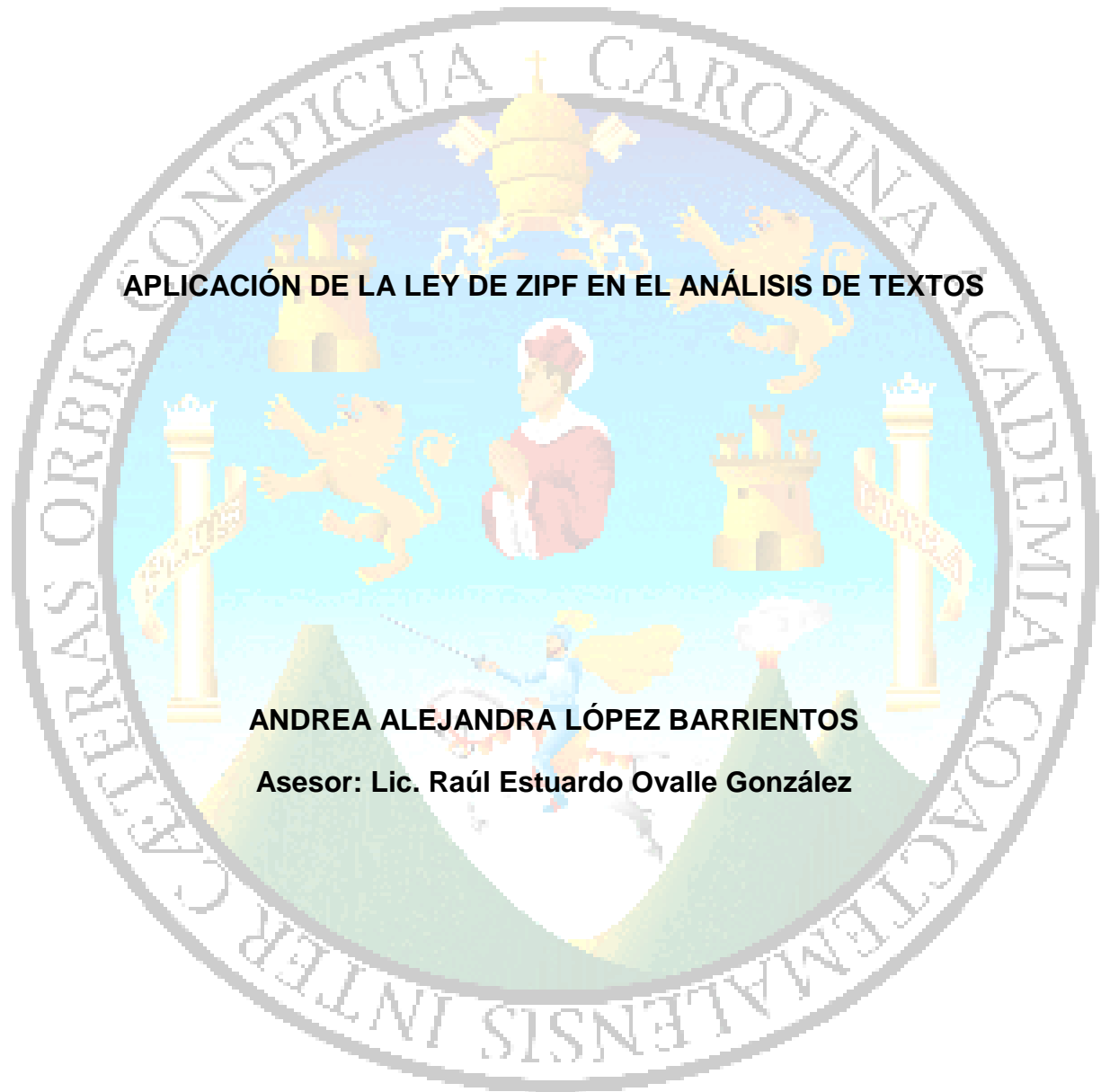


UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
ESCUELA DE CIENCIAS LINGÜÍSTICAS



APLICACIÓN DE LA LEY DE ZIPF EN EL ANÁLISIS DE TEXTOS

ANDREA ALEJANDRA LÓPEZ BARRIENTOS

Asesor: Lic. Raúl Estuardo Ovalle González

TÉCNICO EN TRADUCCIÓN Y CORRESPONDENCIA INTERNACIONAL

GUATEMALA, MAYO DE 2013

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
ESCUELA DE CIENCIAS LINGÜÍSTICAS

Aplicación de la Ley de Zipf en el Análisis de Textos

TRABAJO DE GRADUACIÓN

Presentado Al Consejo Directivo de la
Escuela De Ciencias Lingüísticas

POR

Andrea Alejandra López Barrientos

Asesorada por: Lic. Raúl Estuardo Ovalle González

Al conferírsele el título de

TÉCNICO EN TRADUCCIÓN Y CORRESPONDENCIA INTERNACIONAL

Guatemala, mayo de 2013

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
ESCUELA DE CIENCIAS LINGÜÍSTICAS



Rector

Lic. Carlos Estuardo Gálvez Barrios

Consejo Directivo

DIRECTOR:	Ing. José Humberto Calderón Díaz
SECRETARIA ACADÉMICA:	Dra. Evelyn Carolina Masaya Anleu
REPRESENTANTE DE LOS DOCENTES:	Lic. Raúl Estuardo Ovalle González
REPRESENTANTE DE LOS DOCENTES:	Licda. María Argelia Estrada Vázquez
REPRESENTANTE DE LOS ESTUDIANTES:	Bachiller Claudia Lorena Palacios Osorio
REPRESENTANTE DE LOS ESTUDIANTES:	Perito Contador Hans Ludvin del Cid López



Guatemala, 17 de Julio del 2013

**Ref. Secretaría Académica
No. 001-2013**

TESINA TITULADA:

“Aplicación de la Ley de Zipf en el
Análisis de Textos”.

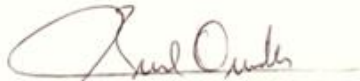
DESARROLLADA POR LA ESTUDIANTE:


Andrea Alejandra López Barrientos
Carné: 201013614

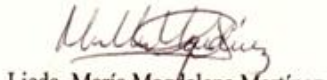
EVALUADA POR LOS PROFESIONALES:


Lic. Raúl Estuardo Ovalle González
Licda. Vaglia Linares Domínguez
Licda. María Magdalena Martínez

Las Autoridades y los Asesores de la Escuela de Ciencias Lingüísticas, hacen constar que ha cumplido con las Normas y Reglamentos de la Escuela No Facultativa de Ciencias Lingüísticas de la Universidad de San Carlos de Guatemala.


Lic. Raúl Estuardo Ovalle González
ASESOR


Licda. Vaglia Lisseth Linares Domínguez
EXAMINADORA


Licda. María Magdalena Martínez
EXAMINADORA


Dra. Evelyn Carolina Masaya Anleu
SECRETARIA ACADEMICA

IMPRIMASE


Ing. Agr. José Humberto Calderón Díaz
DIRECTOR



DEDICATORIA

- A Dios: Por haberme brindado sabiduría y salud para lograr mis objetivos y por haber puesto en mi camino a aquellas personas que me han acompañado en mi período de estudio.
- A mis padres: Por darme la vida, por creer en mí y apoyarme en todo momento y por su constante motivación que me ha llevado a ser la persona que soy.
- A mi hermano: Por estar conmigo y apoyarme siempre.
- A mis abuelos: Por quererme y apoyarme siempre.
- A mis amigos: Por compartir conmigo buenos y malos momentos.
- A mi asesor: Por su apoyo y motivación para la elaboración de esta tesina y por impulsar el desarrollo de mi formación profesional.
- A: La Escuela de Ciencias Lingüísticas de la Universidad de San Carlos de Guatemala por facilitar mi oportunidad de desarrollo profesional.

ÍNDICE

Contenido	Pág.
Resumen.....	i
Introducción	ii
Objetivos	iii
General	iii
Específicos.....	iii
Planteamiento del problema	iv
Justificación	v
Método	vi

CAPÍTULO I

GENERALIDADES

1.1 Definición de Lingüística	7
1.1.1 Objeto de la Lingüística.....	8
1.1.2 Ramas de la Lingüística.....	8
1.1.3 Historia de la Lingüística	9
1.2 Definición de Estadística.....	13
1.2.1 Historia de la Estadística.....	13
1.2.2 Aplicaciones de la Estadística.....	16
1.2.3 División de la Estadística	17
1.3 Ciencias auxiliares y afines de la Lingüística	17
1.4 Estadística Lingüística	18

CAPÍTULO II
GEORGE KINGSLEY ZIPF

Contenido	Pág.
2.1 Antecedentes familiares	21
2.2 Biografía.....	21
2.3 Vida familiar	24
2.4 Lingüística de Zipf	25
2.5 Ley de Zipf	27
2.5.1 Economía Lingüística	27
2.5.2 Frecuencia de aparición de las palabras en un texto	28

CAPÍTULO III
USO DE LA LEY DE ZIPF EN LA LINGÜÍSTICA COMPUTACIONAL

3.1 Importancia de la Estadística en el estudio del lenguaje.....	29
3.2 Definición de Lingüística Computacional	30
3.2.1 Similitud entre la mente humana y los ordenadores	31
3.2.2 Enfoques y métodos de la Lingüística Teórica y la Lingüística Computacional.....	32
3.2.3 Usos de la Lingüística Computacional	33
3.3 Lingüística Computacional y su relación con la Ley de Zipf.....	35
3.3.1 La Indización Automática y la generación de <i>tesauros</i>	36

CAPÍTULO IV

USO DE LA LEY DE ZIPF EN LA LINGÜÍSTICA FORENSE

Contenido	Pág.
4.1 Definición de Lingüística Forense	39
4.1.1 El plagio	40
4.2 Estilo y Estadística	42
4.3 Importancia de la Estadística en la Lingüística Forense	43
4.3.1 Recorrido histórico	43
4.3 La Lingüística Forense y su relación con la Ley de Zipf.....	45

CAPÍTULO V

APLICACIÓN DE LA LEY DE ZIPF EN EL ANÁLISIS DE TEXTOS

5.1 Aplicación de la Ley de Zipf en el análisis de un texto turístico	47
5.2 Aplicación de la Ley de Zipf en el análisis de un artículo de revista	50
5.3 Aplicación de la Ley de Zipf en el análisis de una fábula	52
5.4 Análisis.....	54
Conclusiones	59
Recomendaciones	61
Bibliografía	63
Anexos	71
Texto No.1	71
Tabla No.1.	73
Texto No .2.....	79

Contenido	Pág.
Tabla No. 2	81
Texto No. 3	87
Tabla No. 3.....	89
Texto No. 4	97
Tabla No. 4.....	99
Texto No. 5	107
Tabla No. 5.	108
Texto No. 6	111
Tabla No. 6.	112
Glosario.....	115

RESUMEN

La Lingüística, es el estudio científico del lenguaje, por tal razón debe ser estudiada y analizada desde diversos enfoques y puntos de vista, haciendo uso tanto de ciencias afines como de ciencias auxiliares que la nutren y apoyan. Una de las ciencias autónomas que puede contribuir con el estudio de esta disciplina es la Estadística Lingüística, que no es más que la aplicación de métodos estadísticos para la explicación de fenómenos lingüísticos.

El primero en observar que la Estadística podía contribuir al estudio de ciertos fenómenos lingüísticos fue el filólogo George Kingsley Zipf, que planteó un principio denominado “Ley del Mínimo Esfuerzo”, el cual consiste en afirmar que el comportamiento humano se rige por la reducción de esfuerzo, lo que significa que las personas cuando hablan o escriben utilizan mucho más palabras cortas que aquellas más largas. En diferentes textos si se ordenan todas las palabras de acuerdo a su frecuencia decreciente y se les asigna un orden, el producto total de la ocurrencia por el orden arrojará una constante.

Actualmente, los estudios de Zipf tratan de mejorarse y son más utilizados que hace unos cuantos años, gracias a la aparición del *internet* y el auge de las Lingüísticas Computacionales, así como de la Lingüística Forense y puede aplicarse en el análisis de cualquier texto para determinar sus características y particularidades.

INTRODUCCIÓN

Muchas son las ciencias que nutren y apoyan el estudio de la Lingüística, una de ellas es la Estadística, que a pesar de no conocerse a profundidad el aporte que brinda a la Lingüística es muy utilizada en la Lingüística Computacional, Lingüística Forense y análisis de textos, por lo que debe ser aplicada como una herramienta de investigación por todos aquellos que se dedican al estudio del lenguaje humano.

Esta investigación está dividida en cinco capítulos, en los cuales se hace referencia a la utilidad de la Estadística y específicamente de la Ley de Zipf en la investigación lingüística.

El filólogo George Kingsley Zipf planteó un principio denominado “Ley del Mínimo Esfuerzo” que afirma que el comportamiento humano se rige por la reducción de esfuerzo, lo que significa que las personas cuando escriben o hablan utilizan mucho más palabras cortas que aquellas más largas. En diferentes textos si se ordenan todas las palabras de acuerdo a su ocurrencia decreciente y se les asigna un orden, el producto total de la ocurrencia por el orden arrojará una constante.

En el presente trabajo se aplicó la Ley de Zipf a textos en español y a su respectiva traducción en inglés para verificar las ventajas de aplicar métodos de investigación estadística para el análisis de textos y se determinó que es más común que un texto en español sea más denso que en inglés y se verificó que tal como lo planteó Zipf, las palabras con una mayor longitud son menos usuales que aquella de menor longitud.

OBJETIVOS

General

Dar a conocer la aplicación de la Ley de Zipf en el análisis de textos.

Específicos

1. Definir los términos Lingüística, Estadística y Estadística Lingüística.
2. Presentar la vida y el aporte de George Kingsley Zipf.
3. Determinar el uso que la Lingüística Computacional hace de la Ley de Zipf.
4. Especificar la relación entre la Lingüística Forense y la Ley de Zipf.
5. Aplicar la Ley de Zipf en el análisis de un texto.

PLANTEAMIENTO DEL PROBLEMA

Al hablar de Estadística suele pensarse automáticamente en ciencias exactas, aunque en realidad los recursos de investigación que proporciona la Estadística pueden aplicarse en cualquier otra ciencia, pero los estudios de la Estadística para su aplicación a la Lingüística aunque son relativamente nuevos, promueven avances considerables.

Debido a que la relación que existe entre Lingüística y Estadística no es muy conocida, surge la inquietud por investigar sobre este tema, además que ha habido avances en la Estadística Lingüística que deben conocerse, como el principio de George Kingsley Zipf denominado “Ley del Mínimo Esfuerzo” que afirma que en cualquier texto, el producto de la ocurrencia de cada palabra por su orden es constante, aunque la Ley de Zipf se ha utilizado más en Lingüística Computacional y en Lingüística Forense, se establece que es posible e importante aplicar la Estadística en el análisis de textos. Las preguntas de investigación que surgieron fueron:

1. ¿Cómo utilizar la Estadística para el análisis de textos?
2. ¿Qué es: Lingüística, Estadística, Estadística Lingüística?
3. ¿Quién fue y qué aporte brindó George Kingsley Zipf a la Lingüística?
4. ¿Qué uso hacen de la Ley del Mínimo Esfuerzo la Lingüística Computacional y la Lingüística Forense?
5. ¿Cuál es el procedimiento para el análisis de textos utilizando la Ley del Mínimo Esfuerzo?

JUSTIFICACIÓN

Actualmente la Estadística y la Lingüística son vistas como dos ciencias completamente diferentes, la primera relacionada a datos numéricos presentados de forma ordenada y sistemática, mientras que la segunda es una ciencia social encargada del estudio del lenguaje humano, que a primera vista, nada tiene que ver con datos numéricos. Sin embargo, aunque sean dos ciencias que se dedican al estudio de campos diferentes, la Estadística es una herramienta útil para la Lingüística en el análisis de textos.

La necesidad de utilizar la Estadística en el análisis de textos fue objeto de estudio de un lingüista de la universidad de Harvard llamado George Kingsley Zipf, quien estableció a finales de los años 40 un principio denominado “Ley del Mínimo Esfuerzo” que establece que en un texto la ocurrencia de una palabra sigue una distribución que puede aproximarse mediante: $P_n \sim 1/n^a$; donde P_n representa la ocurrencia de una palabra ordenada n -ésima y el exponente a es próximo a 1. Lo cual significa que el segundo elemento (palabra) se repetirá aproximadamente con una frecuencia de 1/2 de la del primero, y el tercer elemento con una frecuencia de 1/3 y así sucesivamente. En la actualidad, los aportes del mencionado lingüista son utilizados en Lingüística Computacional y Lingüística Forense.

La realización de este estudio es necesaria ya que brindará las bases teóricas y prácticas del proceso que Zipf empleó y que es utilizado para el análisis de textos en Lingüística Computacional y Forense para determinar la frecuencia de aparición de las palabras dentro del texto.

El estudio beneficiará a los alumnos del Técnico en Traducción y Correspondencia Internacional de la Escuela de Ciencias Lingüísticas, ya que gracias a él podrán conocer el aporte de la Ley de Zipf a la Lingüística y además su proceso de aplicación en cualquier texto, incluso un texto traducido.

MÉTODO

En el presente trabajo de investigación se utilizó un método de investigación cuantitativo, apoyado por la investigación de tipo documental.

Se utilizó el método cuantitativo para analizar tres textos de manera científica, siguiendo los lineamientos planteados por G. K. Zipf para representar los datos obtenidos por medio de tablas y gráficas que representan la frecuencia de aparición de palabras dentro de los textos.

El primer texto seleccionado fue un texto obtenido de una revista turística guatemalteca, ya que los textos turísticos en su mayoría deben ser traducidos hacia otros idiomas, en este caso se analizó el texto en español e inglés. El segundo texto seleccionado fue un artículo de revista con su respectiva traducción al inglés y el último fue una fábula, también en ambos idiomas. Los textos fueron seleccionados porque era importante aplicar el análisis estadístico en diferentes estilos de textos, con el fin de determinar las características y diferencias de cada uno y las de su respectiva traducción.

CAPÍTULO I

Generalidades

1.1 Definición de Lingüística

Según el diccionario de la Real Academia de la Lengua Española, Lingüística es el “estudio teórico del lenguaje que se ocupa de métodos de investigación y de cuestiones comunes a las diversas lenguas”. (Real Academia Española. 2001. *Diccionario de la Lengua Española*. 22º ed. Consultado en: <http://lema.rae.es/drae/>)

La palabra “Lingüística” se deriva del vocablo francés *linguistique*, que hace referencia a todo aquello que está relacionado con el lenguaje y además a la ciencia que tiene como objeto de estudio a la lengua.

La Lingüística es la ciencia del lenguaje articulado y de su funcionamiento como sistema de comunicación entre los seres humanos, se dedica al estudio del lenguaje humano en toda su extensión y abarca tanto la Gramática como la Filología; se dedica al estudio de los procedimientos para analizar una lengua, para conocer su funcionamiento, su organización, la manera en que se relacionan entre sí los elementos que la componen, la forma de adquisición de una lengua y su función en la sociedad.

La Lingüística es, pues, el estudio científico del lenguaje y del conocimiento que los hablantes poseen de ella y es únicamente el ser humano quien tiene la facultad de comunicarse consigo mismo y con los demás mediante el lenguaje.

Los seres humanos hablan entre sí, es decir, intercambian ideas con los demás seres humanos, pero como se mencionó anteriormente, también hablan consigo mismos, aunque no se expresen palabras, ya que el lenguaje es la base del pensamiento humano. Debido a esto es que el lenguaje es la base de la vida intelectual y social de todo ser humano. Es por ello que las lenguas, que no son

más que formas de manifestar el lenguaje humano, juegan un rol importante en la cultura de cualquier pueblo.

1.1.1 Objeto de la Lingüística

La Lingüística es el estudio científico de las lenguas, aporta los instrumentos para conocer una lengua a profundidad y analizar su función dentro de una sociedad. Además, permite determinar sus componentes y la manera en que se organizan para formar los sistemas que la integran. Sumado a esto, la Lingüística facilita los conocimientos para estudiar y analizar las lenguas de acuerdo con su realidad interna y su relación con la sociedad que la habla.

La Lingüística aporta las herramientas para reconocer el papel que una lengua desempeña en la sociedad y busca las soluciones a sus problemas de uso, enseñanza y desarrollo.

Todo esto quiere decir que el objeto de la Lingüística es el estudio científico de las lenguas, de sus circunstancias a través del tiempo, para comprender su realidad, situación y función.

1.1.2 Ramas de la Lingüística

Dependiendo del objeto o de la razón de estudio, la Lingüística posee diferentes ramas. La primera de ellas es la Lingüística General, que se dedica al análisis de la lengua y los aspectos que proporcionan una visión general del lenguaje y de las lenguas. La Psicolingüística es otra de las ramas de la Lingüística pero ésta es un campo interdisciplinario que se encarga de reunir los fundamentos, tanto de la Psicología como de la Lingüística, con el fin de estudiar los procesos mentales durante la adquisición o el uso del lenguaje humano.

Otra rama es la Lingüística Aplicada, y se hace uso de ella cuando se intenta utilizar los conocimientos de Lingüística, que incluyen el estudio de mecanismos para la definición de sistemas ortográficos, los procesos de comunicación y los medios para mejorar o desarrollar una lengua, mientras que el desarrollo y cambio

de las lenguas durante el paso del tiempo en la humanidad es objeto de estudio de la Lingüística Histórica, también llamada Lingüística Diacrónica o Comparativa.

Existe además, un campo muy importante, éste es la Lingüística Pura que estudia los sistemas generales, es decir, la fonología, la morfología, la sintaxis y la semántica. Otra de sus ramas es la Sociolingüística, que se dedica a estudiar la lengua y la relación o el funcionamiento que ésta tiene dentro de la sociedad.

1.1.3 Historia de la Lingüística

Al igual que el ser humano y cualquier otro campo del conocimiento, la Lingüística tiene una historia. Aunque la Lingüística, como ciencia, es relativamente nueva, sus fundamentos se originaron hace miles de años, ya que el hombre siempre ha tenido curiosidad por el estudio de la lengua y de la comunicación.

Una historia detallada de la Lingüística puede encontrarse en: Yáñez Cossio, C. (2007). *Una Introducción a la Lingüística General*. Sin embargo, para efectos de esta investigación la historia de la Lingüística se dividirá de la siguiente manera:

a) Los griegos

Los primeros registros con los que se cuenta se relacionan con el estudio de la formación y composición de palabras, frases y oraciones de textos sagrados para los occidentales. Los primeros estudios de la lengua los realizaron los griegos a través de varias escuelas, cada una con su propia interpretación de la lengua en el Siglo V A.C.

La primera de estas escuelas es la Sofista, que se interesaba en transformar los estudios empíricos de la lengua en medidas, tal y como se hacía en otras áreas y ciencias. Para ello los sofistas propusieron que las oraciones, frases, entre otros tuvieran la misma longitud. Debido a la dificultad de lograr este objetivo esta posición causó muchas reacciones. Es importante resaltar los aportes que los sofistas brindaron al estudio de la lengua, entre ellos: el estudio de los sonidos, la descomposición de palabras en sílabas y la definición de varios

tipos de categoría para el análisis; además se enfocaron en el origen de las palabras y la naturaleza de la relación que podía existir entre las palabras y los conceptos.

Otra escuela es la de los Analogistas, que creían que la lengua era un acto natural y que por ello existía una conexión natural con las cosas y con la realidad, mientras que los Anomalistas sostenían que la lengua es arbitraria o bien el resultado de un acuerdo entre los hablantes.

b) Los latinos

Entre los latinos o romanos sobresalió el gramático Marcus Terentius Varro. Este último le dio seguimiento a la controversia entre los Anomalistas y Analogistas, escribió un tratado acerca de la lengua latina en la que asumía una posición intermedia entre estos dos tipos de pensamiento. Varro sostenía que la lengua tenía correspondencia con el mundo, pero que también existían irregularidades que se debían a las decisiones humanas. La Gramática latina introdujo el concepto “declinación”, lo que actualmente es conocido como derivación que corresponde a las variantes de una misma raíz.

c) Los indios

El estudio científico de los lenguajes da inicio con los indios, éstos realizaron trabajos importantes relacionados al sánscrito, lengua sagrada en la que estaban escritos sus libros sagrados. Una de las mayores preocupaciones de los indios era la preservación del sánscrito que era considerada como la lengua de los dioses.

Es en Siglo III A.C. cuando el sánscrito, que significa lengua culta, perfecta y correcta, es sustituido por el pácrito que se vierte un manual de Filología con la Gramática del sánscrito. Este trabajo realizado por Panini es importante ya que sienta las bases para la formación de la Lingüística, ya que esta Gramática es uno de los trabajos de descripción de una lengua antigua más sistemáticos y

completos que se conoce y es de esta manera como se marca el inicio de lo que hoy se conoce como Lingüística.

Gracias a este trabajo se aportaron conocimientos para el estudio de las formas orales no registradas por la escritura y también de la relación entre lenguas cuyo origen podía situarse antes del griego, el latín y el sánscrito. Todo esto dio lugar al desarrollo del método comparativo entre las diferentes lenguas habladas en el mundo.

Para la historia de la Lingüística, el sánscrito es importante por dos motivos: el primero es que fue la primera lengua sometida al análisis científico y el segundo motivo es que el estudio que se hizo acerca de esta lengua sentó las bases para la Lingüística Comparativa.

d) La Edad Media

La Lingüística durante la Edad Media se torna más descriptiva y normativa y se continúa con la formación griega y romana estableciéndose el aprendizaje del latín. La Gramática es considerada como un arte, por ejemplo la Etimología tiene un gran auge y se da un importante trabajo de planificación lingüística en los distintos reinos, que culmina con la consolidación de la lengua romance.

Durante esta época la Escuela Monista también desarrolla una tarea lingüística, ya que consideran que cada parte de la oración se caracteriza por representar una parte de la realidad de un modo determinado, de aquí abordan cuestiones sintácticas, morfológicas y semánticas, así como la función metalingüística del lenguaje y el concepto de "significado".

e) El Renacimiento

Durante este período aparece la Gramática del castellano de Antonio de Nebrija, que muestra una intención normativa de la lengua, sin embargo, aparecen gramáticas de lenguas vernáculas e indígenas. Además decae el interés por el estudio del latín.

f) La Ilustración

Durante la Ilustración aparece la Gramática de Port-Royal. Debido al origen del lenguaje y sus relaciones con el pensamiento, este se haya dividido entre hipótesis racionalistas e hipótesis empírico censistas. Muchos autores tratan de distinguir un isomorfismo (la misma forma o características similares) entre lengua, pensamiento y realidad.

g) La Lingüística Moderna

Tiene su comienzo en el Siglo XIX con las actividades de los neogramáticos, que con el descubrimiento del sánscrito pudieron comparar las lenguas y reconstruir una lengua original. Surge la figura de Ferdinand de Saussure que establece a la Lingüística como ciencia a la que a su vez integra la Semiología que forma parte de la Psicología Social.

El aporte de Saussure como padre de esta nueva ciencia fue la distinción entre lengua y habla y la definición del signo lingüístico.

En el Siglo XX el lingüista Noam Chomsky crea el generativismo, que consiste en que la capacidad de aprender una lengua es innata.

Según la escuela generativista, toda propuesta de modelo lingüístico debe adecuarse al problema global del estudio de la mente humana. Tanto la escuela de Chomsky como la de Saussure se plantean como objetivo la descripción y explicación de la lengua como un sistema autónomo. Ambas se enfrentan con la escuela funcionalista, que surgió a finales del siglo XX, ya que esta tiene como premisa que el lenguaje no puede ser estudiado sin tener en cuenta su principal función, que es la comunicación humana.

La Lingüística Generativista y la Funcionalista configuran el panorama de la Lingüística actual, ya que ambas tratan de explicar la naturaleza del lenguaje.

1.2 Definición de Estadística

La palabra “Estadística” suele utilizarse con dos significados distintos, el primero como colección de datos numéricos, sin embargo, este es el significado más vulgar de la palabra “Estadística”; se entiende que estos datos deben estar presentados de forma ordenada y sistemática, pero una información numérica cualquiera no puede precisamente constituir una Estadística, ya que para que pueda ser llamada Estadística debe constituir un conjunto coherente, establecido de forma sistemática y siguiendo un criterio de ordenación.

El segundo significado que se le da a la palabra “Estadística” es como ciencia, en este sentido la Estadística estudia el comportamiento de los fenómenos de masas. Al igual que otras ciencias, busca las características generales de un colectivo y prescinde de las particularidades de cada elemento.

El Diccionario de la Real Academia de la Lengua Española define a la Estadística como la “rama de la matemática que utiliza grandes conjuntos de datos numéricos para obtener inferencias basadas en el cálculo de probabilidades”. (Real Academia Española. 2001. *Diccionario de la Lengua Española*. 22^o ed. Consultado en: <http://lema.rae.es/drae/>)

Podemos decir entonces que la Estadística es la rama de la Matemática que tiene como objetivo recopilar, organizar y procesar información cuantitativa de una determinada característica de la población, series o hechos, que posteriormente se organizan y se presentan gráficamente para que sean analizados y ayuden a tomar decisiones más efectivas.

1.2.1 Historia de la Estadística

Aunque resulte difícil de creer, formas sencillas de Estadística han estado presentes desde los inicios de la civilización.

Una historia detallada de la Estadística puede encontrarse en: Araya Alpízar, C. (2004). *Estadística para Laboratorista Químico*. Sin embargo, para efectos de esta investigación se dirá que en el año 3,000 A.C. en Babilonia se

empezaron a utilizar tablas de arcilla para recopilar datos acerca de la producción agrícola y de los géneros vendidos o intercambiados. Aunque, el inicio de la Estadística puede darse en el antiguo Egipto, en donde los faraones lograron recopilar datos relativos a la población y a la riqueza del país hacia el año 3,050 A.C. Según Heródoto, el mencionado registro de riqueza y población fue hecho para preparar la construcción de las pirámides, además Ramsés II hizo un censo de las tierras para verificar un nuevo reparto.

Es importante resaltar dentro de la historia de la Estadística que en los Libros Bíblicos también, por ejemplo en el Libro de los Números, se hace referencia a datos estadísticos, el primero contiene dos censos de la población israelita y el segundo describe las riquezas que poseían las tribus judías.

Asimismo, los chinos llevaban registros numéricos mucho antes del 2,000 A.C., éstos efectuaron censos hace más de cuarenta siglos, mientras que los griegos efectuaron censos periódicamente con el fin de cobrar impuestos y otros fines sociales y militares. Datos históricos revelan que se realizaron 69 censos para calcular los impuestos, determinar los derechos de voto y ponderar la potencia guerrera.

No obstante, fueron los romanos los maestros de la organización política, quienes mejor supieron emplear los recursos de la Estadística. Estos realizaban censos de población cada cinco años y sus funcionarios debían anotar nacimientos, defunciones y matrimonios, sin dejar de lado los recuentos de ganado y de las riquezas de los pueblos conquistados. Es tan notorio el avance que realizaron los romanos en materia de Estadística que se sabe que para el nacimiento de Jesucristo se llevaba a cabo un empadronamiento de la población.

Durante los mil años que le siguieron a la caída del imperio romano se realizaron pocas actividades estadísticas, excepto las relacionadas a las tierras y propiedades de la Iglesia. Durante el siglo IX se llevaron a cabo en Francia algunos censos de siervos. En Inglaterra, Guillermo recopiló el Libro del Gran

Catastro que era un documento de la propiedad, extensión y valor de las tierras de Inglaterra.

A pesar de que en Inglaterra y en Francia se intentó revivir la técnica romana, durante la edad media los métodos estadísticos permanecieron casi en el olvido.

Durante los siglos XV, XVI y XVII personajes como Leonardo da Vinci, Nicolás Copérnico, Galileo Galilei, René Descartes, entre otros, realizaron grandes operaciones utilizando el método científico, de manera que cuando fueron creados los Estados Nacionales y surgió el comercio internacional, ya existía un método capaz de aplicarse a los datos económicos.

En el año 1,532 en Inglaterra se empezaron a registrar las defunciones debido al temor de una peste. En la misma época en Francia se les exigió a los clérigos registrar los nacimientos, defunciones y matrimonios. A finales de la década de 1,500, a raíz de una peste, el gobierno inglés comenzó a publicar los decesos semanalmente. Esta costumbre continuó hasta que en 1,662 el capitán John Graunt utilizó documentos que abarcaban treinta años y efectuó predicciones sobre el número de muertes que habría sobre las proporciones de nacimientos de varones y mujeres que se podía esperar. Este último fue un innovador esfuerzo en el análisis estadístico.

En el año 1,540, Sebastian Muster realizó una compilación estadística de los recursos nacionales de Alemania y aportó indicaciones concretas acerca de métodos de observación y análisis cuantitativo de la teoría estadística.

En el siglo XVII se mostró interés por la Estadística Demográfica debido a la especulación sobre si la población aumentaba, disminuía o permanecía estática.

En la época moderna los mencionados métodos fueron puestos en práctica por algunos monarcas que tenían necesidad de conocer las riquezas y el potencial humano de sus naciones. En 1,961 se llevó a cabo el primer empleo de datos numéricos para fines políticos, el cual estuvo a cargo de Gapar Neumman. Este último se propuso destruir la idea de que en los años que terminaban en siete

moría más gente y para lograrlo revisó en los archivos parroquiales de la ciudad de Breslau. Al terminar de revisar los archivos logró demostrar que en esos años no fallecen más personas. Actualmente sus cálculos sirven como base para las tablas de mortalidad utilizados en todas las compañías de seguros.

Durante los siglos XVII y XVIII se desarrolló la teoría de probabilidades, aunque solo eran utilizados en juegos de azar, no fue hasta el siglo XVIII cuando se utilizó en problemas científicos.

La palabra “Estadística” fue acuñada por el profesor Godofredo Achenwall. Etimológicamente la palabra viene del italiano *statista* que quiere decir estadista. Es Jaques Quételet quien aplica la Estadística a las ciencias sociales, para resolver la aplicación del principio de promedios y de la variabilidad de los fenómenos sociales.

Durante el año 1,800 al 1,820 se desarrollan dos conceptos fundamentales para la teoría estadística; la teoría de los errores de observación y la teoría de los mínimos cuadrados. Aunque los progresos más recientes en Estadística se refieren al desarrollo del cálculo de probabilidades.

En la actualidad la Estadística se ha convertido en una ciencia muy útil que se utiliza para describir con exactitud valores de datos, ya sean, políticos, sociales, económicos, entre otros y a su vez también funciona como herramienta para relacionar y analizar dichos datos.

1.2.2 Aplicaciones de la Estadística

La Estadística puede ser aplicada en cualquier ciencia siempre que se requiera o se necesite el manejo y el análisis de datos, es decir, la información extraída de la realidad, así que puede aplicarse Estadística en campos muy variados, su aplicación puede llevarse a cabo tanto en Ciencias de la Naturaleza, Economía y Sociología como en Medicina, Ciencias de la Salud y por ende en Lingüística.

1.2.3 División de la Estadística

La Estadística es un conjunto de datos numéricos que permiten la representación y análisis de un colectivo. La Estadística se divide en dos grandes ramas:

- Estadística descriptiva, que son todos los métodos estadísticos que intentan describir las características más importantes de un conjunto. También se dedica al estudio de la forma de obtención y descripción de los datos relevantes.
- Estadística inferencial, generaliza los resultados de una muestra a toda la población, mediante el planteamiento de hipótesis, esto quiere decir que pretende conocer cómo se distribuye una variable en la población, para poder extraer conclusiones generalizadas.

1.3 Ciencias auxiliares y afines de la Lingüística

Como ya se mencionó, el lenguaje es exclusivamente humano y por ende se encuentra implicado en varios aspectos de la vida del hombre, es por ello que puede ser estudiado desde diversas perspectivas, todas ellas muy válidas para tener un conocimiento más pleno de esta facultad eminentemente humana.

Son muchas las ciencias que apoyan y tienen una estrecha relación con la Lingüística, que es la ciencia que se dedica al estudio del lenguaje. Gracias a esto la Lingüística posee ciencias auxiliares y afines. Consecuentemente, es de vital importancia saber cuál es la diferencia entre una ciencia auxiliar y una ciencia afín. La diferencia radica en que una ciencia auxiliar es una ciencia autónoma, aquella que funciona como soporte o bien que presta un servicio a otra ciencia para que cumpla con todas sus metas y objetivos brindándole elementos esenciales para su estudio, mientras que una ciencia afín guarda una relación de intercambio, aportándole sus métodos y conclusiones a la otra ciencia para que la nutran o la apoyen.

Entonces se puede decir que las ciencias auxiliares de la Lingüística tienen sus propios objetivos y métodos pero le prestan un servicio, mediante su aporte de los elementos que le son útiles, tal como la Estadística; y las ciencias afines que mantienen un intercambio con la Lingüística, ya que la apoyan brindándole sus métodos y conclusiones.

Las ciencias auxiliares de la Lingüística son muy numerosas, entre ellas podemos citar la Fisiología, Acústica, Lógica, Psicología y una que es relativamente más reciente dentro de los estudios lingüísticos, la Estadística; y dentro de las ciencias afines podemos encontrar a la Literatura, Estilística, Retórica, Filosofía e Historia.

1.4 Estadística Lingüística

A primera vista pareciera improbable que la Estadística y la Lingüística guarden alguna relación entre sí, ya que la primera está relacionada con datos numéricos que se presentan de una forma ordenada y sistemática, mientras que la Lingüística se encarga del estudio del lenguaje humano. Sin embargo, aunque ambas ciencias presenten características muy diferentes, la Estadística es una herramienta sumamente útil para la Lingüística.

Antes de asegurar que ambas ciencias no se relacionan entre sí, se debe recordar que una de las características del lenguaje humano es que este se basa en secuencias y estas secuencias pueden arrojar un número o números que posteriormente pueden ser utilizados por la Estadística para ordenar o aclarar fenómenos del lenguaje.

La Estadística Lingüística no es más que la aplicación de los métodos estadísticos para la explicación de fenómenos lingüísticos, podría decirse que la Estadística es una ciencia auxiliar de la Lingüística porque por medio de sus objetivos y métodos le brinda un servicio. Los recursos estadísticos pueden aplicarse en la investigación de cualquier campo de las ciencias, entre ellas las ciencias sociales incluyendo la Lingüística.

Se ha abarcado ya la historia tanto de la Estadística como de la Lingüística y queda establecido que ambas han acompañado a la humanidad desde hace miles de años, aunque cada una como ciencia independiente y no haciendo referencia a la utilidad que una tiene de la otra. Cuando ya ambas estaban completamente establecidas y eran reconocidas como ciencia, se empezó entonces a utilizar a la Estadística como herramienta para la investigación lingüística que promueve avances significativos, ya que se hace posible establecer en el lenguaje determinaciones cuantitativas y permite a los investigadores estudiar los rasgos de ciertos fenómenos lingüísticos y sus relaciones. La Estadística permite demostrar que el lenguaje responde en muchos aspectos a mecanismos que se adaptan a reglas que sirven para conseguir la correcta transmisión de información, esto desarrollaría la teoría de la información y los lenguajes computacionales.

La Estadística permite abordar cuestiones de estilo, la longitud y frecuencia de las palabras e incluso establecer la probabilidad de que aparezca una sílaba en el seno de una lengua. Aunque el principal objetivo de estudio de la Estadística en la investigación lingüística es la palabra, esto ha sido objeto de interés desde hace muchos años, por ejemplo, los estudiosos del Talmud contaban las palabras y las ideas de la obra para averiguar con qué frecuencia aparecían formas poco usuales.

La frecuencia de aparición de palabras se asocia a George Kingsley Zipf, ya que él estableció un principio que afirma que en cualquier texto, el producto de la ocurrencia de cada palabra por su orden es constante.

Después de que Zipf formulara este principio, la Lingüística moderna cuenta con una gran cantidad de material proveniente de la observación y descripción de hechos, relaciones, leyes y comparaciones que han esparcido la capacidad de tratamiento de los problemas por métodos tradicionales y han elevado la metodología basada en modelos estadísticos.

Todos estos estudios sobre Lingüística que hace la Estadística han hecho surgir a la Estadística Lingüística que sirve como solución a problemas aún más

nuevos que la misma Estadística Lingüística, como la localización en *internet* y determinar la autoría de un texto mediante su análisis.

CAPÍTULO II

George Kingsley Zipf

La biografía detallada de G. K. Zipf puede encontrarse en la revista científica *Glottometrics* 3, To Honor G. K. Zipf (2002).

2.1 Antecedentes familiares

George Kingsley Zipf, hijo de Oscar Robert Zipf y de Maria Luisa Bogardus Zipf tenía descendencia alemana, ya que su abuelo Frederick Sebastian Zipf era originario de un pueblo de Alemania. La familia era eminentemente católica romana y como costumbre participaban en el coro de la iglesia, además eran dueños de la cervecería local. Su abuelo llegó a los Estados Unidos de Alemania alrededor del año 1,850, se casó con una muchacha que también tenía descendencia alemana, llamada Ottilia con quien tuvo varios hijos, incluyendo a Oscar Robert Zipf, el padre de George Kingsley Zipf, que nació en Kankakee, Illinois. Robert Zipf asistió a la escuela de leyes de la de la Universidad de Michigan de donde se graduó. Rápidamente empezó a ejercer su profesión y se casó con María Luisa Bogardus

Maria Luisa Bogardus, madre de Zipf, era hija de Charles Bogardus y de Hannah Wittaker Pells. Bogardus nació en Nueva York el 10 de septiembre de 1,865.

Zipf tuvo tres hermanos; Oscar Robert, Charles y Theodore. El primero sirvió en la fuerza aérea durante la Primera y Segunda Guerra Mundial, Charles se desempeñó en Freeport como médico durante toda su vida y su hermano menor, Theodore murió víctima de la gran pandemia de gripe de 1,918.

2.2 Biografía

El lingüista y filólogo estadounidense George Kingsley Zipf nació el 7 de enero de 1,902 y era hijo de Oscar Robert Zipf y Maria Luisa Bogardus Zipf. Asistió a la secundaria pública Freeport ubicada en la ciudad de Nueva York, posteriormente estudió en la Universidad de Harvard de donde se graduó en

1,924 con el reconocimiento *summa cum laude* debido a su desempeño académico. Después de haberse graduado viajó a la República Federal de Alemania para estudiar en la Universidad de Berlín y fue ahí donde surgió su idea de estudiar al lenguaje como un fenómeno natural. Al término de sus estudios en Alemania regresó a los Estados Unidos y obtuvo su Doctorado en Filología Comparativa en la Universidad de Harvard. Presentó la tesis "*Relative Frequency as Determinant of Phonetic Change*" (Zipf, 1,929). Después de obtener su doctorado empezó a trabajar como catedrático de alemán en la Universidad de Harvard, pero al mismo tiempo continuaba realizando sus investigaciones lingüísticas, lo que le permitió en 1935 publicar su primer libro: "*The Psycho-Biology of Language*" (Zipf, 1,965).

Después de la publicación de su primer libro Zipf amplió su interés acerca de varios fenómenos sociales, lo que dio lugar a la publicación de su segundo libro denominado "*National Unity and Disunity*" (Zipf, 1,941).

Quiso trasladarse a Washington para trabajar durante la Segunda Guerra Mundial pero, terminó quedándose en Harvard en donde cambió la metodología de enseñanza de alemán e hizo énfasis en el aprendizaje de vocabulario, lo que permitió a sus estudiantes que al final del año pudieran leer alemán con la ayuda de un diccionario. Para Zipf esto representaba un avance en la enseñanza de idiomas, así que presentó su método acompañándolo de las razones que lo justificaban y recomendaban.

Además de siempre haber sido un excelente estudiante, Zipf también logró ser un buen catedrático que se preocupaba por el aprendizaje de sus estudiantes y por la investigación lingüística, así que sus éxitos como catedrático fueron sumamente elogiados. Debido a que esto sucedió durante la Segunda Guerra Mundial, Zipf sabía que los estudiantes quizá no tendrían mucho tiempo para estudiar alemán en Harvard, sino que se unirían a las fuerzas armadas, lo que lo impulsó a aplicar este método de enseñanza a sus estudiantes.

Zipf, apasionado por sus investigaciones continuó investigando sobre las relaciones sociales, pero esta vez ayudado por su hijo Robert, quien aplicaba cálculos estadísticos manuales ya que aún no existían computadoras.

Aunque Zipf tenía descendencia alemana, sus estudios e interés acerca del alemán nada tenían que ver con eso porque en su familia y hogar no hablaban alemán sino inglés. Igualmente, se interesó por el trabajo de George Meredith, un novelista y poeta inglés que había estudiado en Alemania, el fruto del interés de Zipf en Meredith terminó con su publicación de "*New Facts in the Early Life of George Meredith*" (Zipf, 1,938) que trataba sobre cómo la vida de las personas se incorporaba en los procesos y sistemas.

El 20 de junio de 1,931 Zipf contrajo matrimonio con Joyce Waters Brown proveniente de Massachusetts y descendiente de cuatro gobernadores de la colonia, Waters al igual que Zipf fue una estudiante destacada que trabajaba en la Oficina de la Asociación de Estudiantes de Harvard. Ambos tuvieron cuatro hijos, el mayor Robert, la segunda Katharine, la tercera Joyce y el último Henry, el haber tenido tantos hijos resultó en que Zipf y Waters aún hoy en día tengan descendencia.

Después de su matrimonio con Waters, Zipf se mudó a Cambridge, después a Duxbury y por último a Newton, en donde se establecieron para que sus hijos asistieran a la escuela.

Zipf dio múltiples conferencias en la universidad gracias a sus investigaciones, no solo de la enseñanza de idiomas sino también en las ciencias sociales, lo que le permitió ser conocido y respetado por la comunidad académica. Zipf continuó con sus conferencias hasta el momento de su trágica muerte, después de haber obtenido una beca Guggenheim, que es entregada únicamente a profesionales avanzados en diferentes áreas del saber.

Para ganar una beca Guggenheim, en aquel entonces los candidatos debían someterse a un examen físico y fue en el proceso que a Zipf se le diagnosticó cáncer. Se le intervino quirúrgicamente, pero el cáncer estaba muy

avanzado, así que regresó después de un mes a su casa hasta que falleció el 25 de septiembre de 1,950 y sus cenizas fueron enterradas en el cementerio de Mayflower en Duxbury.

2.3 Vida familiar

Durante el tiempo que Zipf residió en Duxbury trabajó desde su casa, pero tomaba el tren dos o tres veces por semana y viajaba a Cambridge para trabajar y dar clases en la universidad. Él y su familia vivían cómodamente, lo demuestra su casa que estaba construida en un terreno de siete hectáreas y media en medio de un bosque. Debido a las condiciones de vida que observaba en el resto de ciudadanos estadounidenses y a la gran propiedad de la que era propietario, durante la guerra crió pollos, patos y cerdos, asimismo tenía un huerto en donde sembraba y cosechaba varios vegetales y frutas, Zipf hacía trabajar a toda la familia en el huerto aunque no todos sentían agrado de hacerlo. No conforme con su huerto y la crianza de animales, la familia además tenía un viñedo y un gran jardín cerca de su casa. Aunque utilizan mucha tierra para realizar todas estas labores quedaba mucho bosque de pinos de donde con su hijo mayor, Zipf seleccionaba y talaba un árbol cada navidad para decorar su casa.

Años después la familia se mudó a Newton a una casa de verano cerca de la playa, pero Zipf solo pudo disfrutarla durante dos o tres veranos antes de su fallecimiento. En esta nueva propiedad cerca del mar, la familia tenía un bote en el que solían navegar, cuando Zipf no estaba ocupado escribiendo su obra principal, "*Human Behavior and the Principle of Least Effort*" (Zipf, 1,972), en el cual trabajó arduamente durante mucho tiempo, en el tercer piso de su casa de Newton que convirtió en oficinas. Su obra la trabajaba principalmente con libros de idiomas, libros de alemán pero basados en diversos temas, incluyendo un libro de la gramática sánscrita.

Zipf era una persona sumamente entusiasta, original y cordial, actitudes que lo hicieron muy popular entre sus estudiantes, a quienes involucraba en su trabajo estadístico, por eso frecuentemente en sus publicaciones mencionaba los

nombres de aquellos que lo habían ayudado a recopilar la información y a hacer la estadística.

2.4 Lingüística de Zipf

Durante el siglo pasado, además de que se estudió a la Lingüística utilizando muchas otras ciencias, con el objetivo de explicar ciertos fenómenos se hicieron muchos descubrimientos, investigaciones y avances relacionados a esta ciencia, que dieron como fruto el nacimiento de nuevas disciplinas, entre ellas, la Estadística Lingüística, que apareció gracias a los estudios de G. K. Zipf. Sin embargo, en la evolución de la Lingüística se han dado controversias y anomalías porque la mayoría de las ciencias persiguen un único objetivo, y es que cada una de las ciencias afines lo que intentan es describir solamente una parte del lenguaje, es decir, que intenta dividir a la lengua en varios estudios como la Morfología, la Sintaxis, la Fonología o su dimensión espacial y social. Esto quiere decir que los métodos que se utilizaron para su estudio solo describían y estudiaban al lenguaje superficialmente y se dejó de lado el estudio de las relaciones dinámicas que caracterizan al lenguaje humano. Aquí es donde entra el trabajo de G. K. Zipf, ya que hasta el momento no se estaba preparado para enfrentarse a cuestiones que nada tuvieran que ver con la Gramática o la Semántica, como la frecuencia o la longitud de las unidades lingüísticas.

Esto sucedía porque no se creía que el lenguaje pudiera ser medido, pero G. K. Zipf fue el primero en aplicar métodos estadísticos para el análisis lingüístico.

Seguramente, antes de que Zipf empezara a aplicar Estadística a textos, se intentó explicar de muchas maneras los fenómenos lingüísticos, pero éstos no eran suficientes para lograr satisfacer las demandas de la Lingüística que se dice es el estudio científico del lenguaje.

Durante el siglo XX, en pleno auge del Estructuralismo, Zipf se atrevió a estudiar la dinámica interna de la lengua y descubrió los mecanismos responsables de los fenómenos lingüísticos que otras ciencias afines y auxiliares de la Lingüística no habían podido identificar. Al identificar ese “algo” que cuidaba

el movimiento del lenguaje, por supuesto fue víctima de críticas, pero nadie continuó investigando el tema, ya que no se pensaba que la Estadística pudiera aplicarse a fenómenos lingüísticos porque el lenguaje se origina en la mente de los seres humanos, lo que lo convierte en algo impredecible.

En todas las escuelas lingüísticas había objeciones en contra del principio propuesto por Zipf, pero estableció una gran inquietud sobre la relevancia que éste tenía en los estudios lingüísticos. Aunque su descubrimiento provocó muchas críticas, Zipf y su principio, con el paso de los años se hacen cada vez más populares y se trata de mejorar y aplicar sus métodos, pero son los lingüistas cuantitativos los que más se esfuerzan por sistematizar sus ideas.

La idea de Zipf que se extendió entre los científicos se refiere únicamente a que la frecuencia de las entidades dependen de su rango, pero a Zipf no le interesaba la convergencia sino únicamente estudiaba el conjunto de datos, que según él se podían encontrar con cierta frecuencia, pero más frecuentemente aquellos que no siguen la curva de potencia simple. Así que se le plantearon dos retos a los científicos: primero encontrar una curva o distribución de probabilidad para obtener el rango y el segundo era resolver el fundamento de la regularidad de esos elementos, es decir, el por qué se utilizan entidades lingüísticas de una clase en dosis tan bien proporcionadas.

En la Lingüística, Zipf establece que el rango se presenta automáticamente en el proceso de la comunicación y que eso controla la percepción y la decodificación del mensaje, lo que se debe posiblemente al almacenamiento de palabras en el cerebro y, aunque Zipf lo que quería era explicar el fenómeno de aparición de palabras en el discurso, su principio fue utilizado en otras ciencias.

La importancia del descubrimiento de la ley de frecuencias es importante para otras ciencias, pero para ninguna tan importante como para la Lingüística. Zipf no se detuvo después de establecer su ley de frecuencias sino que unió la ley de frecuencias con otras propiedades del lenguaje: frecuencia y duración, frecuencia y polisemia, entre otros. Cabe mencionar, que a pesar de sus valiosos

descubrimientos, Zipf no era un matemático, en su tiempo la teoría de sistemas tenía poca relevancia, pero intentó buscar el apoyo que la Lingüística podía encontrar en la Estadística, a esto es a lo que posteriormente a sus descubrimientos se le llamó “Lingüística de Zipf”.

Gracias a todas las conexiones que estableció entre su ley de frecuencias y otros fenómenos lingüísticos, Zipf propuso su muy famoso principio llamado “Ley del Menor Esfuerzo”.

2.5 Ley de Zipf

Si bien es cierto que el lenguaje se produce dentro de la mente del ser humano y que por ende es impredecible, Zipf aplicó un método estadístico que le permitiera medirlo, mediante la frecuencia de aparición de una palabra y su orden. En el año 1,949, Zipf publicó *Human Behavior and the Principle of Last Effort*, en éste libro Zipf sostenía que el comportamiento humano se rige por la reducción de esfuerzo. Esto quiere decir que, el ser humano tiende a economizar palabras para lograr la comunicación y de esa manera realizar un esfuerzo menor al decodificar el mensaje.

2.5.1 Economía Lingüística

La Economía Lingüística fue el objeto de estudio de Zipf y gracias a sus estudios determinó cinco principios dinámicos del habla, que son:

- Dinamismo de un sistema fonético, referido a que la dificultad de una comunidad lingüística de pronunciar un fonema era inversamente proporcional a la frecuencia con la que aparece el fonema.
- Ley generalizada de abreviación, se refiere a que no únicamente los fonemas se enfrentan a la relación inversamente proporcional, sino que se aplica a todos los elementos del habla. Esto quiere decir, que mientras más larga es la palabra menos veces se va a utilizar, de esta manera se puede explicar la relación entre longitud y frecuencia que planteó Zipf.

- Efecto de la analogía, se refiere a que los cambios analógicos reducen el número de clases que se desarrollan al momento de hablar.
- Distribución de frecuencias de las palabras, este es el principio que más interesa en este trabajo, ya que se estudió mediante análisis estadísticos.
- La distribución de los significados de las palabras, que es un estudio de los diferentes significados de las palabras y que se relacionan con la frecuencia de uso de las mismas.

Así pues, se puede decir que el ser humano recurre al menor esfuerzo para comunicarse, tanto de forma oral como de manera escrita. Esto no quiere decir que el ser humano haga un trabajo menor sino que el esfuerzo que realiza al hablar es menor.

2.5.2 Frecuencia de aparición de las palabras en un texto

Zipf en la década de los 40 realizó una serie de estudios empíricos para demostrar su Ley del Mínimo Esfuerzo, que defiende el principio de que las personas cuando escriben utilizan mucho más las palabras cortas que aquellas más largas.

En textos de diversa índole se puede observar que si todas las palabras se ordenan de acuerdo a su frecuencia u ocurrencia decreciente y se les asigna un orden, el producto total de la ocurrencia por el orden arrojará una constante, así: $f \times r = c$ en donde f es la ocurrencia (frecuencia), r el orden y c la constante y determinará que la palabra que está después de la que tiene la mayor frecuencia tendrá la mitad de la frecuencia de la primera y así sucesivamente. Zipf, el precursor del tratamiento estadístico en fenómenos lingüísticos mediante los resultados de esta ecuación comprobó su principio del menor esfuerzo.

CAPÍTULO III

Uso de la Ley de Zipf en la Lingüística Computacional

3.1 Importancia de la Estadística en el estudio del lenguaje

Es importante destacar que en una ciencia social como la Lingüística puede aplicarse la Estadística, porque no se ha utilizado extensamente y porque cuando se habla de Estadística suele pensarse en una colección de datos numéricos que nada tienen que ver con el lenguaje humano. Pero en la cotidianidad se puede observar la utilidad de Estadística en diferentes ámbitos de la vida humana; esto se traduce en que se está usando a la Estadística como herramienta para explicar y darle solución a muchos problemas de la humanidad, ¿por qué no utilizarla entonces para explicar fenómenos lingüísticos?

Como se mencionó al inicio de este trabajo, han existido formas de Estadística primitiva desde los inicios de la civilización. Muchas civilizaciones antiguas la utilizaron para estudiar y tomar decisiones respecto a las poblaciones, los impuestos, los índices de mortalidad, entre otros; lo que significa que la Estadística además de haber acompañado los estudios de civilizaciones antiguas también fue útil para explicar y darle solución a problemas de la vida del ser humano, hasta que con el paso de los años se consolida como una ciencia.

Después de la consolidación de la Estadística como ciencia, se va utilizando para la explicación y solución de problemas de otras ciencias como la Física, pero es muchos años después que se aplica en los estudios de la lengua, lo que permite que se hagan avances considerables en Lingüística. La utilización de Estadística en la investigación lingüística permite establecer en el habla determinaciones cuantitativas y además permite fijar rasgos de ciertos fenómenos lingüísticos y sus relaciones. La Estadística permitió demostrar que el lenguaje humano corresponde a mecanismos que se ajustan a reglas fijas para conseguir la correcta transmisión del mensaje y de la información. Este hecho resulta crucial para el desarrollo de los lenguajes computacionales.

En los estudios de Estadística Lingüística el objeto de estudio es la palabra, como se abordó en el tema de la Estadística Lingüística, los primeros en aplicar empíricamente Estadística a textos fueron los estudiosos del Talmud, que contaban palabras de la Tora con el fin de determinar su frecuencia, o sea, la cantidad de veces que aparece una forma inusual en un texto, pero cuando se habla de frecuencia de aparición de palabras se asocia con el filólogo G. K. Zipf, ya que estableció una ley para este fenómeno, que afirma que en cualquier texto, el producto de la frecuencia de cualquier palabra por el rango es constante.

Hay que recordar que la Estadística recaba información de colectivos, poblaciones, entre otros, que comparten alguna característica y una de ellas puede ser la lengua. Sin importar si se trata de Lingüística Teórica o Computacional, la importancia de la aplicación de métodos estadísticos a la Lingüística puede darse en la adquisición del lenguaje porque durante este proceso se dan muchas variaciones, en el cambio lingüístico porque con el paso del tiempo una lengua puede ir evolucionando, en la variación lingüística que puede ser tanto intralingüística como extralingüística, es decir, el grupo de lenguas y sus variaciones dialectales y en la competencia frente a la acción porque se dice que los aspectos interesantes del lenguaje son los conocimientos homogéneos que poseen todos los hablantes de una lengua, aunque en este caso los lingüistas defienden la postura de que los métodos estadísticos no ayudan a estudiar la competencia, pero la contraparte piensa que estudiar únicamente la competencia es limitar el estudio del lenguaje porque hay aspectos de la actuación que son lingüísticos.

3.2 Definición de Lingüística Computacional

La Lingüística Computacional o Procesamiento del Lenguaje Natural (PLN) es una ciencia interdisciplinar que se dedica tanto al estudio de los sistemas de comunicación como a los sistemas que se emplean para la comprensión de lenguas naturales. Estudia las lenguas naturales y los ordenadores, su objetivo es crear modelos computacionales del lenguaje que estén tan bien detallados que

permitan crear programas informáticos que realicen tareas en las que intervenga el lenguaje natural.

Entonces se puede decir que, la Lingüística Computacional es una ciencia que estudia, tanto las lenguas naturales, como los ordenadores para crear programas informáticos que simulen la capacidad lingüística que es exclusivamente humana.

Para la Lingüística Computacional, el lenguaje es una acción en la cual un emisor y un receptor procesan una información de acuerdo a su conocimiento homogéneo del mundo, es decir, conocimientos que ambos comparten. Aquí es donde entran las tareas de la Lingüística Computacional que son explicar el funcionamiento del proceso lingüístico y de las estructuras cognitivas, para ello se debe crear un modelo de la competencia de los hablantes, que es el conocimiento que los hablantes poseen de su lengua y de la actuación lingüística que además incluya factores extralingüísticos. El modelo debe ser matemático porque en los fenómenos lingüísticos se puede aplicar variedad de modelos matemáticos que arrojarán respuestas acerca del fenómeno que se estudia.

Los modelos matemáticos que se empleen para estudiar los fenómenos lingüísticos pueden ser simbólicos y probabilísticos. Los simbólicos tienen su origen en la teoría de conjuntos y la lógica matemática, están compuestos de símbolos y de reglas y su fin es reflejar la estructura lógica del lenguaje; mientras que los probabilísticos, que fueron desarrollados a partir de la Teoría de la Información y la Estadística, explican las lenguas como un colectivo de situaciones que se presentan con cierta frecuencia.

3.2.1 Similitud entre la mente humana y los ordenadores

La Lingüística Computacional surgió de la similitud que presenta la mente humana y los ordenadores, esta radica en que ambos tienen la capacidad de procesar información mediante conocimientos previos y nuevos que ayudan a tomar decisiones, aunque uno lo hace de forma natural y el otro de forma automática. Esto permite que se puedan explicar cantidad de fenómenos

lingüísticos, porque un ordenador permite hacer pruebas de los modelos de la mente humana. La mayor similitud radica en que ambos tienen la capacidad de procesar entidades con un significante y un significado y, en base a eso, tomar una decisión.

3.2.2 Enfoques y métodos de la Lingüística Teórica y la Lingüística Computacional

Tanto la Lingüística Teórica como la Lingüística Computacional comparten el hecho de que ambas pretenden estudiar y explicar fenómenos lingüísticos, pero sus enfoques y métodos difieren los unos de los otros.

a. Enfoques

La Lingüística Teórica gira en torno a la competencia (conocimiento que los hablantes poseen de una lengua), en los principios gramaticales y en el descubrimiento de la teoría gramatical. La Lingüística ha tratado de explicar todos los fenómenos lingüísticos partiendo de estos tres factores, pero la Lingüística Computacional se encarga de desarrollar programas informáticos para tratar una lengua natural. La dificultad que presentan es que estos programas no abarcan todas las disciplinas del conocimiento humano sino que tienen una aplicación limitada a campos como el legal, el científico, técnico, etc. Últimamente, la Lingüística Computacional ha intentado tener un proceso total del lenguaje, es decir que pretende acceder tanto a la competencia como a la actuación Lingüística mediante la utilización de otras disciplinas como la Psicología y la Lógica, pero ésta es realmente una dificultad porque el lenguaje humano es impredecible y está influenciado por los conocimientos extralingüísticos que los hablantes poseen.

b. Métodos

Para la obtención de datos lingüísticos, la Lingüística Teórica recurre a una introspección para explicar los fenómenos lingüísticos que desea estudiar, pero actualmente la utilización de ordenadores está consiguiendo mejores avances en materia de investigación lingüística.

En la Lingüística Computacional en cambio, se estudian todas las construcciones que deben ser codificadas en el programa para que éste las reconozca, pero utilizando la competencia del hablante y recurriendo a los datos empíricos que le proporciona la Lingüística Teórica. Básicamente estas dos disciplinas se complementan, ya que las dos estudian el lenguaje pero desde diferentes enfoques.

3.2.3 Usos de la Lingüística Computacional

La Lingüística Computacional es una ciencia que tiene muchas áreas de aplicación, éstas son:

a. Sistemas que tratan de procesar las lenguas naturales

- **Traducción automática:** es una de las tareas más difíciles con las que se enfrenta la Lingüística Computacional. La traducción automática es la utilización de sistemas informáticos para la realización de traducciones de una lengua a otra con la intervención y sin la intervención humana. Lo que pretenden los sistemas de traducción automática es prescindir de la intervención humana para traducir textos completos.
- **Recuperación y extracción de información:** lo que se pretende con estas dos tareas es tratar la información que está almacenada en bases de datos. La recuperación de la información se dedica a tomar la consulta de un usuario y elegir entre la base de datos aquellos textos que mejor responden a las condiciones planteadas por el usuario, por su parte la extracción de información lee grandes cantidades de texto para después reconocer la información relevante contenida en dichos textos y trasladarla a un formato para que su recuperación sea sin dificultad, por ejemplo, escanear documentos escritos para llenar una base de datos con esa información.
- **Interfaces hombre-máquina:** es la relación del usuario con el ordenador, su objetivo es que el usuario pueda entender al ordenador en su lenguaje natural para que acceda a todas las funciones para las que ésta fue diseñada.

b. Sistemas que ayudan en las tareas lingüísticas

- **Herramientas de análisis textual:** son útiles para determinar la frecuencia de aparición de palabras dentro de un texto para distinguir concordancias, realizar estadísticas y realizar el análisis textual que es tarea del lingüista.
- **Herramientas para el manejo de corpus:** estas aplicaciones permiten etiquetar las palabras del corpus según su categoría sintáctica, que permite disponer de una colección de datos lingüísticos bien estructurados.
- **Bases de datos lexicográficas:** la Lexicografía es una rama de la Lexicología que se encarga del vocabulario de una lengua natural y de sus variaciones y que permite la composición de diccionarios, es por ello que es uno de los campos que más puede beneficiarse de la gran capacidad de almacenamiento de los ordenadores. Por medio de las bases de datos lexicográficas puede trasladarse un diccionario normal a un soporte magnético y también permite la creación de diccionarios mucho más completos que los convencionales. Las bases de datos lexicográficas también permiten crear bases de datos terminológicas que son de suma utilidad para traductores sobre información especializada sobre temas específicos.

Este grupo de herramientas es sumamente útil en la tarea de traductores y lingüistas, ya que les facilita algunas tareas propias del campo en el que se desenvuelven.

c. Programas de ayuda a la escritura y composición textual

- **Correctores ortográficos:** cualquier usuario está bien familiarizado con estos correctores ortográficos, ya que todos los ordenadores cuentan con una herramienta dirigida a la revisión del texto, que se encarga de detectar y hasta de corregir errores ortográficos.
- **Correctores sintácticos y de estilo:** son parecidos a los correctores ortográficos pero éstos en lugar de detectar errores ortográficos detectan errores de concordancia entre oraciones y hasta errores de estilo, para ello llevan a cabo un importante análisis sintáctico.

d. Enseñanza asistida por ordenador

Esta aplicación tiene diferentes ámbitos de acción, uno de ellos son los programas educativos que pretenden la enseñanza de idiomas extranjeros, estas herramientas pretenden procesar las lenguas naturales para su adquisición, para lograr su objetivo hacen uso de la combinación de texto, imágenes y voz.

3.3 Lingüística Computacional y su relación con la Ley de Zipf

Para iniciar a describir la relación de la ley de Zipf con la Lingüística Computacional es importante definir primero qué es el análisis documental. El análisis documental es una actividad que tiene como fin responder adecuadamente a las necesidades de información de los usuarios.

La ley de Zipf se relaciona con el análisis documental porque dicha ley es una herramienta estadística que puede utilizarse en los procesos de identificación de términos representativos y el filtrado para describir concretamente el contenido de un documento por medio de la jerarquización de aquellos términos que sean más significativos para que la recuperación de la información sea más fácil.

El análisis documental es muy importante en la recuperación y consulta de información. La relación de la ley de Zipf con la Lingüística Computacional se da en el tratamiento automático de textos, ya que su ley es la que más trascendencia ha tenido en el tratamiento del lenguaje como un fenómeno cuantitativo. Los trabajos de Zipf fueron pioneros en la Lingüística Computacional para el análisis estadístico del lenguaje natural, pero todos sus trabajos fueron hechos en la era pre-computacional así que sus descubrimientos fueron aplicados después de los años 50, cuando empezó el auge de la teoría de la información.

Zipf se dedicó inicialmente a demostrar algunas regularidades en el lenguaje, considerando que algunos fenómenos lingüísticos podían ser estudiados estadísticamente, ya que el lenguaje era un proceso en el que se podía aplicar métodos científicos que eran más utilizados en las ciencias exactas.

La forma en que la Ley de Zipf o “Ley del Mínimo Esfuerzo” se difundió en el ámbito de la ciencia de la información es que si en un texto se cuenta la frecuencia de aparición o uso de una palabra y después se ordena de forma descendente y se le asigna el orden 1 a la más frecuente y así sucesivamente, para después multiplicar el orden por la ocurrencia el resultado será constante. En la escala logarítmica la gráfica de esta distribución es una línea recta cuya pendiente es negativa y forma un ángulo de 45° con el eje de las abscisas, que es la coordenada X en un plano cartesiano que representa la distancia horizontal de un punto al eje vertical o eje Y.

El comportamiento de las palabras está establecido en la ley empírica de Zipf, que consiste en el conteo de palabras dentro de textos relativamente grandes, para después asociarles un orden de acuerdo a sus frecuencias empezando por aquella que tenga la frecuencia más alta o la que más veces se repita. Con su ley, Zipf logró plantear una nueva manera de ver la distribución de uso de las palabras. La naturaleza de la ley de Zipf no es clara, pero describe el comportamiento de las palabras en el texto.

La aplicación de la ley de Zipf en el análisis documental cumple dos funciones, la primera es la recuperación de la información y la segunda la indización automática, que es la descripción del contenido de un documento.

3.3.1 La Indización Automática y la generación de *tesauros*

Al utilizar la implementación de sistemas informáticos y estadísticos se puede generar automáticamente *tesauros*, que en el ámbito de las ciencias de la información se refiere a una lista de términos que sirven para representar el contenido de un documento con el fin de normalizar la terminología para lograr que el canal de acceso y la comunicación con los usuarios de las unidades de información sea más fácil.

A pesar de que se han logrado avances significativos en la generación automática de *tesauros* aún es indispensable la intervención del ser humano. La generación de los *tesauros* está comprendida por dos etapas: la identificación y

adquisición de componentes representativos de un dominio y la obtención de relaciones entre componentes.

La primera etapa mencionada se divide a su vez en varias etapas, siendo estas: el análisis léxico que es la etapa automática en la que se buscan términos que representen un significado colectivo, el tratamiento de palabras vacías (aquellas con menor frecuencia como artículos, conjunciones, adverbios, entre otros), el tratamiento de términos flexionados (aquellos que se relacionan morfológicamente entre sí, es decir aquellos con variaciones de género, tiempo y número), el tratamiento de palabras compuestas, el filtrado de términos, la obtención de relaciones jerárquicas y la agrupación de clases. La Ley de Zipf es utilizada en el filtrado de términos, que empieza una vez, los términos representativos hayan sido identificados para proceder a su filtrado mediante el sistema de filtrado IDF o Indización Estadística de Términos de Frecuencia, ya que está basada en la ley propuesta por G. K. Zipf que establece que las palabras con mayor frecuencia son palabras vacías, pero las menos frecuentes reflejan el estilo y riqueza de vocabulario del autor, las que aparecen en la zona media son las que representan al documento.

CAPÍTULO IV

Uso de la Ley de Zipf en la Lingüística Forense

4.1 Definición de Lingüística Forense

Se sabe que la Lingüística es el estudio científico del lenguaje, pero poco se conoce acerca de la Lingüística Forense que gran auge tiene en la determinación de la autoría de textos.

La Asociación Internacional de Lingüistas Forenses (IAFL) define a la Lingüística Forense como la “interfaz entre el Derecho y el Lenguaje” (IAFL. Citado en Estopá, 2006, p. 71), es decir, todos aquellos aspectos en los que el Derecho y la Lingüística se relacionan. La Lingüística Forense es la disciplina de la Lingüística que utiliza todo tipo de evidencia lingüística en ámbitos como: la identificación de hablantes de una variedad lingüística, estilo o registro, análisis de identificación de falsificación de firmas, determinación de la autoría de textos, el discurso legal y judicial y la interpretación y la traducción legal y judicial.

En realidad se dan dos definiciones de Lingüística Forense, pero en el presente trabajo de investigación la definición que se utilizará es aquella que la restringe a la utilización de evidencia lingüística en juicios con el fin de determinar la autoría de textos.

La Lingüística Forense, al igual que otras áreas del conocimiento, utiliza varias técnicas que se derivan de nuevas tecnologías y, al igual que otras áreas, utiliza herramientas informáticas y técnicas estadísticas que permiten un mejor desempeño y mejores resultados de esta disciplina lingüística.

Durante las décadas de los 60, 70 y 80, en países como Estados Unidos y Canadá, profesionales dedicados a la investigación de crímenes solicitaron la intervención de lingüistas como expertos en Derecho y Lengua; aunque esta práctica profesional fue aislada y sus méritos casi nunca fueron publicados.

En los años 90 se produjo una serie de cambios que reclamaban una nueva área temática, es así como surge la Lingüística Forense. Este hecho se originó gracias a tres fenómenos: el primero la demanda de servicios de Lingüística Forense, en segundo lugar; la publicación de textos de Lingüística Forense; y por último, la creación de la *International Association of Forensic Phonetics* (IAFP) y de la *International Association of Forensic Linguistics*. Todo esto hizo de la Lingüística Forense una nueva disciplina de la Lingüística.

Como se mencionó anteriormente, la Lingüística Forense permite determinar quién es el autor de un texto, para ello los lingüistas forenses analizan aspectos como la forma y el estilo, que usualmente no es importante para las personas, ya que éstas se interesan más por saber si el texto fue escrito a mano, con lápiz o bolígrafo o si está en formato digital. Además, cualquier persona puede notar si en el texto aparecen faltas ortográficas, conocer el vocabulario utilizado y entender el texto si su autor utilizó una forma ordenada de escribir. Pero para un lingüista forense, dentro de un texto hay infinidad de marcas lingüísticas propias de cada persona que permiten determinar si el texto fue copiado o determinar también quién es el autor de un texto.

Los lingüistas forenses pueden determinar si un texto es original o ha sido plagiado porque saben diferenciar las marcas lingüísticas, ya que cuando una persona escribe un texto hace uso de su idiolecto (forma de expresarse propia de cada individuo), lo que lo convierte en un texto único y que pone de manifiesto el estilo de su autor.

4.1.1 El plagio

El plagio es definido por la Real Academia de la Lengua Española como “copiar en lo sustancial obras ajenas, dándolas como propias”. (Real Academia Española. 2001. *Diccionario de la Lengua Española*. 22^o ed. Consultado en: <http://lema.rae.es/drae/>)

El plagio es un préstamo que se hace de una idea o de un texto pero con plena conciencia de que está siendo copiado y consiste en copiar la idea de otro

autor, copiar el texto que expresa tal idea y la copia, tanto de la idea, como del texto original. Actualmente, gracias a la opción *copy and paste* de los programas de computadora el plagio es una situación en la que cualquiera puede caer.

Se pueden dar varios tipos de plagio, entre ellos se pueden mencionar:

- **Plagio de ideas:** consiste en copiar la idea de otro autor sin darle la debida mención al mismo.
- **Plagio palabra por palabra:** consiste en copiar una parte del texto y se considera plagio cuando no se hace referencia de la fuente de donde fue extraído.
- **Plagio de fuentes:** se refiere a incluir las referencias bibliográficas que otro autor ha indicado en su trabajo sin siquiera haberlas consultado.
- **Plagio de autoría:** se da cuando se presume que un texto es de autoría propia cuando en realidad la autoría es de alguien más.

De los cuatro tipos de plagio mencionados, los más difíciles de determinar son los de ideas y los de fuentes. Además de los tipos de plagio también existe una clasificación para las clases de copia que pueden hacerse de un texto, entre ellas: la copia exacta que no incluye ningún tipo de modificación del original, la copia modificada, el plagio traducido y la paráfrasis. A decir verdad, los lingüistas forenses son los únicos que están académicamente preparados para determinar si un texto fue plagiado o no porque conocen y ponen atención a las marcas lingüísticas, es decir, el estilo de cada autor que no puede ser descubierto por cualquier lector a simple vista.

El papel de estos profesionales de la lengua, en el caso del plagio está relacionado con su papel dentro de los juicios con el fin de determinar si dos o más textos son originales o fueron plagiados. Para realizar su trabajo de una manera más eficiente, los lingüistas forenses hacen uso de herramientas informáticas y técnicas estadísticas, una de ellas es la aplicación *CopyCatch Gold* que trata con textos de diferente índole y procesa rápidamente datos para verificar la similitud de los textos, como el vocabulario y resultados netamente estadísticos.

La evidencia lingüística cuantitativa que presenta el lingüista en su reporte le da validez a su investigación y de igual manera a las pruebas que presenta en el juicio.

4.2 Estilo y Estadística

En la investigación estilística, se dan dos vertientes: la Estilística Literaria y la Estilística Lingüística. La diferencia entre ambas radica en que la literaria se dirige hacia el lenguaje literario y la lingüística se centra en el lenguaje y sus registros y variedades.

Se ha mencionado ya, que cuando una persona escribe imprime sus propias marcas, es decir, hace uso de su idiolecto, que no es más que la forma particular de cada persona de utilizar el lenguaje. La Lingüística Forense estudia las propiedades estadísticas del lenguaje, es decir, de la forma en que cada individuo utiliza la lengua. Lo que específicamente estudia la Lingüística Forense es el estilo de cada individuo, lo que se llama “Estilística” que se refiere a la búsqueda de patrones estadísticos repetitivos. El lingüista forense estudia textos de dos autores y en ellos cada uno de los autores pone de manifiesto su propio estilo, mediante la utilización de construcciones sintácticas, formas léxicas o la repetición de ideas con determinada frecuencia. Es aquí en donde puede aplicarse un análisis estadístico para determinar si las frecuencias de uso difieren significativamente o no.

Una de las características más sorprendentes del uso que cada persona hace de su idiolecto, es la regularidad estadística con la que controlan las construcciones sintácticas, formas léxicas o la repetición de ideas con cierta frecuencia, pero son utilizadas inconscientemente, así que el estilo de cada individuo puede ser analizado estadísticamente para determinar la autoría de un texto, porque es imposible que dos personas escriban exactamente igual.

4.3 Importancia de la Estadística en la Lingüística Forense

4.3.1 Recorrido histórico

Desde el principio hubo quienes se interesaron sobre los enfoques estadísticos y científicos de la Lingüística Forense, uno de ellos fue Augustus de Morgan, profesor de Matemática de la Universidad de Londres. De Morgan le escribió una carta al Reverendo Heald, quien se interesaba por conocer la autoría de las cartas bíblicas atribuidas a San Pablo y en la carta le sugería que contara una gran cantidad de palabras en el primer libro de Heródoto y que contara todas las letras; luego dividiera el segundo número entre el primero y ese resultado le arrojaría el promedio de letras por palabra en ese libro. Le sugirió hacer lo mismo en el segundo libro, porque seguramente tendría una aproximación. Si el primer libro daba un total de 5,624 letras por palabra, el segundo libro tendría 5,619 pero seguramente si aplicaba el mismo procedimiento a los libros de Tucídides le daría un promedio más alto de letras por palabra, este fenómeno era atribuido por de Morgan a que ambos tenían diferentes estilos para escribir. Entonces si lo que quería descubrir era la autoría de las cartas bíblicas que se le atribuían a San Pablo que empezaban con 5,428 letras por texto y los hebreos tenían un promedio de 5,516 letras por texto, ésta marcada diferencia en el número de letras por texto podrían dar una evidencia sobre quién era el autor de dichas cartas.

Sin lugar a dudas, de Morgan tuvo una brillante idea al aplicar esos métodos estadísticos para verificar la autoría de textos, pero no se profundizó en sus estudios, hasta que 30 años después Thomas Corwin Mendenhall publicó sus ideas sobre la autoría de textos mediante el promedio de la longitud de las palabras.

Udney Yule, un estadístico de la Universidad de Cambridge, se interesó en Estadística y por el tema de la autoría de textos. Yule se percató de la longitud de significado de una oración como una marca viable, pero sus estudios no fueron tomados en cuenta. Probablemente sus estudios no fueron hechos con gran precisión debido a que no era un lingüista, así que aplicó la Estadística pura

porque si hubiera sido un experto de la lengua hubiera sabido que hay diferencias significantes entre la longitud de una oración en el discurso oral y en el escrito, ya que la longitud de las oraciones con significado se reduce cuando la formalidad aumenta. Entonces, la longitud de las oraciones depende del género y el registro de cada persona, sin embargo, esto no significa que el uso de Estadística en la determinación de autorías no haya dado resultados provechosos.

A mediados de la década de los 80 se dio el auge de la Lingüística Computacional, que guarda estrecha relación con la Lingüística Forense, ya que su objetivo es la determinación de autoría por medio de la búsqueda de algoritmos que permitan determinar el autor de un texto. Para algunos lingüistas computacionales el estilo individual de cada autor es axiomático, es decir que, cada texto tiene la marca particular de su autor.

Una importante aplicación de Estadística la hace Morton, ya que propone analizar un texto utilizando el método de la frecuencia acumulada, asegurando que la forma en las que las personas hablan o escriben no presentan cambios significativos y que durante el discurso oral o escrito se ponen de manifiesto ciertas características únicas que forman un patrón que puede estudiarse mediante el uso de frecuencias acumuladas.

En Lingüística Forense se hace uso de la Estadística para medir probabilidades, tal como lo hacen muchas otras ciencias, aunque los lingüistas forenses conocen muy poco acerca de los beneficios que les podría traer el uso de Estadística en sus investigaciones. El primer paso que un lingüista forense debe dar en su investigación es reunir toda la evidencia, eso consiste en reunir los textos que va a analizar; así que si lo que se quiere descubrir es quién escribió un texto, se deben reunir, tanto textos escritos por el probable autor, como de otros probables autores y luego, aplicarles un análisis estadístico.

4.3 La Lingüística Forense y su relación con la Ley de Zipf

Anteriormente se mencionó que cada persona al hablar o escribir hace uso de su idiolecto, este hecho representa una característica sorprendente porque puede entonces aplicarse Estadística para determinar la regularidades de las que hace uso cada individuo al escribir.

Los porcentajes estadísticos de aparición de aspectos como la sintaxis, el léxico, entre otros, son utilizados por los individuos de forma inconsciente, pero aparecen las mismas regularidades en cualquier texto lo suficientemente extenso como para hacerle un análisis estadístico. Estas regularidades fueron objeto de estudio del filólogo estadounidense G. K. Zipf. Este último estableció la existencia de una relación matemática constante entre el número de apariciones de una palabra en un texto y su orden, además postuló la existencia de una relación inversamente proporcional entre la longitud de una palabra y su frecuencia. Para Zipf estos fenómenos se dan al margen del autor, de la lengua o del tema.

En el lenguaje se ha detectado una estabilidad en el porcentaje de uso de determinados rasgos lingüísticos y la Estilometría, también conocida como Estilística Estadística, basándose en ese porcentaje de uso puede determinar la datación cronológica o la autoría de textos. Para ello se analiza el material léxico de los textos bajo la hipótesis de que cada individuo plasma al escribir su propio estilo. Se utilizará el siguiente caso para ejemplificar el procedimiento: “supongamos que se sabe que un texto T es obra o de X o de Y, pero no se sabe de quién con total certeza. Si el texto T se hallara en soporte digital, podríamos descubrir sin ninguna dificultad cuál es el porcentaje P de apariciones de la palabra *hasta* con respecto al número total de palabras del texto. Se trataría, a continuación, de averiguar el valor del mismo dato en las obras completas de X y en las obras completas de Y, y de comprobar cuál de los dos se encuentra más cercano de P. Con un conjunto bien establecido de comparaciones similares, podríamos llegar a conclusiones estadísticamente válidas sobre el grado de probabilidad de cada una de las dos hipótesis de autoría”. (Gómez, 1,999, p. 165)

En la Lingüística Forense, determinar autorías a partir de técnicas estadísticas ha sido objeto de estudio desde los inicios de la disciplina, pero recientemente con el auge de la Lingüística Computacional es mucho más fácil la aplicación de métodos estilométricos en el análisis de textos.

CAPÍTULO V

Aplicación de la Ley de Zipf en el análisis de textos

5.1 Aplicación de la Ley de Zipf en el análisis de un texto turístico

Los textos turísticos poseen la particularidad de que pueden, y muchas veces deben, ser traducidos hacia otros idiomas. En este caso se aplicó la ley de Zipf para el análisis de un texto, el cual es un folleto turístico en español y el de su respectiva traducción al inglés, utilizado por el INGUAT (2011), dichos textos pueden encontrarse en los anexos, como texto No. 1 y texto No. 2.

La primera diferencia que puede notarse en ambos textos es su densidad, es decir, la relación entre la extensión de ambos textos; el que se encuentra escrito en español tiene un total de 435 palabras, mientras que el mismo texto en inglés cuenta con 419 palabras, lo que representa una diferencia de 16 palabras menos en el texto traducido al inglés.

Para analizar textos según la ley de Zipf, primero se toma un texto y se analiza la cantidad de ocurrencias de las palabras que conforman el texto y luego se ordenan de forma decreciente de acuerdo a su frecuencia, lo que permitió observar que las palabras que tienen los ordenes más altos poseen las frecuencias más altas, esto significa que, la frecuencia de aparición de una palabra en un texto es inversamente proporcional a su orden. En este texto en español la palabra que se encuentra en el orden 1 es la preposición “*de*”, ya que su ocurrencia de aparición en el texto que contiene 435 palabras es de 40, mientras que en el mismo texto en inglés que contiene 419 palabras, la que más ocurrencias posee es el artículo “*the*”. Dentro del mismo texto en español la palabra que tiene el último orden es “*multiculturales*” que solamente fue utilizada una vez en todo el texto, en el texto en inglés la palabra que tiene el último orden es “*representatives*”. Lo anterior permite observar que las palabras que poseen

los órdenes más altos siempre tienen una longitud más corta, esta particularidad puede observarse, tanto en los textos en español, como en los textos en inglés.

Estas características de los textos fueron apreciadas por G. K. Zipf y a ello se debe la elaboración de la Ley del Mínimo Esfuerzo, que especifica que, mientras menor sea el vocabulario utilizado, mayor será la ocurrencia de palabras en los primeros órdenes.

Otra particularidad que puede observarse es que el tema del texto aparece dentro de los órdenes medios, en este caso, en el texto en español la palabra clave, que es "*Guatemala*" aparece el orden 10 y en el texto en español tiene el orden 4.

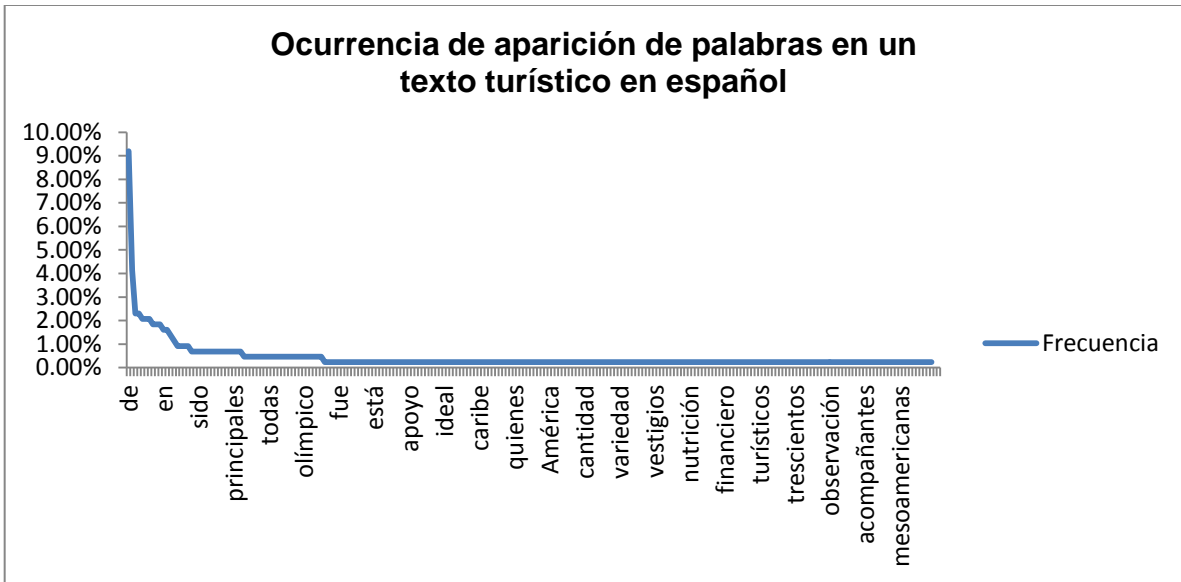


Gráfico No. 1: Ocurrencia de aparición de palabras en texto turístico en español. INGUAT. Guatemala corazón del mundo Maya, el destino ideal para reuniones. Pág. 11

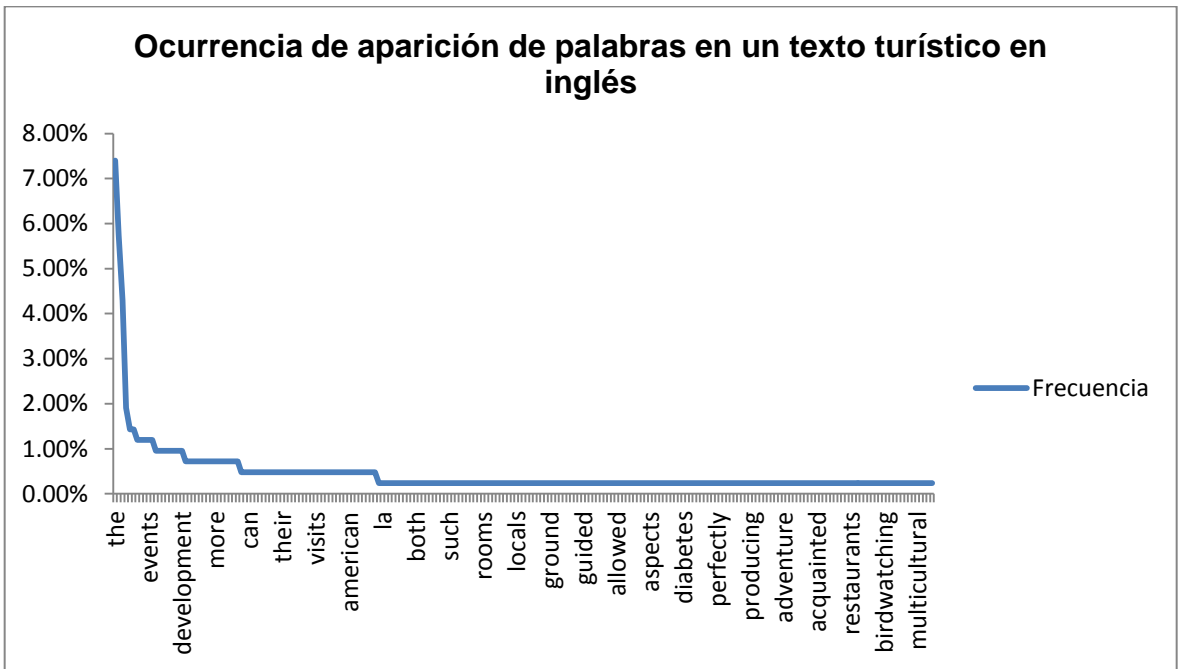


Gráfico No. 2: Ocurrencia de aparición de palabras en texto turístico en inglés. INGUAT. Guatemala heart of the Mayan world, an ideal destination for meetings. Why Guatemala? Pág. 12

5.2 Aplicación de la Ley de Zipf en el análisis de un artículo de revista

El segundo texto que se analizó utilizando la Ley de Zipf es un artículo de una revista electrónica bilingüe, escrito por Lanzito, C. (2007) y pueden encontrarse en el anexo como Texto No. 3 y Texto No. 4. En este texto en español, la palabra que representa el orden más alto es nuevamente la preposición “*de*”, y este mismo fenómeno se aprecia en el mismo texto pero en inglés, ya que la palabra que más ocurrencias presenta es el artículo “*the*”.

La densidad del texto en español es de nuevo más alta que la del texto en inglés, ya que en español posee 503 palabras y en inglés 461, la diferencia es de 42 palabras, así que fácilmente podríamos pensar que en español utilizamos más palabras que en inglés, tanto al hablar como al escribir, por supuesto dependerá del tipo de texto y de su formalidad, hasta el momento se han analizado dos cuyo carácter es informativo.

Además, las palabras que tienen los últimos órdenes son las que mayor longitud presentan en los dos textos, en español “*probablemente*” y en inglés “*uncharacteristic*”, esto significa que las palabras más largas fueron utilizadas únicamente una vez dentro de todo el texto, esto representa el principio de economía lingüística propuesto por Zipf que no se da exclusivamente en inglés sino también en español.

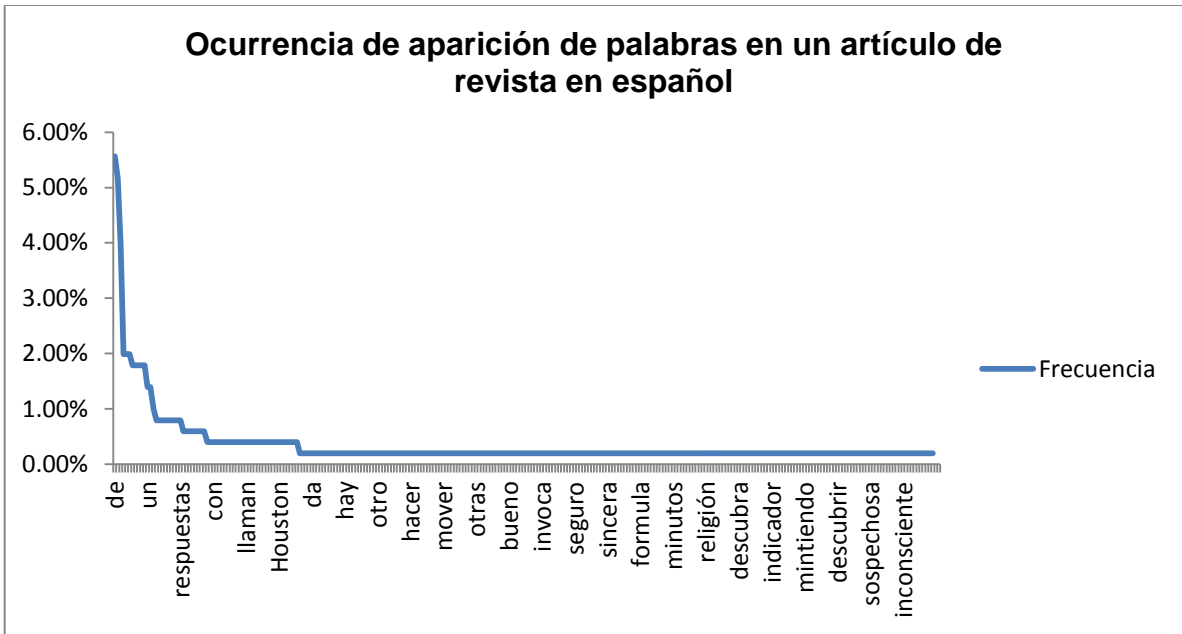


Gráfico No.3: Ocurrencia de aparición de palabras en un artículo de revista en español. Lanzito, C. (2012). Señales de que le mienten. AARP The Magazine.

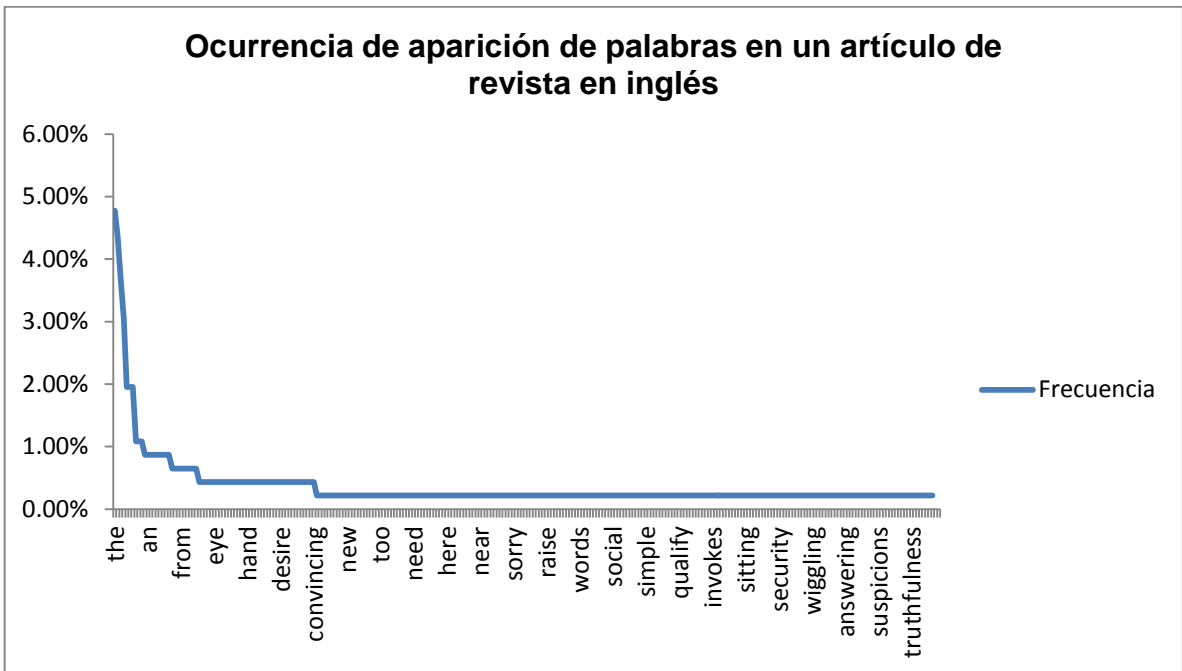


Gráfico No.4: Ocurrencia de aparición de palabras en un artículo de revista en inglés. Lanzito, C. (2012). Signals that you're being lied to. AARP The Magazine.

5.3 Aplicación de la Ley de Zipf en el análisis de una fábula

El texto que se analizó es la fábula de “La Zorra y las Uvas” de Esopo (2011) y puede encontrarse en español e inglés en el anexo como Texto No.5 y Texto No.6. Este texto en español cuenta con una densidad de 187 palabras mientras que en inglés tiene 198 palabras. El hecho que el texto en inglés posea una mayor cantidad de palabras que el texto en español representa una característica interesante que puede atribuirse al tipo de texto, al estilo de quien lo haya traducido y al hecho de que el texto en inglés no presenta contracciones, a diferencia de los dos primeros textos analizados. Las contracciones en inglés son una forma fácil de unir dos palabras, por ejemplo “*it’s*” que es la contracción para las palabras “*it*” e “*is*” y que al final se contabilizó en el análisis como una sola palabra. La utilización de las contracciones depende de la formalidad del texto y del estilo propio de escritura de cada autor, ésta herramienta de economía lingüística puede utilizarse en inglés pero no en español, lo que puede contribuir a que un texto sea más o menos denso. En el análisis de este texto en español la palabra que tiene la mayor ocurrencia es la conjunción “*y*”, mientras que en inglés la que más veces se repite es el artículo “*the*”. Nuevamente, las que tienen una longitud más corta tienen los órdenes más bajos, mientras que en ambos idiomas, los órdenes más altos están representados por palabras con mayor longitud, así en español con la palabra “*menosprecian*” y en inglés “*underestimate*”. Estos análisis permiten determinar que mientras en español la palabra que se utiliza más frecuentemente es la preposición “*de*”, en inglés, sin embargo no se utiliza más frecuentemente su equivalente lingüístico, ya que la palabra que más se repite en inglés que es el artículo “*the*”.

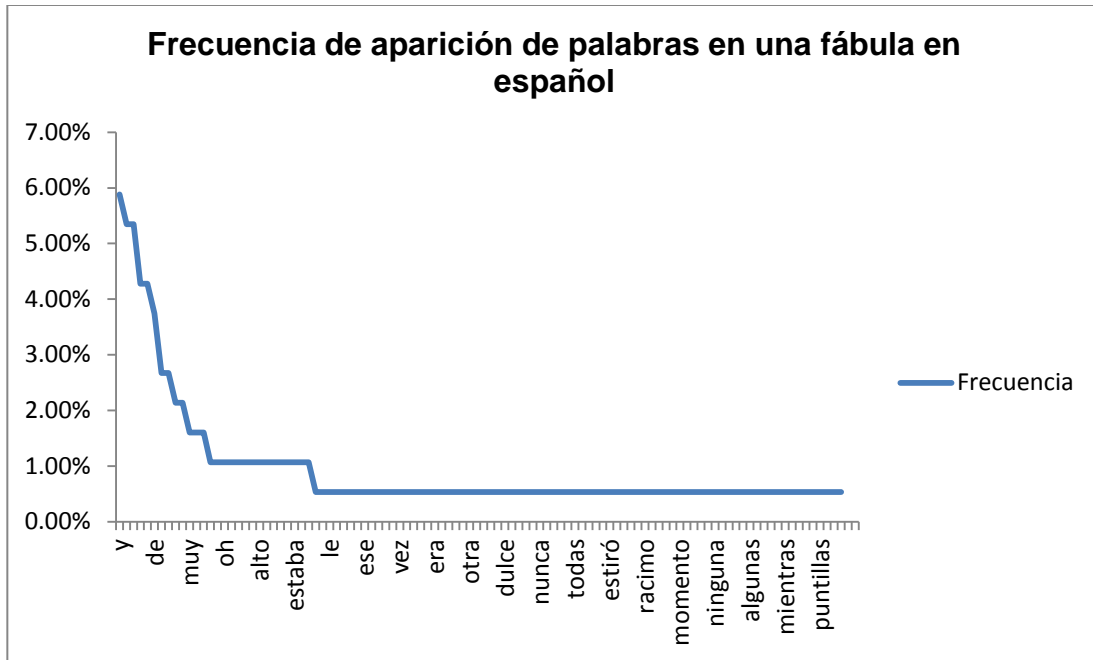


Gráfico No.5: Ocurrencia de aparición de palabras en una fábula en español. Esopo. La zorra y las uvas.

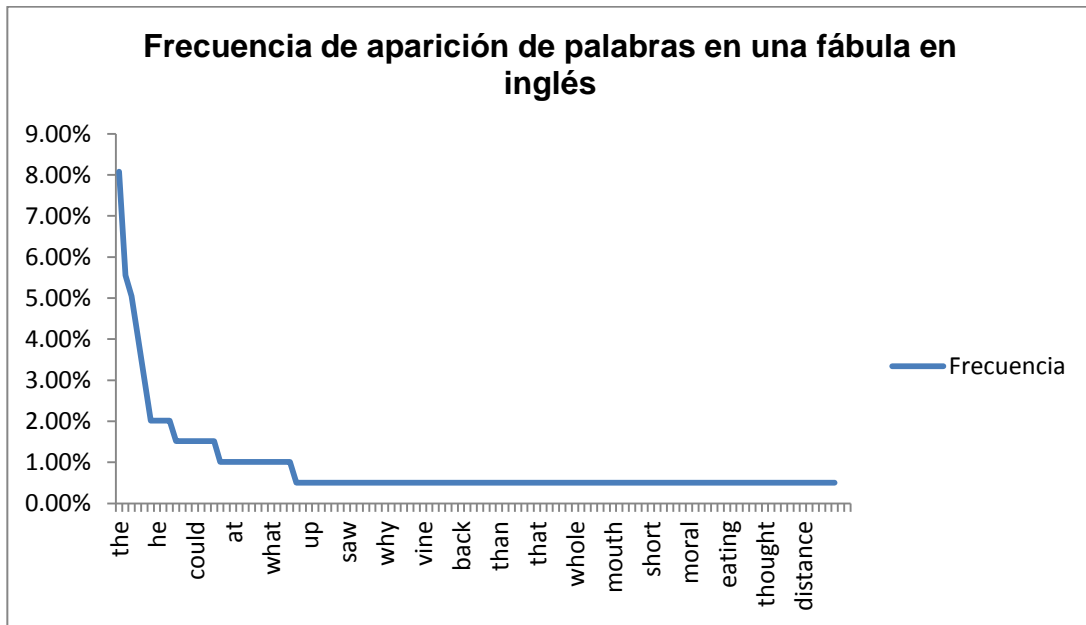


Gráfico No.6: Ocurrencia de aparición de palabras en una fábula en inglés. Esopo. The fox and the grapes.

5.4 Análisis

Después de haber aplicado la Ley de Zipf en el análisis de textos, tanto a textos en español, como al de sus respectivas traducciones, se notaron varias características que los diferencian, éstas son:

- **Densidad:** la primera característica que pudo notarse en los textos es la densidad, que no es más que la relación entre la extensión de ambos textos. En los textos en español, se tiende a utilizar muchas más palabras para transmitir exactamente el mismo mensaje que transmite el texto en inglés, esto quiere decir, que en inglés se concentra en pocas palabras una cantidad mayor de información. Este fenómeno se manifestó en los primeros dos textos analizados estadísticamente, pero en el último texto ocurrió lo contrario. La diferencia radica en el tipo de texto y también en el estilo, tanto del autor, como del traductor del texto.

La diferencia entre la densidad de un mismo texto en español e inglés, puede originarse por el uso de contracciones, que en inglés son muy comunes, ya que es una combinación usualmente utilizada para acortar dos palabras, haciendo uso del apóstrofe. Dichas contracciones suelen utilizarse de manera informal, así que, mientras más formal sea el texto, menos probabilidades existen que estas formas lingüísticas sean utilizadas.

Otro factor que puede marcar una densidad mayor en textos escritos en inglés es el uso de pronombres, ya que en inglés no puede omitirse el pronombre, mientras que en español puede omitirse gracias a que el verbo conjugado da la idea acerca de quien se habla en una oración, a esto en español se le llama sujeto tácito.

- **Longitud de las palabras:** por longitud de una palabra se entiende el número de caracteres que esta contiene, así por ejemplo, la longitud de la palabra “*teléfono*” es ocho. Cuando se habla de longitud de palabras se asocia al lingüista G. K. Zipf ya que para este fenómeno, él propuso un principio al que denominó “Ley del Mínimo Esfuerzo”, que consiste en que en cualquier texto

son más comunes las palabras que tienen una longitud más corta que aquellas que tienen una más larga. Este fenómeno quedó evidenciado en las gráficas mostradas en páginas anteriores, ya que en español es mucho más común utilizar la preposición “*de*” y la conjunción “*y*”, mientras que en inglés el artículo “*the*” y la conjunción “*and*” son las más utilizadas y éstas tienen una longitud muy corta en relación con las que menos se utilizan en los textos analizados, como por ejemplo “*multiculturales*” y “*representatives*”.

El uso más frecuente de palabras cortas se debe a la necesidad del ser humano de transmitir un mensaje de forma más clara y eficiente, esta propiedad del lenguaje es universal, es decir, que ocurre lo mismo tanto en español, como en inglés, italiano, francés y todos los idiomas. La utilización de palabras con una mayor longitud también depende del léxico y riqueza del vocabulario de una persona, ya que mientras más palabras se conocen, menos necesidad habrá de repetir una palabra que ya ha sido utilizada dentro de un texto, aunque la riqueza de vocabulario no es esencialmente indispensable para la correcta transmisión de un mensaje.

- **Posición de las palabras en la gráfica:** al observarse las gráficas correspondientes a los textos en inglés y español se nota que los equivalentes lingüísticos de las palabras no ocupan la misma posición, esto se debe a que las palabras fueron colocadas conforme a su aparición dentro del texto.

El hecho que las palabras no aparezcan en las mismas posiciones dentro del texto se debe a que ambos idiomas (español e inglés) tienen diferencias y particularidades, un ejemplo claro puede ser la utilización de artículos. En español hay muchos más artículos definidos que en inglés, ya que este último posee solamente uno “*the*”, así que en español se contabilizaron todos los artículos definidos encontrados dentro del texto como palabras independientes, mientras que en inglés se hace uso y se contabilizó solamente un artículo definido. Otra diferencia entre las gramáticas de ambos idiomas que fue determinante para que sucediera este fenómeno es la utilización del adjetivo, ya que en español el

adjetivo usualmente se coloca después del sustantivo, a diferencia del inglés, porque éste se coloca antes del sustantivo.

Debido a lo anteriormente expuesto es que los equivalentes lingüísticos no ocupan la misma posición dentro de las gráficas y tablas del presente trabajo.

- **Posición de palabras clave:** una palabra clave podría definirse como aquella que da la idea acerca del tema de un texto o bien que lo resume.

Las palabras clave dentro del texto, utilizando la ley de Zipf se pueden localizar inmediatamente después de aquellas que poseen los primeros órdenes, así por ejemplo en el primer texto analizado estadísticamente la palabra clave es “Guatemala” y tiene el orden 10 y en el texto traducido ocupa el orden 4, esta palabra resume el tema central del texto, que es un folleto turístico guatemalteco en el que se trata de invitar a turistas extranjeros que visiten el país. Aunque la palabra “Guatemala” tiene una mayor longitud, ocupa los órdenes 10 y 4 porque hace referencia al tema del texto, este fenómeno puede apreciarse en cualquier texto si se aplica el proceso propuesto por G. K. Zipf para comprobar la Ley del Mínimo Esfuerzo y las palabras clave pueden localizarse dentro de la tabla si se buscan aquellas palabras con mayor longitud en los órdenes más altos.

- **La Ley de Zipf como herramienta para la Traducción:** es importante establecer el uso que la traducción puede hacer de la ley de Zipf, que es un método estadístico que ayuda a comprender el comportamiento de los textos.

Es de vital importancia determinar que la Estadística es una herramienta de investigación y que la investigación está ligada al trabajo de traducción. Para ser un traductor no basta con conocer dos idiomas, ya que no se traduce un texto solamente mediante su transposición a otro idioma. La traducción es un proceso para el que se necesita un conocimiento profundo de los idiomas y además de las culturas que entran en juego. Para traducir se necesita realizar un análisis del significado de un texto, tal significado puede extraerse de las palabras clave que gracias a la ley de Zipf son mucho más fáciles de encontrar, ya que se hallan inmediatamente después de las palabras que presentan las mayores ocurrencias.

Muchas son las cualidades que un traductor debe poseer, pero imprescindible resulta que sea un investigador y ¿por qué no?, que pueda aplicar métodos de investigación estadística para que mediante ellos conozca mejor el estilo del autor, el tipo de texto y el papel que cada unidad lingüística desempeña en un texto.

Tan importante puede ser la Estadística en traducción que las recientes memorias de traducción incluyen herramientas de cálculo estadístico que sirven para comparar textos, así pues la ley de Zipf, que primero mide la densidad de un texto, puede asistir a un traductor a realizar el cálculo del costo total de una traducción, identificar las palabras clave del texto que le servirán para expresar el mismo significado del texto fuente en el texto meta, apoyarlo a determinar la forma en la que utiliza las palabras, así como auxiliarlo para definir qué tan amplio es su vocabulario y además impulsarlo a corregir problemas como los vicios de dicción tan comunes al momento de hablar y por ende al escribir.

Debido a todo esto es que actualmente los procesos de investigación estadística deben formar parte del quehacer de todos aquellos que trabajan con Lingüística, entre ellos los traductores.

CONCLUSIONES

1. Se ha dado a conocer la Ley de Zipf, también llamada “Ley del Mínimo Esfuerzo” es una ley estadística empírica que sostiene que el ser humano tiende a economizar palabras para lograr la comunicación, y que dicha ley puede aplicarse a cualquier tipo de texto.
2. Se ha definido a la Lingüística como el estudio científico del lenguaje que se manifiesta únicamente en los seres humanos, a la Estadística como la rama de la matemática que recopila, organiza y procesa información para ayudar en la toma de decisiones y a la Estadística Lingüística como una disciplina relativamente nueva que aplica métodos estadísticos para la explicación de fenómenos lingüísticos.
3. Se presentó la vida y el aporte del filólogo estadounidense George Kingsley Zipf a la Lingüística y la nueva Estadística Lingüística.
4. Se determinó que la Lingüística Computacional hace uso de la ley de Zipf en el proceso de identificación de términos representativos de un texto para responder adecuadamente a las necesidades de los usuarios.
5. Se especificó que la Lingüística Forense y la ley de Zipf guardan estrecha relación en la determinación de autorías, ya que a un texto puede aplicársele métodos estadísticos para determinar marcas lingüísticas propias de cada persona.
6. Se aplicó la ley de Zipf en el análisis de tres textos de diversa índole y estilo en inglés y español, y se determinó que es importante que los traductores apliquen este proceso para calcular el costo de una traducción e identificar palabras clave del texto con el fin de transmitir el mismo significado del texto fuente en el texto meta.

RECOMENDACIONES

1. Que los traductores que deseen realizar un análisis más profundo del texto fuente y también del texto meta, realicen un análisis estadístico como el propuesto por G. K. Zipf.
2. Que los traductores vean en la Estadística una herramienta que puede contribuir con la realización de un mejor trabajo y no crean que es una ciencia autónoma que nada tiene que ver con fenómenos lingüísticos.
3. Motivar a los estudiantes y traductores para que apliquen métodos estadísticos para el análisis de textos.
4. Incentivar al lector a que conozca acerca de la Ley de Zipf y se interese por investigar temas relacionados con el análisis estadístico de textos.
5. Que en la Escuela de Ciencias Lingüísticas de la Universidad de San Carlos de Guatemala se siga impartiendo el curso de Estadística para que los estudiantes conozcan la relación que tienen la Lingüística y la Estadística y los beneficios de realizar análisis estadístico de textos.

BIBLIOGRAFÍA

1. Altman, G. (2002). Zipfian Linguistics. *Glottometrics* 3. 150 páginas.
Consultado: [26 de septiembre, 2010] En:
<http://www.arteuna.com/talleres/lab/ediciones/libreria/Glottometrics-zipf.pdf>
2. Araya Alpizar, C. (2004). *Estadística para Laboratorista Químico*. [En línea].
Consultado: [24 de septiembre, 2012]. Disponible en:
http://books.google.com.gt/books?id=libtfIBQLMoC&pg=PA1&lpg=PA1&dq=historia+de+la+estad%C3%ADstica&source=bl&ots=gk8seYCq_l&sig=T98RHloJQE8Nwt-JlWfceWMqxyw&hl=es&sa=X&ei=UQ1iUJWrFomC8ASu7YHgDQ&ved=0CEAQ6wEwAw#v=onepage&q=historia%20de%20la%20estad%C3%ADstica&f=false
3. Arellano, F. *Historia de la Lingüística Tomo I Desde sus Orígenes Hasta el Siglo XIX*. [En línea]. Consultado: [24 de septiembre, 2012]. Disponible en:
<http://books.google.com.gt/books?id=4WNtwWHBfc8C&printsec=frontcover&dq=historia+de+la+ling%C3%BC%C3%ADstica&source=bl&ots=bOLFKA2lvF&sig=DDMWwT130NCXowvrVzV4bwMhTLg&hl=es&sa=X&ei=iAViULfIN42o8gTJ04GABg&ved=0CCsQ6wEwAA#v=onepage&q=historia%20de%20la%20ling%C3%BC%C3%ADstica&f=false>
4. Barrios Calmaestra, L. (2005). Descartes 2D. *Distribuciones Unidimensionales*.
Consultado: [25 de septiembre, 2012.] Disponible en:
http://recursostic.educacion.es/descartes/web/materiales_didacticos/unidimensional_lbarrios/index.htm

5. Barrón Cedeño A., Vila M., y Rosso P. *Detección Automática de Plagio: de la copia exacta a la paráfrasis*. Universidad Politécnica de Valencia. Departamento de Lingüística. [En línea]. Consultado: [9 de octubre, 2012]. Disponible en: http://users.dsic.upv.es/~proso/resources/BarronEtAl_JLF10.pdf

6. Bonnet Jerez, J.L. *Lecciones de Estadística: Estadística Descriptiva y Probabilidad*. [En línea]. Consultado: [24 de septiembre, 2012]. Disponible en: http://books.google.com.gt/books?id=5WZG98WLPnQC&printsec=frontcover&dq=lecciones+de+estadistica&source=bl&ots=t_4e3o3rPE&sig=oMf ktrwrT_wA33LYIfKaAq63IJU&hl=es&sa=X&ei=BkJiUNyJNoa8gS1jYDgCA&ved=0CCwQ6wEwAA#v=onepage&q=lecciones%20de%20estadistica&f=false

7. Diaz Nosty, B. (2012). George Kingsley Zipf. *Infoamérica 9*. Consultado: [26 de septiembre, 2012] En: <http://www.infoamerica.org/teoria/zipf1.htm>

8. Esopo. (2007). La Zorra y las Uvas. [Recuperado el 25 de noviembre de 2012] disponible en: <http://albalearning.com/audiolibros/samaniego/f1-4-06zorrayucas.html>

9. Estopá R., Paz M., Borrás L., Cabré M., Cicres J., DeCesaris J., Yzaguirre L., Freixa J., Lorente M., y Turrel M. (2006). *Hablamos de Lengua con Niños y Niñas*. Barcelona: Consejo Social de la Universitat Pompeu Fabra. *Procesamiento de la Impropiedad Estilística y la Legibilidad*. España: Universidad de Vigo. [En línea]. Consultado: [9 de octubre, 2012]. Disponible en: <http://books.google.com.gt/books?id=ejqmM6GzbUwC&pg=PA71&lpg=P>

A71&dq=linguistica+forense+estilo+y+estad%C3%ADstica&source=bl&ots=Jnea1_7ezZ&sig=AiLk54IBU2RLoeN_h4FQ7ONEOE4&hl=es#v=onepage&q=linguistica%20forense%20estilo%20y%20estad%C3%ADstica&f=false

10. García Mejía, A. Estudios de Lengua y Estadística. *Los Métodos Estadísticos y la Investigación Lingüística. Breve Recorrido Histórico*. Consultado: [25 de septiembre, 2012] En: <http://es.scribd.com/doc/14595890/Los-metodos-estadisticos-y-la-investigacion-linguistica-Breve-recorrido-historico>
11. Gómez Guinovart, J. (1999). *Bases Lingüísticas y Computacionales del Procesamiento de la Impropiedad Estilística y la Legibilidad*. España: Universidad de Vigo. [En línea]. Consultado: [9 de octubre, 2012]. Disponible en: dialnet.unirioja.es/descarga/articulo/227029.pdf
12. González, C.M. (2007). *La recuperación de la información en el siglo XX, revisión y aplicación de aspectos de la lingüística cuantitativa y la modelización matemática de la información*. Tesina de licenciatura, Universidad Nacional de La Plata, La Plata Argentina. <http://www.memoria.fahce.unlp.edu.ar/tesis/te.350/te.350.pdf>
13. Gonzalvo, G. (1978). *Diccionario de Metodología Estadística*. [En línea]. Consultado: [22 de octubre, 2012]. Disponible en: http://books.google.com.gt/books?id=cb41dkURoHMC&printsec=frontcover&dq=explicaci%C3%B3n+para+principiantes+de+%22abscisa%22+estadística&source=bl&ots=pd36uo5nkN&sig=R3Q_zakznesBAyqntb72LawNcdg&hl=es&sa=X&ei=dcCFUN2UGYGm9ATQkIDwBQ&ved=0CDEQ6wEwAQ#v=onepage&q&f=false

14. Guerrero Peña, A., Buitrago, M.V. y, Curieses, M. A. (2007). *Estadística Básica* [En línea]. Consultado: [25 de septiembre, 2012]. Disponible en: <http://books.google.com.gt/books?id=pLF16E8nF6gC&pg=PA6&lpg=PA6&dq=Guerrero+Pe%C3%B1a,+Adriana,+Buitrago,+Mar%C3%ADa+Victoria,+Curieses,+Mar%C3%ADa+de+los+Angeles,+Estad%C3%ADstica+B%C3%A1sica,+Instituto+Tecnol%C3%B3gico+Metropolitano,+primera+edici%C3%B3n,+septiembre+de+2007,+Colombia.&source=bl&ots=mOphuTGoSC&sig=hkITvwgFZQSVDuLvagl5Mf8d9cU&hl=es&sa=X&ei=BkBiUN67Doas9ATKpYC4DQ&ved=0CB0Q6AEwAA#v=onepage&q&f=false>
15. INGUAT, (2011). *Guatemala Corazón del Mundo Maya, El Destino Ideal para Reuniones. Por qué Guatemala.* Págs. 11-12.
16. Lanzito, C. (2007). *AARP The Magazine.* Señales de que le mienten. [Recuperado el 25 de noviembre de 2012] disponible en: <http://www.aarp.org/espanol/hogar-familia-amigos/info-08-2012/senales-metiras-infidelidad.html>
17. López Casuso, R. (2006). *Cálculo de Probabilidades e Inferencia Estadística con Tópicos de Econometría.* [En línea]. Consultado: [24 de septiembre, 2012]. Disponible en: <http://books.google.com.gt/books?id=qWwR4jP8LlgC&pg=PA1&lpg=PA1&dq=que+es+estad%C3%ADstica&source=bl&ots=5OBvn7BCSq&sig=h376tuMoi2nWtXInCtFJGuKjA8&hl=es&sa=X&ei=gNxgUJjTB6SF0QGXPoCAAw&ved=0CDIQ6wEwAQ#v=onepage&q=que%20es%20estad%C3%ADstica&f=false>

18. Martínez Celdrán, E. (2002). *Lingüística Teoría y Aplicaciones*. [En línea]. Consultado: [24 de septiembre, 2012] Disponible en: http://books.google.com.gt/books?id=gXSveYBqSYUC&pg=PA1&lpg=PA1&dq=que+es+ling%C3%BC%C3%ADstica&source=bl&ots=6Au7WT_1nG&sig=_1d2Heez7tGI7_Xv29dSJR-D9k&hl=es&sa=X&ei=4ptgULvVA8ig0gHGroCoAQ&ved=0CCwQ6wEwAA#v=onepage&q=que%20es%20ling%C3%BC%C3%ADstica&f=false
19. Martínez, L. La Ley de Zipf y sus Aplicaciones en el Análisis Documental: *Análisis Automático de Contenido y la Generación de Tesauros*. Consultado: [4 octubre, 2012]. Disponible en: <https:// analisisdocumental2011.wikispaces.com/Ley+de+Zipf+y+AD>
20. Matemáticas Básicas. Estadística Descriptiva. *Definición y Clasificación de Variables*. PDF. Consultado: [25 de septiembre, 2012] En: http://www.fca.unam.mx/docs/apuntes_matematicas/34.%20Estadistica%20Descriptiva.pdf
21. Moreno Sandoval, A. (1996). *Textos de Apoyo Lingüística Computacional*. [En línea]. Consultado: [4 de octubre, 2012]. Disponible en: <http://www.textosenlinea.com.ar/academicos/Linguistica%20computacional.pdf>
22. Niño Rojas, V.M. (2007). *Fundamentos de Semiótica y Lingüística*. [En línea]. Consultado: [24 de septiembre, 2012]. Disponible en: <http://books.google.com.gt/books?id=qpj9xJH3ZhcC&pg=PA139&lpg=PA139&dq=ciencias+auxiliares+de+la+linguistica&source=bl&ots=dcBvH6NziV&sig=Mb0A2KJQkiWcCSsLOIIIxiaZfYk&hl=es&sa=X&ei=crc2UMrFC>

MXy6wHFhYHYCA&ved=0CC0Q6wEwAA#v=onepage&q=ciencias%20auxiliares%20de%20la%20linguística&f=false

23. Olson, J. (2004). *Forensic Linguistics, An Introduction to Language, Crime and the Law*. En línea]. Consultado: [9 de octubre, 2012]. Disponible en: http://books.google.com.gt/books?id=i3399LFSzqQC&pg=PA18&lpg=PA18&dq=statistics+in+forensic+linguistics&source=bl&ots=0GwpUhOLcp&sig=hjnNBOK1bfvcWcS_Xc_DUQGgvRg&hl=es&sa=X&ei=FXd0UKvKKI7K9gT2kYDYAw&ved=0CCYQ6AEwAA
24. Paredes Duarte, J. (2008). Pragmalingüística. *El Principio de la Economía Lingüística*. Consultado: [3 de octubre, 2012] En: revistas.uca.es/index.php/pragma/article/view/10
25. Prun, C., Zipf, R. (2002). Biographical Notes on G.K.Zipf. *Glottometrics* 3. 150 páginas. Consultado: [26 de septiembre, 2012] En: <http://www.arteauna.com/talleres/lab/ediciones/libreria/Glottometrics-zipf.pdf>
26. Real Academia Española. (2004). *Diccionario de la Lengua Española*. (22ª Edición). [En línea]. Consultado: [19 de octubre, 2012]. Disponible en: <http://lema.rae.es/drae/>
27. Rodrigues, J.M. (2009). La lingüística del corpus y el análisis gramatical del español: propuesta del tratamiento de las preposiciones del español como especificadores semánticos. Tesis doctoral, Universidad de Valladolid, Valladolid España. <http://uvadoc.uva.es/bitstream/10324/109/1/TESIS36-091216.pdf>

28. Ruiz Muñoz, D. Biblioteca Virtual eumed.net. *Manual de Estadística*. Consultado: [25 de septiembre, 2012] En: <http://www.eumed.net/cursecon/libreria/drm/0.htm>
29. Rull García, C. (2006-2011). Verba. *Historia de la Lingüística*. Consultado: [25 de septiembre, 2012] En: <https://sites.google.com/site/carlesrull/historia>
30. Turrel, M.T (2005). *Lingüística Forense, Lengua y Derecho: Conceptos, Métodos y Aplicaciones*. Barcelona: Documenta Universitaria. [En línea]. Consultado [9, octubre, 2012]. Disponible en: http://books.google.com.gt/books?id=2MEBUef6fQoC&pg=PA249&lpg=PA249&dq=introducci%C3%B3n+a+la+ling%C3%BC%C3%ADstica+forense&source=bl&ots=fnE-DsJwg5&sig=zv18DNyC5_SQ6nwhBuz69_Nf4Jl&hl=es&sa=X&ei=KCJzUOXWI4WDrAGSrIC4Bw&ved=0CCwQ6wEwAA#v=onepage&q=introducci%C3%B3n%20a%20la%20ling%C3%BC%C3%ADstica%20forense&f=false
31. Yáñez Cossio, C. (2007). *Una Introducción a la Lingüística General*. [En línea]. Consultado: [24 de septiembre, 2012]. Disponible en: <http://books.google.com.gt/books?id=wpOmpupxuh8C&printsec=frontcover&dq=introducci%C3%B3n+a+la+ling%C3%BCistica&source=bl&ots=LlNyk1kV22&sig=h8qZgKROOlwVHMnoIVWL6R9V0s0&hl=es&sa=X&ei=KaxgULHOOJLC9QTJmIGgCA&ved=0CF4Q6wEwCA#v=onepage&q=introducci%C3%B3n%20a%20la%20ling%C3%BCistica&f=false>

ANEXOS

Texto No.1: ¿Por qué Guatemala?

Guatemala, corazón del mundo Maya

Guatemala es cuna de una de las civilizaciones mesoamericanas precolombinas más importantes de la historia -la cultura maya-. Su desarrollo económico, político, religioso y social fue trascendental, el que aún hoy sorprende a propios y extraños. Gran cantidad de vestigios en zonas arqueológicas perfectamente restauradas han sido visitados por miles de personas de distintas partes del mundo, quienes han quedado maravillados por el testimonio dejado por esta cultura. El Corazón del Mundo Maya está disponible para ser visitado y admirado por los asistentes a congresos, convenciones y viajes de incentivo, siendo una excelente y única alternativa de programa complementario para los eventos.

Experiencia en desarrollo de eventos mundiales

Guatemala ha sido sede de importantes eventos de alcance mundial, entre ellos la Asamblea de Gobernadores del Banco Interamericano de Desarrollo, donde los asistentes de más de trescientos países debatieron durante varios días sobre los principales programas de apoyo financiero para el continente americano; la Ciento Diecinueve Sesión del Comité Olímpico Internacional (COI), organismo que reúne a delegados de más de doscientos países en el que se toman importantes decisiones sobre el futuro del movimiento olímpico internacional; la Tercera Conferencia Mundial de la Organización Internacional del Café, que reúne a los principales países productores de café; y el Onceavo Congreso Centroamericano y del Caribe de Endocrinología, Metabolismo, Nutrición y Diabetes, evento que congrega a importantes médicos de esta región continental. Estas experiencias, entre otras, han permitido que Guatemala se especialice en eventos de alcance mundial por el nivel de servicios que demandan.

Infraestructura especializada en el desarrollo de eventos

Guatemala cuenta con más de doscientos salones, teatros, museos, mercados, plazas, espacios para eventos y centros de convenciones integrados a hoteles, así como cadenas internacionales de hoteles. Una amplia variedad de restaurantes, importantes proveedores especializados en servicios para eventos, operadores profesionales de congresos, convenciones y viajes de incentivos. Transporte terrestre y conectividad aérea a las principales ciudades de Estados Unidos, México, centro y sur América. Todas estas instalaciones y servicios han sido probados en eventos internacionales.

Extraordinarios atractivos turísticos naturales y culturales

Guatemala tiene una vasta gama de opciones para diversión y entretenimiento que están al alcance tanto del congresista como de su acompañante. Entre las opciones encontramos: visita a volcanes, museos, mercados de artesanías, ciudades coloniales como la antigua Guatemala (declarada Patrimonio Cultural de la Humanidad por la UNESCO), turismo de aventura, parques temáticos, observación de aves, compras y visita a comunidades para adentrarse y conocer los aspectos multiculturales de la región, todas ellas con recorridos guiados. Estas opciones son un complemento ideal para el programa de congresistas y acompañantes.

Tabla No.1: ordenación de ocurrencia de palabras de texto turístico en español.

Orden	Palabra	Ocurrencia	Fr	Fa	Constante¹
1	de	40	9.20%	9.20%	40
2	y	18	4.14%	13.33%	36
3	la	10	2.30%	15.63%	30
4	el	10	2.30%	17.93%	40
5	a	9	2.07%	20.00%	45
6	del	9	2.07%	22.07%	54
7	que	9	2.07%	24.14%	63
8	para	8	1.84%	25.98%	64
9	eventos	8	1.84%	27.82%	72
10	Guatemala	8	1.84%	29.66%	80
11	en	7	1.61%	31.26%	77
12	por	7	1.61%	32.87%	84
13	los	6	1.38%	34.25%	78
14	importantes	5	1.15%	35.40%	70
15	una	4	0.92%	36.32%	60
16	más	4	0.92%	37.24%	64
17	han	4	0.92%	38.16%	68
18	desarrollo	4	0.92%	39.08%	72
19	las	3	0.69%	39.77%	57
20	como	3	0.69%	40.46%	60
21	sido	3	0.69%	41.15%	63
22	maya	3	0.69%	41.84%	66
23	estas	3	0.69%	42.53%	69
24	mundo	3	0.69%	43.22%	72
25	entre	3	0.69%	43.91%	75
26	países	3	0.69%	44.60%	78
27	alcance	3	0.69%	45.29%	81
28	mundial	3	0.69%	45.98%	84
29	opciones	3	0.69%	46.67%	87
30	servicios	3	0.69%	47.36%	90
31	principales	3	0.69%	48.05%	93
32	convenciones	3	0.69%	48.74%	96
33	internacional	3	0.69%	49.43%	99
34	su	2	0.46%	49.89%	68
35	se	2	0.46%	50.34%	70
36	con	2	0.46%	50.80%	72

¹ C = Orden x Ocurrencia

37	esta	2	0.46%	51.26%	74
38	café	2	0.46%	51.72%	76
39	sobre	2	0.46%	52.18%	78
40	reúne	2	0.46%	52.64%	80
41	todas	2	0.46%	53.10%	82
42	viajes	2	0.46%	53.56%	84
43	región	2	0.46%	54.02%	86
44	visita	2	0.46%	54.48%	88
45	museos	2	0.46%	54.94%	90
46	corazón	2	0.46%	55.40%	92
47	cultura	2	0.46%	55.86%	94
48	hoteles	2	0.46%	56.32%	96
49	ciudades	2	0.46%	56.78%	98
50	mercados	2	0.46%	57.24%	100
51	olímpico	2	0.46%	57.70%	102
52	programa	2	0.46%	58.16%	104
53	congresos	2	0.46%	58.62%	106
54	asistentes	2	0.46%	59.08%	108
55	doscientos	2	0.46%	59.54%	110
56	internacionales	2	0.46%	60.00%	112
57	es	1	0.23%	60.23%	57
58	ha	1	0.23%	60.46%	58
59	al	1	0.23%	60.69%	59
60	un	1	0.23%	60.92%	60
61	fue	1	0.23%	61.15%	61
62	aún	1	0.23%	61.38%	62
63	hoy	1	0.23%	61.61%	63
64	ser	1	0.23%	61.84%	64
65	COI	1	0.23%	62.07%	65
66	así	1	0.23%	62.30%	66
67	sur	1	0.23%	62.53%	67
68	son	1	0.23%	62.76%	68
69	cuna	1	0.23%	62.99%	69
70	gran	1	0.23%	63.22%	70
71	está	1	0.23%	63.45%	71
72	sede	1	0.23%	63.68%	72
73	días	1	0.23%	63.91%	73
74	gama	1	0.23%	64.14%	74
75	aves	1	0.23%	64.37%	75

76	zonas	1	0.23%	64.60%	76
77	miles	1	0.23%	64.83%	77
78	única	1	0.23%	65.06%	78
79	ellos	1	0.23%	65.29%	79
80	banco	1	0.23%	65.52%	80
81	apoyo	1	0.23%	65.75%	81
82	toman	1	0.23%	65.98%	82
83	otras	1	0.23%	66.21%	83
84	nivel	1	0.23%	66.44%	84
85	aérea	1	0.23%	66.67%	85
86	tanto	1	0.23%	66.90%	86
87	tiene	1	0.23%	67.13%	87
88	vasta	1	0.23%	67.36%	88
89	están	1	0.23%	67.59%	89
90	ellas	1	0.23%	67.82%	90
91	ideal	1	0.23%	68.05%	91
92	donde	1	0.23%	68.28%	92
93	dejado	1	0.23%	68.51%	93
94	social	1	0.23%	68.74%	94
95	partes	1	0.23%	68.97%	95
96	varios	1	0.23%	69.20%	96
97	ciento	1	0.23%	69.43%	97
98	sesión	1	0.23%	69.66%	98
99	comité	1	0.23%	69.89%	99
100	futuro	1	0.23%	70.11%	100
101	caribe	1	0.23%	70.34%	101
102	evento	1	0.23%	70.57%	102
103	cuenta	1	0.23%	70.80%	103
104	plazas	1	0.23%	71.03%	104
105	amplia	1	0.23%	71.26%	105
106	México	1	0.23%	71.49%	106
107	centro	1	0.23%	71.72%	107
108	UNESCO	1	0.23%	71.95%	108
109	siendo	1	0.23%	72.18%	109
110	propios	1	0.23%	72.41%	110
111	quienes	1	0.23%	72.64%	111
112	quedado	1	0.23%	72.87%	112
113	durante	1	0.23%	73.10%	113
114	tercera	1	0.23%	73.33%	114

115	onceavo	1	0.23%	73.56%	115
116	médicos	1	0.23%	73.79%	116
117	salones	1	0.23%	74.02%	117
118	teatros	1	0.23%	74.25%	118
119	centros	1	0.23%	74.48%	119
120	cadenas	1	0.23%	74.71%	120
121	América	1	0.23%	74.94%	121
122	Antigua	1	0.23%	75.17%	122
123	turismo	1	0.23%	75.40%	123
124	parques	1	0.23%	75.63%	124
125	compras	1	0.23%	75.86%	125
126	conocer	1	0.23%	76.09%	126
127	guiados	1	0.23%	76.32%	127
128	historia	1	0.23%	76.55%	128
129	político	1	0.23%	76.78%	129
130	extraños	1	0.23%	77.01%	130
131	cantidad	1	0.23%	77.24%	131
132	personas	1	0.23%	77.47%	132
133	visitado	1	0.23%	77.70%	133
134	admirado	1	0.23%	77.93%	134
135	asamblea	1	0.23%	78.16%	135
136	congreso	1	0.23%	78.39%	136
137	diabetes	1	0.23%	78.62%	137
138	congrega	1	0.23%	78.85%	138
139	demandan	1	0.23%	79.08%	139
140	espacios	1	0.23%	79.31%	140
141	variedad	1	0.23%	79.54%	141
142	probados	1	0.23%	79.77%	142
143	diversión	1	0.23%	80.00%	143
144	volcanes	1	0.23%	80.23%	144
145	cultural	1	0.23%	80.46%	145
146	aventura	1	0.23%	80.69%	146
147	aspectos	1	0.23%	80.92%	147
148	económico	1	0.23%	81.15%	148
149	religioso	1	0.23%	81.38%	149
150	sorprende	1	0.23%	81.61%	150
151	vestigios	1	0.23%	81.84%	151
152	visitados	1	0.23%	82.07%	152
153	distintas	1	0.23%	82.30%	153

154	incentivo	1	0.23%	82.53%	154
155	excelente	1	0.23%	82.76%	155
156	mundiales	1	0.23%	82.99%	156
157	programas	1	0.23%	83.22%	157
158	americano	1	0.23%	83.45%	158
159	organismo	1	0.23%	83.68%	159
160	delegados	1	0.23%	83.91%	160
161	nutrición	1	0.23%	84.14%	161
162	permitido	1	0.23%	84.37%	162
163	terrestre	1	0.23%	84.60%	163
164	naturales	1	0.23%	84.83%	164
165	declarada	1	0.23%	85.06%	165
166	humanidad	1	0.23%	85.29%	166
167	temáticos	1	0.23%	85.52%	167
168	testimonio	1	0.23%	85.75%	168
169	disponible	1	0.23%	85.98%	169
170	debatieron	1	0.23%	86.21%	170
171	financiero	1	0.23%	86.44%	171
172	continente	1	0.23%	86.67%	172
173	decisiones	1	0.23%	86.90%	173
174	diecinueve	1	0.23%	87.13%	174
175	movimiento	1	0.23%	87.36%	175
176	integrados	1	0.23%	87.59%	176
177	operadores	1	0.23%	87.82%	177
178	incentivos	1	0.23%	88.05%	178
179	transporte	1	0.23%	88.28%	179
180	atractivos	1	0.23%	88.51%	180
181	turísticos	1	0.23%	88.74%	181
182	culturales	1	0.23%	88.97%	182
183	artesanías	1	0.23%	89.20%	183
184	coloniales	1	0.23%	89.43%	184
185	patrimonio	1	0.23%	89.66%	185
186	adentrarse	1	0.23%	89.89%	186
187	recorridos	1	0.23%	90.11%	187
188	restauradas	1	0.23%	90.34%	188
189	alternativa	1	0.23%	90.57%	189
190	experiencia	1	0.23%	90.80%	190
191	trescientos	1	0.23%	91.03%	191
192	conferencia	1	0.23%	91.26%	192

193	productores	1	0.23%	91.49%	193
194	metabolismo	1	0.23%	91.72%	194
195	continental	1	0.23%	91.95%	195
196	especialice	1	0.23%	92.18%	196
197	proveedores	1	0.23%	92.41%	197
198	congresista	1	0.23%	92.64%	198
199	acompañante	1	0.23%	92.87%	199
200	encontramos	1	0.23%	93.10%	200
201	observación	1	0.23%	93.33%	201
202	comunidades	1	0.23%	93.56%	202
203	complemento	1	0.23%	93.79%	203
204	maravillados	1	0.23%	94.02%	204
205	gobernadores	1	0.23%	94.25%	205
206	organización	1	0.23%	94.48%	206
207	experiencias	1	0.23%	94.71%	207
208	restaurantes	1	0.23%	94.94%	208
209	conectividad	1	0.23%	95.17%	209
210	congresistas	1	0.23%	95.40%	210
211	acompañantes	1	0.23%	95.63%	211
212	precolombinas	1	0.23%	95.86%	212
213	arqueológicas	1	0.23%	96.09%	213
214	perfectamente	1	0.23%	96.32%	214
215	especializada	1	0.23%	96.55%	215
216	profesionales	1	0.23%	96.78%	216
217	Estados Unidos	1	0.23%	97.01%	217
218	instalaciones	1	0.23%	97.24%	218
219	trascendental	1	0.23%	97.47%	219
220	civilizaciones	1	0.23%	97.70%	220
221	mesoamericanas	1	0.23%	97.93%	221
222	complementario	1	0.23%	98.16%	222
223	interamericano	1	0.23%	98.39%	223
224	endocrinología	1	0.23%	98.62%	224
225	especializados	1	0.23%	98.85%	225
226	centroamericano	1	0.23%	99.08%	226
227	infraestructura	1	0.23%	99.31%	227
228	extraordinarios	1	0.23%	99.54%	228
229	entretenimiento	1	0.23%	99.77%	229
230	multiculturales	1	0.23%	100.00%	230

Nota: INGUAT. Guatemala Corazón del Mundo Maya, el Destino Ideal para Reuniones. Pág.11

Texto No.2: Why Guatemala?

Guatemala; heart of the Mayan world

Guatemala is the birthplace of one of the most important pre-Columbian Mesoamerican civilizations of history –the Mayan Culture-. Its economical, political, religious, and social development was trascendental, which still amazes both locals and foreigners. A great quantity of traces of perfectly restored archeological areas have been visited by thousands of people from different parts of the world, who have been astounded by the testimony left by this culture. The Heart of the Mayan World can be visited and admired by attendants of congresses, conventions, and incentive trips, as being an excellent and unique complementary program alternative for the events.

Experience in the development of the world events

Guatemala has been host of important international events, amongst them the Governor Assambley of the Inter-American Development Bank (IDB), where the attendants, from more than thirty countries, debated during several days about the main financial support programs for the American Continent. The One Hundred Nineteenth International Olympic Committee Session, agency that gathers representatives of more than two hundred countries, in which important decisions about the future of The International Olympic Movement are taken. The Third World Conference of the International Coffee Organization, which gathers the main coffee producing countries; and the Eleventh Central American and Caribbean Endocrinology, Metabolism, Nutrition, and Diabetes Congress, event that gathers important doctors of this continental region. These experiences, among others, have allowed Guatemala to specialize in events with global impact because of the service that is demanded.

Specialized event development infrastructure

Guatemala has more than two hundred conference rooms, theaters, museums, market places, public squares, event areas, and hoted integrated

convention centers as well as important international hotel chains. It also has a variety of restaurants, specialized event service providers, professional congress, conventions, and incentive trips operators, ground transportation and flight connectivity to the main cities of the United States, Mexico, Central and South America. All these facilities and services have been tested in international events.

Extraordinary natural and cultural touristic attractions

Guatemala has a vast range of amusement and entertainment options, which will be within reach of the delegates as well as of their accompanying guests. Among the options that can be found are the following: visits to volcanoes, museums, handicraft market places, theme parks and colonial cities such as La Antigua Guatemala, adventure tourism, birdwatching, shopping, and guided visits to communities, which is an opportunity to get acquainted with multicultural aspects of the region. These options are an ideal complement for the delegates' program and their accompanying guests.

Tabla No.2: ordenación de ocurrencia de palabras de texto turístico en inglés.

Orden	Palabra	Ocurrencia	Fr	Fa	Constante
1	the	31	7.40%	7.40%	31
2	of	24	5.73%	13.13%	48
3	and	18	4.30%	17.42%	54
4	Guatemala	8	1.91%	19.33%	32
5	as	6	1.43%	20.76%	30
6	international	6	1.43%	22.20%	36
7	to	5	1.19%	23.39%	35
8	world	5	1.19%	24.58%	40
9	which	5	1.19%	25.78%	45
10	events	5	1.19%	26.97%	50
11	important	5	1.19%	28.16%	55
12	in	4	0.95%	29.12%	48
13	by	4	0.95%	30.07%	52
14	has	4	0.95%	31.03%	56
15	have	4	0.95%	31.98%	60
16	been	4	0.95%	32.94%	64
17	that	4	0.95%	33.89%	68
18	event	4	0.95%	34.84%	72
19	development	4	0.95%	35.80%	76
20	a	3	0.72%	36.52%	60
21	and	3	0.72%	37.23%	63
22	be	3	0.72%	37.95%	66
23	is	3	0.72%	38.66%	69
24	are	3	0.72%	39.38%	72
25	for	3	0.72%	40.10%	75
26	main	3	0.72%	40.81%	78
27	than	3	0.72%	41.53%	81
28	more	3	0.72%	42.24%	84
29	mayan	3	0.72%	42.96%	87
30	these	3	0.72%	43.68%	90
31	options	3	0.72%	44.39%	93
32	gathers	3	0.72%	45.11%	96
33	countries	3	0.72%	45.82%	99
34	hundred	3	0.72%	46.54%	102

35	two	2	0.48%	47.02%	70
36	one	2	0.48%	47.49%	72
37	can	2	0.48%	47.97%	74
38	from	2	0.48%	48.45%	76
39	this	2	0.48%	48.93%	78
40	with	2	0.48%	49.40%	80
41	well	2	0.48%	49.88%	82
42	areas	2	0.48%	50.36%	84
43	trips	2	0.48%	50.84%	86
44	about	2	0.48%	51.31%	88
45	hotel	2	0.48%	51.79%	90
46	their	2	0.48%	52.27%	92
47	heart	2	0.48%	52.74%	94
48	among	2	0.48%	53.22%	96
49	visited	2	0.48%	53.70%	98
50	coffee	2	0.48%	54.18%	100
51	region	2	0.48%	54.65%	102
52	market	2	0.48%	55.13%	104
53	cities	2	0.48%	55.61%	106
54	places	2	0.48%	56.09%	108
55	visits	2	0.48%	56.56%	110
56	guests	2	0.48%	57.04%	112
57	culture	2	0.48%	57.52%	114
58	program	2	0.48%	58.00%	116
59	olympic	2	0.48%	58.47%	118
60	central	2	0.48%	58.95%	120
61	congress	2	0.48%	59.43%	122
62	service	2	0.48%	59.90%	124
63	museums	2	0.48%	60.38%	126
64	american	2	0.48%	60.86%	128
65	incentive	2	0.48%	61.34%	130
66	delegates	2	0.48%	61.81%	132
67	attendants	2	0.48%	62.29%	134
68	conference	2	0.48%	62.77%	136
69	conventions	2	0.48%	63.25%	138
70	specialized	2	0.48%	63.72%	140
71	accompanying	2	0.48%	64.20%	142
72	it	1	0.24%	64.44%	72

73	la	1	0.24%	64.68%	73
74	why	1	0.24%	64.92%	74
75	its	1	0.24%	65.16%	75
76	was	1	0.24%	65.39%	76
77	who	1	0.24%	65.63%	77
78	IDB	1	0.24%	65.87%	78
79	all	1	0.24%	66.11%	79
80	get	1	0.24%	66.35%	80
81	most	1	0.24%	66.59%	81
82	both	1	0.24%	66.83%	82
83	left	1	0.24%	67.06%	83
84	host	1	0.24%	67.30%	84
85	them	1	0.24%	67.54%	85
86	bank	1	0.24%	67.78%	86
87	days	1	0.24%	68.02%	87
88	also	1	0.24%	68.26%	88
89	vast	1	0.24%	68.50%	89
90	will	1	0.24%	68.74%	90
91	such	1	0.24%	68.97%	91
92	still	1	0.24%	69.21%	92
93	great	1	0.24%	69.45%	93
94	parts	1	0.24%	69.69%	94
95	being	1	0.24%	69.93%	95
96	ideal	1	0.24%	70.17%	96
97	where	1	0.24%	70.41%	97
98	taken	1	0.24%	70.64%	98
99	third	1	0.24%	70.88%	99
100	rooms	1	0.24%	71.12%	100
101	south	1	0.24%	71.36%	101
102	range	1	0.24%	71.60%	102
103	reach	1	0.24%	71.84%	103
104	found	1	0.24%	72.08%	104
105	theme	1	0.24%	72.32%	105
106	parks	1	0.24%	72.55%	106
107	social	1	0.24%	72.79%	107
108	amazes	1	0.24%	73.03%	108
109	locals	1	0.24%	73.27%	109
110	traces	1	0.24%	73.51%	110

111	people	1	0.24%	73.75%	111
112	unique	1	0.24%	73.99%	112
113	thirty	1	0.24%	74.22%	113
114	during	1	0.24%	74.46%	114
115	agency	1	0.24%	74.70%	115
116	future	1	0.24%	74.94%	116
117	others	1	0.24%	75.18%	117
118	ground	1	0.24%	75.42%	118
119	global	1	0.24%	75.66%	119
120	impact	1	0.24%	75.89%	120
121	public	1	0.24%	76.13%	121
122	chains	1	0.24%	76.37%	122
123	flight	1	0.24%	76.61%	123
124	Mexico	1	0.24%	76.85%	124
125	tested	1	0.24%	77.09%	125
126	within	1	0.24%	77.33%	126
127	guided	1	0.24%	77.57%	127
128	history	1	0.24%	77.80%	128
129	admired	1	0.24%	78.04%	129
130	amongst	1	0.24%	78.28%	130
131	debated	1	0.24%	78.52%	131
132	several	1	0.24%	78.76%	132
133	support	1	0.24%	79.00%	133
134	session	1	0.24%	79.24%	134
135	doctors	1	0.24%	79.47%	135
136	allowed	1	0.24%	79.71%	136
137	because	1	0.24%	79.95%	137
138	squares	1	0.24%	80.19%	138
139	centers	1	0.24%	80.43%	139
140	variety	1	0.24%	80.67%	140
141	America	1	0.24%	80.91%	141
142	natural	1	0.24%	81.15%	142
143	Antigua	1	0.24%	81.38%	143
144	tourism	1	0.24%	81.62%	144
145	aspects	1	0.24%	81.86%	145
146	quantity	1	0.24%	82.10%	146
147	restored	1	0.24%	82.34%	147
148	governor	1	0.24%	82.58%	148

149	assambly	1	0.24%	82.82%	149
150	financial	1	0.24%	83.05%	150
151	programs	1	0.24%	83.29%	151
152	movement	1	0.24%	83.53%	152
153	eleventh	1	0.24%	83.77%	153
154	diabetes	1	0.24%	84.01%	154
155	demanded	1	0.24%	84.25%	155
156	services	1	0.24%	84.49%	156
157	theater	1	0.24%	84.73%	157
158	cultural	1	0.24%	84.96%	158
159	colonial	1	0.24%	85.20%	159
160	shopping	1	0.24%	85.44%	160
161	political	1	0.24%	85.68%	161
162	religious	1	0.24%	85.92%	162
163	perfectly	1	0.24%	86.16%	163
164	thousands	1	0.24%	86.40%	164
165	different	1	0.24%	86.63%	165
166	astounded	1	0.24%	86.87%	166
167	testimony	1	0.24%	87.11%	167
168	excellent	1	0.24%	87.35%	168
169	continent	1	0.24%	87.59%	169
170	committee	1	0.24%	87.83%	170
171	decisions	1	0.24%	88.07%	171
172	producing	1	0.24%	88.31%	172
173	Caribbean	1	0.24%	88.54%	173
174	nutrition	1	0.24%	88.78%	174
175	providers	1	0.24%	89.02%	175
176	operators	1	0.24%	89.26%	176
177	touristic	1	0.24%	89.50%	177
178	amusement	1	0.24%	89.74%	178
179	following	1	0.24%	89.98%	179
180	volcanoes	1	0.24%	90.21%	180
181	adventure	1	0.24%	90.45%	181
182	birthplace	1	0.24%	90.69%	182
183	economical	1	0.24%	90.93%	183
184	foreigners	1	0.24%	91.17%	184
185	congresses	1	0.24%	91.41%	185
186	experience	1	0.24%	91.65%	186

187	nineteenth	1	0.24%	91.89%	187
188	metabolism	1	0.24%	92.12%	188
189	specialize	1	0.24%	92.36%	189
190	acquainted	1	0.24%	92.60%	190
191	integrated	1	0.24%	92.84%	191
192	convention	1	0.24%	93.08%	192
193	facilities	1	0.24%	93.32%	193
194	handicraft	1	0.24%	93.56%	194
195	complement	1	0.24%	93.79%	195
196	alternative	1	0.24%	94.03%	196
197	continental	1	0.24%	94.27%	197
198	experiences	1	0.24%	94.51%	198
199	restaurants	1	0.24%	94.75%	199
200	attractions	1	0.24%	94.99%	200
201	communities	1	0.24%	95.23%	201
202	opportunity	1	0.24%	95.47%	202
203	Mesoamerican	1	0.24%	95.70%	203
204	organization	1	0.24%	95.94%	204
205	professional	1	0.24%	96.18%	205
206	connectivity	1	0.24%	96.42%	206
207	United States	1	0.24%	96.66%	207
208	birdwatching	1	0.24%	96.90%	208
209	pre-Columbian	1	0.24%	97.14%	209
210	civilizations	1	0.24%	97.37%	210
211	trascendental	1	0.24%	97.61%	211
212	archeological	1	0.24%	97.85%	212
213	complementary	1	0.24%	98.09%	213
214	endocrinology	1	0.24%	98.33%	214
215	extraordinary	1	0.24%	98.57%	215
216	entertainment	1	0.24%	98.81%	216
217	multicultural	1	0.24%	99.05%	217
218	Inter-American	1	0.24%	99.28%	218
219	infrastructure	1	0.24%	99.52%	219
220	transportation	1	0.24%	99.76%	220
221	representatives	1	0.24%	100.00%	221

INGUAT. Guatemala Heart of the Mayan World, An Ideal Destination for Meetings. Why Guatemala?
Pág.12

Texto No. 3: Señales de que le mienten

5 señales de alerta y 5 consejos que lo ayudarán a descubrir la verdad.

Por: Christina Lanzito | fuente: AARP | 23 de agosto de 2012

"Qué pena, pero tengo que trabajar hasta tarde de nuevo".

"Todo lo que necesita para conseguir esta oferta es darnos su número de Seguro Social".

"Esto no tomará más de unos minutos".

¿Sospecha que alguien le miente? No podemos leer las mentes —lo que probablemente es bueno en la mayoría de los casos— pero según un nuevo libro titulado *Spy the Lie*, (Descubra la mentira), hay formas fáciles de determinar si lo están engañando. Los autores aseguran que si aprende a reconocer ciertos indicios físicos y verbales, puede ser tan acertado en su evaluación como un examen de polígrafo.

Ellos están capacitados para hablar del tema. Philip Houston, Michael Floyd y Susan Carnicero son exagentes de la CIA (Agencia Central de Inteligencia de Estados Unidos) que han condensado los conocimientos e instintos adquiridos tras años de arrestar a supuestos maleantes. Entérese de cinco "señales de alerta" que pueden despertar sospechas si alguien le estaría mintiendo.

1. La persona repite su pregunta o hace afirmaciones que no son respuestas, como "esa es una buena pregunta" o "me alegro que me haya preguntado eso". Estas respuestas pueden indicar un intento de ganar tiempo mientras se formula una respuesta verosímil. Igualmente sospechosa es una pausa poco característica antes de contestar lo que generalmente se considera una pregunta sencilla ("¿Lo han despedido alguna vez de un trabajo?").
2. Da respuestas que son —según denominan los autores— "afirmaciones persuasivas". Son respuestas cuyo fin no es transmitir información sino influenciar la percepción que tiene usted de éste. Si le pregunta: "¿Te llevaste

mi billetera?", puede que responda: "Yo no haría tal cosa" o "Soy una persona honrada". Cuando la respuesta debería ser un simple "No".

3. Invoca la religión ("Juro sobre la Biblia" o "Dios es mi testigo".) Los psicólogos llaman a esto "vestir la mentira". Es otra versión de la afirmación persuasiva. En otras palabras, se justifica demasiado.
4. Se cubre o se toca la boca o los ojos. Una reacción natural cuando se dice una mentira es el deseo de ocultarla. El interlocutor puede que se lo demuestre, y de forma literal, poniéndose la mano sobre la boca o cerca de ella. Taparse los ojos puede indicar el deseo inconsciente de no ver la reacción del otro a la mentira.

Se debe señalar que, a diferencia de lo que comúnmente se cree, la falta de contacto visual no es un indicador confiable de que alguien miente. Una persona mentirosa que ha aprendido a mirar a los demás a los ojos —porque todos asocian eso con la veracidad— puede que logre hacer mejor contacto visual que una persona sincera, dice Houston.

5. No puede quedarse quieto. Una persona que tiene las piernas cruzadas, por ejemplo, puede que empiece a mover un pie. Otra indicación es retorcer o frotar las manos, junto con lo que llaman "gestos de acicalamiento", como ajustarse la ropa, mirarse las uñas o rascarse. "El sistema nervioso está tratando de disipar la ansiedad", dice Houston.

Tabla No.3: ordenación de ocurrencia de palabras de un artículo de revista en español.

Orden	Palabra	Ocurrencia	Fr	Fa	Constante
1	de	28	5.57%	5.57%	28
2	que	26	5.17%	10.74%	52
3	la	20	3.98%	14.71%	60
4	a	10	1.99%	16.70%	40
5	es	10	1.99%	18.69%	50
6	una	10	1.99%	20.68%	60
7	o	9	1.79%	22.47%	63
8	lo	9	1.79%	24.25%	72
9	se	9	1.79%	26.04%	81
10	no	9	1.79%	27.83%	90
11	los	9	1.79%	29.62%	99
12	un	7	1.39%	31.01%	84
13	puede	7	1.39%	32.41%	91
14	persona	5	0.99%	33.40%	70
15	y	4	0.80%	34.19%	60
16	el	4	0.80%	34.99%	64
17	si	4	0.80%	35.79%	68
18	le	4	0.80%	36.58%	72
19	son	4	0.80%	37.38%	76
20	las	4	0.80%	38.17%	80
21	mentira	4	0.80%	38.97%	84
22	pregunta	4	0.80%	39.76%	88
23	respuestas	4	0.80%	40.56%	92
24	en	3	0.60%	41.15%	72
25	su	3	0.60%	41.75%	75
26	dice	3	0.60%	42.35%	78
27	ojos	3	0.60%	42.94%	81
28	como	3	0.60%	43.54%	84
29	cinco	3	0.60%	44.14%	87
30	alguien	3	0.60%	44.73%	90
31	señales	3	0.60%	45.33%	93
32	mi	2	0.40%	45.73%	64
33	me	2	0.40%	46.12%	66
34	con	2	0.40%	46.52%	68
35	eso	2	0.40%	46.92%	70

36	han	2	0.40%	47.32%	72
37	del	2	0.40%	47.71%	74
38	ser	2	0.40%	48.11%	76
39	boca	2	0.40%	48.51%	78
40	otra	2	0.40%	48.91%	80
41	esto	2	0.40%	49.30%	82
42	para	2	0.40%	49.70%	84
43	pero	2	0.40%	50.10%	86
44	deseo	2	0.40%	50.50%	88
45	llaman	2	0.40%	50.89%	90
46	sobre	2	0.40%	51.29%	92
47	tiene	2	0.40%	51.69%	94
48	están	2	0.40%	52.09%	96
49	según	2	0.40%	52.49%	98
50	nuevo	2	0.40%	52.88%	100
51	visual	2	0.40%	53.28%	102
52	cuando	2	0.40%	53.68%	104
53	pueden	2	0.40%	54.08%	106
54	miente	2	0.40%	54.47%	108
55	alerta	2	0.40%	54.87%	110
56	Houston	2	0.40%	55.27%	112
57	autores	2	0.40%	55.67%	114
58	indicar	2	0.40%	56.06%	116
59	contacto	2	0.40%	56.46%	118
60	reacción	2	0.40%	56.86%	120
61	respuesta	2	0.40%	57.26%	122
62	afirmaciones	2	0.40%	57.65%	124
63	e	1	0.20%	57.85%	63
64	ha	1	0.20%	58.05%	64
65	yo	1	0.20%	58.25%	65
66	te	1	0.20%	58.45%	66
67	da	1	0.20%	58.65%	67
68	fin	1	0.20%	58.85%	68
69	pie	1	0.20%	59.05%	69
70	por	1	0.20%	59.24%	70
71	ver	1	0.20%	59.44%	71
72	soy	1	0.20%	59.64%	72
73	tal	1	0.20%	59.84%	73

74	vez	1	0.20%	60.04%	74
75	esa	1	0.20%	60.24%	75
76	CIA	1	0.20%	60.44%	76
77	tan	1	0.20%	60.64%	77
78	hay	1	0.20%	60.83%	78
79	lie	1	0.20%	61.03%	79
80	the	1	0.20%	61.23%	80
81	spy	1	0.20%	61.43%	81
82	más	1	0.20%	61.63%	82
83	éste	1	0.20%	61.83%	83
84	está	1	0.20%	62.03%	84
85	ropa	1	0.20%	62.23%	85
86	debe	1	0.20%	62.43%	86
87	cree	1	0.20%	62.62%	87
88	leer	1	0.20%	62.82%	88
89	otro	1	0.20%	63.02%	89
90	ella	1	0.20%	63.22%	90
91	mano	1	0.20%	63.42%	91
92	toca	1	0.20%	63.62%	92
93	Dios	1	0.20%	63.82%	93
94	juro	1	0.20%	64.02%	94
95	cosa	1	0.20%	64.21%	95
96	sino	1	0.20%	64.41%	96
97	cuyo	1	0.20%	64.61%	97
98	poco	1	0.20%	64.81%	98
99	haya	1	0.20%	65.01%	99
100	hacer	1	0.20%	65.21%	100
101	años	1	0.20%	65.41%	101
102	tras	1	0.20%	65.61%	102
103	tema	1	0.20%	65.81%	103
104	unos	1	0.20%	66.00%	104
105	esta	1	0.20%	66.20%	105
106	todo	1	0.20%	66.40%	106
107	pena	1	0.20%	66.60%	107
108	uñas	1	0.20%	66.80%	108
109	junto	1	0.20%	67.00%	109
110	manos	1	0.20%	67.20%	110
111	mover	1	0.20%	67.40%	111

112	mejor	1	0.20%	67.59%	112
113	hacer	1	0.20%	67.79%	113
114	cerca	1	0.20%	67.99%	114
115	todos	1	0.20%	68.19%	115
116	demás	1	0.20%	68.39%	116
117	mirar	1	0.20%	68.59%	117
118	logre	1	0.20%	68.79%	118
119	falta	1	0.20%	68.99%	119
120	forma	1	0.20%	69.18%	120
121	cubre	1	0.20%	69.38%	121
122	otras	1	0.20%	69.58%	122
123	haría	1	0.20%	69.78%	123
124	usted	1	0.20%	69.98%	124
125	antes	1	0.20%	70.18%	125
126	pausa	1	0.20%	70.38%	126
127	ganar	1	0.20%	70.58%	127
128	estas	1	0.20%	70.78%	128
129	buena	1	0.20%	70.97%	129
130	ellos	1	0.20%	71.17%	130
131	libro	1	0.20%	71.37%	131
132	casos	1	0.20%	71.57%	132
133	bueno	1	0.20%	71.77%	133
134	mentes	1	0.20%	71.97%	134
135	tarde	1	0.20%	72.17%	135
136	hasta	1	0.20%	72.37%	136
137	tengo	1	0.20%	72.56%	137
138	gestos	1	0.20%	72.76%	138
139	frotar	1	0.20%	72.96%	139
140	quieto	1	0.20%	73.16%	140
141	porque	1	0.20%	73.36%	141
142	vestir	1	0.20%	73.56%	142
143	Biblia	1	0.20%	73.76%	143
144	invoca	1	0.20%	73.96%	144
145	simple	1	0.20%	74.16%	145
146	alguna	1	0.20%	74.35%	146
147	tiempo	1	0.20%	74.55%	147
148	alegro	1	0.20%	74.75%	148
149	repite	1	0.20%	74.95%	149

150	hablar	1	0.20%	75.15%	150
151	examen	1	0.20%	75.35%	151
152	formas	1	0.20%	75.55%	152
153	tomarán	1	0.20%	75.75%	153
154	social	1	0.20%	75.94%	154
155	seguro	1	0.20%	76.14%	155
156	número	1	0.20%	76.34%	156
157	darnos	1	0.20%	76.54%	157
158	oferta	1	0.20%	76.74%	158
159	verdad	1	0.20%	76.94%	159
160	disipar	1	0.20%	77.14%	160
161	sistema	1	0.20%	77.34%	161
162	mirarse	1	0.20%	77.53%	162
163	empiece	1	0.20%	77.73%	163
164	ejemplo	1	0.20%	77.93%	164
165	piernas	1	0.20%	78.13%	165
166	sincera	1	0.20%	78.33%	166
167	asocian	1	0.20%	78.53%	167
168	taparse	1	0.20%	78.73%	168
169	señalar	1	0.20%	78.93%	169
170	literal	1	0.20%	79.13%	170
171	natural	1	0.20%	79.32%	171
172	versión	1	0.20%	79.52%	172
173	testigo	1	0.20%	79.72%	173
174	debería	1	0.20%	79.92%	174
175	honrada	1	0.20%	80.12%	175
176	trabajo	1	0.20%	80.32%	176
177	formula	1	0.20%	80.52%	177
178	intento	1	0.20%	80.72%	178
179	estaría	1	0.20%	80.91%	179
180	central	1	0.20%	81.11%	180
181	agencia	1	0.20%	81.31%	181
182	físicos	1	0.20%	81.51%	182
183	ciertos	1	0.20%	81.71%	183
184	aprende	1	0.20%	81.91%	184
185	fáciles	1	0.20%	82.11%	185
186	mayoría	1	0.20%	82.31%	186
187	podemos	1	0.20%	82.50%	187

188	minutos	1	0.20%	82.70%	188
189	consejos	1	0.20%	82.90%	189
190	mienten	1	0.20%	83.10%	190
191	ansiedad	1	0.20%	83.30%	191
192	tratando	1	0.20%	83.50%	192
193	nervioso	1	0.20%	83.70%	193
194	rascarse	1	0.20%	83.90%	194
195	retorcer	1	0.20%	84.10%	195
196	cruzadas	1	0.20%	84.29%	196
197	quedarse	1	0.20%	84.49%	197
198	palabras	1	0.20%	84.69%	198
199	religión	1	0.20%	84.89%	199
200	responda	1	0.20%	85.09%	200
201	llevaste	1	0.20%	85.29%	201
202	sencilla	1	0.20%	85.49%	202
203	mientras	1	0.20%	85.69%	203
204	entérese	1	0.20%	85.88%	204
205	arrestar	1	0.20%	86.08%	205
206	acertado	1	0.20%	86.28%	206
207	verbales	1	0.20%	86.48%	207
208	indicios	1	0.20%	86.68%	208
209	aseguran	1	0.20%	86.88%	209
210	descubra	1	0.20%	87.08%	210
211	titulado	1	0.20%	87.28%	211
212	sospecha	1	0.20%	87.48%	212
213	necesita	1	0.20%	87.67%	213
214	ayudarán	1	0.20%	87.87%	214
215	trabajar	1	0.20%	88.07%	215
216	ajustarse	1	0.20%	88.27%	216
217	veracidad	1	0.20%	88.47%	217
218	aprendido	1	0.20%	88.67%	218
219	mentirosa	1	0.20%	88.87%	219
220	confiable	1	0.20%	89.07%	220
221	indicador	1	0.20%	89.26%	221
222	demuestre	1	0.20%	89.46%	222
223	ocultarla	1	0.20%	89.66%	223
224	demasiado	1	0.20%	89.86%	224
225	justifica	1	0.20%	90.06%	225

226	billetera	1	0.20%	90.26%	226
227	denominan	1	0.20%	90.46%	227
228	despedido	1	0.20%	90.66%	228
229	considera	1	0.20%	90.85%	229
230	contestar	1	0.20%	91.05%	230
231	verosímil	1	0.20%	91.25%	231
232	mintiendo	1	0.20%	91.45%	232
233	sospechas	1	0.20%	91.65%	233
234	despertar	1	0.20%	91.85%	234
235	maleantes	1	0.20%	92.05%	235
236	supuestos	1	0.20%	92.25%	236
237	instintos	1	0.20%	92.45%	237
238	exagentes	1	0.20%	92.64%	238
239	polígrafo	1	0.20%	92.84%	239
240	reconocer	1	0.20%	93.04%	240
241	engañando	1	0.20%	93.24%	241
242	conseguir	1	0.20%	93.44%	242
243	descubrir	1	0.20%	93.64%	243
244	indicación	1	0.20%	93.84%	244
245	comúnmente	1	0.20%	94.04%	245
246	diferencia	1	0.20%	94.23%	246
247	poniéndose	1	0.20%	94.43%	247
248	persuasiva	1	0.20%	94.63%	248
249	afirmación	1	0.20%	94.83%	249
250	psicólogos	1	0.20%	95.03%	250
251	percepción	1	0.20%	95.23%	251
252	información	1	0.20%	95.43%	252
253	transmitir	1	0.20%	95.63%	253
254	sospechosa	1	0.20%	95.83%	254
255	igualmente	1	0.20%	96.02%	255
256	preguntado	1	0.20%	96.22%	256
257	adquiridos	1	0.20%	96.42%	257
258	condensado	1	0.20%	96.62%	258
259	evaluación	1	0.20%	96.82%	259
260	determinar	1	0.20%	97.02%	260
261	influenciar	1	0.20%	97.22%	261
262	persuasivas	1	0.20%	97.42%	262
263	capacitados	1	0.20%	97.61%	263

264	inteligencia	1	0.20%	97.81%	264
265	inconsciente	1	0.20%	98.01%	265
266	interlocutor	1	0.20%	98.21%	266
267	generalmente	1	0.20%	98.41%	267
268	acicalamiento	1	0.20%	98.61%	268
269	característica	1	0.20%	98.81%	269
270	conocimientos	1	0.20%	99.01%	270
271	Michael Floyd	1	0.20%	99.20%	271
272	probablemente	1	0.20%	99.40%	272
273	Estados Unidos	1	0.20%	99.60%	273
274	Philip Houston	1	0.20%	99.80%	274
275	Susan Carnicero	1	0.20%	100.00%	275

Lanzito, C. (2007). Señales de que le mienten. AARP The Magazine. [Recuperado el 25 de noviembre de 2012] Disponible en: <http://www.aarp.org/espanol/hogar-familia/familia-amigos/info-08-2012/senales-mentiras-infidelidad.html>

Texto No.4: Signals that you're being lied to

5 telltale signs — plus 5 tips for getting to the truth

"Sorry, I have to work late again tonight."

"All we need to qualify you for this offer is your Social Security number."

"This isn't going to take any more than a few minutes."

Do you suspect someone is lying to you? You can't read anyone's mind — which is probably a good thing in most cases — but, according to a new book called *Spy the Lie*, there are easy ways to determine whether you're probably being duped. The authors claim that if you know the physical and verbal tics to look and listen for, you can be as accurate as a polygraph test.

They should know. Philip Houston, Michael Floyd and Susan Carnicero are former CIA officers who've condensed the wisdom they've gleaned from decades of busting bad guys. Here are five red flags that should raise your suspicions that perhaps you're being deceived.

1. The person repeats your question, or offers nonanswer statements such as "That's a good question" or "I'm glad you asked that." These responses might signal an attempt to buy time while formulating a believable response. Similarly suspicious is an uncharacteristic pause before answering what you consider a straightforward question ("Have you ever been fired from a job?").
2. You get answers that are what the authors call "convincing statements." They're responses that aren't meant to convey information but rather to influence your perception of the speaker. If you ask, "Did you take my wallet?" you might hear, "I wouldn't do something like that," or "I am an honest person." You were expecting a simple "no."
3. The speaker invokes religion ("I swear on the Bible" or "As Allah is my witness"). Psychologists call this "dressing up the lie." It's another form of a convincing statement. In other words, they doth protest too much.

4. A hand covers or touches the speaker's mouth or eyes. A natural reaction while telling a lie is a desire to hide it. The speaker might display this in literal form by placing a hand near or over the mouth. Obscuring the eyes might indicate the speaker's unconscious desire to be shielded from the interviewer's reaction to the lie.

Also note that, contrary to conventional wisdom, a lack of eye contact is not a reliable indicator of lying. A deceptive person who forces a direct gaze because it's universally associated with truthfulness could have better eye contact than an honest person, Houston says.

5. The speaker can't stay still. Someone sitting with one leg crossed over the other, for example, might start wiggling a foot. Wringing or rubbing hands is another giveaway, along with grooming gestures, such as adjusting clothes, inspecting nails or scratching. "The nervous system is trying to dissipate the anxiety," Houston says.

Tabla No.4: ordenación de ocurrencia de palabras de un artículo de revista en inglés.

Orden	Palabra	Ocurrencia	Fr	Fa	Constante
1	the	22	4.77%	4.77%	22
2	a	20	4.34%	9.11%	40
3	to	17	3.69%	12.80%	51
4	you	14	3.04%	15.84%	56
5	is	9	1.95%	17.79%	45
6	or	9	1.95%	19.74%	54
7	that	9	1.95%	21.69%	63
8	as	5	1.08%	22.78%	40
9	of	5	1.08%	23.86%	45
10	might	5	1.08%	24.95%	50
11	I	4	0.87%	25.81%	44
12	an	4	0.87%	26.68%	48
13	for	4	0.87%	27.55%	52
14	lie	4	0.87%	28.42%	56
15	are	4	0.87%	29.28%	60
16	your	4	0.87%	30.15%	64
17	this	4	0.87%	31.02%	68
18	person	4	0.87%	31.89%	72
19	speaker	4	0.87%	32.75%	76
20	in	3	0.65%	33.41%	60
21	and	3	0.65%	34.06%	63
22	five	3	0.65%	34.71%	66
23	from	3	0.65%	35.36%	69
24	with	3	0.65%	36.01%	72
25	have	3	0.65%	36.66%	75
26	being	3	0.65%	37.31%	78
27	you're	3	0.65%	37.96%	81
28	question	3	0.65%	38.61%	84
29	do	2	0.43%	39.05%	58
30	if	2	0.43%	39.48%	60
31	be	2	0.43%	39.91%	62
32	my	2	0.43%	40.35%	64
33	but	2	0.43%	40.78%	66
34	eye	2	0.43%	41.21%	68
35	it's	2	0.43%	41.65%	70
36	take	2	0.43%	42.08%	72

37	than	2	0.43%	42.52%	74
38	good	2	0.43%	42.95%	76
39	know	2	0.43%	43.38%	78
40	they	2	0.43%	43.82%	80
41	such	2	0.43%	44.25%	82
42	what	2	0.43%	44.69%	84
43	call	2	0.43%	45.12%	86
44	form	2	0.43%	45.55%	88
45	hand	2	0.43%	45.99%	90
46	eyes	2	0.43%	46.42%	92
47	over	2	0.43%	46.85%	94
48	says	2	0.43%	47.29%	96
49	can't	2	0.43%	47.72%	98
50	lying	2	0.43%	48.16%	100
51	other	2	0.43%	48.59%	102
52	mouth	2	0.43%	49.02%	104
53	should	2	0.43%	49.46%	106
54	wisdom	2	0.43%	49.89%	108
55	honest	2	0.43%	50.33%	110
56	desire	2	0.43%	50.76%	112
57	someone	2	0.43%	51.19%	114
58	authors	2	0.43%	51.63%	116
59	another	2	0.43%	52.06%	118
60	contact	2	0.43%	52.49%	120
61	Houston	2	0.43%	52.93%	122
62	probably	2	0.43%	53.36%	124
63	reaction	2	0.43%	53.80%	126
64	speaker's	2	0.43%	54.23%	128
65	responses	2	0.43%	54.66%	130
66	statements	2	0.43%	55.10%	132
67	convincing	2	0.43%	55.53%	134
68	we	1	0.22%	55.75%	68
69	am	1	0.22%	55.97%	69
70	no	1	0.22%	56.18%	70
71	on	1	0.22%	56.40%	71
72	up	1	0.22%	56.62%	72
73	it	1	0.22%	56.83%	73
74	by	1	0.22%	57.05%	74
75	I'm	1	0.22%	57.27%	75

76	all	1	0.22%	57.48%	76
77	few	1	0.22%	57.70%	77
78	new	1	0.22%	57.92%	78
79	spy	1	0.22%	58.13%	79
80	can	1	0.22%	58.35%	80
81	CIA	1	0.22%	58.57%	81
82	bad	1	0.22%	58.79%	82
83	red	1	0.22%	59.00%	83
84	buy	1	0.22%	59.22%	84
85	job	1	0.22%	59.44%	85
86	get	1	0.22%	59.65%	86
87	ask	1	0.22%	59.87%	87
88	did	1	0.22%	60.09%	88
89	too	1	0.22%	60.30%	89
90	not	1	0.22%	60.52%	90
91	who	1	0.22%	60.74%	91
92	any	1	0.22%	60.95%	92
93	one	1	0.22%	61.17%	93
94	leg	1	0.22%	61.39%	94
95	lied	1	0.22%	61.61%	95
96	plus	1	0.22%	61.82%	96
97	tips	1	0.22%	62.04%	97
98	work	1	0.22%	62.26%	98
99	late	1	0.22%	62.47%	99
100	need	1	0.22%	62.69%	100
101	more	1	0.22%	62.91%	101
102	read	1	0.22%	63.12%	102
103	most	1	0.22%	63.34%	103
104	book	1	0.22%	63.56%	104
105	easy	1	0.22%	63.77%	105
106	ways	1	0.22%	63.99%	106
107	tics	1	0.22%	64.21%	107
108	look	1	0.22%	64.43%	108
109	test	1	0.22%	64.64%	109
110	guys	1	0.22%	64.86%	110
111	here	1	0.22%	65.08%	111
112	glad	1	0.22%	65.29%	112
113	time	1	0.22%	65.51%	113
114	been	1	0.22%	65.73%	114

115	ever	1	0.22%	65.94%	115
116	hear	1	0.22%	66.16%	116
117	like	1	0.22%	66.38%	117
118	were	1	0.22%	66.59%	118
119	doth	1	0.22%	66.81%	119
120	much	1	0.22%	67.03%	120
121	hide	1	0.22%	67.25%	121
122	near	1	0.22%	67.46%	122
123	also	1	0.22%	67.68%	123
124	note	1	0.22%	67.90%	124
125	lack	1	0.22%	68.11%	125
126	gaze	1	0.22%	68.33%	126
127	stay	1	0.22%	68.55%	127
128	foot	1	0.22%	68.76%	128
129	mind	1	0.22%	68.98%	129
130	isn't	1	0.22%	69.20%	130
131	signs	1	0.22%	69.41%	131
132	truth	1	0.22%	69.63%	132
133	sorry	1	0.22%	69.85%	133
134	again	1	0.22%	70.07%	134
135	offer	1	0.22%	70.28%	135
136	going	1	0.22%	70.50%	136
137	which	1	0.22%	70.72%	137
138	thing	1	0.22%	70.93%	138
139	cases	1	0.22%	71.15%	139
140	there	1	0.22%	71.37%	140
141	duped	1	0.22%	71.58%	141
142	claim	1	0.22%	71.80%	142
143	flags	1	0.22%	72.02%	143
144	raise	1	0.22%	72.23%	144
145	asked	1	0.22%	72.45%	145
146	these	1	0.22%	72.67%	146
147	while	1	0.22%	72.89%	147
148	pause	1	0.22%	73.10%	148
149	before	1	0.22%	73.32%	149
150	fired	1	0.22%	73.54%	150
151	meant	1	0.22%	73.75%	151
152	swear	1	0.22%	73.97%	152
153	Bible	1	0.22%	74.19%	153

154	Allah	1	0.22%	74.40%	154
155	words	1	0.22%	74.62%	155
156	forces	1	0.22%	74.84%	156
157	could	1	0.22%	75.05%	157
158	still	1	0.22%	75.27%	158
159	start	1	0.22%	75.49%	159
160	hands	1	0.22%	75.70%	160
161	along	1	0.22%	75.92%	161
162	nails	1	0.22%	76.14%	162
163	who've	1	0.22%	76.36%	163
164	that's	1	0.22%	76.57%	164
165	aren't	1	0.22%	76.79%	165
166	social	1	0.22%	77.01%	166
167	number	1	0.22%	77.22%	167
168	called	1	0.22%	77.44%	168
169	verbal	1	0.22%	77.66%	169
170	former	1	0.22%	77.87%	170
171	offers	1	0.22%	78.09%	171
172	signal	1	0.22%	78.31%	172
173	convey	1	0.22%	78.52%	173
174	wallet	1	0.22%	78.74%	174
175	rather	1	0.22%	78.96%	175
176	listen	1	0.22%	79.18%	176
177	simple	1	0.22%	79.39%	177
178	covers	1	0.22%	79.61%	178
179	direct	1	0.22%	79.83%	179
180	better	1	0.22%	80.04%	180
181	system	1	0.22%	80.26%	181
182	trying	1	0.22%	80.48%	182
183	they've	1	0.22%	80.69%	183
184	they're	1	0.22%	80.91%	184
185	signals	1	0.22%	81.13%	185
186	getting	1	0.22%	81.34%	186
187	tonight	1	0.22%	81.56%	187
188	qualify	1	0.22%	81.78%	188
189	minutes	1	0.22%	82.00%	189
190	suspect	1	0.22%	82.21%	190
191	whether	1	0.22%	82.43%	191
192	gleaned	1	0.22%	82.65%	192

193	decades	1	0.22%	82.86%	193
194	busting	1	0.22%	83.08%	194
195	perhaps	1	0.22%	83.30%	195
196	repeats	1	0.22%	83.51%	196
197	attempt	1	0.22%	83.73%	197
198	answers	1	0.22%	83.95%	198
199	invokes	1	0.22%	84.16%	199
200	witness	1	0.22%	84.38%	200
201	protest	1	0.22%	84.60%	201
202	touches	1	0.22%	84.82%	202
203	natural	1	0.22%	85.03%	203
204	telling	1	0.22%	85.25%	204
205	display	1	0.22%	85.47%	205
206	literal	1	0.22%	85.68%	206
207	placing	1	0.22%	85.90%	207
208	indicate	1	0.22%	86.12%	208
209	because	1	0.22%	86.33%	209
210	sitting	1	0.22%	86.55%	210
211	crossed	1	0.22%	86.77%	211
212	example	1	0.22%	86.98%	212
213	rubbing	1	0.22%	87.20%	123
214	reliable	1	0.22%	87.42%	214
215	clothes	1	0.22%	87.64%	215
216	nervous	1	0.22%	87.85%	216
217	anxiety	1	0.22%	88.07%	217
218	anyone's	1	0.22%	88.29%	218
219	wouldn't	1	0.22%	88.50%	219
220	telltale	1	0.22%	88.72%	220
221	security	1	0.22%	88.94%	221
222	physical	1	0.22%	89.15%	222
223	accurate	1	0.22%	89.37%	223
224	officers	1	0.22%	89.59%	224
225	deceived	1	0.22%	89.80%	225
226	response	1	0.22%	90.02%	226
227	consider	1	0.22%	90.24%	227
228	religion	1	0.22%	90.46%	228
229	dressing	1	0.22%	90.67%	229
230	shielded	1	0.22%	90.89%	230
231	contrary	1	0.22%	91.11%	231

232	wiggling	1	0.22%	91.32%	232
233	wringing	1	0.22%	91.54%	233
234	giveaway	1	0.22%	91.76%	234
235	grooming	1	0.22%	91.97%	235
236	gestures	1	0.22%	92.19%	236
237	according	1	0.22%	92.41%	237
238	determine	1	0.22%	92.62%	238
239	polygraph	1	0.22%	92.84%	239
240	condensed	1	0.22%	93.06%	240
241	nonanswer	1	0.22%	93.28%	241
242	similarly	1	0.22%	93.49%	242
243	answering	1	0.22%	93.71%	243
244	influence	1	0.22%	93.93%	244
245	something	1	0.22%	94.14%	245
246	expecting	1	0.22%	94.36%	246
247	statement	1	0.22%	94.58%	247
248	obscuring	1	0.22%	94.79%	248
249	conventional	1	0.22%	95.01%	249
250	indicator	1	0.22%	95.23%	250
251	deceptive	1	0.22%	95.44%	251
252	adjusting	1	0.22%	95.66%	252
253	dissipate	1	0.22%	95.88%	253
254	suspicious	1	0.22%	96.10%	254
255	formulating	1	0.22%	96.31%	255
256	believable	1	0.22%	96.53%	256
257	suspicious	1	0.22%	96.75%	257
258	perception	1	0.22%	96.96%	258
259	associated	1	0.22%	97.18%	259
260	inspecting	1	0.22%	97.40%	260
261	scratching	1	0.22%	97.61%	261
262	information	1	0.22%	97.83%	262
263	unconscious	1	0.22%	98.05%	263
264	universally	1	0.22%	98.26%	264
265	truthfulness	1	0.22%	98.48%	265
266	interviewer's	1	0.22%	98.70%	266
267	Michael Floyd	1	0.22%	98.92%	267
268	psychologists	1	0.22%	99.13%	268
269	Philip Houston	1	0.22%	99.35%	269
270	Susan Carnicero	1	0.22%	99.57%	270

271	straightforward	1	0.22%	99.78%	271
272	uncharacteristic	1	0.22%	100.00%	272

Lanzito, C. (2007). Signals That You're Being Lied To. AARP The Magazine. [Recuperado el 25 de noviembre de 2012] Disponible en: <http://www.aarp.org/espanol/hogar-familia/familia-amigos/info-08-2012/senales-mentiras-infidelidad.html>

Texto No.5: La zorra y las uvas

Era una tarde muy soleada y calurosa. Una zorra, que había estado cazando todo el día, estaba muy sedienta.

“Cómo me gustaría encontrar agua”, pensó la zorra.

En ese momento vio un racimo de uvas grandes y jugosas colgando muy alto de una parra. Las uvas parecían maduras y llenas de zumo.

“¡Oh, oh!” dijo la zorra mientras la boca se le hacía agua. “ El zumo dulce de uva sacia mi sed!”.

La zorra se puso de puntillas y se estiró todo lo alto que pudo, pero las uvas estaban fuera de su alcance.

No queriendo abandonar, la zorra tomó impulso para alcanzar las uvas. Fue inútil, no pudo alcanzar las uvas.

La zorra saltó y brincó una y otra vez pero no pudo alcanzar las uvas en ninguna ocasión. Al final la zorra estaba más sedienta y cansada que nunca.

“¡Qué tonta soy!” dijo la zorra con rabia. “Las uvas están verdes y no se pueden comer. De todas maneras, ¿para qué las querría?”

Y así se marchó la zorra.

Moraleja: Algunas personas desdeñan y menosprecian lo que no pueden tener.

Tabla No.5: ordenación de ocurrencia de palabras de una fábula en español.

Orden	Palabra	Ocurrencia	Fr	Fa	Constante
1	y	11	5.88%	5.88%	11
2	la	10	5.35%	11.23%	20
3	zorra	10	5.35%	16.58%	30
4	las	8	4.28%	20.86%	32
5	uvas	8	4.28%	25.13%	40
6	de	7	3.74%	28.88%	42
7	se	5	2.67%	31.55%	35
8	no	5	2.67%	34.22%	40
9	una	4	2.14%	36.36%	36
10	que	4	2.14%	38.50%	40
11	muy	3	1.60%	40.11%	33
12	pudo	3	1.60%	41.71%	36
13	alcanzar	3	1.60%	43.32%	39
14	el	2	1.07%	44.39%	28
15	en	2	1.07%	45.45%	30
16	oh	2	1.07%	46.52%	32
17	lo	2	1.07%	47.59%	34
18	qué	2	1.07%	48.66%	36
19	todo	2	1.07%	49.73%	38
20	agua	2	1.07%	50.80%	40
21	alto	2	1.07%	51.87%	42
22	zumo	2	1.07%	52.94%	44
23	dijo	2	1.07%	54.01%	46
24	pero	2	1.07%	55.08%	48
25	para	2	1.07%	56.15%	50
26	estaba	2	1.07%	57.22%	52
27	pueden	2	1.07%	58.29%	54
28	sedienta	2	1.07%	59.36%	56
29	me	1	0.53%	59.89%	29
30	un	1	0.53%	60.43%	30
31	le	1	0.53%	60.96%	31
32	mi	1	0.53%	61.50%	32
33	su	1	0.53%	62.03%	33
34	al	1	0.53%	62.57%	34
35	día	1	0.53%	63.10%	35
36	ese	1	0.53%	63.64%	36
37	vio	1	0.53%	64.17%	37

38	uva	1	0.53%	64.71%	38
39	sed	1	0.53%	65.24%	39
40	fue	1	0.53%	65.78%	40
41	vez	1	0.53%	66.31%	41
42	así	1	0.53%	66.84%	42
43	soy	1	0.53%	67.38%	43
44	con	1	0.53%	67.91%	44
45	más	1	0.53%	68.45%	45
46	era	1	0.53%	68.98%	46
47	cómo	1	0.53%	69.52%	47
48	boca	1	0.53%	70.05%	48
49	puso	1	0.53%	70.59%	49
50	tomó	1	0.53%	71.12%	50
51	otra	1	0.53%	71.66%	51
52	había	1	0.53%	72.19%	52
53	pensó	1	0.53%	72.73%	53
54	parra	1	0.53%	73.26%	54
55	hacía	1	0.53%	73.80%	55
56	dulce	1	0.53%	74.33%	56
57	sacia	1	0.53%	74.87%	57
58	fuera	1	0.53%	75.40%	58
59	saltó	1	0.53%	75.94%	59
60	final	1	0.53%	76.47%	60
61	nunca	1	0.53%	77.01%	61
62	tonta	1	0.53%	77.54%	62
63	rabia	1	0.53%	78.07%	63
64	están	1	0.53%	78.61%	64
65	comer	1	0.53%	79.14%	65
66	todas	1	0.53%	79.68%	66
67	tener	1	0.53%	80.21%	67
68	tarde	1	0.53%	80.75%	68
69	estado	1	0.53%	81.28%	69
70	llenas	1	0.53%	81.82%	70
71	estiró	1	0.53%	82.35%	71
72	alcance	1	0.53%	82.89%	72
73	impuso	1	0.53%	83.42%	73
74	inútil	1	0.53%	83.96%	74
75	brincó	1	0.53%	84.49%	75
76	racimo	1	0.53%	85.03%	76

77	verdes	1	0.53%	85.56%	77
78	marchó	1	0.53%	86.10%	78
79	soleada	1	0.53%	86.63%	79
80	cazando	1	0.53%	87.17%	80
81	momento	1	0.53%	87.70%	81
82	grandes	1	0.53%	88.24%	82
83	jugosas	1	0.53%	88.77%	83
84	maduras	1	0.53%	89.30%	84
85	estaban	1	0.53%	89.84%	85
86	ninguna	1	0.53%	90.37%	86
87	ocasión	1	0.53%	90.91%	87
88	cansada	1	0.53%	91.44%	88
89	maneras	1	0.53%	91.98%	89
90	querría	1	0.53%	92.51%	90
91	algunas	1	0.53%	93.05%	91
92	calurosa	1	0.53%	93.58%	92
93	gustaría	1	0.53%	94.12%	93
94	colgando	1	0.53%	94.65%	94
95	parecían	1	0.53%	95.19%	95
96	mientras	1	0.53%	95.72%	96
97	moraleja	1	0.53%	96.26%	97
98	personas	1	0.53%	96.79%	98
99	desdeñan	1	0.53%	97.33%	99
100	encontrar	1	0.53%	97.86%	100
101	puntillas	1	0.53%	98.40%	101
102	queriendo	1	0.53%	98.93%	102
103	abandonar	1	0.53%	99.47%	103
104	menosprecian	1	0.53%	100.00%	104

Esopo. (2007). La Zorra y las Uvas. [Recuperado: 25 de noviembre de 2012] Disponible en: <http://albalearning.com/audiolibros/samaniego/f1-4-06zorrayuvas.html>

Texto No.6: The fox and the grapes

It was a very hot and sunny afternoon. A fox, which had been hunting the whole day, was very thirsty.

"How I wish there was some water," the fox thought to himself.

Just then, he saw bunches of fat and juicy grapes hanging from a vine above his head. The grapes looked ripe and ready to burst with juice.

"Oh, my! Oh, my!" the fox said as his mouth began to water. "Sweet grape juice, quench my thirst!"

The fox stood on tiptoe and stretched as high as he could, but the grapes were out of his reach.

Not about to give up, the fox walked back a short distance and took a running leap at the grapes. Again, he could not reach the grapes.

The fox jumped and leapt, again and again, but each time he could not reach the grapes. Until, at last, the fox was tired and thirstier than ever.

"What a fool I am!" said the fox furiously. "These grapes are sour and not fit for eating. Why would I want them anyway?" With that, the fox walked away.

Moral: Some people disdain and underestimate what they cannot have.

Tabla No.6: ordenación de ocurrencia de palabras de una fábula en inglés.

Orden	Palabra	Ocurrencia	Fr	Fa	Constante
1	the	16	8.08%	8.08%	16
2	and	11	5.56%	13.64%	22
3	fox	10	5.05%	18.69%	30
4	grapes	8	4.04%	22.73%	32
5	a	6	3.03%	25.76%	30
6	to	4	2.02%	27.78%	24
7	he	4	2.02%	29.80%	28
8	was	4	2.02%	31.82%	32
9	not	4	2.02%	33.84%	36
10	I	3	1.52%	35.35%	30
11	his	3	1.52%	36.87%	33
12	reach	3	1.52%	38.38%	36
13	could	3	1.52%	39.90%	39
14	again	3	1.52%	41.41%	42
15	my	3	1.52%	42.93%	45
16	as	3	1.52%	44.44%	48
17	of	2	1.01%	45.45%	34
18	oh	2	1.01%	46.46%	36
19	at	2	1.01%	47.47%	38
20	but	2	1.01%	48.48%	40
21	very	2	1.01%	49.49%	42
22	said	2	1.01%	50.51%	44
23	some	2	1.01%	51.52%	46
24	with	2	1.01%	52.53%	48
25	what	2	1.01%	53.54%	50
26	water	2	1.01%	54.55%	52
27	juice	2	1.01%	55.56%	54
28	walked	2	1.01%	56.57%	56
29	it	1	0.51%	57.07%	29
30	on	1	0.51%	57.58%	30
31	up	1	0.51%	58.08%	31
32	am	1	0.51%	58.59%	32
33	hot	1	0.51%	59.09%	33
34	had	1	0.51%	59.60%	34
35	day	1	0.51%	60.10%	35
36	how	1	0.51%	60.61%	36
37	saw	1	0.51%	61.11%	37

38	fat	1	0.51%	61.62%	38
39	out	1	0.51%	62.12%	39
40	are	1	0.51%	62.63%	40
41	fit	1	0.51%	63.13%	41
42	for	1	0.51%	63.64%	42
43	why	1	0.51%	64.14%	43
44	been	1	0.51%	64.65%	44
45	wish	1	0.51%	65.15%	45
46	just	1	0.51%	65.66%	46
47	then	1	0.51%	66.16%	47
48	from	1	0.51%	66.67%	48
49	vine	1	0.51%	67.17%	49
50	head	1	0.51%	67.68%	50
51	ripe	1	0.51%	68.18%	51
52	high	1	0.51%	68.69%	52
53	were	1	0.51%	69.19%	53
54	give	1	0.51%	69.70%	54
55	back	1	0.51%	70.20%	55
56	took	1	0.51%	70.71%	56
57	leap	1	0.51%	71.21%	57
58	each	1	0.51%	71.72%	58
59	time	1	0.51%	72.22%	59
60	last	1	0.51%	72.73%	60
61	than	1	0.51%	73.23%	61
62	ever	1	0.51%	73.74%	62
63	fool	1	0.51%	74.24%	63
64	sour	1	0.51%	74.75%	64
65	want	1	0.51%	75.25%	65
66	them	1	0.51%	75.76%	66
67	that	1	0.51%	76.26%	67
68	away	1	0.51%	76.77%	68
69	they	1	0.51%	77.27%	69
70	have	1	0.51%	77.78%	70
71	sunny	1	0.51%	78.28%	71
72	which	1	0.51%	78.79%	72
73	whole	1	0.51%	79.29%	73
74	there	1	0.51%	79.80%	74
75	juicy	1	0.51%	80.30%	75
76	above	1	0.51%	80.81%	76

77	ready	1	0.51%	81.31%	77
78	burst	1	0.51%	81.82%	78
79	mouth	1	0.51%	82.32%	79
80	began	1	0.51%	82.83%	80
81	sweet	1	0.51%	83.33%	81
82	grape	1	0.51%	83.84%	82
83	stood	1	0.51%	84.34%	83
84	about	1	0.51%	84.85%	84
85	short	1	0.51%	85.35%	85
86	leapt	1	0.51%	85.86%	86
87	until	1	0.51%	86.36%	87
88	tired	1	0.51%	86.87%	88
89	these	1	0.51%	87.37%	89
90	would	1	0.51%	87.88%	90
91	moral	1	0.51%	88.38%	91
92	looked	1	0.51%	88.89%	92
93	quench	1	0.51%	89.39%	93
94	thirst	1	0.51%	89.90%	94
95	tiptoe	1	0.51%	90.40%	95
96	jumped	1	0.51%	90.91%	96
97	eating	1	0.51%	91.41%	97
98	anyway	1	0.51%	91.92%	98
99	people	1	0.51%	92.42%	99
100	cannot	1	0.51%	92.93%	100
101	hunting	1	0.51%	93.43%	101
102	thirsty	1	0.51%	93.94%	102
103	thought	1	0.51%	94.44%	103
104	himself	1	0.51%	94.95%	104
105	bunches	1	0.51%	95.45%	105
106	hanging	1	0.51%	95.96%	106
107	running	1	0.51%	96.46%	107
108	disdain	1	0.51%	96.97%	108
109	distance	1	0.51%	97.47%	109
110	afternoon	1	0.51%	97.98%	110
111	stretched	1	0.51%	98.48%	111
112	thirstier	1	0.51%	98.99%	112
113	furiously	1	0.51%	99.49%	113
114	underestimate	1	0.51%	100.00%	114

La Zorra y las Uvas. [Recuperado: 25 de noviembre de 2012] Disponible en: <http://albalearning.com>

GLOSARIO

Abscisa	Coordenada X en un plano cartesiano rectangular, que representa la distancia horizontal de un punto al eje vertical o eje Y. El denominado eje de las abscisas representa al eje de coordenadas horizontal.
Análisis documental	Actividad que tiene como fin la descripción de un documento en sus partes esenciales para su identificación y recuperación para responder a las necesidades de información de los usuarios.
Axiomático	Incontrovertible, evidente.
Ciencia afín	Ciencia que guarda una relación de intercambio con otra, aportándole sus métodos y conclusiones para apoyarla.
Ciencia auxiliar	Ciencia autónoma que funciona como soporte y que presta un servicio a otra ciencia para que cumpla con todos sus objetivos.
Competencia	Conocimientos homogéneos que los hablantes poseen de una lengua.
Constante	Variable matemática que tiene un valor fijo en un determinado proceso o cálculo.
Curva	Línea continua de una dimensión que representa gráficamente la magnitud de un fenómeno según los factores que va tomando una de sus variables.
Estadística Lingüística	Aplicación de métodos estadísticos para el estudio y explicación de fenómenos lingüísticos.

Estadística	Rama de la Matemática que recopila, organiza y procesa información cuantitativa de una determinada característica de la población, que posteriormente es organizada y presentada gráficamente para que los resultados sean analizados y ayuden en la toma de decisiones.
Estilística	Disciplina científica que estudia los procedimientos expresivos o el uso artístico que las personas hacen del habla.
Estilometría	Disciplina desarrollada como herramienta en el análisis estadístico de características cuantificables para tratar problemas de autoría.
Etimología	Especialidad lingüística que se dedica al estudio del origen de las palabras.
Filología	Ciencia que se dedica al estudio de la lengua y la Literatura de un pueblo a través de los textos escritos.
Fonología	Rama de la Lingüística que se dedica al estudio de los elementos fónicos de una lengua.
Frecuencia acumulada	Suma de todas las frecuencias absolutas anteriores a ella.
Frecuencia/Ocurrencia	Número de veces que un elemento se repite dentro de un intervalo determinado.
Gramática	Ciencia que tiene como objeto el estudio de los elementos de una lengua y sus combinaciones.
Idiolecto	Variedades del habla que pertenecen exclusivamente a un sólo individuo.

Indización Automática	Técnica del análisis documental que se utiliza para representar y describir el contenido de los documentos, mediante palabras clave o vocabularios controlados, que permiten guiar al usuario en la recuperación de los documentos que necesita.
Isomorfismo	Que tiene la misma forma o que tiene características similares.
Lengua	Sistema de comunicación verbal y casi siempre escrito, propio de una comunidad humana.
Lenguaje	Capacidad que poseen los seres humanos para comunicarse mediante el uso de diversos signos.
Léxico	Vocabulario, conjunto de las palabras de un idioma, o de las que pertenecen al uso de una región, a una actividad determinada, a un campo semántico, etc.
Lexicografía	Rama de la Lexicología encargada de la elaboración de diccionarios.
Lexicología	Estudio de las unidades léxicas de una lengua y de las relaciones sistemáticas que se establecen entre ellas.
Lingüística Aplicada	Rama de la Lingüística que incluye el estudio de mecanismos para la definición de sistemas ortográficos, procesos de comunicación y medios para desarrollar una lengua.
Lingüística Comparada	Estudio de las semejanzas y diferencias entre las distintas lenguas para establecer sus familias y orígenes comunes.

Lingüística Computacional	Ciencia interdisciplinar que se dedica al estudio de los sistemas de comunicación y los sistemas que se emplean para la comprensión de lenguas naturales.
Lingüística Forense	Interfaz entre el Derecho y la Lingüística que permite la utilización de pruebas periciales en juicios.
Lingüística General	Rama de la Lingüística que se dedica al análisis de la lengua y los aspectos que proporcionan una visión general del lenguaje y de las lenguas.
Lingüística Histórica	Rama de la Lingüística que se encarga del estudio del desarrollo y cambio de las lenguas durante el paso del tiempo en la humanidad.
Lingüística Pura	Rama de la Lingüística que estudia los sistemas generales, tales como: fonología, morfología, sintaxis y semántica.
Lingüística	Estudio científico del lenguaje humano y del conocimiento que los hablantes poseen de su lengua.
Metalingüística	Estudio de la relación del lenguaje humano con la cultura.
Morfología	Parte de la Gramática que se dedica al estudio de la estructura de las palabras.
Plagio	Préstamo que se hace de una idea o texto con plena conciencia de que está siendo copiado y sin hacer ninguna referencia a su autor.
Probabilidad	Cualidad o posibilidad fundada de que algo pueda pasar.

Psicolingüística	Rama de la Lingüística que se encarga de reunir fundamentos de Psicología y Lingüística que permitan estudiar los procesos mentales durante la adquisición y el uso del lenguaje.
Rango	Amplitud de la variación de un fenómeno entre un límite menor y uno mayor claramente especificados.
Semántica	Estudio del significado, sentido o interpretación de los signos lingüísticos y de sus combinaciones.
Semiología	Ciencia que se encarga del estudio de los signos en la vida social.
Sintaxis	Parte de la Gramática que se dedica al estudio de la coordinación de las palabras en las oraciones para expresar conceptos.
Sociolingüística	Rama de la Lingüística que se dedica al estudio de la lengua y su relación o el funcionamiento de ésta dentro de la sociedad.
Texto	Unidad lingüística o conjunto coherente de enunciados orales o escritos que tiene una intención comunicativa.