividad de las mismas debe servir como una guía para los diseñadores de programas académicos de ESL (*English as a Second Language*) y ESP (*English for Specific Purposes*) así como libros de texto dedicados a la enseñanza del vocabulario académico del área de medicina.

Finalmente para el área de Agricultura, Martínez et al. (2009), desarrollaron una lista dedicada a cubrir las necesidades de los hablantes no nativos del inglés que por necesidades académicas y profesionales debían leer y escribir en esta lengua. Su metodología consistió en utilizar la AWL de Coxhead (2000) y compararla con un corpus especializado en Agricultura conformado a partir de diversas publicaciones científicas de esta disciplina. El análisis cuantitativo realizado por los autores arrojó una lista altamente restringida de solamente 92 familias de palabras requeridas para el área de Agricultura, comparadas con las 570 familias de palabras propuestas por la AWL; mientras que el análisis cualitativo reveló que no solamente algunas palabras académicas tienen significados y comportamientos específicos del área de Agricultura, sino que algunas palabras provenientes de la AWL tienen un significado diferente en el corpus de Agricultura. De acuerdo con los autores, estos resultados apoyan la necesidad de elaborar listas académicas especializadas por disciplina, las cuales deben además incorporar términos, frases y colocaciones académicas necesarias para la construcción del conocimiento científico por áreas de investigación específicas.

Sin embargo, todas las listas mencionadas han sido principalmente desarrolladas y aplicadas en la enseñanza de vocabulario académico en inglés o para la elaboración de programas de enseñanza de ESL o ESP. El objetivo de esta investigación es el de utilizar la metodología para la construcción de listas de vocabulario especializado y aplicarla a la elaboración de un instrumento de decisión léxica que permita evaluar el nivel de reconocimiento de vocabulario

académico especializado de los estudiantes nativo hablantes del español que realizan estudios universitarios, ya sea en Nutrición o en Odontología. Se espera que los resultados obtenidos no sólo sean aplicables a la evaluación de los estudiantes de dichas disciplinas, sino también a los distintos programas de estudio de las instituciones educativas que participaron en esta investigación.

## 6.2 Frecuencia

En general, la frecuencia juega un rol importante en la adquisición tanto de primera como de segunda lengua. La hipótesis más común sostiene que entre más a menudo una entrada esté disponible para el hablante, ésta será adquirida más fácil y rápidamente (Dell, 1988; Kwon, 2005; McClelland y Franz, 1986).

Cuando se realizaron los primeros estudios relacionados con la influencia de la frecuencia sobre la adquisición de una lengua, la interpretación convencional fue asociada fuertemente con el conductismo. Según la perspectiva conductista, entre más exposición se tiene a una forma, es más probablemente imitar su uso correctamente. Más tarde los nativistas describieron el papel de la frecuencia como el detonador del dispositivo natural de adquisición de la lengua (Chomsky, 1981).

En 1973 Brown investigó los posibles determinantes que afectan la adquisición de primera lengua. Entre los factores mencionados por el autor se encuentran la frecuencia de *input* en el discurso de los padres, la complejidad semántica y la complejidad gramatical de las formas meta. El autor concluyó que no existía evidencia suficiente para afirmar que la frecuencia en el discurso de los padres influenciaba la adquisición de las formas. Por el contrario, en el caso de la adquisición de segundas lenguas, Larsen-Freeman (1976) encontró correlación entre el orden de adquisición de ciertos morfemas en inglés y la frecuencia de uso de esos morfemas en las muestras de discurso espontáneo de sus estudiantes. Su estudio concluyó que la frecuencia del *input* en el discurso del nativo hablante es el factor principal para la adquisición de morfemas en el discurso de los estudiantes de segunda lengua.

Los resultados de la investigación de Larsen-Freeman (1976) y otros estudios similares llevados a cabo en años posteriores (Dulay y Burt, 1978; Kessler y Idar, 1979) provocaron el interés de nuevos investigadores por analizar la influencia de la frecuencia en diversos aspectos específicos de la lengua. En el caso de la fonética y fonología, se ha encontrado que tanto las palabras de alta como las de baja frecuencia facilitan el aumento de la producción fonológica de los niños que adquieren su primera lengua, sin embargo son las palabras de alta frecuencia las que proveen los patrones fonotácticos generales de la lengua (Gierut et al., 1999).

En 1990 Marslen-Wilson publicó su modelo del reconocimiento del discurso oral. El modelo sostiene que el cerebro humano utiliza el fonema inicial de una palabra para activar toda una gama de palabras que inician con el mismo fonema. Al obtener más información, esta gama de palabras se reduce hasta el punto de identificar la palabra utilizada por el hablante antes de que este termine de emitirla. Marslen-Wilson sostiene que esta activación es mucho más rápida y productiva en el caso de los fonemas que pertenecen a palabras de alta frecuencia. En el año 2001, Frisch et al., llevaron a cabo un estudio en el cual solicitaron a un grupo de nativos hablantes del inglés que analizaran una serie de no palabras de la misma lengua y que las calificaran en una escala del 1 al 7 de acuerdo a la posibilidad de cada una de ellas de ser palabras reales. Los autores concluyeron que las no palabras que obtuvieron mayor puntaje en cuanto a esta posibilidad, fueron aquellas que compartían patrones fonéticos similares a los de palabras reales del inglés de más alta frecuencia.

En cuanto al léxico, se ha encontrado evidencia suficiente para afirmar que tanto el reconocimiento como la producción de las palabras guardan una proporción directa con la frecuencia de ocurrencia de las mismas. Existe un efecto de frecuencia de palabra en la velocidad y precisión de los procesos de reconocimiento y producción léxica, y estos efectos son comprobables en niños y adultos tanto en primeras como en segundas lenguas (Ellis, 2002; Kirsner, 1994).

En el caso de la lectura, las palabras de más alta frecuencia son leídas con mayor rapidez que las palabras de baja frecuencia (Balota y Chumbley, 1984; Forster y Chambers, 1973). De igual manera, las palabras de alta frecuencia son deletreadas y escritas con mayor precisión que las de baja frecuencia (Barry y Seymour, 1988).

En lo que respecta a la relación de la frecuencia con la adquisición de la morfología se ha comprobado que los nativos hablantes de lengua inglesa que son expuestos a una serie de verbos irregulares en inglés en su forma infinita, tardan significativamente menos tiempo en producir la forma del pasado de los verbos cuando el pretérito es una forma de alta frecuencia, independientemente de que las raíces sean de alta o baja frecuencia (Prasada et al., 1990; Seidenberg y Bruck, 1990). En experimentos llevados a cabo con lenguas artificiales, se ha comprobado que existe gran influencia de la frecuencia de *input* en la producción de formas del plural de pseudo sustantivos y de formas del pasado de pseudo verbos independientemente de que estos sean regulares o irregulares (Ellis y Schmidt, 1997). Incluso se ha comprobado la influencia de la frecuencia en la producción de formas morfológicas regulares en niños con trastorno de lenguaje (Oetting y Rice, 1993).

También se ha encontrado influencia de la frecuencia en relación al uso de frases idiomáticas. De la misma forma en que los hablantes aprenden secuencias comunes de componentes subléxicos, también aprenden secuencias comunes de palabras o fórmulas que resultan de colocaciones léxicas frecuentes (Ellis, 2002; Sinclair 1991). Las expresiones idiomáticas de más alta frecuencia dentro de la lengua, son más fáciles de comprender debido a que sus significados figurativos se han vuelto familiares y con el uso y el tiempo, se han vuelto más salientes que los significados literales de las mismas (Giora, 1997; Nippold y Rudzinski, 1993). La importancia del uso de frases idiomáticas de alta frecuencia en la adquisición de lengua ha sido ampliamente investigada en estudios relacionados con rutinas y patrones establecidos en la lengua (Hakuta, 1974), discurso idiomático (Wong-Fillmore, 1976), oraciones memorizadas y raíces lexicalizadas (Pawley y Syder, 1983), fórmulas (Ellis, 1994), secuencias en adquisición en segunda lengua (Ellis,

2001), manejo del discurso (Dörnyei y Kormos, 1998; Tannen, 1987), registro (Biber y Finegan, 1994), estilo (Brewster, 1999) y patrones léxicos y colocaciones (Carter, 1998; Lewis, 1993; Schmitt, 2000).

En el área de comprensión de lenguaje, diversos estudios psicolingüísticos han demostrado que los nativos hablantes de una lengua resuelven las ambigüedades del lenguaje utilizando un proceso de orden lógico de significado basado en el uso más frecuente del significado de verbos, preposiciones y sustantivos ambiguos (MacDonald, 1994; MacDonald et al., 1994; Trueswell, 1996).

Los resultados de estos estudios, que demuestran la influencia de la frecuencia en los diferentes tipos de procesamiento del lenguaje, han tenido implicaciones profundas en las teorías de adquisición y aprendizaje de lenguas, por lo que hoy en día se ha fortalecido una corriente de enseñanza de vocabulario con tendencia a enseñar y evaluar las palabras de más alta frecuencia de la lengua (Laufer y Nation, 1999; Nation, 1993). En esta investigación en particular, los criterios de frecuencia de ocurrencia dentro de los dos corpus especializados fueron aplicados no sólo para la selección de ítems meta (Ver 7.4.3), sino para la elaboración de las palabras aplicadas a los participantes identificando los prefijos y bases más frecuentes de cada disciplina (Ver 7.4.4).

### 6.3 Terminología

La terminología se define como el proceso de compilar, describir, procesar y presentar los términos relacionados con un campo específico en una o varias lenguas. (Cabré, 1998). Su objetivo es el de optimizar la comunicación entre los especialistas y profesionales de una disciplina al proveer herramientas para estandarizar el vocabulario que requieren en su campo.

La terminología como la conocemos hoy en día tiene sus orígenes en el desarrollo de las ciencias naturales de los siglos XVIII y XIX, cuando químicos, botánicos y zoólogos manifestaron la necesidad de nombrar los nuevos conceptos

científicos desarrollados a partir de sus crecientes investigaciones. Ya para el siglo XX nuevas ciencias como la ingeniería y la computación requirieron no sólo de reglas para la generación de nuevos términos, sino de establecer acuerdos sobre los términos para referirse a un concepto específico dentro de una comunidad científica internacional.

A partir de entonces los rápidos cambios tecnológicos en los campos de la información y la comunicación crearon la necesidad de desarrollar nuevas formas de comunicación que previamente no existían, junto con los conceptos y términos que constantemente tienen que ser actualizados. Las ciencias computacionales son uno de los grandes impulsores de la terminología en la actualidad, al proveer las herramientas y recursos para desarrollarse rápida y efectivamente.

Al tener sus orígenes en múltiples disciplinas y ciencias, existe divergencia en relación a si la terminología constituye una disciplina en sí. Mientras que para algunos teóricos la terminología es una práctica relacionada con el estudio de las ciencias establecidas, para otros constituye una disciplina científica independiente que, aunque tiene fundamentos básicos en el estudio de otros campos de estudio, ha logrado reformular y sintetizar estos fundamentos para lograr construir así su propio campo de estudio (Cabré, 1998).

Actualmente, la terminología es concebida como tres conceptos diferentes:

- A) Los principios y bases conceptuales que rigen el estudio de los términos
- B) Las bases teóricas para el desarrollo de la lexicografía
- C) Todos los términos que se utilizan en una disciplina de estudio en particular

Así pues, aunque aún existe divergencia entre que es o no una disciplina en sí, la terminología puede ser considerada un campo con sus propios principios teóricos, y sus propios objetivos y aplicaciones. Si bien los conceptos que constituyen sus bases teóricas no son originales, puede concebirse como una disciplina que agrupa diferentes conceptos de disciplinas afines como son la lingüística, la lógica, la ontología y las ciencias de la información.



### 6.3.1 La Terminología y la lexicología

A pesar de que hay muchas similitudes entre la terminología y la lexicología (ambas estudian las palabras, tienen principios tanto teóricos como aplicados, y se encuentran altamente relacionadas con la elaboración de diccionarios) estas disciplinas cuentan con importantes diferencias. Para algunos teóricos la terminología forma parte de la lexicología, mientras que para otros constituyen dos aproximaciones distintas. Entre los aspectos que claramente diferencian a ambas se mencionan (Cabré, 1992):

#### A) El campo de estudio:

Mientras que la lexicología se dedica al análisis y descripción de la competencia léxica de los hablantes (su repertorio de palabras, las reglas de formación de las mismas, y el conocimiento lingüístico y enciclopédico relacionada con cada una de ellas) la terminología se enfoca en el estudio de las palabras que pertenecen a un campo de estudio en particular (como la química, la física o la ingeniería).

#### B) La unidad básica de estudio:

La lexicología estudia las palabras, mientras que la terminología estudia los términos. Una palabra puede ser definida como una unidad descrita por una serie de características lingüísticas y que tiene la propiedad de referirse a un elemento de la realidad. Mientras que un término es una unidad con características lingüísticas pero que se refiere a un uso específico en un campo específico. Desde el punto de vista lingüístico, no parece haber mucha diferencia entre palabra y término, ya que ambos son unidades de significado que poseen una estructura fonética, gráfica, morfológica y semántica. Sin embargo existen algunas diferencias, por ejemplo: mientras que las palabras (objeto de estudio de la lexicología) pueden poseer cualquier categoría gramatical: sustantivos, artículos, verbos, etc., los términos regularmente sólo se refieren a sustantivos, y en

algunos casos a verbos. Las palabras poseen un uso pragmático y diferencian a un hablante de otro en relación al uso comunicativo que hacen de ellas, mientras que los términos adquieren significado sólo en un contexto específico. Para los propósitos prácticos de esta investigación, se utilizan ambos conceptos, palabra y término para referirse a las unidades léxicas utilizadas en el desarrollo y aplicación de la metodología propuesta.

### C) Los objetivos

La lexicología estudia las palabras con el propósito de evaluar la competencia de un hablante, la terminología estudia los términos para poder establecer un punto de referencia sobre los conceptos de mundo real, identificando sólo segmentos de un campo de especialización.

### D) La metodología aplicada

La lexicología utiliza conceptos teóricos para confirmar o rechazar hipótesis a partir de muestras del discurso de los hablantes, mientras que la terminología no explica el comportamiento humano, solamente establece las bases para los términos que construyen una disciplina de estudio.

Ahora bien, se debe considerar que los términos constituyen un subcomponente del léxico de una lengua, y que la competencia lingüística de un hablante no puede excluir el lenguaje especializado. Desde este punto de vista, la terminología forma parte de la lingüística aplicada, ya que toma elementos de la morfología, la lexicología y la semántica y los aplica a un número limitado de conceptos.

#### 6.3.2 La terminología y la lingüística

Si bien la terminología cuenta con sus propias bases teóricas y metodológicas, es bien sabido que muchas de ellas provienen de la lingüística. Por ejemplo, la aplicación del concepto de frecuencia (Ver 6.2) para la selección

de los términos que deben ser incluidos en un diccionario especializado tiene sus orígenes en los estudios realizados por la lingüística descriptiva. Si este mismo diccionario incluyese la pronunciación de los términos debe hacerse uso de la fonética. Si el término a definir es polisémico, ya que cuenta con diferentes significados en diferentes disciplinas (Ver 5.1.1) la terminología puede hacer uso de la lingüística histórica para determinar el significado central del término, sus orígenes (los cuales incluyen también aplicaciones etimológicas) y el desarrollo de sus diferentes significados.

Sin embargo, también la terminología puede aplicarse a la lingüística. La terminología ha precedido diversos estudios de lingüística aplicada. El desarrollo de herramientas de procesamiento originalmente creadas para ser utilizadas por lexicógrafos y terminólogos actualmente impulsa estudios de lingüística de corpus que permiten a los lingüistas comprobar sus hipótesis sobre el comportamiento de una o más lenguas, o facilitan los procesos para estandarizar el vocabulario y convenciones técnicas utilizadas en diferentes disciplinas (Ver 5.3.3).

Tradicionalmente, las aplicaciones de la terminología han consistido en el desarrollo de diccionarios especializados, glosarios, estudios de lingüística histórica, traducciones, estudios sociolingüísticos e incluso diseño de planes de estudio de enseñanza de segundas lenguas (ESP) y poco se han explorado sus aplicaciones en otras áreas como son la adquisición de primera lengua o la evaluación de léxico. La metodología que se propone a continuación se enmarca en esta tendencia multidisciplinar y es una invitación a abrir nuevas puertas para la terminología. El experimento desarrollado (en el sentido de trabajo empírico y también en el sentido de trabajo innovador) combina procedimientos propios de la terminología con técnicas de la lingüística de corpus y hallazgos en la medición de lenguas, todo ello con vistas a aplicarse en la evaluación del conocimiento en la educación superior.

## 7. METODOLOGÍA

A continuación se detalla el procedimiento para la construcción y procesamiento de los dos corpus utilizados en el estudio, se describe la metodología para el diseño de las pruebas aplicadas y se describen los grupos y participantes tanto piloto como experimentales.

### 7.1 Criterios generales para la construcción de los corpus

Para cubrir los objetivos planteados por este estudio fue necesario compilar y procesar dos corpus académicos especializados, uno dedicado a la Nutrición y el otro a la Odontología. En ambos casos, los textos incluidos fueron seleccionados aplicando los criterios propuestos por Sinclair (2005) considerando su representatividad del área de especialización y su disponibilidad en formato electrónico. En la siguiente tabla se especifican los criterios seguidos para la construcción de ambos corpus:

<b>Criterios</b>	<b>Selección</b>
<b>Dominio</b>	Académico
<b>Modo</b>	Escrito
<b>Tipo de Texto</b>	Artículo Científico
<b>Idioma</b>	Español
<b>Origen</b>	Iberoamérica
<b>Fecha de los Textos</b>	De 1999 a 2011

**Tabla 4 Criterios para la construcción de los Corpus, basados en Sinclair (2005)**

Como puede apreciarse en la Tabla 4, los corpus pertenecen al dominio o tipo académico, los textos incluidos se encontraban en formato escrito, el tipo de texto

al que pertenecen es el de artículo científico, en idioma español de Iberoamérica, pertenecientes al periodo comprendido entre 1999 y el año 2011.

Todos los artículos científicos que se incluyeron contaban con la estructura IMRD (Swales, 1990): *Introducción – Metodología – Resultados – Discusión*. Solamente fueron incluidos aquellos textos escritos por nativo hablantes del español, siguiendo los criterios propuestos por Wood (2001) según los cuales el primer autor debe tener nombres característicos de la comunidad hispano hablante y estar afiliado a una institución académica cuya lengua oficial sea el español. Los resúmenes, tablas, diagramas, referencias bibliográficas, agradecimientos y apéndices fueron excluidos del corpus.

## 7.2 Construcción del corpus de Nutrición

Siguiendo los criterios mencionados en la sección anterior, se seleccionaron dos publicaciones científicas dedicadas a la Nutrición y disponibles electrónicamente a través del Sistema de Bibliotecas de la Universidad Autónoma de Querétaro:

A) *Anales Venezolanos de Nutrición*: con fecha del 2001 al 2009 del cual se incluyeron los artículos que cumplían con la estructura IMRD incluidos en los 17 números disponibles en línea, y

B) *Archivos Latinoamericanos de Nutrición*, con fecha del 2000 al 2009 y de la cual se incluyeron los artículos IMRD de los 37 números disponibles en línea.

Una vez integrados los textos, éstos fueron convertidos a formato de Texto (.txt)- ya que originalmente se encontraban en formato para publicación (.pdf) - y procesados manualmente para la eliminación de resúmenes, tablas, diagramas, referencias bibliográficas, agradecimientos y apéndices. Finalmente el corpus para el área de Nutrición quedó constituido por un total de 1, 026, 107 tokens y 31, 638 tipos.

### 7.3 Construcción del corpus de Odontología

Siguiendo los mismos criterios, se seleccionaron 5 publicaciones científicas dedicadas a la Odontología y disponibles electrónicamente a través del Sistema de Bibliotecas de la Universidad Autónoma de Querétaro:

- A) Acta Odontológica Venezolana, con fecha del 2007 al 2009 del cual se incluyeron los artículos que cumplieran con la estructura IMRD incluidos en los 3 números disponibles en línea
- B) Avances en Estomatología, con fecha del 2003 al 2010 del cual se incluyeron los artículos que cumplieran con la estructura IMRD incluidos en los 8 números disponibles en línea
- C) Medicina Oral, Patología Oral y Cirugía Bucal, con fecha del 2004 al 2005 del cual se incluyeron los artículos que cumplieran con la estructura IMRD incluidos en los 4 números disponibles en línea
- D) Revista Cubana de Estomatología, con fecha del 2006 al 2010 del cual se incluyeron los artículos que cumplieran con la estructura IMRD incluidos en los 4 números disponibles en línea, y
- E) Revista Odontológica Mexicana, con fecha del 2009 al 2011 del cual se incluyeron los artículos que cumplieran con la estructura IMRD incluidos en los 3 números disponibles en línea

Una vez integrados los textos, también fueron convertidos a formato de Texto (.txt) y procesados para la eliminación de tablas, agradecimientos, índices y figuras, tras lo cual el corpus para el área de Nutrición quedó constituido por un total de 2, 513, 647 tokens y 104, 834 tipos.

### 7.4 Procesamiento de los corpus

#### 7.4.1 Etiquetado

El etiquetado de ambos corpus se llevó a cabo con la herramienta *TreeTagger* (Schmid, 1994). Este programa es una herramienta para el etiquetado de textos el cual proporciona una etiqueta de categoría sintáctica (*part of the speech*) y un lema por palabra. *TreeTagger* fue desarrollado en el *Institute for Computational Linguistics of the University of Stuttgart* y ha sido aplicado exitosamente en diversas lenguas como son el alemán, inglés, francés, italiano, español, búlgaro, ruso, griego, portugués, chino, swahili y latín. Al ingresar los textos que se desean etiquetar en formato de texto (.txt), se obtiene un sistema de etiquetado que muestra el término, la etiqueta POS y el lema correspondiente (e.g. *representa VLfin representar*).

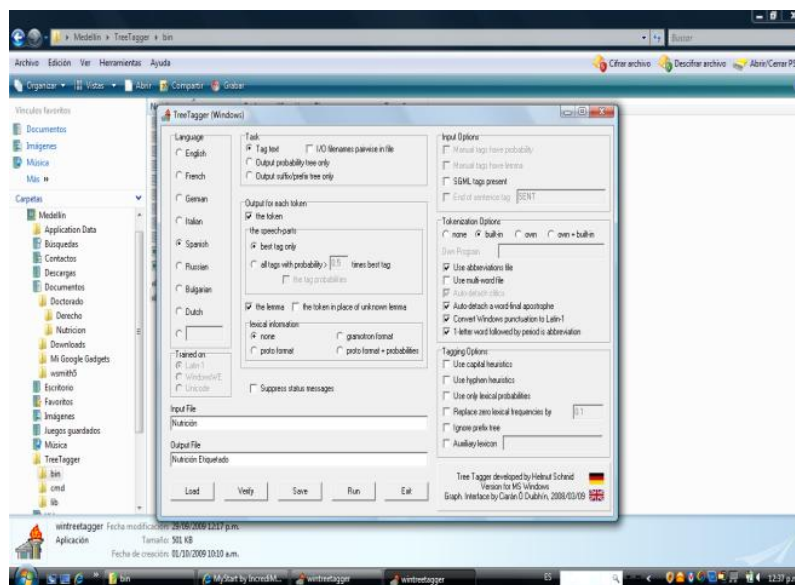


Figura 1. Etiquetado del corpus. Pantalla de ingreso de datos del Programa TreeTagger

Word	Grammatical Category	Lemmatized Form
La	ART	e
obesidad	NC	obesidad
en	PREP	en
todos	QU	todo
sus	PPO	suyo
grados	NMEA	grado
,	CM	,
desde	PREP	desde
leve	ADJ	leve
a	PREP	a
mórbida	NC	<unknown>
,	CM	,
representa	VLFin	representar
un	ART	un
problema	NC	problema
de	PREP	de
creciente	ADJ	creciente
magnitud	NC	magnitud
y	CC	y
sus	PPO	suyo
tasas	NC	tasa
de	PREP	de
prevalencia	NC	<unknown>
han	VHFIn	haber
aumentado	VLadj	aumentar

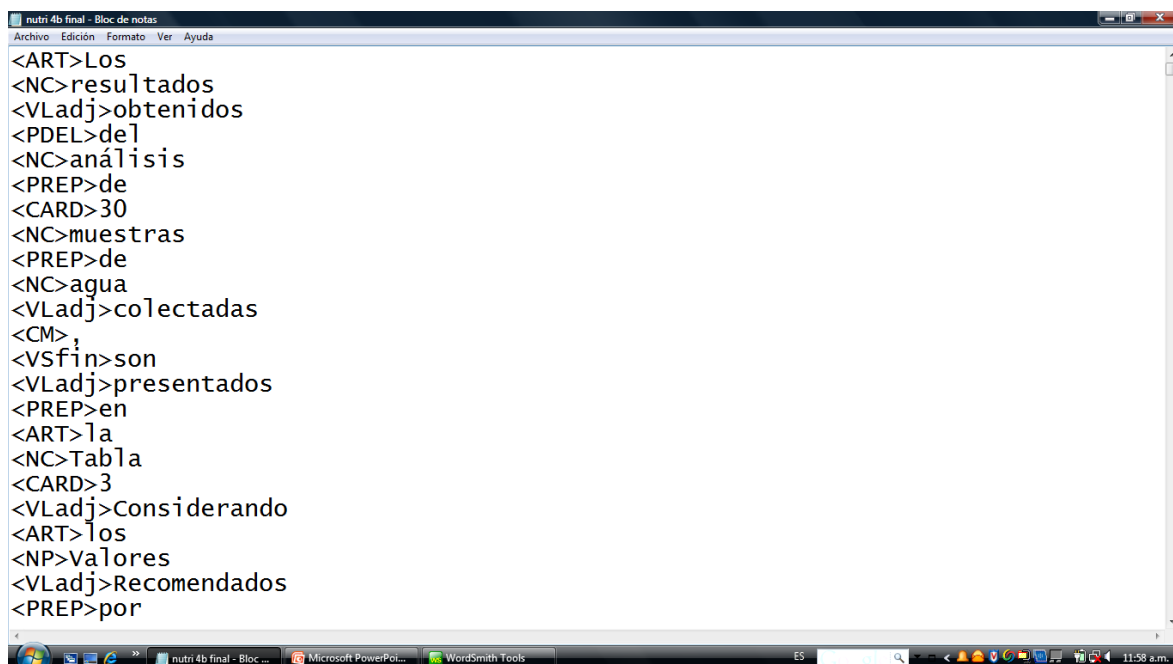
**Figura 2 Etiquetado del corpus. Pantalla de resultado de etiquetado del Programa TreeTagger**

La Figura 1 muestra la pantalla de la herramienta *TreeTagger* en la que se introdujo el corpus completo y se solicitó el etiquetado en español para la totalidad del texto. Las categorías gramaticales están previamente establecidas por el programa y no pueden ser modificadas por el usuario. En la Figura 2 puede observarse el resultado del proceso de etiquetado, el cual arroja una lista de términos aislados de acuerdo a su orden de aparición en el texto. El resultado se presenta en tres columnas: la primera dedicada al término, la segunda que indica la categoría gramatical del mismo, y la tercera correspondiente a la lematización. Lematizar es concentrar en un único lema las formas de una palabra variable, por ejemplo la forma sustantiva plural “muestras” puede lematizarse en la forma sustantiva singular “muestra” por razones de economía para el manejo de un corpus. Este procedimiento puede realizarse de forma automática, con herramientas como el *TreeTagger* y muchas más disponibles actualmente. Sin embargo, es necesario verificar cuidadosamente los procesos de lematización automática, ya que generalmente la polisemia impide que sean muy precisos. Por ejemplo, la herramienta de lematización es incapaz de distinguir si la palabra “muestras” mencionada en el ejemplo anterior es en realidad un sustantivo plural o pertenece a la conjugación de la segunda persona singular del verbo “mostrar”.



Para el caso específico de esta investigación, el proceso de lematizado primero se realizó automáticamente, pero después tuvo que revisarse y realizarse de forma manual, ya que los programas como el *TreeTagger* incluyen un rango muy limitado de lemas, es decir, incluyen sólo vocabulario muy general de una lengua como “grupo” o “alimentar”, y los términos especializados como “corticoidal” o “adenoides” no están considerados.

Tras este primer paso de etiquetado, ambos corpus fueron nuevamente procesados por medio de *macros* en Word y eliminación de columnas en Excel para obtener el formato de etiquetado aceptado por la herramienta *WordSmith Tools*, Versión 5.0 (Scott, 2009) (Ver 7.4.2), eliminando la columna de lema debido a que no era necesaria para la realización de este análisis.



```
nutri 4b final - Bloc de notas
Archivo Edición Formato Ver Ayuda
<ART>Los
<NC>resultados
<VLadj>obtenidos
<PDEL>del
<NC>análisis
<PREP>de
<CARD>30
<NC>muestras
<PREP>de
<NC>agua
<VLadj>colectadas
<CM>
<VSfin>son
<VLadj>presentados
<PREP>en
<ART>la
<NC>Tabla
<CARD>3
<VLadj>Considerando
<ART>los
<NP>Valores
<VLadj>Recomendados
<PREP>por
```

**Figura 3 Etiquetas y términos procesados para el análisis del Corpus**

La Figura 2 muestra el etiquetado del corpus en el formato aceptado por *WordSmith Tools*, el cual debe incluir la categoría gramatical entre picoparéntesis (< >) y después el término sin espacios ni marcas de puntuación.

#### 7.4.2 Herramientas para el análisis de los corpus

Como se mencionó anteriormente, el programa utilizado para el análisis de los corpus fue *WordSmith Tools, Versión 5.0*, el cual consiste en una serie de tres herramientas que permiten estudiar el comportamiento de las palabras dentro de un texto:

- A) *WordList*, que permite elaborar listas e índices de todas las palabras contenidas en el corpus y verlas de forma alfabética o por orden de frecuencia,
- B) *Concordance*, el cual facilita el análisis de palabras determinadas en su contexto y extrae listas de ejemplos de uso de palabras o expresiones dentro del corpus, y
- C) *Keywords*, herramienta que determina las palabras clave dentro de un corpus, comparándolo con un corpus de referencia, es decir detecta las palabras que ocurren con más frecuencia de lo que se esperaría dentro de un corpus especializado al compararlo con un corpus general o corpus de referencia (Ver 5.3.1). El grado que otorga a esta alta frecuencia se denomina *Keyness*.

The screenshot shows the Wordlist application window. The main window displays a table with the following columns: N, Word, Freq., %, Texts, %, Lemmas, and Set. The data is as follows:

N	Word	Freq.	%	Texts	%	Lemmas	Set
28	CONSUMO	2,993	0.29	1	100.00		
29	NIÑOS	2,784	0.27	1	100.00		
30	FUE	2,740	0.26	1	100.00		
31	ALIMENTOS	2,663	0.26	1	100.00		
32	MAYOR	2,568	0.25	1	100.00		
33	ESTUDIO	2,513	0.24	1	100.00		
34	SON	2,167	0.21	1	100.00		
35	EDAD	2,149	0.21	1	100.00		
36	AÑOS	2,148	0.21	1	100.00		
37	PESO	2,145	0.21	1	100.00		
38	ESTA	2,074	0.20	1	100.00		
39	GRUPO	2,019	0.19	1	100.00		
40	HA	1,934	0.19	1	100.00		
41	SIN	1,809	0.17	1	100.00		
42	NUTRICIONAL	1,766	0.17	1	100.00		
43	SOBRE	1,744	0.17	1	100.00		

Figura 4 Herramientas utilizadas en el procesamiento de los corpus, Wordlist

The screenshot shows the Concordance application window. The main window displays a table with the following columns: N, Cluster, Freq., Length, and Related. The data is as follows:

N	Cluster	Freq.	Length	Related
1	NIVELES DE COLESTEROL	51	3	
2	LOS NIVELES DE	42	3	
3	DE COLESTEROL TOTAL	40	3	
4	COLESTEROL TOTAL Y	36	3	
5	DE COLESTEROL Y	30	3	
6	DEL COLESTEROL TOTAL	22	3	
7	EN EL COLESTEROL	20	3	
8	DE COLESTEROL EN	18	3	
9	EL COLESTEROL TOTAL	18	3	
10	DE COLESTEROL PLASMÁTICO	18	3	
11	COLESTEROL DE LA	17	3	
12	CONCENTRACIÓN DE COLESTEROL	17	3	
13	LA CONCENTRACIÓN DE	15	3	
14	EN LOS NIVELES	14	3	
15	QUE EL COLESTEROL	14	3	
16	DEL COLESTEROL LDL	13	3	

Figura 5 Herramientas utilizadas en el procesamiento de los corpus, Concordance

La Figura 4 muestra un ejemplo de una lista producida por la herramienta WordList utilizando el corpus de Nutrición específicamente elaborado para este

estudio. La lista mostrada a la izquierda de la figura arroja los resultados con base en la frecuencia de ocurrencia de las palabras (Freq.) dentro del corpus. La Figura 5 muestra una búsqueda realizada con la herramienta *Concordance* en el mismo corpus de Nutrición, para ver ejemplos de colocaciones frecuentes del término “colesterol”.

#### 7.4.3 Selección de los términos académicos especializados

La primera selección de términos académicos especializados se llevó a cabo aplicando tres criterios: *ocurrencia especializada*, *frecuencia* y *valor de dispersión*. Para cumplir con el primer criterio, se utilizó la herramienta *Stopword list*, con el propósito de comparar el corpus especializado con un corpus general de más de dos millones de palabras. El objetivo de esta comparación era el de eliminar las palabras de uso general, como son artículos, preposiciones, conjunciones, pronombres y sustantivos no especializados. Tomando como base la proporción de ocurrencias en estas listas especializadas, se establecieron dos rangos de *frecuencia*: las palabras de alta frecuencia, con más de 50 ocurrencias dentro del corpus, y las palabras de baja frecuencia, cuyos rangos de ocurrencia iban de 5 a 49 *tokens*. Los términos con una frecuencia menor a 5 ocurrencias fueron excluidos automáticamente en esta primera selección. La lista completa de términos seleccionados por frecuencias en el corpus de Nutrición se encuentra en la sección de Apéndices (Ver 10.3)

Como siguiente requisito de selección se consideró el *valor de dispersión* de cada una de los términos que cumplieron con los dos criterios mencionados anteriormente. Este valor refleja el grado en el cual una palabra está distribuida uniformemente dentro de un corpus (Oakes, 1998). Para los propósitos de este estudio se estableció que el valor de dispersión mínimo que los términos debían cumplir para poder ser incluidos debía ser de 0.4, es decir, estar distribuidos uniformemente en por lo menos el 40% del corpus.

#### 7.4.4 Generación de no palabras

Como se mencionó anteriormente, el formato de tarea de decisión léxica (*checklist*) que se utilizó en este estudio no solo requiere de una lista de términos del área de especialización evaluada, también necesita una serie de no palabras o pseudo palabras que se utilizan para evaluar si los participantes realmente no están sobrestimando su conocimiento.

Durante la primera etapa de selección se observó que dentro del corpus de Nutrición, el cual fue el primero que se elaboró y se procesó, 722 palabras del total de las 871 seleccionadas eran términos que incluían prefijos de origen griego o latino, lo cual es frecuente en los textos especializados (Cabré, 1998). Dado el alto número de ocurrencias y la alta frecuencia de palabras prefijadas, se determinó que las no palabras de la prueba se generaran utilizando prefijos y bases de alta frecuencia dentro de los corpus, pero en combinaciones inexistentes, por ejemplo: los términos *ateromatosis* y *fucosilación*, ambos muy frecuentes dentro del corpus de Nutrición, están compuestos por un prefijo (*atero-* y *fuco-*) y una base (*-matosis* y *-silación*) pero la combinación *aterosilación* no existe, generando así una no palabra. A continuación se describe más a detalle este proceso:

La primera parte de esta búsqueda consistió en la identificación de prefijos representativos dentro de los corpus. Con este objetivo, se realizó un análisis con la herramienta *Keywords* (Ver 7.4.2) comparando el corpus desarrollado para este estudio, con un corpus de lenguaje periodístico del español conformado por poco más de 2 millones de palabras. Para el corpus de Nutrición el programa identificó 1984 palabras “clave” para el área de Nutrición, dentro de las cuales se encontró una serie de palabras que contenían 53 prefijos de alta frecuencia, e.g. *cito-*, *fito-*, *galacto-*, *gluco-*. (Ver 10.7).

La frecuencia mínima obtenida por cada prefijo se situó en igual o mayor que 5 tipos, siendo *micro-* el prefijo con mayor frecuencia (52 tipos y 938 tokens) y *osteo-* el de menor frecuencia (6 tipos y 21 tokens). Mientras que para el área de Odontología el programa identificó 7465 palabras “clave”, dentro de las cuales se identificaron 44 prefijos de alta frecuencia, e.g. *gingi-*, *perio-*, *silo-*, siendo el más frecuente *odonto-* (32 tipos y 412 tokens) y el de menor frecuencia *tiso-* (6 tipos y

18 tokens) (Ver 10.8) dentro de la misma escala de frecuencias utilizadas para el análisis del corpus de Nutrición. La lista completa de los prefijos más frecuentes dentro del corpus de Nutrición y el corpus de Odontología se encuentra en la sección de Apéndices (Ver 10.7 y 10.8)

Tras la identificación de los prefijos, se realizó una búsqueda en *Concord* (Ver 7.4.2) para cada uno de los prefijos con el objeto de obtener todos los sustantivos que los incluyeran, e.g. *atero*:

**ATERO**GENESIS  
**ATERO**MA  
**ATERO**MATOSIS  
**ATERO**SCLEROSIS  
**ATERO**TROMBÓTICO

Al concluir la búsqueda por concordancias para cada prefijo, se obtuvo una base de datos total de palabras prefijadas. La siguiente parte del análisis consistió en la eliminación de los prefijos dentro de estas listas; esto con el objeto de identificar las bases que con mayor frecuencia se adjuntan a diferentes prefijos, e.g. La raíz *-calórica* se encontró con los siguientes prefijos:

HIPERCALÓRICA  
HIPOCALÓRICA  
ISOCALÓRICA  
NORMOCALÓRICA

Para realizar este análisis, primero se eliminaron los prefijos por medio de la herramienta de “Búsqueda-Sustitución” de Excel, tras lo cual se procesó la lista de palabras en la herramienta *WordList* de *WordSmith Tools*, dando como resultado la lista bases que se combinan más frecuentemente con los prefijos identificados anteriormente, e.g.

-GÉNESIS  
-CALÓRICA  
-LISIS  
-ACTIVIDAD  
-CÍTICA  
-GÉNICA  
-GLICEMIA  
-SACÁRIDO  
-TERAPIA

El resultado de estas combinaciones fue revisado tanto en el Diccionario Médico-Biológico, Histórico y Etimológico disponible en línea, así como en el motor de búsqueda Google, incluyendo en la selección de no palabras sólo aquellos

términos que no arrojaron ningún resultado en la búsqueda. La lista completa de no palabras tanto para la prueba de Nutrición como para la prueba de Odontología se encuentra en la sección de Apéndices (Ver 10.4 y 10.6).

## 7.5 Validación de términos y no palabras con expertos del área de Nutrición y Odontología

Con el objeto de verificar tanto la selección de términos reales como la creación de las no palabras se decidió consultar a un grupo de 3 expertos adscritos a dos diferentes instituciones educativas que cuentan con el programa de Licenciatura en Nutrición y la Licenciatura en Odontología en la ciudad de Querétaro. Todos los especialistas eran profesores universitarios de tiempo completo, poseían estudios de posgrado y eran docentes de por lo menos dos materias impartidas a nivel licenciatura en el área de Nutrición u Odontología y contaban con más de 20 años de experiencia en la práctica profesional de su área de especialidad.

Todos los términos incluidos en cada una de las pruebas fueron seleccionados por lo menos por dos expertos en el área de especialización como “palabras altamente representativas dentro del lenguaje académico de su campo de estudio” y todas las no palabras incluidas en la selección final fueron reconocidas por lo menos por dos de los expertos como “no confusas o ambiguas”. Por ejemplo, la no palabra *edenoma* fue rechazada por 2 expertos debido a que es muy cercana tanto fonológicamente como gráficamente al término real *adenoma*; mientras que la no palabra *baseloide* no existe en español, pero en el inglés existe como *basaloid*, razón por la cual también fue rechazada.

## 7.6 Participantes



### 7.6.1 Grupo Control

El Grupo Control únicamente resolvió la prueba de decisión léxica para el área de Nutrición, ya que ésta fue la primera prueba que se elaboró y el objetivo de la Prueba de Odontología fue únicamente el de corroborar que la metodología propuesta en este estudio fuera aplicable a otras áreas de especialización.

Como Grupo Control se seleccionó un grupo de 40 alumnos que estudiaban el octavo semestre de bachillerato. El rango de edad de los sujetos fluctuaba entre los 16 y 17 años de edad. Todos los estudiantes eran nativos hablantes del español y carecían de conocimiento especializado en el área de Nutrición.

### 7.6.2 Grupo Experimental para el área de Nutrición

En el Grupo Experimental de Nutrición participaron un total de 157 estudiantes que cursaban una Licenciatura en Nutrición en dos instituciones de educación superior mexicanas: La Universidad Autónoma de Querétaro (UAQ) y la Universidad de Valle de México, Campus Querétaro (UVM). En el momento de la aplicación de la prueba, los participantes cursaban segundo, cuarto, sexto y octavo semestre de la carrera, conformando así un total de 8 grupos: un grupo de segundo semestre de la UAQ, y un grupo de segundo semestre de la UVM; un grupo de cuarto semestre de la UAQ y un grupo de cuarto semestre de la UVM; y así respectivamente hasta completar dos grupos más para el sexto semestre y dos grupos más para el octavo.

En el momento de la aplicación de la prueba, 50 de los participantes cursaban el segundo semestre de la licenciatura, 32 pertenecían al 4to semestre, 32 al 6to semestre y 43 cursaban el octavo y último semestre de la carrera. Todos los participantes eran nativos hablantes del español y su rango de edad fluctuaba entre los 17 y 23 años en el momento de aplicación de la prueba. La Tabla 4

detalla el número de estudiantes por semestre que pertenecían a cada una de las instituciones educativas.

<b>Institución</b>				
<b>Educativa</b>	<b>2do Sem.</b>	<b>4to. Sem</b>	<b>6to. Sem.</b>	<b>8vo. Sem.</b>
<b>UAQ</b>	35	14	17	34
<b>UVM</b>	15	18	15	9

**Tabla 5 Número de participantes por institución educativa y por semestre para la Prueba de Nutrición**

#### 7.6.2.1 Grupo Experimental para el área de Nutrición- Maestría

Con el objetivo de verificar que la metodología propuesta no contara con un máximo posible de respuestas acertadas – efecto de techo- el cual fuera fácilmente alcanzado por los estudiantes de los semestres superiores se decidió utilizar un Grupo Experimental de Nutrición que fuesen estudiantes de Posgrado. En este grupo participaron 14 estudiantes del 3er semestre de la Maestría en Nutrición de la Universidad Autónoma de Querétaro.

#### 7.6.3 Grupo Experimental para el área de Odontología

En el Grupo Experimental de Odontología participaron un total de 93 estudiantes de la Licenciatura en Odontología de la Universidad del Valle de México. La prueba no se aplicó en otras instituciones ya que, como se mencionó anteriormente, la Prueba de Odontología se desarrolló únicamente con el

propósito de corroborar que la metodología propuesta en este estudio fuera aplicable a otras áreas de especialización además de la Nutrición.

En el momento de la aplicación de la prueba, 26 de los participantes cursaban el segundo semestre de la licenciatura, 27 pertenecían al 4to semestre, 20 al 6to semestre y 20 cursaban el octavo y último semestre de la carrera. Todos los participantes eran nativo hablantes del español y su rango de edad fluctuaba entre los 17 y 24 años en el momento de aplicación de la prueba.

Institución	2do Sem.	4to. Sem	6to. Sem.	8vo. Sem.
<b>UVM</b>	26	27	20	20

**Tabla 6 Número de participantes por semestre para la prueba de Odontología**

### 7.7 Elaboración y aplicación de las pruebas

A partir de la selección final de los expertos se desarrollaron dos pruebas de decisión léxica de formato *checklist*, una para Nutrición y otra para Odontología, las cuales incluían un total de 200 ítems cada una, de los cuales 60 eran palabras de alta frecuencia, 100 de baja frecuencia y 40 no palabras. Estos ítems fueron combinados de forma aleatoria y organizados en columnas, (Ver 10.1 y 10.2) e.g:

1  anodoncia

2  orofaringe

3  amelocementaria

4  sialofítica

5  dentiplásico

6  inmunohistoquímica

La instrucción que se les dio a los participantes fue la de marcar las palabras que claramente conocieran, dejando sin marcar todas aquellas palabras que no conocieran o que no estuvieran completamente seguros de conocer, además de proporcionar un ejemplo del tipo de respuesta que se solicitaba:

**Instrucciones:**

**Si conoces el significado de la palabra, rellena el recuadro de la izquierda.**

**Si no la conoces o no estás seguro de conocer su significado, deja el recuadro en blanco.**

Ejemplo:  gingivitis  melanoacantoma

La prueba se aplicó en formato impreso en el horario y salón en que los estudiantes regularmente acudían a clases. En promedio la resolución de la prueba requirió de 20 minutos, aunque previamente no se había establecido límite de tiempo.

El formato y contenidos de ambas pruebas, tanto de la desarrollada para el área de Nutrición y así como la de Odontología, se encuentran la sección correspondiente a Apéndices, en los apartados 10.1 y 10.2, respectivamente.

## 8. RESULTADOS Y DISCUSIÓN

A continuación se presentan los resultados obtenidos en cada una de las pruebas. Los datos se muestran por grupos correspondientes al semestre del programa de estudios que cursaban los participantes en el momento de aplicación de la prueba: 2do, 4to, 6to y 8vo. Semestre, Grupo Control ó Maestría.

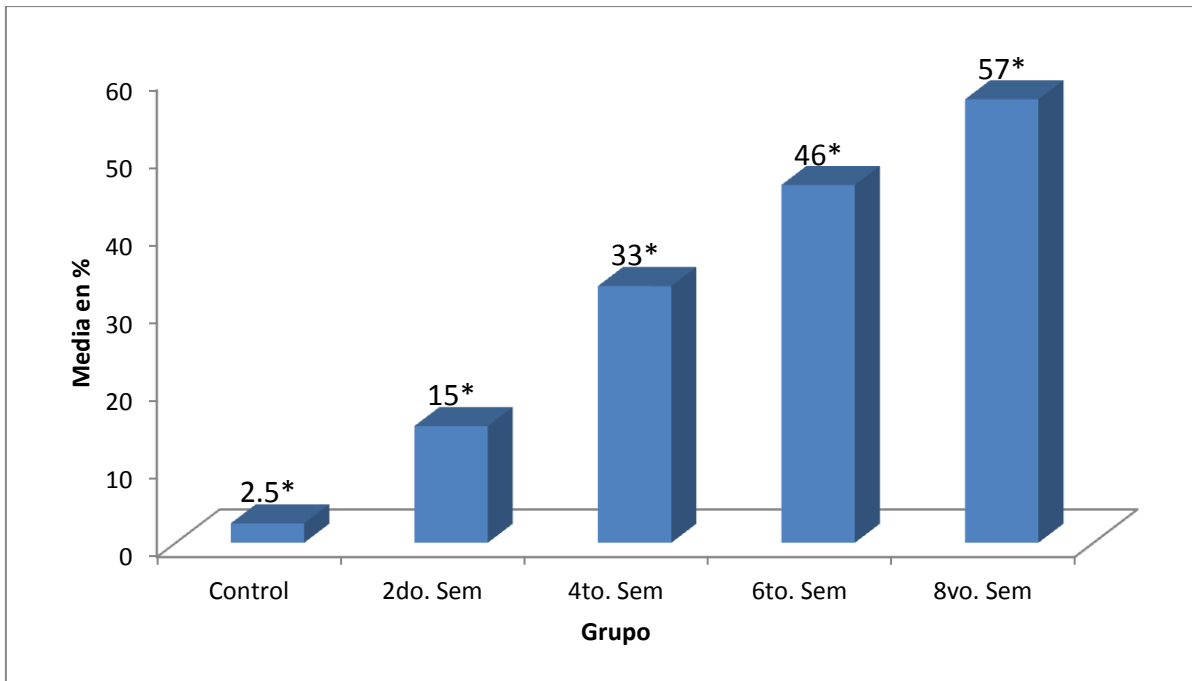
El análisis de los resultados obtenidos en ambas pruebas se llevó a cabo ajustando el porcentaje de respuestas correctas (palabras reales marcadas como conocidas) con el porcentaje de falsas alarmas (no palabras marcadas como conocidas) realizando la sustracción del total de las respuestas correctas menos el total de las falsas alarmas. Es decir, si un participante indicó que conocía el 62% de las palabras reales, y un 10% de las no palabras, la puntuación que obtendría en la prueba sería de 52%. Se optó por esta fórmula para la evaluación de los resultados (*h-fa*) debido a que otras fórmulas propuestas en estudios previos con pruebas de decisión léxica (Ver 5.2.1.1.1) penalizaban fuertemente a los estudiantes conservadores, es decir, aquellos estudiantes que marcaban pocas palabras debido a que no estaban completamente seguros de conocer el significado exacto de los ítems evaluados, arrojando resultados negativos. Por ejemplo, un participante conservador, marcó únicamente 13 palabras reales como conocidas y no marcó ninguna no palabra. Al realizar el análisis estadístico con la fórmula propuesta por Huibregtse et al.(2002), el participante obtuvo una media de -214 (menos doscientos catorce) que no es un resultado válido para una prueba de tamaño de vocabulario – no se puede conocer el menos 200% de una terminología especializada- mientras que aplicando la fórmula de *h-fa* obtenía un puntaje de 13%. Para una discusión detallada sobre modos alternativos de evaluación en este tipo de pruebas ver Beeckmans et al. (2001); Huibregtse et al.(2002); y Mochida y Harrington (2006).

En relación a la confiabilidad de la prueba, se realizó un análisis de consistencia interna de las correlaciones de los ítems (*Alfa de Cronbach*). Para este análisis se consideraron únicamente las 160 palabras de la prueba, excluyendo las no palabras. El análisis para la Prueba de Nutrición arrojó  $\alpha = .987$  mientras que el de la Prueba de Odontología fue de  $\alpha = .977$ . Según Kline (1994), los valores del *Alfa de Cronbach* de .8 o más indican una buena consistencia interna, por lo que los resultados obtenidos indican que la prueba es altamente confiable.

## 8.1 Resultados de la Prueba de Nutrición

### 8.1.1 Licenciatura en Nutrición de la UAQ

La Figura 6 presenta los resultados obtenidos para cada uno de los grupos de la Licenciatura en Nutrición de la Universidad Autónoma de Querétaro y el Grupo Control de estudiantes de Bachillerato. Los datos están presentados por semestre: Grupo Control, 2do, 4to, 6to y 8vo. Semestre. La media en porcentaje representa el promedio de palabras reconocidas obtenido por cada uno de los grupos una vez que se realizó el ajuste de respuestas correctas menos las falsas alarmas.



\*p<.001

**Figura 6 Resultados de la Prueba de Nutrición aplicada a estudiantes de la UAQ + Grupo Control**

Como puede observarse en la figura, el grupo que obtuvo los resultados más bajos fue el Grupo Control, conformado por estudiantes de bachillerato. El hecho de que este grupo obtuviera resultados más bajos en comparación con los resultados de los grupos constituidos por estudiantes de la Licenciatura en Nutrición permite suponer que la prueba puede discriminar entre los grupos que han estado expuestos a un vocabulario especializado y aquellos que no. Más aún, los resultados parecen indicar que la prueba es sensible al tiempo de exposición al vocabulario, ya que la media obtenida por cada grupo se incrementa en relación al semestre cursado por los participantes. Es decir, los estudiantes del cuarto semestre obtuvieron resultados un 18% más altos que los de segundo, los de sexto superaron a los de cuarto en un 13%, y los de octavo lograron un promedio un 11% mayor que el de los de sexto.

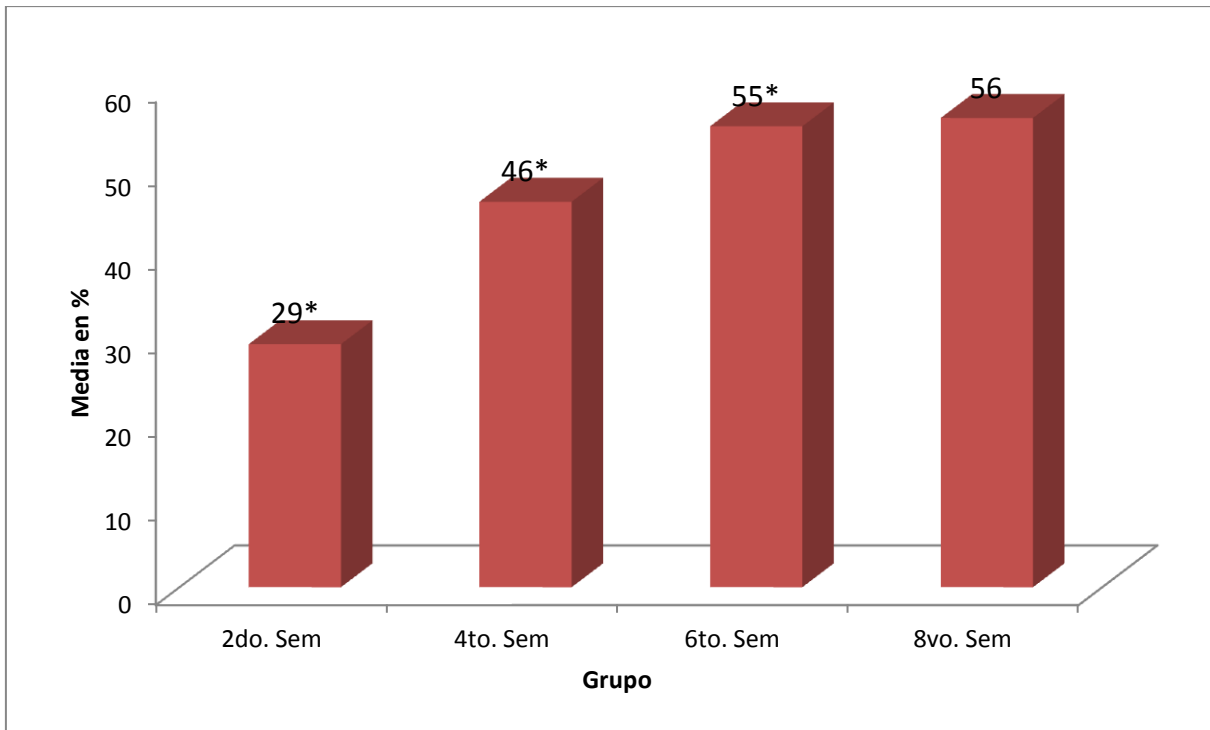
Las pruebas de normalidad (Kolmogorov-Smirnov y Shapiro-Wilks) mostraron que los datos obtenidos para todos los grupos, excepto el Grupo Control, se distribuían normalmente. El hecho de que el Grupo Control no obtuviera datos con comportamientos típicos era un resultado esperable debido a que, como se mencionó anteriormente, al ser estudiantes de bachillerato no habían tenido exposición previa a la terminología especializada del área de Nutrición. Por este motivo, se decidió realizar el análisis de la varianza (ANOVA) tanto incluyendo los datos del Grupo Control, como contando únicamente los datos de los 4 grupos que sí habían obtenidos datos con distribuciones normales.

Tanto el análisis que incluye los datos del Grupo Control como el que no, arrojan resultados que demuestran diferencias altamente significativas:  $F(4,140)=217.03$ ,  $p=.000$  y  $F(3,101)=102.46$ ,  $p=.000$ , respectivamente. Estos datos fueron corroborados con dos pruebas *post hoc* (Bonferroni y LSD), obteniendo siempre un valor de  $p<.000$ . Asimismo, el análisis del efecto del tamaño de la muestra ( $\eta^2$ ) reveló para el análisis que incluye al Grupo Control un valor de .93 y de .87 para el análisis de los 4 grupos de licenciatura.

#### 8.1.2 Licenciatura en Nutrición UVM

La Figura 7 presenta los resultados obtenidos para cada uno de los grupos de la Licenciatura en Nutrición de la Universidad del Valle de México. Al igual que en la sección anterior, los datos están presentados por semestre: 2do, 4to, 6to y 8vo. Semestre. Nuevamente la media en porcentaje representa el promedio obtenido por cada uno de los grupos una vez que se realizó el ajuste de respuestas correctas menos las falsas alarmas.





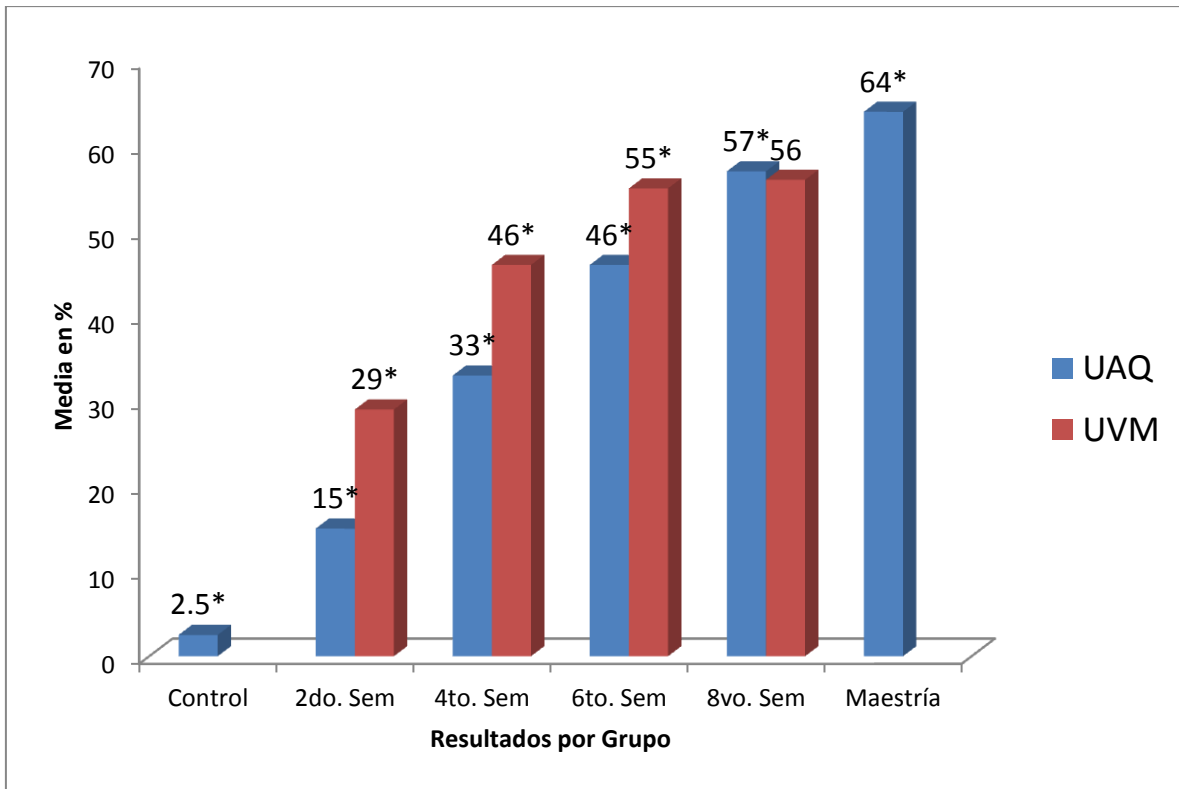
\*p<.001

**Figura 7 Resultados de la Prueba de Nutrición aplicada a estudiantes de la UVM**

Como puede apreciarse en la figura, los resultados obtenidos por los estudiantes de la Licenciatura en Nutrición de la UVM muestran un progreso constante durante los tres primeros años de la carrera y un estancamiento en el último año. Así, los estudiantes del 2do semestre tienen un conocimiento del 29% de las palabras medidas. El aumento entre el 2do y 4to semestre es de 17%, ya que los de 4to. tienen un promedio de 46% de palabras conocidas. Entre 4to y 6to semestre el aumento de vocabulario es de un 9% (casi la mitad que de 2 a 4); y el avance entre el 6to y octavo es de un 1%). Estadísticamente la diferencia entre el promedio de aciertos de los participantes de estos semestres es altamente significativa tal como refleja el Análisis de la Varianza ( $F(3,48)= 217.03$ ,  $p<.001$ ). Únicamente la diferencia entre el 6to. y el 8vo. semestres no resultó significativa.

### 8.1.3 Maestría en Nutrición UAQ

Debido a que la media de los resultados obtenidos por los estudiantes del 8vo. Semestre de la Licenciatura en Nutrición de la UAQ y los resultados del 6to. y 8vo. Semestres de la misma licenciatura en la UVM fueron muy similares (57, 55 y 56 %, respectivamente) se decidió aplicar la prueba a estudiantes de la Maestría en Nutrición de la UAQ con el objeto de explorar la posibilidad de que el formato de la prueba alcanzara el efecto de techo en niveles inferiores al número de aciertos totales que los participantes podían obtener. La media de los resultados obtenidos por los estudiantes de maestría se ubicó 7 puntos por encima del grupo correspondiente al 8vo. semestre de la licenciatura, demostrando que los estudiantes de licenciatura no alcanzaron el puntaje mayor que la prueba es capaz de detectar. A continuación se presenta el comparativo de ambas instituciones, incluyendo los datos obtenidos tras la aplicación de la Prueba de Nutrición a los estudiantes de Posgrado.



\*p<.001

**Figura 8 Resultados de la Prueba de Nutrición aplicada a todos los Grupos UAQ y UVM**

La comparación entre los resultados que arrojan las dos instituciones muestran que el conocimiento inicial es dispar pero queda igualado al finalizar los programas de estudio de la Licenciatura en Nutrición en ambas instituciones educativas: En el semestre 2, los estudiantes de la UVM tienen unos resultados que casi duplican los resultados de los estudiantes de la UAQ (hay 14 puntos entre el 15% de la primera y el 29% de la segunda). El mismo fenómeno se observa en el 4to semestre: los participantes de ambas instituciones aumentan de manera igual (se mantiene una diferencia de 13%; 33% en la UAQ y 46% en la UVM). En el semestre 6, aunque todavía existe una importante diferencia entre ambas instituciones, ésta es menos (sólo de 9%; de 46% en la UAQ a 55% en la

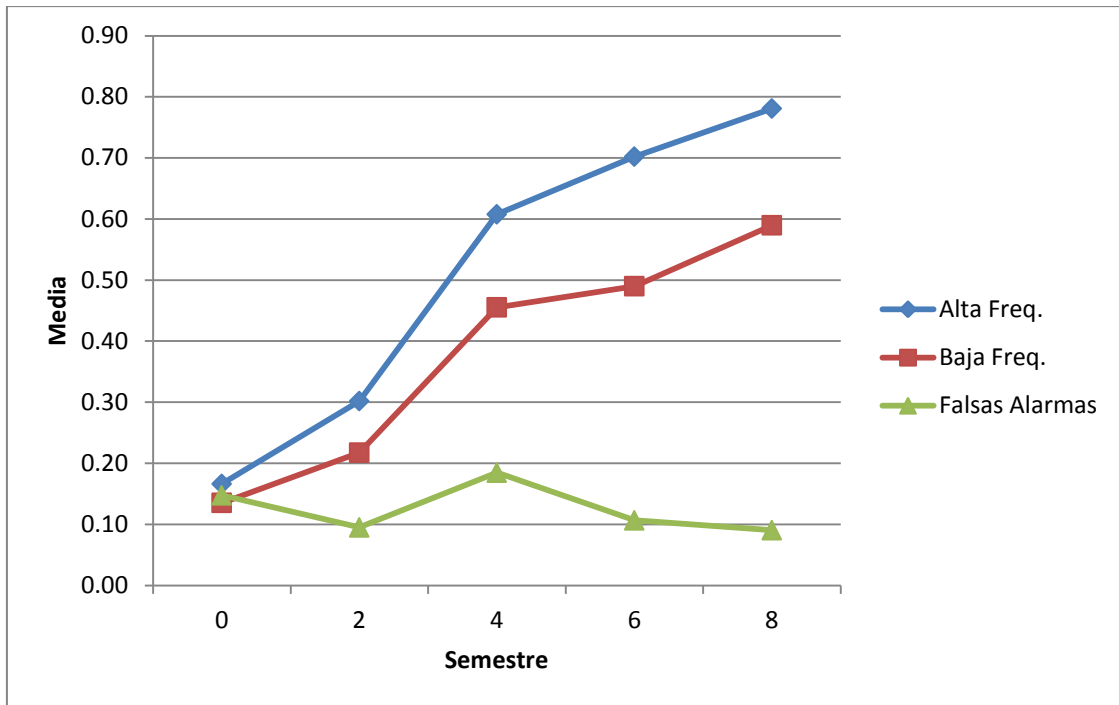
UVM). Finalmente en el 8vo semestre, el nivel de conocimiento de los participantes de la UAQ sigue aumentando mientras que el de la UVM se estanca, lo cual hace que se igualen (57% para la primera institución y 56% para la segunda).

Una lectura que resume los resultados expuestos hasta ahora sería que los participantes de la UAQ presentan durante los semestres 2, 4 y 6 un rezago de un año en comparación con la UVM (el cual finalmente desaparece en el último año). Esto puede deberse a muy diversos factores, entre ellos que una de las instituciones es privada mientras que la otra es pública, lo cual influye en los antecedentes escolares de los alumnos de cada institución y en el número de estudiantes en cada grupo. Además, mientras que el programa de Licenciatura en Nutrición de la UAQ contaba ya con varias generaciones de egresados, el mismo programa en la UVM todavía no contaba con titulados al momento de la aplicación de la prueba, siendo los estudiantes del 8vo. Semestre los pioneros en el programa.

#### 8.1.4 Análisis por frecuencia de los términos y no palabras de la Prueba de Nutrición

Tras realizar el análisis estadístico por grupos, se procedió a realizar el análisis por frecuencia de palabras, es decir, el análisis de los términos de alta frecuencia y de baja frecuencia, para posteriormente culminar con el análisis de no palabras. La Prueba de Nutrición incluía 60 términos de alta frecuencia dentro de corpus, 100 palabras de baja frecuencia y un total de 40 no palabras.

Para realizar este análisis se sumaron los resultados de ambas instituciones, UAQ y UVM incluyendo el Grupo Control, los dos grupos de 2do. Semestre, los dos grupos de 4to. Semestre y los dos grupos de 6to. y 8vo. Semestre. A continuación se presentan los resultados obtenidos por semestres para este análisis:



**Figura 9 Análisis por frecuencia de los términos y no palabras**

Como puede observarse en la figura, el Grupo Control no pudo discriminar entre las palabras de alta y baja frecuencia, y las no palabras. Esto es un resultado esperado ya que, como se mencionó anteriormente, este grupo no contaba con conocimiento previo en el área de Nutrición. También se puede apreciar cómo a partir del 2do. Semestre, los estudiantes de la Licenciatura en Nutrición marcan como conocidas las palabras de alta y baja frecuencia, mientras que manifiestan desconocer las no palabras. Es interesante observar cómo las palabras de más alta y baja frecuencia aumentan significativamente su porcentaje conforme avanza los semestres de estudio en el programa de la licenciatura, mientras que las falsas alarmas (no palabras marcadas como palabras reales) obtienen siempre los porcentajes más bajos.

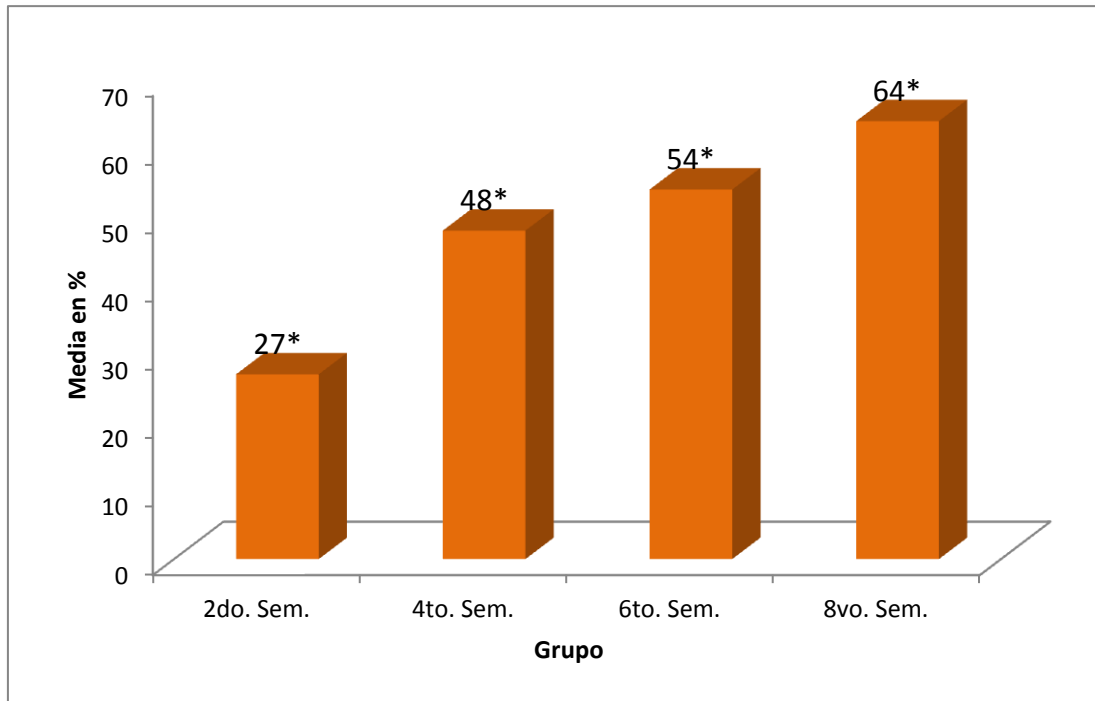
Posteriormente se realizó análisis de consistencia interna de las correlaciones de los ítems para ambas pruebas para comprobar si variar el orden de presentación de los mismos o sustituir los términos intercambiándolos por otros

que hayan cumplido con los mismos criterios de selección no afectaría los resultados obtenidos. El resultado de este análisis indicó un alto nivel de consistencia interna validando nuevamente la confiabilidad de la prueba.

Estos resultados permiten suponer que conforme avanza su educación universitaria, los estudiantes de la Licenciatura en Nutrición no sólo incrementan el conocimiento del vocabulario especializado de su área, sino que son capaces de discriminar con más precisión entre las palabras reales y las no palabras, ya que para obtener mejores resultados en la prueba es necesario no sólo marcar como conocidas un mayor número de palabras reales de la prueba, sino identificar las no palabras y no marcarlas como conocidas. Es decir, los falsos positivos (no palabras marcadas como reales) de los estudiantes de 4 semestre (que son un 18% de las no palabras) se reducen a un 40% en el semestre 6 (11% de las no palabras) y hasta un 50% en el semestre 8 (9% de las no palabras), y las desviaciones estándar se reducen aproximadamente en las mismas proporciones.

## 8.2 Resultados de la Prueba de Odontología

La Figura 10 presenta los resultados obtenidos para cada uno de los grupos de la Licenciatura en Odontología de la Universidad del Valle de México. Los datos están presentados por semestre: 2do, 4to, 6to y 8vo. Semestre. La media en porcentaje representa el promedio obtenido por cada uno de los grupos una vez que se realizó el ajuste de respuestas correctas menos las falsas alarmas.

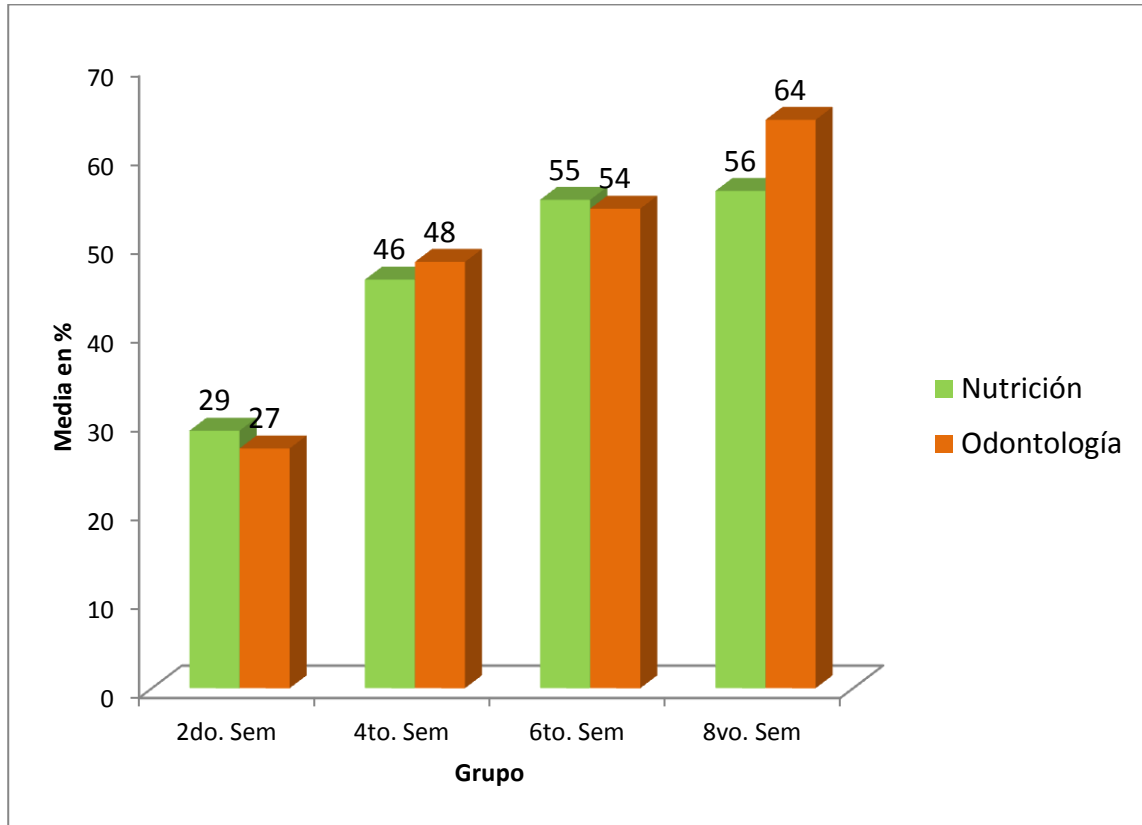


\*p<.001

**Figura 10 Resultados de la Prueba de Odontología aplicada a estudiantes de la UVM**

Como puede apreciarse en la figura, nuevamente el grupo que obtuvo los resultados más bajos fue el correspondiente al 2do. Semestre, con 27% de aciertos, y al igual que en el caso de la Prueba de Nutrición, este porcentaje se incrementa en el 4to, 6to y 8vo semestre. Entre el 2do y el 4to semestre hay una diferencia de 21 puntos porcentuales (del 27 al 48%, la más notable). La diferencia entre el 4to y el 6to semestre es de 6%; de 48 a 54%. Y la diferencia entre el 6to y 8vo semestre presenta tan sólo un incremento del 10%, de 54 al 64%. El Análisis de la Varianza para los 4 grupos muestra diferencia significativa ( $F(3,89) = 60.08$ ,  $p < .000$ ). El análisis de consistencia interna de las correlaciones de los ítems (*Alfa de Cronbach*) arrojó también en este caso un resultado que demuestra la alta confiabilidad de la prueba ( $\alpha = .977$ ), lo cual confirma que la metodología propuesta puede ser aplicada a diferentes áreas de conocimiento o disciplinas científicas. Incluso, aún a pesar de que los resultados no son idénticos para cada grupo, bien puede realizarse un comparativo entre los distintos programas de estudio

evaluados en esta investigación: Nutrición y Odontología. Dado que la prueba de Odontología únicamente se aplicó en la Universidad del Valle de México, se realizará sólo la comparación en dicha institución.



**Figura 11 Comparativo de los resultados de la Prueba de Nutrición y de Odontología en la UVM**

En la Figura 11 se puede apreciar como los resultados, a pesar de provenir de dos pruebas distintas, son comparables en función a que presentan una relación similar a la proporción del incremento del reconocimiento de vocabulario especializado de acuerdo al número de semestres cursados en el programa de estudios (excepto el 8vo. semestre). La elaboración y aplicación de pruebas para más disciplinas científicas dentro de la misma institución, por ejemplo Fisioterapia, o Ciencias del Deporte comparadas con los resultados obtenidos por esta investigación podrían tal vez establecer rangos para cada semestre, con lo cual



podría tal vez evaluarse el desempeño de los estudiantes ya sea por grupo o incluso a nivel individual.

Los resultados obtenidos hasta el momento incluso pueden ya abrir temas de discusión respecto al desempeño del grupo de 8vo. semestre de Nutrición: ¿Acaso debieron obtener puntuaciones similares a las obtenidas por los estudiantes de Odontología? ¿Por qué las puntuaciones de los estudiantes del 8vo. Semestre de Odontología coinciden exactamente con las puntuaciones de los estudiantes de Maestría de Nutrición? El hecho de que los estudiantes de Nutrición fueran la generación pionera en su programa de estudio ¿influyó en sus resultados? Los estudiantes de Nutrición de las siguientes generaciones ¿obtendrán resultados similares a los obtenidos en esta investigación por los estudiantes de 8vo. de Odontología?

Otras posibles comparaciones que se podrían realizar basándonos en estos resultados serían las referentes a las evaluaciones institucionales (los métodos de enseñanza, el efecto de cambios en los planes de estudios, etc.). En realidad, los resultados obtenidos por este estudio permiten ya especular sobre las causas de las disparidades entre las instituciones basándonos en datos. Por ejemplo, habrá quien argumente que las notables diferencias entre instituciones (con excepción del último semestre) se deban al hecho de que los resultados obtenidos en la Prueba de Nutrición de los alumnos de la Universidad de Valle de México – que supera en mucho el promedio de los alumnos de la Universidad Autónoma de Querétaro- es debido a que la primera es una institución privada. También sería posible explicar la similitud de los resultados del octavo semestre por factores como el tamaño del grupo del 8vo semestre de la UVM (sólo 14 estudiantes), ó al hecho de que este grupo fuera la primera generación en cursar el programa completamente, etc. Sin embargo, puede decirse que la metodología aquí presentada proporciona datos que permiten abrir debates más fundamentados

sobre los aspectos que se consideren oportunos para los responsables de las instituciones y los programas.

## 9. CONCLUSIONES

El objetivo general de esta investigación fue el de proponer una metodología para la elaboración de pruebas de evaluación de vocabulario académico especializado en español para estudiantes universitarios nativo hablantes del español. Esta metodología tiene sus bases en la evaluación de vocabulario, la lingüística de corpus y en la terminología.

En relación a la evaluación de vocabulario, se presentaron conceptos básicos sobre la definición y conceptualización del vocabulario como constructo, se definió el concepto de vocabulario académico utilizado en esta investigación y se presentaron las bases teóricas sobre la evaluación del vocabulario. Finalmente se introdujo el concepto de las pruebas de decisión léxica con formato de *checklist* y se justificó su aplicación en este estudio debido a que cuenta con elementos valiosos que coinciden con los objetivos de esta investigación.

En lo que respecta a la lingüística de corpus se definió el concepto de corpus lingüístico, se mencionaron los distintos tipos de corpus y su aplicación en diversos estudios relacionados con la lingüística y ciencias afines. También se establecieron las consideraciones generales para la construcción de corpus, debido a que para realizar la presente investigación fue necesario construir dos de ellos.

Finalmente se definió a la terminología y se presentó un breve resumen sobre su evolución y crecimiento, así como su contribución para la generación de listas especializadas de vocabulario académico, las cuales fueron la base para la generación de la metodología propuesta.

Esta metodología básicamente se puede resumir en los siguientes pasos: 1) generar un corpus especializado para la disciplina que se desee evaluar, 2) utilizar diversas herramientas de procesamiento de corpus para obtener términos representativos del área de estudio, 3) utilizar las herramientas de procesamiento

de corpus para obtener bases y prefijos altamente frecuentes dentro de la disciplina, 4) elaborar no palabras (o distractores) a partir de la combinación de las bases y prefijos seleccionados 5) seleccionar los términos meta y las no palabras con la ayuda de las herramientas de procesamiento de corpus y de especialistas en la disciplina a evaluar, 6) elaborar una prueba de decisión léxica con formato de *checklist* que nos permita evaluar de forma rápida y objetiva el conocimiento del vocabulario académico de la disciplina seleccionada.

El objetivo particular de la investigación fue el de reportar los resultados obtenidos en la aplicación de dos pruebas de evaluación de vocabulario académico especializado en español elaboradas según los parámetros de la metodología propuesta, una para el área de Nutrición y la segunda para el área de Odontología. Los resultados se analizaron y se reportaron en función de las siguientes variables:

- a) Disciplina de estudio, Nutrición y Odontología, respectivamente, incluyéndose un Grupo Control que respondió a la prueba de Nutrición.
- b) Tiempo de exposición al vocabulario especializado, el cual se reportó de acuerdo al número de semestres cursados dentro del programa de estudio de la Licenciatura en Nutrición y la Licenciatura en Odontología, 2do, 4to, 6to y 8vo. Semestre.
- c) Institución, que corresponde a las dos instituciones académicas a las que pertenecían los participantes de este estudio: la Universidad Autónoma de Querétaro y la Universidad del Valle de México, Campus Querétaro.

En relación a estos resultados podemos resumir que, los alumnos del Grupo Control que no fueron expuestos previamente a la terminología especializada- en este caso el vocabulario académico meta- son los participantes que obtuvieron los resultados más bajos dentro de la prueba de Nutrición, ya que no fueron capaces de discriminar entre las palabras reales que pertenecen a la disciplina y las no palabras que fueron generadas como distractores. Los alumnos de la Licenciatura

en Nutrición de la UAQ sí pudieron identificar la terminología especializada de su área de estudio, obteniendo resultados escalonados y progresivos por semestre, es decir, de acuerdo con el número de semestres que habían cursado en el programa de estudio. Así pues, los estudiantes del 8vo. Semestre fueron los que obtuvieron resultados más altos tras la aplicación de la prueba, al reconocer las palabras reales como terminología y marcar el menor número de no palabras como reales, seguidos por los resultados de los estudiantes del 6to, 4to, y 2do. Semestre, respectivamente.

Los resultados obtenidos por los alumnos de la Licenciatura en Nutrición de la UAQ fueron superados por los resultados de los alumnos de la Maestría en Nutrición de la misma institución, lo cual nos permite suponer que el conocimiento del vocabulario académico especializado continúa aumentando en proporción al tiempo de exposición al mismo.

Sin embargo, los resultados obtenidos para la misma prueba aplicada a los estudiantes de la Licenciatura en Nutrición de la UVM no siguen el mismo patrón. Si bien, en el 2do., 4to, y 6to Semestres si se obtienen resultados progresivos y escalonados, hay diferencias importantes con los resultados obtenidos en la UAQ. En primer lugar, los alumnos del 2do, 4to y 6to. Semestre de la UVM obtienen resultados que prácticamente duplican a los resultados de los mismos semestres en la UAQ. Mientras que los resultados del 6to. y 8vo. Semestre de la UVM son prácticamente iguales, es decir, parece no haber incremento del vocabulario en el último año de estudios del programa de la Licenciatura en Nutrición de la UVM. Pero al final, si comparamos los resultados de los grupos del 8vo. Semestre de ambas instituciones, éstos se igualan, a pesar de haber sido tan diferentes durante los primeros semestres.

Finalmente, la prueba de Odontología que se desarrolló con el objetivo de comprobar la metodología propuesta y su aplicación a otras disciplinas de estudio,

nuevamente obtuvo resultados escalonados y progresivos, siendo los estudiantes de 2do. Semestre los que obtuvieron los resultados más bajos, mismos que aumentaron según el número de semestres cursados en el programa.

A continuación se responde a las Hipótesis planteadas al inicio de este reporte:

*H1. A través de la lingüística de corpus y de diversas herramientas para el procesamiento de corpus es posible seleccionar una serie de palabras clave para una disciplina científica específica.*

La aplicación de la lingüística de corpus y algunas de herramientas de procesamiento sí nos permitieron obtener una lista robusta de terminología en el área meta, tanto para Nutrición como para Odontología. Esto fue confirmado por los expertos que consultamos, ya que más del 90% de los términos que se extrajeron del corpus fueron seleccionados por los expertos como representativos y frecuentes.

Creemos que en gran medida, esto se debe a los criterios para la conformación de corpus. El contar con artículos científicos publicados por revistas internacionales reconocidas nos permitió en cierta medida garantizar la pertinencia de los términos seleccionados en relación a su representatividad. Los resultados de frecuencia arrojados por las herramientas de procesamiento de corpus también nos permitieron establecer un rango para la utilización de los términos, ya que a pesar de desconocer el vocabulario académico relativo a cada área de especialización, fuimos capaces de identificar gran parte de su terminología e reconocerla como básica en la formación de un profesionista en el área.

*H2a. El conocimiento del vocabulario académico especializado aumenta según el tiempo de exposición al mismo.*

*H2b. Los estudiantes que tienen mayor tiempo de exposición a un vocabulario académico especializado obtienen mejor puntuación en una prueba de reconocimiento léxico.*

Los resultados obtenidos en esta investigación permiten apoyar esta hipótesis. Los estudiantes de los niveles superiores obtuvieron en promedio resultados más altos en las pruebas de decisión léxica aplicadas que aquellos estudiantes que tenían menor o ningún tiempo de exposición al léxico especializado. Si bien, esta afirmación pareciera poder soportarse por el sentido común, algunos de los resultados obtenidos en este estudio plantean más preguntas relacionadas con el tema: ¿Por qué los alumnos del 6to. y 8vo. Semestre de Nutrición de la UVM obtienen los mismos resultados? ¿Por qué al inicio de sus estudios de licenciatura los estudiantes de Nutrición de la UVM y la UAQ tienen resultados tan diferentes? ¿Por qué estos resultados se igualan al final de su formación como licenciados? ¿Es el tiempo de exposición el único factor que influye en este conocimiento? ¿Cuáles podrían ser los otros factores que influyen? ¿Los maestros? ¿Las lecturas que realizan? ¿Sus antecedentes familiares? ¿Las prácticas profesionales?

*H3. Una prueba de decisión léxica es capaz de determinar cuantitativamente el tiempo de exposición a un vocabulario académico especializado.*

En relación a esta hipótesis podemos concluir que una prueba de decisión léxica elaborada con criterios rigurosos para la selección de las palabras y para la elaboración de no palabras sí es capaz de determinar cuantitativamente la diferencia entre diversos grupos expuestos a un vocabulario académico especializado. Sin embargo, se debe planear muy cuidadosamente el uso de este formato y considerar diversos factores como son:

- El número de ítems aplicados en la prueba: una prueba muy larga es tediosa y aburrida para los participantes, lo cual necesariamente repercutiría en los resultados, pero una prueba demasiado breve no arrojaría resultados que pudieran ser considerados significativos.
- La motivación de los participantes: en México, los estudiantes no están acostumbrados a participar en investigaciones ni a resolver pruebas con propósitos científicos. Lamentablemente, en esta investigación fue necesario decirles a los participantes que los resultados de las pruebas se reportarían a sus profesores y coordinadores, ya que no se confiaba en la disposición de los estudiantes de responder seria y concienzudamente a la prueba si esta no tenía repercusiones.
- La fórmula estadística que se utilizará para analizar los resultados: el método de análisis, por lo menos en esta investigación, arrojó resultados muy diversos, al grado de presentar resultados en porcentajes negativos (e.g. -200%). El utilizar una u otra fórmula propuesta por los teóricos e investigadores en el área necesariamente tendrá variaciones, y en consecuencia resultados diferentes.

*H4. El nivel de conocimiento de vocabulario académico especializado puede ser comparado aún entre distintas disciplinas de estudio.*

Para poder corroborar esta hipótesis es necesario realizar pruebas para más disciplinas y aplicarlas a un gran número de participantes. Sin embargo, los resultados obtenidos en esta investigación permiten suponer que esta comparación podría realizarse. Aparentemente, hay un rango de respuestas esperado para cada nivel (para cada semestre en el caso de esta investigación) y ese rango guarda una proporción directa con el rango de respuestas esperado tanto en el nivel superior como en el inferior, esto lo vimos al comparar los resultados de ambas pruebas, Nutrición y Odontología en la UVM, pero este rango puede presentar inmensas variaciones, como lo vimos en los resultados de



los estudiantes de 2do. Semestre de Nutrición de la UAQ y la UVM cuya variación fue prácticamente del doble. Así pues, y como se mencionó anteriormente, muchas más pruebas deben ser elaboradas y analizadas para poder establecer estos rangos, por lo que esta hipótesis no puede ser corroborada por esta investigación.

*H5. No existen diferencias significativas entre el nivel de reconocimiento de vocabulario académico especializado de los estudiantes de una misma disciplina que estudian en distintas instituciones educativas.*

Con base en los resultados de esta investigación, esta hipótesis no puede ser corroborada pero tampoco puede ser descartada. Por un lado, tenemos los resultados de los estudiantes del 2do, 4to, y 6to Semestre de Nutrición, los cuales varían significativamente en cada institución, pero por otro lado, si comparamos los resultados de ambos grupos de 8vo. Semestre son prácticamente iguales. Este hecho más que responder a una pregunta, genera nuevas preguntas: ¿Es la función de una universidad el unificar el conocimiento? A pesar de las diferencias en los antecedentes escolares ¿realmente contamos con profesionistas los cuales comparten un nivel de conocimiento? Las instancias que rigen nuestros programas educativos- la Secretaría de Educación Pública y las organizaciones de profesionistas y organismos certificadores- ¿realmente están cumpliendo cabalmente con su trabajo al aceptar o no un programa o al certificar o no a un profesionista? ¿Hay un rango en el mínimo de conocimiento aceptable para poder obtener un título profesional? ¿Cuál es ese mínimo? ¿Qué o quién determina ese mínimo? Si comparamos estos resultados con instituciones de otros países ¿obtendremos resultados similares? ¿Pruebas como las propuestas por esta metodología podrían formar parte de una acreditación profesional como el Examen CENEVAL?

*H6 Los estudiantes con un mayor tiempo de exposición a un vocabulario académico especializado son capaces de distinguir entre el léxico más frecuente de su disciplina y el léxico que no pertenece a su área de conocimiento.*

Basándonos en los resultados obtenidos en el análisis por ítems podemos afirmar que los estudiantes que han estado expuestos durante un mayor tiempo a un léxico especializado desarrollan una mayor capacidad para detectar aquellos términos que no pertenecen a su disciplina de estudio. Mientras que el Grupo Control no pudo diferenciar entre las palabras de Nutrición y las no palabras generadas como distractores, los alumnos de 2do, 4to, 6to y 8vo semestre obtuvieron resultados nuevamente progresivos y escalonados, los cuales indican su capacidad para reconocer cada vez más las palabras muy frecuentes de su área, y descartar las no palabras.

Creemos que estos resultados además respaldan la metodología propuesta para la generación de no palabras, ya que éstas cumplieron su función como distractores, pero pudieron ser identificadas por los grupos especializados. Por poner un ejemplo mencionaremos un error ocurrido en la prueba de Odontología: Mientras que el grupo de expertos seleccionó como pertinente la no palabra *orofaringe* en el momento de la construcción de la prueba se cometió un error y la escritura de la no palabra fue modificada por *orogaringe*, un cambio que como se podrá apreciar modifica únicamente una consonante. Sin embargo, esta no palabra no obtuvo una sola marca, es decir, ningún participante de ningún semestre la marcó como palabra real. Esto nos permite suponer que otras metodologías propuestas para la generación de no palabras, las cuales sustituyen únicamente vocales o únicamente consonantes de palabras reales pueden no ser tan eficientes como la construcción de no palabras a partir de prefijos y bases de alta frecuencia.

En relación a las dificultades que se presentaron a lo largo de esta investigación podemos mencionar algunas relacionadas con la construcción y procesamiento de los corpus. En primer lugar, es muy difícil determinar el tamaño ideal de un corpus especializado. Elaborar un corpus demasiado robusto presenta la ventaja de arrojar más datos, brindar más ejemplos y permitir determinar con mayor seguridad los rangos de frecuencia a utilizar, sin embargo, representa diversas dificultades como son: 1) tener acceso a las revistas especializadas de forma gratuita, 2) encontrar las publicaciones especializadas más utilizadas en la disciplina, 3) tener acceso a los textos en español y en formato electrónico, 4) identificar las variaciones léxicas que se pueden presentar en diferentes países hispano hablantes, etc.

Además, el construir y procesar un corpus demasiado robusto representa más lentitud en el análisis de los datos y una alta inversión de tiempo por parte del investigador, ya que aunque las herramientas para el procesamiento de corpus cada vez están más desarrolladas, la eliminación de tablas, resúmenes, agradecimientos, índices, bibliografías, etc. aún tuvo que ser ejecutada de forma manual. Igualmente, la selección de términos y no palabras, aún a pesar de haberse realizado siguiendo estrictos procedimientos, tuvo que ser confirmada por expertos, es decir, aunque las herramientas de procesamiento utilizadas sean muy útiles y avanzadas, siempre se requerirá del ojo del investigador y del experto dentro del trabajo de la lingüística de corpus.

Por otra parte, las herramientas para el procesamiento de corpus utilizadas en esta investigación tienen un costo por licencia, lo cual representa una inversión que no siempre está considerada por los investigadores. Además, llegar a comprender y utilizar adecuadamente todos los elementos de las herramientas de procesamiento requiere de una capacitación especializada y muchas horas de práctica que no siempre arrojan los resultados esperados.

Así mismo, se debe considerar que cada disciplina presenta un comportamiento diferente en relación a su léxico especializado, lo cual necesariamente influirá en el tamaño y procesamiento del corpus. Para el caso específico de esta investigación se detectó que el área de Nutrición cuenta con un número más reducido de términos especializados en comparación con el área de Odontología. Este fue un descubrimiento muy interesante, ya que afectó directamente la selección de los ítems aplicados en la prueba. Mientras que en el corpus de Nutrición encontramos pocas palabras muy frecuentes, en el corpus de Odontología encontramos muchas más palabras pero con rangos de frecuencia mucho menores. Incluso ampliando a más del doble el tamaño del corpus de Odontología en relación con el corpus de Nutrición, fue difícil encontrar 60 palabras de Odontología que tuvieran rangos de frecuencia por encima de los 50 *tokens* y estuvieran distribuidas uniformemente en más del 40% del texto.

Además durante la selección de las disciplinas a evaluar encontramos más diferencias en cuanto al léxico especializado y la forma de argumentación científica de cada área. Mientras que en las ciencias naturales como la Biología, Química, Nutrición, Medicina, etc., sí encontramos terminología especializada en forma de palabras individuales e.g. odontopediatría, adenoma, etc, en otras disciplinas como la Ingeniería y Derecho son pocos los términos especializados que se encontraron en los corpus consultados. Se encontró que tanto la Ingeniería como el Derecho, más que contar con términos cuentan con paratérminos, es decir, construcciones de dos a más palabras que representan conceptos especializados en el área. Por ejemplo: *teoría de cuerdas*, para la Ingeniería y *órgano de decisión*, para el Derecho, ciencias para las cuales la metodología propuesta no puede ser aplicada, sino adaptada, por ejemplo generando una serie de ítems que en lugar del formato *checklist* utilizaran un formato de opción múltiple para la construcción de paratérminos, e.g:

*El órgano \_\_\_\_\_ con funciones administrativas electorales*  
*constitucional*  
*consultivo*  
*sancionador*  
*de decisión*  
*de instancia*

Otro aspecto problemático relacionado con esta investigación fue la validez de la prueba. Mientras que estadísticamente fue posible comprobar su confiabilidad, no nos fue posible comprobar su validez. Un intento fallido fue el de encontrar una correlación entre los resultados de la prueba aplicada y los promedios de las calificaciones de los participantes, pero lamentablemente los datos no se correlacionaron de ninguna forma. Una posible explicación a este fenómeno es que el rango de calificaciones que otorgamos a los alumnos va de 0 a 10, o de NA (No Aprobatorio) a A (Aprobatorio). Pero en la escala del 0 al 10 ¿realmente qué significa un 7 o un 8? ¿Implica conocimiento en el área? ¿Suficiencia? ¿Capacidad? ¿Trabajo en equipo? Una posible solución a este problema sería el de correlacionar los resultados de un examen CENEVAL con los obtenidos en la prueba, lo cual implicaría tener acceso a una guía muestra y aplicarla a los mismos grupos. Sin embargo, esto no sería posible para el área de Nutrición, ya que no existe el examen CENEVAL para esta disciplina.

También se debe tener en cuenta que este instrumento no mide la profundidad de conocimiento de vocabulario, ya que no es capaz de establecer si el participante conoce el significado preciso y el uso correcto de los ítems evaluados. Únicamente mide el vocabulario receptivo. También es importante resaltar que la prueba fue desarrollada con el objetivo de discriminar el nivel de conocimiento de grupos pertenecientes a diferentes semestres de un programa de estudios, y que en esta investigación no fue utilizado para evaluar el conocimiento de vocabulario de los participantes a nivel individual, para lo cual se podrían que llevar a cabo

estudios de tipo longitudinal. En este mismo sentido, deben considerarse las repercusiones que los resultados de la prueba propuesta podrían tener a nivel individual. No sería ético evaluar el conocimiento y desempeño de un estudiante o profesional de una disciplina únicamente a partir de la aplicación de una prueba de decisión léxica, por lo que nuevamente se resalta el hecho de que el instrumento propuesto debe aplicarse como una herramienta más dentro de una batería de pruebas, no como una prueba independiente.

Sin embargo, consideramos que la metodología propuesta presenta múltiples aspectos positivos. En primer lugar, están los criterios para la selección de los ítems. El utilizar artículos científicos de revistas especializadas indexadas nos brinda acceso a vocabulario académico pertinente y actual. Utilizar herramientas de procesamiento de corpus nos brinda además la posibilidad de establecer cuáles de estos términos son más o menos frecuentes, cuáles tienen un rango muy restringido de uso y cuáles definitivamente pertenecen al vocabulario básico de una disciplina de estudio. Finalmente, acudir a un grupo de expertos en el área imprime un aspecto más humano a la selección de ítems. Si bien resulta necesario hacer uso de las herramientas automatizadas disponibles, siempre es indispensable aplicar el conocimiento de tantos años de aquellos especialistas que no sólo son expertos en la disciplina, sino que además están también dedicados a la enseñanza de su campo de estudio, y conocen mejor que nadie a los estudiantes, sus libros de texto, sus prácticas profesionales y sus necesidades.

En segundo lugar, está la metodología planteada para la generación de no palabras. Consideramos que el uso de prefijos y raíces frecuentes garantiza que los distractores o no palabras compartirán las estructuras fonotácticas características de los términos especializados de un área, pero no serán capaces de engañar a los participantes que realmente hayan estado expuestos al léxico especializado.

Finalmente, está la sencillez del formato de la prueba de decisión léxica. Es un instrumento que arroja resultados cuantitativos de forma práctica y objetiva. Al

ser un simple listado con únicamente 200 ítems, puede ser elaborado muy rápidamente una vez que se cuenta con el corpus especializado. Esto a su vez permite generar diversas versiones de una misma prueba, variando el orden de presentación de los ítems, incluyendo o excluyendo algunas palabras, etc. La aplicación de la prueba también resulta muy sencilla, ya que las instrucciones son claras y sencillas, y la resolución de la prueba no requiere demasiado esfuerzo por parte de participante, ni toma más allá de unos cuantos minutos. Además es una prueba que puede calificarse muy rápidamente, ya sea de forma manual o automatizando el formato. Si bien, el formato no es original de esta investigación, su aplicación en L1 y su uso para la evaluación del vocabulario académico especializado sí lo son.

Resumiendo, los resultados obtenidos en esta investigación permiten suponer que el método para la selección de los términos y para la creación de no palabras permitieron elaborar un instrumento capaz de discriminar de una manera práctica, rápida y objetiva entre grupos que han sido expuestos a una terminología especializada. La confiabilidad de la prueba reside en los criterios establecidos para la selección de los ítems (Estructura IMRD, frecuencia de palabra, índice de dispersión, validación de expertos).

La aplicación de este tipo de instrumento permitiría comparar el conocimiento de vocabulario especializado de los estudiantes de diferentes programas en diversas instituciones educativas, ya que la metodología utilizada para la elaboración de estas pruebas podría extenderse a otras áreas del conocimiento como la Química, la Biología, o las ciencias sociales e incluso permitir una comparación de los diversos programas educativos dentro de una misma institución.

Además, la sencillez del formato de *checklist* permitiría desarrollar formatos de la prueba con un mayor grado de automatización (formato electrónico, revisión automática, combinación aleatoria de ítems, etc.) que podrían facilitar el tiempo de

evaluación. Éste es un tema fundamental en las áreas de evaluación, sobre todo en instituciones en las que debe evaluarse a grandes cantidades de alumnos en un tiempo breve y con una máxima optimización de recursos.

Es así como este trabajo presenta una humilde pero metodológicamente estricta contribución a la evaluación del vocabulario académico especializado. Como se mencionó anteriormente, este tipo de propuestas cuentan con una posible aplicación en los procesos de evaluación de la educación media superior, encaminada a la mejora de la calidad de la educación mediante evaluaciones válidas, confiables y pertinentes, que contribuyan a la toma de decisiones fundamentadas.



## BIBLIOGRAFÍA CITADA

Aitchison, J. (1987). *Words in the mind: An introduction to the mental lexicon*. Oxford: Basil Blackwell.

Anderson, R. C., y Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (ed.), *Comprehension and teaching: Research reviews*. Newark, DE: International Reading Association.

Anderson, R. C. y Freebody, P. (1983). Reading Comprehension and the assessment and acquisition of word knowledge. *Language Research* 2: 231-256.

Baker, M. (1988). Sub-technical vocabulary and the ESP teacher: An analysis of some rhetorical items in medical journal articles. *Reading in a Foreign Language*, 4(2): 91–105.

Balota, D. A y Chumbley, J.L (1994). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology. Perception and Performance*, 10: 340-357.

Bank of English Corpus (2007). HarperCollins Publishers and the University of Birmingham. <http://www.titania.bham.ac.uk/>

Barry, C. y Seymour, P. (1988). Lexical priming and sound-to-spelling contingency effects in nonword spelling. *Quarterly Journal of Experimental Psychology*, 40 (A): 5-40.

Beeckmans, R., Eyckmans, J., Janssens, V., Dufranne, M., & Van de Velde, H. (2001). Examining the Yes/No vocabulary test: some methodological issues in theory and practice. *Language Testing* 18 (3): 235 -274.

Bertram, R., Baayen, R.H. y Schreuder, R. (2000) Effects of family size for complex words, *Journal of Memory and Language* 42: 390-405.

Biber, D. y Finegan, E. (1994). *Sociolinguistic perspectives on register*. Oxford University Press. Oxford.

Bloor, M. (1998). Variations in the methods sections of research articles across disciplines: The case of fast and slow text. In P. Thompson (ed.), *Issues in EAP writing research and instruction*. Reading, UK: Centre for Applied Language Studies, The University of Reading.

Brewster, J. (1999). Exploring gendered talk: Some effects on interactional style: In A. Tosi y C. Leung (eds.) *Rethinking language education*. CILT. London.

Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.

Carter, R. (1998). *Vocabulary: Applied linguistic perspectives* (2<sup>nd</sup> Ed.) Routledge. London.

Cabré, M. T. (1998). *Terminology: Theory, methods and applications*. Philadelphia. PA, John Benjamins.

Cabré, M. Teresa. (1992). La formación de la terminología: uniformización y diversidad'. *Actas del Coloquio Iberoamericano sobre la enseñanza de la terminología*. Granada: Publicaciones de la Universidad de Granada.

Campion, M.E. y Elley, W. (1971). *An Academic Vocabulary List*. Wellington: NZCER.

CENEVAL. Centro Nacional de Evaluación para la Educación Superior, A.C.  
<http://www.ceneval.edu.mx>

Chapelle, C. A. (1998). Analysis of interaction sequences in computer-assisted language learning. (Research Issues). *TESOL Quarterly*, 32 (4),753-757.

Chen, Q., & Ge, G.C. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purpose*, 26: 502-514.

Chomsky, N. (1981). *Lectures on government and binding*. Foris. Dordrecht, the Netherlands.

Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC)  
Laboratorio de Lingüística Informática, Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia, Universidad Autónoma de Madrid  
<http://www.llf.uam.es/ESP/Corlec.html>

Cowan, J. R. (1974). Lexical and syntactic research for the design of EFL reading materials. *TESOL Quarterly*, 8(4): 389–399.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2): 213–238.

Coxhead, A., y Nation, P. (2001). The specialized vocabulary of English for academic purposes. En J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* Cambridge University Press: 252–267.

REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) [en línea]. Corpus de referencia del español actual. <http://www.rae.es>

Dell, G.S. (1988). The retrieval of phonological forms in production: Test of predictions from a connectionist model. *Journal of Memory and Language*, 27:124-142.

DeKeyser, R. (2007) The future of practice. En R. DeKeyser (ed.), *Practicing in a second language: Perspectives from applied linguistics and cognitive psychology*. New York: Cambridge University Press.

DeRocher, J.E. (1973). *The Counting of Words: A Review of the History, Techniques and Theory of Word Counts with Annotated Bibliography*. Syracuse University Research Corp. New York.

Diccionario Médico-Biológico, Histórico y Etimológico. URL: <http://www.dicciomed.es> [fecha acceso: 12.08.10 y 27.05.11]

Dörnyei, Z. y Kormos, J. (1998). Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in Second Language Acquisition*, 20: 349-385.

Dulay, H. C., y Burt, M. K. (1978). Some remarks on creativity in language acquisition. In W. C. Ritchie (ed.), *Second language acquisition research: Issues and implications*. Academic Press. New York.

Ellis, N.C. (1994). Implicit and explicit processes in language acquisition: An introduction. In N. Ellis (ed.) *Implicit and explicit learning in languages*. Academic Press. San Diego.

Ellis, N.C. y Schmidt, R. (1997). Morphology and longer-distance dependencies : Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition*, 19: 145-171.

Ellis, N.C. (2001). Memory for language. In P. Robinson (ed.) *Cognition and second language instruction*. Cambridge University Press. New York.

Ellis, N. C (2002). Frequency Effects in Language Processing. *Studies in Second Language Acquisition*. Cambridge University Press. 24:143-188.

Farrell, P. (1990) *A Lexical Analysis of the English of Electronics and a Study of Semi-technical Vocabulary*. Dublin: Trinity College.

Flowerdew, J. (1993). Concordancing as a tool in course design. *System*, 21(2): 231–244.

Forster, K. y Chambers, S. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 12: 627-635.

Francis, W. N. y Kucera, H. (1982). *Frequency Analysis of English Usage*. Houghton Mifflin Company, Boston.

Fries, C.C. y Traver, A. A. (1960). *English Word Lists*. Ann Arbor: George Wahr.

Frisch, S. F., Large, N. R., Zawaydeh, B. y Pisoni, D.B. (2001). Emergent phonotactic generalization in English and Arabic. In J. Bybee & P. Hopper (eds.) *Proceedings of the Symposium on Frequency Effects and Emergent Grammar*. Benjamins. Amsterdam.

Ghadessy, P. (1979). Frequency counts, word lists, material preparation: A new approach. *English Teaching Forum*, 17: 24–27.

Gierut, J., Morrisette, M., y Champion, A. (1999). Lexical constraints in phonological acquisition. *Journal of Child Language*, 26: 261-294.

Giora, R. (1997). Understanding figurative and literal language. *Cognitive Linguistics*, 7 (1): 183- 206.

Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning*, 24: 287-298.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21: 303-317.

Hernández, R., Fernández, C. y Baptista, P. (2003). *Metodología de la Investigación*. (Tercera Edición) McGraw Hill.

Holmes, R. (1997). Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes*, 16(4): 321–337.

Huibregtse, I., Admiraal, W., y Meara, P. (2002). Scores on a yes-no vocabulary test: Correction for guessing and response style. *Language testing* 19 (3): 227-245.

Hunston, S. (2002) *Corpora in applied Linguistics*. Cambridge University Press.

Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics*, 20(3): 341–367.

Hyland, K. (2002). Specificity revisited: How far should we go now? *English for Specific Purposes*, 21: 385–395.

Hyland, K. (2006). *English for Academic Purposes. An Advanced resource book*. London: Routledge.

Hyland, K., y Tse, P. (2007). Is there an “academic vocabulary”? *TESOL Quarterly*, 41(2): 235–253.

Jones, M. y Haywood, S. (2004) Facilitating the acquisition of formulaic sequences: An exploratory study in EAP context. En Schmitt, N. (ed.) *Formulaic sequences*. Amsterdam: John Benjamins.

Jones, S., Paradis, C., Murphy, L. M. y Wilners, C. (2007). Googling for opposites: a web-based study of antonym canonicity, *Corpora* 2 (2): 129–55.

Kessler, C., e Idar, I. (1979). Acquisition of English by a Vietnamese mother and child. *Working Papers on Bilingualism*, 18: 66-79.

Kilgarriff, A. y Grefenstette, G. (2003). Introduction to the special issue on the Web as Corpus’, *Computational Linguistics* 29 (3): 333–47.

Kirsner, K. (1994). Implicit processes in second language learning. In N. Ellis (ed.) *Implicit and explicit learning of languages*. Academic Press. San Diego, CA.

Kline, P. (1994) *The handbook of psychological testing*. London: Routledge.

Kuehn, P. (1996) *Assessment of academic literacy skills: Preparing minority and limited English proficient (LEP) students for post-secondary education*. Fresno, CA: California State University (ERIC Document Reproduction Service No. ED415498).

Kwon, F.Y (2005). The “Natural Order” of Morpheme Acquisition: A Historical Survey and Discussion of Three Putative Determinants. *Columbia University Working Papers in TESOL & Applied Linguistics*.

Lam, J. (2001). A study of semi-technical vocabulary in computer science texts, with special reference to ESP teaching and lexicography. (Research reports, Vol. 3). Hong Kong: Language Centre, Hong Kong University of Science and Technology.

Larsen-Freeman, D. E. (1976). An explanation for the morpheme acquisition order of second language learners. *Language Learning*, 26: 125-134.

Laufer, B. y Nation, P. (1999) A vocabulary size test of controlled productive ability. *Language Testing* 16 (1): 33 – 51.

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

Lynn, R. W. (1973). Preparing word lists a suggested method. *RELC Journal*, 4(1): 25–32.

MacDonald, M.C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9: 157-201.

MacDonald, M.C., Pearlmutter, N.L. y Seidenberg, M.S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101: 676-703.

Marslen-Wilson, W.D. (1990). Activation, competition and frequency in lexical access. In G. t. M. Altmann (ed.) *Cognitive models of speech processing*. MIT Press. Cambridge, MA.



Martínez, I., Beck, S.C., y Panza, C.B. (2009) Academic vocabulary in agriculture research articles: A corpus-based study, *English for Specific Purposes*, 28(3):183-198.

McClelland, J.L. y Franz, S.K. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18: 1-86.

McEnery, T. y Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice* Cambridge Textbooks in Linguistics. CUP.

Meara, P. (1996). The dimensions of lexical competence, en Brown, G., Malmkjær, K. y Williams, J. (eds.), *Performance and competence in second language acquisition*. Cambridge University Press.

Meara, P. y Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing* 4 (2): 142-154.

Meara, P. y Wolter, B. (2004). Beyond vocabulary depth. *Angles on the English Speaking World*, 4: 85-96.

Mochida, A. y Harrington, M. (2006) The Yes/no test as a measure of receptive vocabulary knowledge. *Language Testing* 23 (1): 73-98.

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, 25(2): 235–256.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston: Heinle & Heinle.

Nation, P. (1993). Measuring readiness for simplified material: a test of the first 1,000 words of English. En M.L. Tickoo (ed.) *Simplification: Theory and Application*. RELC Anthology Series No. 31, 193-203.

Nation, P. y Waring, R. (1997). Vocabulary size, text coverage and word lists en Schmitt, Norbert y Michael McCarthy (eds). Vocabulary description, acquisition and pedagogy en Cambridge Language Teaching Library: CUP

Nation, P. (2007) Fundamental issues in modeling and assessing vocabulary knowledge. En Daller, Helmut; Milton, James; and Treffers-Daller, Jeanine (eds). Modelling and Assessing Vocabulary Knowledge. Cambridge University Press.

Nippold, M.A. y Rudzinski, M. (1993). Familiarity and transparency in idiom explanation : A developmental study of children and adolescent. Journal of Speech and Hearing Research, 36: 728-737.

Oakes, M.P. (1998). Statistics for Corpus Linguistics. Edinburgh Textbooks in Empirical Linguistics. Edinburgh University Press.

Oetting, J.B. y Rice, M.L. (1993). Plural acquisition in children with specific language impairment. Journal of Speech and Hearing Research, 36: 1236-1248.

Paquot, M. (2007). Towards a productively-oriented academic word list. In J. Walinski, K. Kredens, & S. Gozdz-Roszkowski (eds.), Corpora and ICT in language studies. PALC 2005. Lodz studies in language. Frankfurt am Main: Peter Lang.

Pawley, A. y Syder, F.H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J.C. Richards & R.W. Schmitd (eds.) Language and communication. Longman. London.

Prasada, S., Pinker, S. y Snyder, W. (1990). Some evidence that irregular forms are retrieved from memory but regular forms are rule-governed. Paper presented at the 31<sup>st</sup> meeting of the Psychonomic Society, New Orleans, LA.

Praninskas, J. (1972). American University Word list. London: Longman.

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review*, 56 (2): 282-307.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*. 52 (3):513-536.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge, *Language Testing* 10: 355–71.

Read, J. (2005). *Assessing Vocabulary*. Cambridge: CUP

Read, J., y Chapelle, C (2001). A framework for second language vocabulary assessment. *Language Testing*, 18: 1-32.

Read, J. (2004) Plumbing the depths: How should the construct of vocabulary knowledge be defined? En Bogards, P y B. Laufer. (eds.) *Vocabulary in a Second Language: selection, acquisition and testing*. Amsterdam, John Benjamins.

Richards, J.C., Platt, y. Weber, H. (1985) *Longman Dictionary of Applied Linguistics*, London:Longman.

Samraj, B. (2002). Disciplinary variation in abstracts: The case of wildlife behaviour and conservation biology. En J. Flowerdew (ed.), *Academic discourse*. London: Longman.

Seidenberg, M.S. y Bruck, M. (1990). Consistency effects in the generation of past tense morphology. Paper presented at the 31<sup>st</sup> meeting of the Psychonomic Society, New Orleans. LA.

Scott, M. (2009). WordSmith Tools. Lexical Analysis Software. Liverpool.

Schmid, H. (1994): Probabilistic Part-of-Speech Tagging Using Decision Trees. Proceedings of International Conference on New Methods in Language Processing, Manchester, UK.

Schmitt, N. (2000). Vocabulary in Language teaching. Cambridge University Press. New York.

Sims, V.M. (1929) The reliability and validity of four types of vocabulary test. Journal of Educational Research 20: 91-96.

Sinclair, J. (1991). Corpus, concordance, collocation. Oxford University Press. Oxford.

Sinclair, J. (2005). Corpus and text – Basic principles, en M. Wynne (ed.), Developing linguistic corpora: A guide to good practice. Oxford: Oxbow Books: 1–16.

Tannen, D. (1987). Repetition and variation as spoken formulaicity in conversation. Language, 63: 574-605.

Tilley, H.C. (1936) A technique for determining the relative difficulty of word meanings among elementary school children. Journal of Experimental Education 5: 61-64.

Trimble, L. (1985). *English for science and technology: a discourse approach*. Cambridge: Cambridge University Press.

Trueswell, J.C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, 35: 566-585.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics* 22: 217–234.

Vives, G. & Meara, P. (1994). *The Development of a Measure of Lexical Organization: the Association Vocabulary Test*. University of Wales, Swansea: Unpublished PhD thesis.

Wang, J., Liang, S. y Ge, G. (2008) Establishment of a Medical Academic Word List. *English for Specific Purposes* 27: 442–458.

Wesche, M. y Paribakht, T.S. (1996). Assessing second language vocabulary knowledge: depth versus breadth, *Canadian Modern Language Review*, 53: 13–40.

West, M. (1953) *A General Service List of English Words*, Longman, London.

Wong-Fillmore, L. (1976). *The second time around*. Stanford University Report. Stanford, CA.

Wood, A. (2001). International scientific English: The language of research scientists around the world. En J. Flowerdew & M. Peacock (eds.), *Research perspectives on English for academic purposes*. Cambridge: Cambridge University Press.

Yang, H. (1986). A new technique for identifying scientific/technical terms and describing scientific texts. *Literary and Linguistic Computing*, 1(2): 93–103.

Xue, G., & Nation, P. (1984). A university word list. *Language Learning and Communication*, 3(2): 215–219.

Zimmerman, J., Broder, P. K., Shaughnessy, J. J., y Underwood, B. J. (1977). A recognition test of vocabulary using signal-detection measures, and some correlates of word and nonword recognition. *Intelligence* 1: 5-31.

## 10. APÉNDICES

### 10.1 Prueba de Nutrición



UNIVERSIDAD ATÓNOMA DE QUERÉTARO

AMG2010a

Nombre: \_\_\_\_\_ Fecha: \_\_\_\_\_

Semestre: \_\_\_\_\_ Edad: \_\_\_\_\_

#### Instrucciones:

**Si conoces el significado de la palabra, rellena el recuadro de la izquierda.**

**Si no la conoces o no estás seguro de conocer su significado, deja el recuadro en blanco.**

Ejemplo:	<input checked="" type="checkbox"/>	antibiótico	<input type="checkbox"/>	quimiotripsina			
<input type="checkbox"/>	1 osteogénesis	<input type="checkbox"/>	2 hipoclorito	<input type="checkbox"/>	3 biodisponibilidad	<input type="checkbox"/>	4 glucolípidos
<input type="checkbox"/>	5 hipoalimentación	<input type="checkbox"/>	6 sulfonamidas	<input type="checkbox"/>	7 tromboglobinómetro	<input type="checkbox"/>	8 neuromodulador
<input type="checkbox"/>	9 hematopoyesis	<input type="checkbox"/>	10 fitocubital	<input type="checkbox"/>	11 hipotonía	<input type="checkbox"/>	12 hiperkalemia
<input type="checkbox"/>	13 microcitosis	<input type="checkbox"/>	14 gliceraldehido	<input type="checkbox"/>	15 fucocorticoides	<input type="checkbox"/>	16 hiperbetalipoproteinemia
<input type="checkbox"/>	17 osmótica	<input type="checkbox"/>	18 normocalórica	<input type="checkbox"/>	19 hepatotoxicidad	<input type="checkbox"/>	20 hidroxicoalamina
<input type="checkbox"/>	21 trombogénicos	<input type="checkbox"/>	22 galactosuero	<input type="checkbox"/>	23 glucosidasa	<input type="checkbox"/>	24 tetrafructosa
<input type="checkbox"/>	25 isotónicas	<input type="checkbox"/>	26 aterogénica	<input type="checkbox"/>	27 ortocefalia	<input type="checkbox"/>	28 hipoglucemia
<input type="checkbox"/>	29 microcítica	<input type="checkbox"/>	30 aminoglobulina	<input type="checkbox"/>	31 hidroglobulina	<input type="checkbox"/>	32 inmuoestimulación
<input type="checkbox"/>	33 macrobióticos	<input type="checkbox"/>	34 aterosclerosis	<input type="checkbox"/>	35 lipídica	<input type="checkbox"/>	36 multifructosa
<input type="checkbox"/>	37 oligoenzimático	<input type="checkbox"/>	38 polimorfismo	<input type="checkbox"/>	39 homocistinuria	<input type="checkbox"/>	40 osteoenzimático
<input type="checkbox"/>	41 hipotiroidismo	<input type="checkbox"/>	42 lactato	<input type="checkbox"/>	43 hipoactividad	<input type="checkbox"/>	44 citobiótico

- |                                              |                                               |                                                  |                                                    |
|----------------------------------------------|-----------------------------------------------|--------------------------------------------------|----------------------------------------------------|
| <input type="checkbox"/> 45 monocitos        | <input type="checkbox"/> 46 hematológica      | <input type="checkbox"/> 47 oligoglicérido       | <input type="checkbox"/> 48 galactosa              |
| <input type="checkbox"/> 49 hemólisis        | <input type="checkbox"/> 50 gastrocólico      | <input type="checkbox"/> 51 hipoinsulinemia      | <input type="checkbox"/> 52 hemodilución           |
| <input type="checkbox"/> 53 metabolito       | <input type="checkbox"/> 54 glucosuria        | <input type="checkbox"/> 55 ateromatosa          | <input type="checkbox"/> 56 glucosadas             |
| <input type="checkbox"/> 57 hipoxia          | <input type="checkbox"/> 58 termogénesis      | <input type="checkbox"/> 59 microalbuminuria     | <input type="checkbox"/> 60 hiperhomocisteinemia   |
| <input type="checkbox"/> 61 hiperglucídica   | <input type="checkbox"/> 62 tiroduodenal      | <input type="checkbox"/> 63 heterocíclico        | <input type="checkbox"/> 64 bioimpedancia          |
| <input type="checkbox"/> 65 ateroma          | <input type="checkbox"/> 66 polipéptido       | <input type="checkbox"/> 67 osteomalacia         | <input type="checkbox"/> 68 hiperbilirrubinemia    |
| <input type="checkbox"/> 69 micronutriente   | <input type="checkbox"/> 70 macrófago         | <input type="checkbox"/> 71 ateroesclerótica     | <input type="checkbox"/> 72 macrocítica            |
| <input type="checkbox"/> 73 hipolípida       | <input type="checkbox"/> 74 heterocigoto      | <input type="checkbox"/> 75 glucoproteína        | <input type="checkbox"/> 76 fosfácido              |
| <input type="checkbox"/> 77 normopeso        | <input type="checkbox"/> 78 lactolípidos      | <input type="checkbox"/> 79 microtúbulos         | <input type="checkbox"/> 80 inmunomodulación       |
| <input type="checkbox"/> 81 hemaglutininas   | <input type="checkbox"/> 82 neurotoxicidad    | <input type="checkbox"/> 83 fitoquímicos         | <input type="checkbox"/> 84 hipocalcemia           |
| <input type="checkbox"/> 85 hiperfagia       | <input type="checkbox"/> 86 isovitamínico     | <input type="checkbox"/> 87 glucogenogénesis     | <input type="checkbox"/> 88 homoclorito            |
| <input type="checkbox"/> 89 aminofátido      | <input type="checkbox"/> 90 citotóxica        | <input type="checkbox"/> 91 fibroglicemia        | <input type="checkbox"/> 92 microflora             |
| <input type="checkbox"/> 93 fosfaditilcolina | <input type="checkbox"/> 94 cloropéptido      | <input type="checkbox"/> 95 fosfátidos           | <input type="checkbox"/> 96 patogénesis            |
| <input type="checkbox"/> 97 hepatoxilina     | <input type="checkbox"/> 98 inmunosilación    | <input type="checkbox"/> 99 gluconeogénesis      | <input type="checkbox"/> 100 hipovitaminosis       |
| <input type="checkbox"/> 101 macrófaga       | <input type="checkbox"/> 102 lipofílica       | <input type="checkbox"/> 103 lactoferrina        | <input type="checkbox"/> 104 lipogénicas           |
| <input type="checkbox"/> 105 clorosacárido   | <input type="checkbox"/> 106 leucopenia       | <input type="checkbox"/> 107 hipercolesterolemia | <input type="checkbox"/> 108 insulino pigmentación |
| <input type="checkbox"/> 109 inmunonutrición | <input type="checkbox"/> 110 tiroiditis       | <input type="checkbox"/> 111 aminopeptidasa      | <input type="checkbox"/> 112 galactosemia          |
| <input type="checkbox"/> 113 polimérica      | <input type="checkbox"/> 114 normoglucémicas  | <input type="checkbox"/> 115 hematocrito         | <input type="checkbox"/> 116 lactosacárido         |
| <input type="checkbox"/> 117 lipotrombosis   | <input type="checkbox"/> 118 glucocorticoides | <input type="checkbox"/> 119 microvelocidades    | <input type="checkbox"/> 120 hemicelulosa          |
| <input type="checkbox"/> 121 isoflavonas     | <input type="checkbox"/> 122 fitosteroles     | <input type="checkbox"/> 123 hidroelectrolítica  | <input type="checkbox"/> 124 normocolesterolémicos |
| <input type="checkbox"/> 125 glucosilada     | <input type="checkbox"/> 126 fibrina          | <input type="checkbox"/> 127 lipoprotéica        | <input type="checkbox"/> 128 glicerofosfato        |
| <input type="checkbox"/> 129 trombocrito     | <input type="checkbox"/> 130 lipogenasa       | <input type="checkbox"/> 131 hepatoporosis       | <input type="checkbox"/> 132 glucémico             |
| <input type="checkbox"/> 133 isoflavononas   | <input type="checkbox"/> 134 hiperproteica    | <input type="checkbox"/> 135 oligómeros          | <input type="checkbox"/> 136 gastroduodenal        |



- |                                               |                                              |                                                 |                                                      |
|-----------------------------------------------|----------------------------------------------|-------------------------------------------------|------------------------------------------------------|
| <input type="checkbox"/> 137 fibrinolisis     | <input type="checkbox"/> 138 hepatomegalia   | <input type="checkbox"/> 139 isocalórica        | <input type="checkbox"/> 140 gluconato               |
| <input type="checkbox"/> 141 isoglucemia      | <input type="checkbox"/> 142 fibrinógeno     | <input type="checkbox"/> 143 hemoglicemia       | <input type="checkbox"/> 144 hiperlipidemia          |
| <input type="checkbox"/> 145 inmunoglobulina  | <input type="checkbox"/> 146 hidroquinona    | <input type="checkbox"/> 147 osmorregulación    | <input type="checkbox"/> 148 hepatopatías            |
| <input type="checkbox"/> 149 hiperinsulinemia | <input type="checkbox"/> 150 quimioisotópica | <input type="checkbox"/> 151 glucoesclerólisis  | <input type="checkbox"/> 152 quimioanálisis          |
| <input type="checkbox"/> 153 hiposódico       | <input type="checkbox"/> 154 normocítica     | <input type="checkbox"/> 155 polimorfonucleares | <input type="checkbox"/> 156 hemínico                |
| <input type="checkbox"/> 157 trombófagos      | <input type="checkbox"/> 158 hemoferrina     | <input type="checkbox"/> 159 heparina           | <input type="checkbox"/> 160 oligoelemento           |
| <input type="checkbox"/> 161 glucogenólisis   | <input type="checkbox"/> 162 hepatocitos     | <input type="checkbox"/> 163 gastrina           | <input type="checkbox"/> 164 glicematosi             |
| <input type="checkbox"/> 165 osteoclástica    | <input type="checkbox"/> 166 hipokalemia     | <input type="checkbox"/> 167 hidropesía         | <input type="checkbox"/> 168 hipotensión             |
| <input type="checkbox"/> 169 lactoglobulina   | <input type="checkbox"/> 170 osmolaridad     | <input type="checkbox"/> 171 citocromo          | <input type="checkbox"/> 172 hemoconcentración       |
| <input type="checkbox"/> 173 trombogénesis    | <input type="checkbox"/> 174 aterosclerosis  | <input type="checkbox"/> 175 intragástricamente | <input type="checkbox"/> 176 osteoblástica           |
| <input type="checkbox"/> 177 fibrinogénesis   | <input type="checkbox"/> 178 polifenoles     | <input type="checkbox"/> 179 hipoalbuminemia    | <input type="checkbox"/> 180 leucotrienos            |
| <input type="checkbox"/> 181 polipeptídica    | <input type="checkbox"/> 182 glucostomía     | <input type="checkbox"/> 183 isomerasa          | <input type="checkbox"/> 184 glucósido               |
| <input type="checkbox"/> 185 fibroenzimático  | <input type="checkbox"/> 186 hiperamonemia   | <input type="checkbox"/> 187 microbiota         | <input type="checkbox"/> 188 fucogenos               |
| <input type="checkbox"/> 189 ateromatosis     | <input type="checkbox"/> 190 hiponatremia    | <input type="checkbox"/> 191 posprandial        | <input type="checkbox"/> 192 hipercolesterolemiantes |
| <input type="checkbox"/> 193 biomarcador      | <input type="checkbox"/> 194 glucemia        | <input type="checkbox"/> 195 hiperaminoaciduria | <input type="checkbox"/> 196 neurodegeneración       |
| <input type="checkbox"/> 197 aterogenicidad   | <input type="checkbox"/> 198 microbiana      | <input type="checkbox"/> 199 poliuria           | <input type="checkbox"/> 200 tromboembolia           |

## 10.2 Prueba de Odontología



UNIVERSIDAD ATÓNOMA DE QUERÉTARO

AMG2010a

Nombre: \_\_\_\_\_ Fecha: \_\_\_\_\_

Semestre: \_\_\_\_\_ Edad: \_\_\_\_\_

### Instrucciones:

Si conoces el significado de la palabra, rellena el recuadro de la izquierda.

**Si no la conoces o no estás seguro de conocer su significado, deja el recuadro en blanco.**

Ejemplo:



gingivitis



melanoacantoma

- |                                                |                                                |                                             |                                             |
|------------------------------------------------|------------------------------------------------|---------------------------------------------|---------------------------------------------|
| 1 <input type="checkbox"/> anodoncia           | 2 <input type="checkbox"/> inmunoplásico       | 3 <input type="checkbox"/> pericoronaritis  | 4 <input type="checkbox"/> odontocalcina    |
| 5 <input type="checkbox"/> orogaringe          | 6 <input type="checkbox"/> parotídea           | 7 <input type="checkbox"/> osteogénico      | 8 <input type="checkbox"/> basaloide        |
| 9 <input type="checkbox"/> amelocementaria     | 10 <input type="checkbox"/> carcinogénesis     | 11 <input type="checkbox"/> interproximales | 12 <input type="checkbox"/> bicortical      |
| 13 <input type="checkbox"/> sialofítica        | 14 <input type="checkbox"/> cariogénica        | 15 <input type="checkbox"/> odontoma        | 16 <input type="checkbox"/> tartrectomía    |
| 17 <input type="checkbox"/> dentiplásico       | 18 <input type="checkbox"/> idiopática         | 19 <input type="checkbox"/> estomatognático | 20 <input type="checkbox"/> osteoloide      |
| 21 <input type="checkbox"/> inmunohistoquímica | 22 <input type="checkbox"/> pleomórfico        | 23 <input type="checkbox"/> periapical      | 24 <input type="checkbox"/> perioectomia    |
| 25 <input type="checkbox"/> criocirugía        | 26 <input type="checkbox"/> fibrinólisis       | 27 <input type="checkbox"/> tisurales       | 28 <input type="checkbox"/> exóstosis       |
| 29 <input type="checkbox"/> mucogioma          | 30 <input type="checkbox"/> trabeculado        | 31 <input type="checkbox"/> anamnesis       | 32 <input type="checkbox"/> trabepatogenia  |
| 33 <input type="checkbox"/> bacteriemia        | 34 <input type="checkbox"/> gingivoestomatosis | 35 <input type="checkbox"/> agenesia        | 36 <input type="checkbox"/> osteoblástica   |
| 37 <input type="checkbox"/> escamocelular      | 38 <input type="checkbox"/> estomatocalcina    | 39 <input type="checkbox"/> queratinocitos  | 40 <input type="checkbox"/> prostaglandinas |
| 41 <input type="checkbox"/> gingivectomía      | 42 <input type="checkbox"/> eritema            | 43 <input type="checkbox"/> ortodiasis      | 44 <input type="checkbox"/> fibrodentinoma  |
| 45 <input type="checkbox"/> epidermoide        | 46 <input type="checkbox"/> trabeloide         | 47 <input type="checkbox"/> osteoclastos    | 48 <input type="checkbox"/> meningofacial   |
| 49 <input type="checkbox"/> coronoidectomia    | 50 <input type="checkbox"/> carcinomórfico     | 51 <input type="checkbox"/> linfangioma     | 52 <input type="checkbox"/> eosinófilos     |

- 53  etiopatogénico
- 54  cariadisis
- 55  apical
- 56  colutorios
- 57  protésica
- 58  treponema
- 59  sialosis
- 60  túbulo
- 61  citofítica
- 62  sialosinusales
- 63  tinción
- 64  hemostasia
- 65  corticotomía
- 66  osteorradionecrosis
- 67  cariofítica
- 68  liquen
- 69  eritroplasia
- 70  tisular
- 71  hidroxiapatita
- 72  tetraciclina
- 73  catecolaminas
- 74  temporomandibular
- 75  apósito
- 76  tumefacción
- 77  endolitis
- 78  ortognático
- 79  queratoquiste
- 80  recidiva
- 81  parestesia
- 82  tuberosidad
- 83  gingivimatosas
- 84  citocromo
- 85  osteítis
- 86  colgajo
- 87  sialoadenitis
- 88  periomórfico
- 89  exudado
- 90  exéresis
- 91  eritrocromía
- 92  histopatológico
- 93  mucocutáneas
- 94  mucogingival
- 95  maxiloma
- 96  mucomatoso
- 97  etiosomático
- 98  adenopatías
- 99  oclusal
- 100  clindamicina
- 101  carúncula
- 102  frenectomía
- 103  sialometría
- 104  peroxidasa
- 105  histiocitosis
- 106  mecanotransducción
- 107  periamigdalinas
- 108  adenocarcinoma
- 109  radicular
- 110  plasminógeno
- 111  perioadenosis
- 112  periodontopatógenos
- 113  mesiobucal
- 114  articaína
- 115  mesiodistal
- 116  edentodiasis
- 117  aloplásticos
- 118  gingivocalcimatosis
- 119  prevotella
- 120  dentinarias
- 121  piógeno
- 122  candidiasis
- 123  cervicofacial
- 124  estomatofítico
- 125  oligodoncia
- 126  pogonión
- 127  citoquinas
- 128  dentoalveolares
- 129  isquemia
- 130  tetrahidrofurano
- 131  protombina
- 132  hemangioma
- 133  metastásica
- 134  linfadenocalcina
- 135  micrognatismo
- 136  queilitis
- 137  alveolitis
- 138  estomatitis
- 139  cavitarias
- 140  gingival
- 141  edentolismo
- 142  ortopantomografía
- 143  ameloblastoma
- 144  implantoplástico
- 145  plexiforme
- 146  anatomopatológico
- 147  doxiciclina
- 148  endotelio

- |                                                |                                                 |                                               |                                               |
|------------------------------------------------|-------------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| 149 <input type="checkbox"/> tisosinusales     | 150 <input type="checkbox"/> dentícola          | 151 <input type="checkbox"/> orotraqueal      | 152 <input type="checkbox"/> etiológico       |
| 153 <input type="checkbox"/> mioepiteliales    | 154 <input type="checkbox"/> bruxismo           | 155 <input type="checkbox"/> antegonial       | 156 <input type="checkbox"/> maloclusion      |
| 157 <input type="checkbox"/> benzodicepinas    | 158 <input type="checkbox"/> patognomónico      | 159 <input type="checkbox"/> endodónticos     | 160 <input type="checkbox"/> nasoendoscopia   |
| 161 <input type="checkbox"/> mucoepidermoide   | 162 <input type="checkbox"/> supragingival      | 163 <input type="checkbox"/> bucosinusales    | 164 <input type="checkbox"/> rinomanometría   |
| 165 <input type="checkbox"/> dehiscencia       | 166 <input type="checkbox"/> pediculada         | 167 <input type="checkbox"/> dentina          | 168 <input type="checkbox"/> pseudomonas      |
| 169 <input type="checkbox"/> dentoplasia       | 170 <input type="checkbox"/> citomegalovirus    | 171 <input type="checkbox"/> traqueocalcinoma | 172 <input type="checkbox"/> equimosis        |
| 173 <input type="checkbox"/> periagingival     | 174 <input type="checkbox"/> leptoprosopo       | 175 <input type="checkbox"/> glositis         | 176 <input type="checkbox"/> herpetiforme     |
| 177 <input type="checkbox"/> sialosomático     | 178 <input type="checkbox"/> inmunoflouescencia | 179 <input type="checkbox"/> clorhexidina     | 180 <input type="checkbox"/> septal           |
| 181 <input type="checkbox"/> crevicular        | 182 <input type="checkbox"/> etiopafítica       | 183 <input type="checkbox"/> exodoncia        | 184 <input type="checkbox"/> clavulánico      |
| 185 <input type="checkbox"/> osteointegración  | 186 <input type="checkbox"/> prostodoncia       | 187 <input type="checkbox"/> maxilofítica     | 188 <input type="checkbox"/> bucoadenosis     |
| 189 <input type="checkbox"/> condilar          | 190 <input type="checkbox"/> adenoidectomía     | 191 <input type="checkbox"/> exofítica        | 192 <input type="checkbox"/> linfadenolingual |
| 193 <input type="checkbox"/> osteoprogenitoras | 194 <input type="checkbox"/> estroma            | 195 <input type="checkbox"/> odontalgias      | 196 <input type="checkbox"/> adenocalcina     |
| 197 <input type="checkbox"/> periodontitis     | 198 <input type="checkbox"/> eritrocélular      | 199 <input type="checkbox"/> odontoblástica   | 200 <input type="checkbox"/> unilocular       |

### 10.3 Selección de Palabras Reales del Corpus de Nutrición por Frecuencia de Palabra

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
LÍPIDO	386	GLICÉMICA(O)	90
GLUCOSA	357	OSTEOPOROSIS	90
FIBRA	328	BIOQUÍMICO	86
METABÓLICA	325	LACTOBACILOS	83
PATOLOGÍA	312	POSTPRANDIAL	78
METABOLISMO	259	BIOQUÍMICA	77
LACTANTE	243	HIPONATREMIA	75
MICRONUTRIENTE	230	OLIGOSACÁRIDO	75
AMINOÁCIDO(S)	222	GLUCÓGENO	75
POLIFENOLES	221	NEURONAL	74
HEMOGLOBINA	186	HIPERCOLESTEROLEMIA	73
HIPERTENSIÓN	183	HOMOGENEO	73
BIODISPONIBILIDAD	161	MICROORGANISMO	73
HEPÁTICA(S)	148	CITOCINA	72
LIPÍDICA	140	HEMATOLÓGICA	70
HOMOCISTEINA	137	METABOLITO	70
LIPOPROTEINA	135	FITOESTEROLES	68
POLIAMINA	127	ISÓMERO	68
MACRONUTRIENTE	126	FOSFATO	67
GASTROINTESTINAL	116	LÁCTICO	67
GLICEMIA	115	PATOLÓGICO	67
BIOLÓGICO	108	BACTERIANA	67
GÁSTRICA (O)	107	LACTOSA	63
LÁCTEA	100	NEUROLÓGICO	63

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
MONOINSATURADA	62	GLICEROL	35
HEMÍNICO	61	MICROBIOLÓGICO	35
INMUNOLÓGICO	60	MICROBIOTA	34
INTRAUTERINA	55	LEUCOCITO	32
HIDRÓLISIS	53	MICROBIANA	32
SULFATO	52	HIPERGLICEMIA	29
ISOMERÍA	52	LIPEMIA	29
OSTEOPENIA	51	ATEROESCLEROSIS	27
INTRACELULAR	51	HIPOLIPEMIANTE	26
HOMOGENEIDAD	51	BIOPRESERVACIÓN	22
TIROIDEA	51	FOSFATASA	22
HEMATOCRITO	50	HIPERACTIVIDAD	20
ULTRARESISTENCIA	50	TERMOGÉNESIS	20
PATÓGENA	48	BACTERIOCINA	19
HIDROGENACIÓN	48	BIOSÍNTESIS	19
CLOROGÉNICO	47	FITOESTANOLES	19
HIDRÓGENO	47	FITOQUÍMICOS	19
HIDROXILO	46	GLUCOGENOSIS	19
ATEROGÉNICA	44	NEUROTRASMISOR	19
CITOLOGÍA	44	HETEROGÉNEO	19
LEUCINA	43	HIPOGLICEMIA	19
MACRÓFAGO	40	TIROIDES	19
POSTNATAL	40	HIPERINSULINEMIA	18
ATEROSCLEROSIS	37	HIPOGLUCEMIA	18

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
POLÍMERO	18	GLICOSAMINOGLICANOS	12
CLORURO	17	HIPERTROFIA	12
FIBRINOLÍTICA	17	LEUCOTRIENOS	12
HIPOCALÓRICA	17	HIPOTALAMO	12
ISÓTOPO	17	HIPOTAURINA	12
ATEROSCLERÓTICA	16	POLISACÁRIDO	12
GLICOSILACIÓN	16	ATEROGENESIS	11
MICROFLORA	16	CITOMETRÍA	11
BIOPSIA	15	HÉMICO	11
HIPOVITAMINOSIS	15	HIPOALBUMINEMIA	11
ISOLEUCINA	15	MONOCITOS	11
GALACTOSA	14	BIOMECÁNICA	10
GLICÓSIDO	14	HEPATOMEGALIA	10
POSTMENOPAUSIA	14	NEUROGÉNESIS	10
POSTMENOPÁUSICA	14	LIPOLÍTICAS	10
LIPOSOLUBLES	14	TROMBOSIS	10
HIPOCOLESTEROLEMICO	14	ISOTÓPICA	10
MACROSÓMICO	14	MACROMOLÉCULA	10
PATOGÉNESIS	14	MACROSOMÍA	10
BIOPSIOSOCIAL	13	ORTOREXIA	10
CITOCROMO	13	OSMÓTICA	10
LIPÓLISIS	13	POLIDEXTROSA	10
TIROSINA	13	AMINO	9
ATEROMA	12	BIOACTIVO	9

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
FIBROSIS	9	BIOÉTICA	7
MICROESTRUCTURA	9	FITHEMAGLUTENINAS	7
HIDROXICINÁMICOS	9	GLUCOCORTICOIDES	7
NORMOPESO	9	GLUCONEOGÉNESIS	7
MULTIVARIADA	9	HEMATOENCEFÁLICA	7
TROMBOXANO	9	NEUROTRANSMISIÓN	7
INTRAVENOSO	9	HIDROLASA	7
ISOFLAVONAS	9	HIDROSOLUBLE	7
MONOGLICÉRIDO	9	POSTOPERATORIA	7
OLIGÓMEROS	9	LIPOOXIGENASA	7
POLIMERIZACIÓN	9	LACTOALBÚMINA	7
BIOMÉDICA	8	MONOCLONAL	7
FITOESTRÓGENOS	8	OLIGOELEMENTO	7
FIBRINÓGENO	8	POLICLÍNICO	7
HEMODILUCIÓN	8	AMINAS	6
MICRONUTRIMENTOS	8	ATEROGENICIDAD	6
NEURODEGENERATIVA	8	BACTERICIDA	6
NEUROPATÍA	8	BIOLOGÍA	6
MULTICÉNTRICA	8	BIOMARCADOR	6
OSTEOCALCINA	8	CITOQUINA	6
OSTEOPATÍA	8	FUCOSA	6
HIPOXIA	8	GLUCOLÍTICAS	6
MONÓMEROS	8	HEMÓLISIS	6
POLIFENÓLICA	8	HEPARINA	6



<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
HIDROLÍTICA	6	BACTERIOCINOGÉNICA	5
HIPERCALÓRICA	6	BACTERIOCINOGENICIDAD	5
NORMOTENSA	6	BACTERIÓFAGOS	5
ULTRAPURA	6	BACTERIOLISINAS	5
OSTEOCITOS	6	BACTERIOLÓGICO	5
HIPOZINCEMIA	6	BACTERIOSTÁTICO	5
HIPOTIROIDISMO	6	BACTEROIDES	5
ISOFORMA	6	BIOACCESIBILIDAD	5
LACTASA	6	BIOACTIVADORES	5
LACTATO	6	BIOACUMULATIVO	5
MONOACILPALMITINA	6	BIOANÁLISIS	5
POLIMORFISMO	6	BIOANALISTAS	5
POLINSATURADOS	6	BIOBALANCE	5
RADIOINMUNOENSAYO	6	BIOECOLÓGICOS	5
AMINOACÍDICA	5	BIOENERGÉTICA	5
AMINOACIDOPATÍAS	5	BIOENSAYO	5
AMINOGUANIDINA	5	BIOEQUIVALENCIA	5
AMINOPEPTIDASA	5	BIOESTADÍSTICO	5
AMINOPROPILO	5	BIOFÍSICAMENTE	5
AMINOTRANSFERASAS	5	BIOFLAVONOIDES	5
ATEROESCLERERÓTICA	5	BIÓGENAS	5
ATEROMATOSIS	5	BIOHIDROGENACIÓN	5
ATEROSCLEROSA	5	BIOIMPEDANCIÓMETRO	5
ATEROTROMBOSIS	5	BIÓLOGO	5

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
BIOMODELO	5	CLORAMINAS	5
BIOMOLÉCULAS	5	CLÓRICO	5
BIOPOTENCIA	5	CLORINADA	5
BIOPRESERVANTE	5	CLORO	5
BIOSEGURIDAD	5	CLOROACÉTICO	5
BIOSINTÉTICA	5	CLOROFILA	5
BIOSINTETIZAR	5	CLOROFORMO	5
BIOSOCIAL	5	CLOROPLASTOS	5
BIOTECNOLOGÍA	5	CLOROQUINA	5
BIOTERIO	5	CLOROSIS	5
BIÓTICOS	5	FIBRINA	5
BIOTRANSFORMACIÓN	5	FIBRINOGENÉESIS	5
CITOEESQUELETO	5	FIBRINOLISIS	5
CITOKINA	5	FIBRINOLÍTICOS	5
CITOLÍTICA	5	FIBROBLASTOS	5
CITOLÓGICA	5	FIBROELÁSTICA	5
CITOMEGALOVIRUS	5	FIBROMUSCULAR	5
CITOPLASMÁTICA	5	FIBRONECTINA	5
CITOPLÁSMICA	5	FIBROSAS	5
CITOPROTECCIÓN	5	FITOCROMOBILINA	5
CITOS	5	FITHEMA	5
CITOSOL	5	FITOMEJORAMIENTO	5
CITOTÓXICA	5	FITOSANITARIAS	5
CITOTOXICIDAD	5	FITOSTEROLES	5

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
FOSFATIDILSERINA	5	GASTROSTOMÍA	5
FOSFÁTIDOS	5	GASTROYEYUNOANASTOMOSIS	5
FUCOSIDASAS	5	GASTROYEYUNOSTOMIA	5
FUCOSIDOSIS	5	GLICÉNICA	5
FUCOSILACIÓN	5	GLICERALDEHIDO	5
FUCOSILADA	5	GLICÉRIDOS	5
FUCOSILASA	5	GLICÉRMICOS	5
FUCOSILTRANSFERASA	5	GLICEROFOSFATO	5
GALACTOCEREBRÓSIDOS	5	GLICEROFOSFOCOLINA	5
GALACTOLÍPIDOS	5	GLICOGENESIS	5
GALACTOOLIGOSACÁRIDOS	5	GLICÓGENO	5
GALACTOSEMIA	5	GLICOGENÓLISIS	5
GALACTOSIALIDOSIS	5	GLICOLÍPIDOS	5
GALACTÓSIDOS	5	GLICOLIPROTEÍNA	5
GALACTOSILACIÓN	5	GLICÓLISIS	5
GALACTOSILTRANSFERASA	5	GLICOLÍTICO	5
GALACTURÓNICO	5	GLICOOXIDACIÓN	5
GASTRINA	5	GLICOSIDASAS	5
GASTROCÓLICO	5	GLICOSÍDICAS	5
GASTRODUODENAL	5	GLICOSILANTE	5
GASTROENTERITIS	5	GLICOSILTRANSFERASA	5
GASTROENTEROLOGÍA	5	GLUCOAMILASA	5
GASTROESOFÁGICO	5	GLUCOCONJUGADOS	5
GASTRONOMÍA	5	GLUCOGENOGÉNESIS	5

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
GLUCOLÍPIDOS	5	HEMICUERPO	5
GLUCONATO	5	HEMOAGLUTINIZACIÓN	5
GLUCONEOGENÉTICO	5	HEMOCONCENTRACIÓN	5
GLUCONEOGÉNICO	5	HEMODIÁLISIS	5
GLUCONICO	5	HEMODINÁMICOS	5
GLUCORECEPTORES	5	HEMOPILTRACIÓN	5
GLUCORONIL	5	HEMOGLOBINÓMETRO	5
GLUCOSADAS	5	HEMOGLOBINOPATÍAS	5
GLUCOSIDASA	5	HEMOGRAMA	5
GLUCÓSIDO	5	HEMOLÍTICAS	5
GLUCOSILADA	5	HEMOLIZADO	5
GLUCOSILO	5	HEMORRAGIAS	5
GLUCOSURIA	5	HEMORRÁGICO	5
GLUCURÓNICO	5	HEMOSIDERINA	5
HEMAGLUTENINAS	5	HEMOSTASIA	5
HEMÁTICOS	5	HEMOSTASIS	5
HEMATIMÉTRICOS	5	HEMOSTATÓMETRO	5
HEMATÓFAGOS	5	HEPARINIZADA	5
HEMATOFLUORÓMETRO	5	HEPATITIS	5
HEMATOLOGÍA	5	HEPATOBIAR	5
HEMATOMAS	5	HEPATOCARCINOMA	5
HEMATOPOYESIS	5	HEPATOECTOMIA	5
HEMATOXILINA	5	HEPATOESPLENOMEGALIA	5
HEMICELULOSA	5	HEPATOLOGÍA	5

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
HEPATOPATÍAS	5	HIDROXIESTEROIDE	5
HEPATORRENAL	5	HIDROXIFENILACÉTICO	5
HEPATOTOXICIDAD	5	HIDROXILACIÓN	5
HETEROCEDASTICIDAD	5	HIDROXILADO	5
HETEROCÍCLICO	5	HIDROXIPROLINA	5
HETEROCIGOSIS	5	HIDROXITOLUENO	5
HETEROCIGOTO	5	HIDRUROS	5
HIDROCARBONADA	5	HIDRYCINÁMICOS	5
HIDROCARBONOS	5	HIPERACTIVACIÓN	5
HIDROCEFALIA	5	HIPERAMINOACIDURIA	5
HIDROCLOROTIAZIDA	5	HIPERAMONEMIA	5
HIDRODENSITOMETRÍA	5	HIPERANDROGENISMO	5
HIDROELECTROLÍTICA	5	HIPERBETALIPOPROTEINEMIA	5
HIDROFÍLICA	5	HIPERBILIRRUBINEMIA	5
HIDROFLICA	5	HIPERCALCEMIA	5
HIDROFÓBICO	5	HIPERCALCIURIA	5
HIDROGENA	5	HIPERCOLESTEROLEMIANTES	5
HIDROMETRÍA	5	HIPERCORTISOLISMO	5
HIDROPEROXIDO	5	HIPERFAGIA	5
HIDROPERÓXILO	5	HIPERFENILALANINEMIA	5
HIDROPESÍA	5	HIPERGLICINEMIA	5
HIDROQUINONA	5	HIPERGLUCEMIA	5
HIDROSTÁTICA	5	HIPERGLUCÍDICA	5
HIDROXIBENZÓICO	5	HIPERHIDRATACIÓN	5

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
HIPERLACTACIDEMIA	5	HIPONATREMIA	5
HIPERLACTATEMIA	5	HIPOPIGMENTACIÓN	5
HIPERLAXITUD	5	HIPOPLASIA	5
HIPOACTIVIDAD	5	HIPOPNEA	5
HIPOACUSIA	5	HIPOPROTEINEMIAS	5
HIPOALFALIPOPROTEINEMIA	5	HIPOQUINESIA	5
HIPOALIMENTACIÓN	5	HIPOREXIA	5
HIPOAMINOACIDEMIA	5	HIPOSÓDICO	5
HIPOCALCEMIA	5	HIPOSENSIÓN	5
HIPOCETONEMIA	5	HIPOTERMIA	5
HIPOCETONURIA	5	HIPOTRIGLICERIDÉMICA	5
HIPOCETÓSICA	5	HIPOVENTILACIÓN	5
HIPOCLORITO	5	HIPOVOLEMIA	5
HIPOCROMÍA	5	HOMOCEDASTICIDAD	5
HIPODENSA	5	HOMOCIGOSIS	5
HIPOECOGENICIDAD	5	HOMOCISTEINEMIA	5
HIPOFISIARIO	5	HOMOCISTINURIA	5
HIPOFISIARIOS	5	HOMODÍMERO	5
HIPOFISIS	5	HOMOESTASIS	5
HIPOGEUSIA	5	HOMOFERMENTATIVAS	5
HIPOGRASA	5	HOMOSCEDASTICIDAD	5
HIPOINSULINEMIA	5	INMUNOANÁLISIS	5
HIPOKALEMIA	5	INMUNOCITOQUÍMICA	5
HIPOLACTASIA	5	INMUNODEFICIENTES	5

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
INMUNOENZIMÁTICOS	5	INTRANUCLEAR	5
INMUNOESTIMULACIÓN	5	INTRAOBSERVADOR	5
INMUNOGENÉTICA	5	INTRAPERITONEAL	5
INMUNOGLOBULINA	5	INTRAVARIEDAD	5
INMUNOLOGÍA	5	INTRAVASCULAR	5
INMUNOMODULACIÓN	5	ISOCALÓRICA	5
INMUNONEUTRALIZAR	5	ISOCUMARINAS	5
INMUNONUTRICIÓN	5	ISOELÉCTRICO	5
INMUNOPRECIPITACIÓN	5	ISOENZIMA	5
INMUNORADIOELÉCTRICO	5	ISOFLAVINAS	5
INMUNORRADIOMÉTRICO	5	ISOFLAVONOIDES	5
INMUNORREGULADOR	5	ISOFLAVONONAS	5
INMUNOSUPRESIÓN	5	ISOMERASA	5
INTRAABDOMINAL	5	ISOOSMORALIDAD	5
INTRACEREBRALMENTE	5	ISOPRENOIDE	5
INTRAENSAYO	5	ISOPROPÍLICO	5
INTRAEPITELIAL	5	ISOPROSTANOS	5
INTRAGÁSTRICAMENTE	5	ISORHAMNITINA	5
INTRAHEPÁTICA	5	ISOTERMA	5
INTRAHOSPITALARIA	5	ISOTÉRMICAS	5
INTRALISOSOMAL	5	ISOTÍPICO	5
INTRALOBULAR	5	ISOTÓNICAS	5
INTRALUMINAL	5	ISOVALÉRICA	5
INTRAMEDIDOR	5	LACTOCINA	5

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
LACTODENSÍMETRO	5	LIPOSOLUBILIDAD	5
LACTOFERRINA	5	LIPOTEICOICO	5
LACTOGÉNESIS	5	LIPOXIGENASA	5
LACTOGLOBULINA	5	LIPOXINAS	5
LACTOPEROXIDASA	5	MACROBIÓTICOS	5
LACTOSUERO	5	MACROCÍCLICAS	5
LACTULOSA	5	MACROCÍTICA	5
LEUCOANTOCIANIDINAS	5	MACROCITOSIS	5
LEUCOCINA	5	MACROCONSTITUYENTES	5
LEUCOCITARIA	5	MACROFÁGICA	5
LEUCOCÍTICO	5	MACROGLOBULINA	5
LEUCONOCINA	5	MACROPUS	5
LEUCOPENIA	5	MACROSCÓPICA	5
LIPOCALIBRE	5	MACROVASCULARES	5
LIPODISTROFÍAS	5	METACARPIANOS	5
LIPODOLOGÍA	5	METAFISIARIO	5
LIPOFÍLICA	5	METAFOSFÓRICO	5
LIPOGENASA	5	METALOTIONEINA	5
LIPOGÉNICAS	5	METANÓLICA	5
LIPOICO	5	METASTÁSICO	5
LIPOÍDEA	5	MICROALBUMINURIA	5
LIPOPEROXIDACIÓN	5	MICROALGAS	5
LIPOPEROXIDATIVA	5	MICROBALANZAS	5
LIPOPOLISACÁRIDO	5	MICROBICIDA	5



<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
MICROCEFALIA	5	MONOCOTILEDÓNEAS	5
MICROCIRCULACIÓN	5	MONOCYTOGENES	5
MICROCITOSIS	5	MONOFÁSICA	5
MICROENCAPSULACIÓN	5	MONOFENOLAS	5
MICROGLIA	5	MONOFLORES	5
MICROGRAFÍA	5	MONOFOSFATO	5
MICROLITRO	5	MONOFRECUENCIA	5
MICROMÉTODO	5	MONOGÁSTRICOS	5
MICROMINERAL	5	MONOGRAFÍA	5
MICRONIZADA	5	MONOHIDRATO	5
MICRONODULAR	5	MONOHIDRIBENZOATOS	5
MICROONDAS	5	MONOHIDROXIBENZOICOS	5
MICROPARTICULADO	5	MONOMÉRICA	5
MICROPLACAS	5	MONONUCLEARES	5
MICROPOZOS	5	MONOSATURADA	5
MICROSCOPIA	5	MONOTERPENOS	5
MICROSCOPIO	5	MONOZIGÓTICOS	5
MICROTÚBULOS	5	MONOZIGOTOS	5
MICROVASCULARES	5	MULTICAUSAL	5
MICROVELLOCIDADES	5	MULTICELULARES	5
MONOACILGLICEROLAS	5	MULTICOLINARIEDAD	5
MONOAMINO	5	MULTIDIMENSIONAL	5
MONOBLÁSTICAS	5	MULTIENZIMÁTICO	5
MONOCARBONADA	5	MULTIETÁPICO	5

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
MULTIORGÁNICA	5	NEUROTÓXICOS	5
MULTIPARIDAD	5	NEUOTRÓFICO	5
MULTISISTÉMICO	5	NEUROVEGETATIVA	5
MULTIVARIANTE	5	NORMOCALÓRICA	5
MULTIVITAMÍNICO	5	NORMOCEFALIA	5
NEUROACTIVOS	5	NORMOCÍTICA	5
NEUROAMINIDASA	5	NORMOCOLESTEROLÉMICOS	5
NEUROBIOLÓGICAS	5	NORMOCRÁNEO	5
NEUROBLASTOMA	5	NORMOCRÓMICA	5
NEUROCOGNITIVO	5	NORMODENSA	5
NEURODEGENERACIÓN	5	NORMOFUNCIONANTE	5
NEUROECTODERMA	5	NORMOGLICEMIA	5
NEUROENDOCRINO	5	NORMOGLUCÉMICAS	5
NEUROEPITELIO	5	NORMOGRAMA	5
NEUROFISIOLOGÍA	5	NORMONATRÉMICO	5
NEUROGÉNICA	5	NORMOZINCEMICOS	5
NEUROINFLAMATORIA	5	OLIGODENDROCITOS	5
NEUROMODULADOR	5	OLIGOMÉRICOS	5
NEUROMOTOR	5	OLIGOSACARIDOSIS	5
NEUROPATOLÓGICAS	5	ORTOSTÁTICA	5
NEUROPÉPTIDO	5	OSMOREGULADOR	5
NEUROPROTECTINA	5	OSMORREGULACIÓN	5
NEUROPROTECTOR	5	OSTEOARTICULAR	5
NEUROQUÍMICOS	5	OSTEOBLÁSTICA	5

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
OSTEOCLÁSTICA	5	POLIMÓRFICA	5
OSTEODEGENERATIVAS	5	POLIMORFONUCLEARES	5
OSTEODENSITÓMETRO	5	POLINEUROPATÍAS	5
OSTEOGÉNESIS	5	POLINIZACIÓN	5
OSTEOIDE	5	POLINOMIAL	5
OSTEOLOGÍA	5	POLIOMIELITIS	5
OSTEOMALACIA	5	POLYPEPTÍDICA	5
OSTEOMUSCULAR	5	POLIPÉPTIDO	5
OSTEOPETROSIS	5	POLIPICA	5
OSTEOPORÓTICAS	5	POLISOMNOGRAFÍA	5
OSTEOTENDÍNEOS	5	POLISULFIDOS	5
PATOGÉNICO	5	POLITENOLES	5
PATOGNOMÓNICO	5	POLITERAPIA	5
POLIACRILAMIDA	5	POLIURIA	5
POLIALCOHOL	5	POLIVALENTES	5
POLIAROMÁTICOS	5	POLIVITAMÍNICO	5
POLICATIONES	5	POSMENOPAUSIA	5
POLICATIÓNICA	5	POSPARTO	5
POLICLONAL	5	POSPRANDIAL	5
POLIESPLÉNICO	5	POSTCONCEPCIONAL	5
POLIETÁPICO	5	POSTDOSIS	5
POLIETILENGLICOL	5	POSTEST	5
POLIFENILPROPANO	5	POSTGLUCOSA	5
POLIFENOLOXIDASA	5	POSTINTERVENCIÓN	5

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
POSTPARTO	5	SULFOSALICÍLICO	5
POSTPRANDIAL	5	SULFÓXIDO	5
POSTPRUEBA	5	SULFURADO	5
POSTQUIRÚRGICA	5	SULFÚRICO	5
POSTRANSCRIPCIONALES	5	SULFUROSOS	5
QUIMIOATRACTANTE	5	TERMOESTABILIDAD	5
QUIMIOLUMINISCENCIA	5	TERMOGENINAS	5
QUIMIOPREVENCIÓN	5	TERMONEUTRAL	5
QUIMIOTÁCTICA	5	TERMOREGULADA	5
QUIMIOTERAPÉUTICAS	5	TERMORESISTENCIA	5
QUIMIOTRIPSINA	5	TERMORREGULADOR	5
RADIOACTIVIDAD	5	TERMOTOLERANCIA	5
RADIOACTIVO	5	TETRAACETICO	5
RADIOENSAYO	5	TETRAAMONIO	5
RADIOGRAFÍA	5	TETRABUTILAMONIO	5
RADIOINMUNOANÁLISIS	5	TETRACLORURO	5
RADIOINMUNOMÉTRICOS	5	TETRACÓRICAS	5
RADIOISOTÓPICAS	5	TETRAFOSFATO	5
RADIOISÓTOPO	5	TETRAHIDRO	5
RADIOLÓGICO	5	TETRAHIDROBIOPTERINA	5
RADIOMÉTRICO	5	TETRAMÉRICAS	5
RADIOTERAPIA	5	TETRAMETILCROMAN	5
SULFOAMINOÁCIDO	5	TETRAPOLAR	5
SULFONAMIDAS	5	TETRASULFUROS	5

<b>Palabra</b>	<b>Frec.</b>
TETRATERPENOS	5
TIROIDECTOMÍA	5
TIROSINEMIA	5
TIROTOXICOSIS	5
TROMBOEMBOLIA	5
TROMBOGÉNESIS	5
TROMBOGÉNICOS	5
TROMBOLÍTICOS	5
TROMBOPLASTINA	5
TROMBÓTICOS	5
ULTRACENTRÍFUGA	5
ULTRACENTRIFUGACIÓN	5
ULTRADISTANCIA	5
ULTRAFILTRACIÓN	5
ULTRASONOGRAFÍA	5
ULTRAVIOLETA	5

#### 10.4 No Palabras elaboradas a partir de prefijos y bases del Corpus de Nutrición

<b>No Palabra</b>	<b>No Palabra</b>
AMINOglobulina	GLUCODUODENAL
ATEROICO	GLUCOESCLERÓLISIS
BIOGLICEMIA	GLUCOSTOMÍA
CITOLÍPIDO	HEMATODILUCIÓN
CITOPRENOIDE	HEMATOPLASTINA
CITOSEMIA	HEMOFERRINA
CLOROCITOCINA	HEPATOXIDACIÓN
CLOROPÉPTIDO	HEPATOXILINA
FIBROENZIMÁTICO	HETEROGLUCEMIA
FITOCUBITAL	HIPERFISIS
FITÓLISIS	HOMOCLORITO
FOSFÁCIDO	ISOGLUCEMIA
FOSFATOSILACIÓN	ISOVITAMÍNICO
FUCOCORTICOIDES	LIPOTROMBOSIS
FUCOGENOS	OLIGOENZIMÁTICO
FUCOLÍPIDOS	OLIGOGLICÉRIDO
FUCONUTRIENTE	ORTOGLICEMIA
FUCOPROTEINA	OSTEOENZIMÁTICO
GALACTOBACILOS	QUIMIOANÁLISIS
GALACTODENSÍMETRO	SULFOGÉNICO
GALACTOSUERO	TETRAFRUCTOSA
GASTROGENOSIS	TETRAGLICÉRIDO
GASTROTIROXINA	TIRODUODENAL
GLICEMATOSIS	TROMBOGLOBINÓMETRO

<b>No Palabra</b>	<b>No Palabra</b>
ULTRALACTASIA	ORTOCEFALIA
AMINOFÁTIDO	ORTOCÍTICA
ATEROSILACIÓN	OSTEONEMIA
BACTEROPROTECTOR	QUIMIOISOTÓPICA
CITOBIÓTICO	TROMBODIÁLISIS
CLOROSACÁRIDO	TROMBODILUCIÓN
FIBROGLICEMIA	TROMBÓFAGOS
FITOSACÁRIDO	TROMBOGLUTENINA
GALACTOALBÚMINA	ULTRAFAGIA
GALACTOFERRINA	ULTRAGLUCEMICO
GLICEMIPATÍA	ULTRAVITAMINOSIS
GLICEROLÍTICAS	
GLUCOFAGIA	
HEMATOCALÓRICO	
HEMOGLICEMIA	
HEPATOPOROSIS	
HIDROGLOBULINA	
HOMOCALÓRICO	
INMUNOSILACIÓN	
INSULINOPIGMENTACIÓN	
LACTOLÍPIDOS	
LACTOSACÁRIDO	
LIPÓTIDO	
MULTIFRUCTOSA	

## 10.5 Selección de Palabras Reales del Corpus de Odontología por Frecuencia de Palabra

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
PERIODONTAL	619	RECIDIVA	89
GINGIVAL	354	OCLUSAL	86
PERIODONTITIS	214	EXÉRESIS	84
RADICULAR	205	QUERATINOCITOS	84
EPITELIO	199	PERIAPICAL	83
CARCINOMA	195	LIQUEN	82
ALVEOLAR	144	ODONTOMA	82
ETIOLOGÍA	132	TISULAR	82
CANDIDIASIS	112	OSTEOINTEGRACIÓN	81
HIPERPLASIA	112	TETRACICLINA	81
ALVEOLITIS	101	CLORHEXIDINA	79
OCLUSIÓN	100	ESTOMATITIS	79
GINGIVITIS	100	EXODONCIA	78
VESTIBULAR	99	QUEILITIS	78
EPITELIAL	99	HEMOSTASIA	76
METÁSTASIS	99	PROTÉSICA	76
DENTINA	98	ACETALDEHÍDO	75
PERIODONTO	97	DISTAL	72
XEROSTOMÍA	97	APICAL	71
HISTOPATOLÓGICO	93	DOXICICLINA	69
GRANULOMA	92	ERITEMA	69
MELANOMA	91	SUBGINGIVAL	69
AMELOBLASTOMA	89	CLINDAMICINA	68
COLGAJO	89	FIBROBLASTOS	68



<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
ANAMNESIS	67	EDÉNTULA	17
EPIDERMOIDE	67	PARÓTIDA	17
ANATOMAPATOLÓGICO	66	TINCIÓN	17
TÚBULO	66	ODONTOGÉNICO	16
INMUNOHISTOQUÍMICA	65	APÓFISIS	15
PILOCARPINA	55	ENDOCARDITIS	15
OSTEOBLASTOS	54	EXUDADO	15
OSTEOCITOS	54	OROGARINGE	15
OSTEOCLASTOS	54	AGENESIA	14
EDÉNTULOS	53	BENZODIACEPINAS	14
ETIOLÓGICO	53	CARCINÓGENOS	14
EDENOMA	50	ISQUÉMICA	14
ERITEMATOSA	50	BACTERIEMIA	13
NEUTRÓFILOS	50	CITOQUINAS	13
ENDOTELIALES	49	DEXTRANOS	13
ETIOPATIOGENIA	47	EDENTOLISMO	13
HIDROXIAPATITA	45	ESTOMATOGNÁTICO	13
MALOCLUSION	45	PEDICULADA	13
MUCOGINGIVAL	40	PERIOSTIO	13
PLEOMÓRFICO	39	PREVOTELLA	13
ESTROMA	38	SIALOSIS	13
EXTRUSIÓN	33	CITOMEGALOVIRUS	13
TUMEFACCIÓN	30	CREVICULAR	12
ACINAR	17	QUERATINIZACIÓN	12
ANTÍGENOS	17	ADENOPATÍAS	11

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
ALENDRONATO	11	BUCOSINUSALES	9
ENDOTELIO	11	CARIOGÉNICA	8
EXOFÍTICA	11	ENDODÓNTICOS	8
PARESTESIA	11	INMUNOHISTOQUÍMICOS	8
PLEXIFORME	11	MALATOCITOS	8
POLISOMNOGRAFÍA	11	METASTÁSICA	8
PROTOMBINA	11	ORTOPANTOMOGRAFÍA	8
ADENOCARCINOMA	11	PAROTÍDEA	8
BRUXISMO	10	PENFIGOIDE	8
CITOCROMO	10	PLASMINÓGENO	8
LINFANGIOMA	10	PROTAGLANDINAS	8
OLIGODONCIA	10	PROSTODONCIA	8
OSTEOPROGENITORAS	10	TEMPOROMANDIBULAR	8
PERIOCORONARIO	10	TRANEXÁMICO	8
PERIODONCIA	10	ETIOPATOGÉNICO	8
SUPRAGINGIVAL	10	TREPONEMA	8
TOLUIDINA	10	UNILOCULAR	8
CARCINOGENÉISIS	9	CARÚNCULA	7
COLUTORIOS	9	DENTÍCOLA	7
ESCAMOCELULAR	9	DENTOALVEOLARES	6
ISQUEMIA	9	FIBRINÓLISIS	7
PERIODONTOPATÓGENOS	9	FRENECTOMÍA	7
TISURALES	9	HAMARTOMAS	7
TRABÉCULAS	9	HISTIOCITOSIS	7

<b>Palabra</b>	<b>Frec.</b>	<b>Palabra</b>	<b>Frec.</b>
IDIOPÁTICA	7	CATECOLAMINAS	5
MULTILOCULAR	7	CLAVULÁNICO	5
ODONTALGIAS	7	CORONOIDECTOMIA	5
PIÓGENO	7	DEHISCENCIA	5
PREDNISONA	7	ERITROPLASIA	5
TRABECULADO	7	HERPETIFORME	5
AMELOBLÁSTICO	6	INMUNOFLOURESCENCIA	5
AMELOCEMENTARIA	6	LISOSAMALES	5
ANTINEOPLÁSICOS	6	MICROGNATISMO	5
BASALOIDE	6	MUCOCUTÁNEAS	5
EOSINÓFILOS	6	OSTEÍTIS	5
EQUIMOSIS	6	PATOGNOMÓNICO	5
FLEXICORTICOTOMÍA	6	PERICORONARITIS	5
HEMANGIOMA	6	TUBEROSIDAD	5
HISTIOCITOS	6	ANODONCIA	5
INTERPROXIMALES	6	APÓSITO	5
MECANOTRANSDUCCIÓN	6	ARTICAÍNA	5
OSTEORRADIONECROSIS	6	CAVITARIAS	5
OSTEOBLÁSTICA	6	EXÓSTOSIS	5
OXIMETRÍA	6	GLOSITIS	5
PEROXIDASA	6	IMPLANTODONTIA	5
PSEUDOMONAS	6	MELANOACANTOMA	5
TARTRECTOMÍA	6	MIOEPITELIALES	5
BICORTICAL	5	PERIAMIGDALINAS	5
QUERATOQUISTE	5		

## 10.6 No Palabras elaboradas a partir de prefijos y bases del Corpus de Odontología

<b>No Palabra</b>	<b>No Palabra</b>
	EXOSIMUSALES
ADENOCALCINA	GINGIVIMATOSA
AMELOTOMIA	GINGIVOCALCIMATOSIS
BUCOADENOSIS	IMPLANTOMÓRFICO
CARCINOMÓRFICO	IMPLANTOPLÁSICO
CARIADIASIS	INMUNOGINGIVAL
CARIOFÍTICA	INMUNOPLÁSICO
CARIOLOIDE	LINFADENOCLACINA
CITOFÍTICA	LINFADENOLINGUAL
CORONODONCIA	MAXILODIASIS
DENTIPLÁSICO	MAXILOFÍTICA
DENTODIASIS	MAXILOGIOMA
DENTOLOIDE	MUCOADENOSIS
DENTOPLASIA	MUCOGIOMA
EDENTODIASIS	MUCOMATOSO
EDENTOPLÁSICO	ODONTOCALCINA
ENDOLITIS	ORTODIASIS
ERITROCELULAR	ORTOFÍTICO
ERITROCTROMÍA	OSTEOLOIDE
ERITROGIOMA	PERIAGINGIVAL
ESTOMATOBLÁSTICO	PERIOADENOSIS
ESTOMATOCALCINA	PERIOECTOMIA
ESTOMATOFÍTICO	PERIOMÓRFICO
ETIOPAFÍTICA	PERIOPLASIS
ETIOSOMÁTICO	

---

**No Palabra**

---

SIALOFÍTICA

SIALOPATOGENIA

SIALOSINUSALES

SIALOSOMÁTICO

TISOSINUSALES

TRABELOIDE

TRABEPATOGENIA

TRAQUEOCALCINOMA

---

## 10.7 Prefijos de alta frecuencia en el corpus de Nutrición

<b>Prefijo</b>	<b>Frec.</b>	<b>Prefijo</b>	<b>Frec.</b>
MICRO	938	HETERO	182
LEUCO	908	MULTI	169
CITO	821	BACTE	159
QUIMIO	813	COLORO	147
HIPO	627	NEURO	147
MACRO	624	HEMI	130
MONO	557	FITO	112
POST	539	INTRA	107
HEMA	499	ORTO	94
POLI	425	TROMBO	90
FOSFA	406	GLUCO	81
HOMO	387	LIPO	81
INMUNO	345	FUCO	76
HEPA	297	HEMO	69
NORMO	293	TERMO	64
GASTRO	268	SULFO	56
HIPER	268	HIDRO	50
GLICO	262	ULTRA	49
ATERO	250	TETRA	46
BIO	238	FIBR	45
META	219	OLIGO	43
ISO	214	OSMO	33
LACTO	193	TIRO	27
PATO	192	GLICE	25

<b>Prefijo</b>	<b>Frec.</b>
GALACTO	24
RADIO	24
OSTEO	21

## 10.8 Prefijos de alta frecuencia en el corpus de Odontología

<b>Prefijo</b>	<b>Frec.</b>	<b>Prefijo</b>	<b>Frec.</b>
ODONTO	412	SIALO	79
ADENO	403	ISQUEM	74
CITO	403	BUCO	68
ESTOMATO	400	ORTO	56
GINGIVI	319	IMPLANTO	49
EXO	319	PERIA	47
MUCO	316	LINFADENO	47
DENTO	316	CORONA	45
DENTI	312	GASTRO	45
INMUNO	312	ETIOPA	36
OSTEO	310	EDENTUL	35
ERITEM	256	EDEN	34
MAXILO	234	PAROT	33
GINGIVO	220	PROST	32
PERIO	216	LINFANGEO	29
INMUNO	212	ENDO	27
HIST	188	TRABE	24
CARIO	154	CARCINO	23
FIBRO	116	PERIO	22
CORONO	104	VESTIBULO	22
AMELO	99	ERITRO	19
TRAQUEO	92	TISO	18