PONTIFICIA
UNIVERSIDAD
CATÓLICA DE
VALPARAÍSO

Instituto de Literatura y Ciencias del Lenguaje
Facultad de Filosofía y Educación
Pontificia Universidad Católica de Valparaíso

El léxico del español en la red: detección y análisis de sustantivos en las webs de países

Tesis para optar al grado académico de
Licenciado en Lingüística Aplicada

Alumna: Karem Contreras Cortez
Profesora guía: Irene Renau Araque

Viña del Mar, Junio 2019

Tesis financiada con fondos del proyecto Fondecyt Regular nº 1191204.

Agradecimientos

A mi tutora Irene Renau, por la confianza y dedicación entregada en todo el proceso. Además, al profesor Rogelio Nazar, por su paciencia y ayuda en esta investigación. Personalmente, a ambos por el profesionalismo y entusiasmo por este proyecto.

A mis padres y hermana, por su cariño incondicional y apoyo en toda esta etapa académica que sin duda no habría logrado completar sin ellos.

A mi compañero de aventuras, Miguel por no dejarme caer en los momentos más difíciles y siempre apoyarme en mis ideas.

A mis amigas, por su comprensión y fuerza, en particular a Camila por brindarme conocimiento técnico y apoyo en los últimos momentos.

Finalmente, a mis compañeras y amigas de tesis, por ser un pilar fundamental en esta etapa académica y proceso investigativo que significo una gran meta a cumplir.

Índice

1. Introducción

En la presente tesis se expondrá un trabajo de investigación cuyo principal propósito fue recopilar los sustantivos compartidos que poseen los 19 países hispanohablantes estudiados y aquellos que son específicos de Chile. Estos países son Argentina, Bolivia, Chile, Colombia, Costa Rica, República Dominicana, Cuba, Ecuador, El Salvador, Guatemala, Honduras, México, Nicaragua, Panamá, Paraguay, Perú, Uruguay, Venezuela y España. Si bien se han realizado trabajos similares Juilland y Chang-Rodríguez (1964); Davies, (2006), sin embargo no existen estudios actualizados acerca de esta materia, es decir, actualmente no se tiene conocimiento acerca de cuál es el vocabulario representativo por países o un vocabulario en conjunto.

Un aspecto que también cubre este trabajo es la realización de una clasificación preliminar de carácter semántico de los sustantivos más frecuentes compartidos por los países mencionados. Esta categorización por tipos semánticos se realizó según la CPA Ontology del proyecto *Pattern Dictionary of English Verbs* (Hanks, en progreso). Se detectaron los tipos semánticos de los primeros 200 sustantivos y se clasificaron siguiendo la CPA Ontology.

Uno de los aportes de esta investigación se dirige a la enseñanza-aprendizaje del español, como primera o segunda lengua. Este trabajo detecta un conjunto de vocabulario que es representativo de la lengua española en conjunto, y en particular de cada país, por tanto, muestra la base de lo que se debiera enseñar a un aprendiz de español o hablante extranjero que desea adquirir la lengua española como segunda lengua.

En particular, el trabajo de investigación tuvo como objetivo general detectar los sustantivos en las páginas web de los países hispanohablantes. Se trabajó con el corpus EsTenTen (Kilgariff y Renau, 2013), un corpus relativamente actual, conformado con textos extraídos de páginas web. Este corpus tiene como limitación el bajo control del porcentaje de textos provenientes de la prensa, divulgación, literatura, etc. Por esa razón, existe un límite en el corpus estudiado y por ende en la misma investigación, esta limitación se debe justamente a la utilización de un corpus basado en páginas web. En cuanto a los objetivos específicos de esta investigación, en primer lugar nos propusimos establecer un listado de sustantivos del español extraído del EsTenTen, en general y por países. En segundo lugar, identificar la frecuencia y uso activo de los sustantivos generales del español; en tercer lugar, comparar los sustantivos generales con las frecuencias del uso en los países hispanohablantes.

Este trabajo de investigación se divide en los siguientes apartados: en el apartado 2 se encuentran los antecedentes teóricos que sustentan esta tesis. En él se encuentra la definición de palabra, en primer

lugar considerándola como un signo según lo postulado por Saussure (1916) y posteriormente desde la visión gramatical y léxica (unidades léxicas), además se relaciona al sustantivo dentro de la categoría de unidades léxicas y se detalla la conformación de este (2.1). En el apartado de la morfología del sustantivo se encuentra la composición del sustantivo como tipo de palabra y las clasificaciones que se pueden realizar en cuanto a este según lo que expone preferentemente el Manual de La lengua Española (2010), algunos de esos tipos de sustantivos son contables/incontables, abstractos/ concretos, comunes/ propios, eventivos, etc. (2.2). Luego, se encuentra el apartado de la variación del español (2.3). En este apartado se encontrara el primer acercamiento a la base de esta investigación, es decir lo que es la lengua propiamente tal y enfatizando los países que hablan la lengua Española, prosiguiendo con la concepción de lengua estándar y dialectos, y finalizando con lo que se considera una variación en la lengua (2.3.1).

En el apartado de dialectos y tipos de variaciones (2.3.2) se realiza una descripción de los tipos de variaciones que existen, entre ellos están la diafásica, la cual se concentra en variaciones dependiendo de las situaciones o contexto, la diacrónica contempla variaciones a través del tiempo, en la diastrática se presentan variaciones dependiendo de las clases sociales de los hablantes y finalmente en la variante diatópica se concentra en aquellas variaciones que se establecen según espacios geográficos, en países, regiones, barrios, incluso influyen las etnias o la raza. Es esta última variación la más importante dentro de esta investigación, debido a que se ajusta al objetivo de esta tesis, es por esta misma razón que en el siguiente apartado 2.3.3 se enfatiza en las variaciones diatópicas al nivel de los países hispanohablantes, además de agrupar ciertas similitudes y diferencias según las zonas dialectales.

Finalmente, en el último apartado del marco teórico ene el 2.4 denominado el Español en la red, se profundiza en el uso del Español a nivel mundial en la web, además de definir lo que se considera una red social y como afecta a esta tesis, la cual está compuesta en su corpus por sustantivos extraídos desde la web.

Luego, se da paso al apartado de la metodología, se dará a conocer el tipo de investigación (3.1) y además se expondrán las preguntas que guiaron las investigación (3.2), posteriormente se expondrán los objetivos (3.3) los cuales se dividen en general (3.3.1) y los específicos (3.3.2). Continuando con los materiales y métodos (3.4) los cuales estarán especificados 3.4.1 en los materiales y los procesos metodológicos 3.4.2. Finalmente se dará paso a exponer los resultados (4) y las conclusiones finales (5).

2. Antecedentes teóricos

2.1 Unidad léxica

Según lo expuesto por Saussure (1916), la lengua es considerada como un sistema de signos lingüísticos. El signo lingüístico se entiende como una entidad de dos caras, que posee dos elementos que la conforman, los cuales son inseparables. Saussure (1916) utiliza "la palabra *signo* para designar el conjunto, y reemplazar *concepto* e *imagen acústica* respectivamente con *significado* y *significante*; estos dos últimos términos tienen la ventaja de señalar la oposición que los separa, sea entre ellos dos, sea del total de que forman parte" (93). A partir de lo expuesto, se desprenden los elementos de significado y significante como los componentes del signo lingüístico.

El significante corresponde a la imagen acústica, es decir, a los sonidos percibidos por los seres humanos que cobran sentido al asociarlas a un significado, el cual es asociado a un concepto (una imagen mental de la vida real). A modo de ejemplo, la cadena *p-e-r-r-o* consta de una idea o concepto (significado), que es una idea mental que posee el ser humano correspondiente a un mamífero, animal cuadrúpedo, que ladra, etc.; y un significante, formado por los sonidos representados por las letras. En la figura 1 se aprecia el ejemplo mencionado anteriormente:

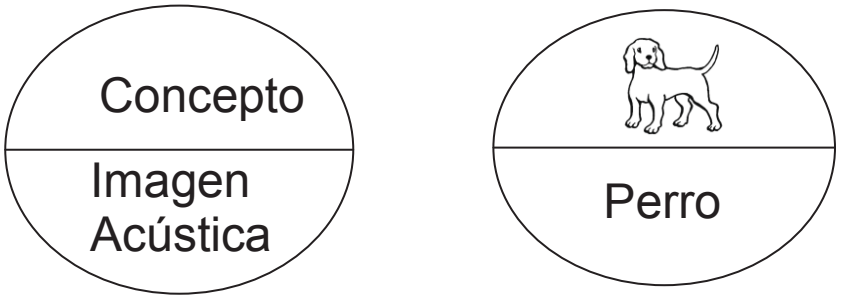


Figura 1: Diagrama del signo lingüístico según Saussure (1916)

Cabe destacar que, si bien los signos poseen, como se ha dicho, un significante y un significado, puede ocurrir que un significante tenga dos significados. A este fenómeno se lo llama *polisemia,* y es definido por Peronard y Gómez (2005) como "fenómeno lingüístico de especial interés para las

disciplinas semánticas. Consiste en que una misma unidad significante o forma de expresión se relaciona con distintos significados o formas de contenido" (162). Por ejemplo, el significante (*perro)* es polisémico, porque significa, según el *Diccionario de la lengua española* (RAE, 2014, acep. 2) "persona despreciable", "Mamífero doméstico de la familia de los cánidos..."(acep.1) entre otras definiciones que varían según el país. Por ejemplo, en el español de Chile, además de estos significados se entiende con este significante a una 'pinza de madera o de plástico para afirmar la ropa a la cuerda en la que se cuelga para secarla'. Cabe destacar, no obstante, que la polisemia no es relevante para esta investigación, que contempla solo el nivel léxico, por lo tanto, no se profundizará en ella.

Esta relación entre los componentes del signo es de carácter arbitrario, lo que quiere decir que no hay una relación causal entre el significante (en este caso *perro*) y el significado (la idea de perro). Saussure (1916) explica, además, que la lengua es impuesta (por generaciones anteriores) y contiene una multiplicidad de signos que son arbitrarios, los cuales pueden ser cuestionados, pero por lo habitual no son modificados con frecuencia. De esta forma, se interpreta que es poco probable que los hablantes de una lengua realicen cambios en ella. Frente a esto, el autor menciona: "…pero en cuanto a la lengua, sistema de signos arbitrarios, (…) y con ella desaparece todo terreno sólido de discusión; no hay motivo alguno para preferir *soeur* a *sister* o a *hermana*, *ochs* a *boeuf* o a *buey*, etcétera" (99). Es decir, no hay algún motivo para escoger un significante con respecto a otro.

Además, Saussure (1916) menciona una característica del significante, que corresponde al carácter lineal: "El significante, por ser de naturaleza auditiva, se desenvuelve en el tiempo únicamente y tiene los caracteres que toma del tiempo: a) representa una extensión, y b) esa extensión es mensurable en una sola dimensión; es una línea" (95). Así, el significante está compuesto básicamente por fonemas, y se desarrolla en una cadena de tiempo; es decir, para establecer la imagen acústica de la palabra es necesario pronunciar ciertos fonemas que componen esta palabra de forma lineal, manteniendo un orden específico. Por ejemplo, para producir la linealidad del significante *gato*, se deberá pronunciar los fonemas en el siguiente orden y continuados: /g/ - /a/- /t/ - /o/.

Otra característica presente en el signo es la inmutabilidad y mutabilidad. Para poder comprender la inmutabilidad, Saussure (1916) toma en consideración cuatro elementos. En primer lugar, está el carácter arbitrario del signo, ya explicado anteriormente. En segundo lugar, es importante considerar la cantidad de signos para la construcción de un sistema, pues esta no tendría límites, pues no sería posible esta construcción con un número limitado de signos. Esto se relaciona con la investigación que se presentará a continuación, pues es imposible detectar todos los sustantivos presentes en la lengua española.

En tercer lugar, está el carácter complejo del sistema. Saussure (1916) considera, como se mencionó antes, que la lengua es un sistema y, por lo tanto, “no se podría concebir un cambio semejante más que con la intervención de especialistas, gramáticos, lógicos, etc.; pero la experiencia demuestra que hasta ahora las injerencias de esta índole no han tenido éxito alguno” (99). En relación a esto, se considera a la sociedad como un grupo que resiste de la inercia a toda innovación lingüística, esto sucede porque “el signo es arbitrario no conoce otra ley que la de la tradición” (Saussure, 1916: 99). Por esta razón, los signos que están establecidos dentro de una sociedad y tiempo continuaran siendo utilizados por otras generaciones.

En cuanto a la mutabilidad, en contraposición a la inmutabilidad, se presenta en relación con los cambios que se establecen entre las dos caras del signo, es decir, significante y significado. El autor nos presenta diferentes transformaciones; por ejemplo, puede haber una mutabilidad del significante y no del significado, una mutabilidad del significado pero no del significante o de ambos. Además, existen mutaciones de signos que se crean o aparecen por la necesidad de la sociedad. Un ejemplo de esto sería el caso del avance de la tecnología, que provoca que se necesiten nuevos signos como *mouse* (‘dispositivo para mover el cursor por un documento informático’), *bloguear* (‘escribir en un blog’), tuitear (‘escribir en Twitter’), *navegar* (existe un cambio semántico, pues ya no se relaciona solamente con ‘trasladarse por el mar en una embarcación’, sino también a ‘consultar páginas de internet’), *descargar* (pasa a significar también ‘extraer documentos o fotos desde la web hacia un dispositivo’), etc.. Estos últimos ejemplos, pertenecen a lo que se conoce como neología, dentro de la cual existen investigaciones interesantes en el área como la de Nazar y Vidal (2008).

Entender la mutabilidad e inmutabilidad del signo es relevante para esta investigación porque, si bien el signo presenta diacronía y sincronía, gracias a la inmutabilidad se pueden analizar los sustantivos presentes en un periodo determinado de tiempo. Sin embargo, un estudio como el presente, debido a la mutabilidad del signo, requerirá de constantes actualizaciones en el futuro.

Como se ha visto, el concepto de signo lingüístico está estrechamente vinculado al de palabra y, en concreto, al de unidad léxica. A continuación, revisaremos algunas aproximaciones a ambos conceptos. El concepto de palabra es visto por algunos autores desde diferentes aristas. Por ejemplo, Piera (2009) lo aborda desde la tradición gramatical de origen grecolatino en que se basan las palabras y como se organizan: “se organizaban a su vez en paradigmas morfológicos, como las declinaciones de los elementos nominales o las conjugaciones de los verbos” (27). Dicho de otra manera, esta tradición se enfoca en que la palabra posee una característica morfosintáctica, es decir, puede ser descompuesta y se compone raíz, afijos o sufijos.

Un ejemplo de lo mencionado anteriormente es la variación en la flexión del verbo *cantar* y la formación de lo que se conoce como flexión verbal. Respecto de esta flexión verbal, Torner (2008) menciona la importancia de la información que aportan los afijos: “Se distinguen las nociones de persona (primera, segunda o tercera), número (singular o plural), tiempo (presente, pasado o futuro), modo (indicativo o subjuntivo) y aspecto (perfectivo e imperfectivo)” (33). Por esta razón, se entiende que del verbo *cantar* existe una variación en relación con el número y el tiempo, según lo que expone este autor. En la misma línea, Piera (2009) considera que la palabra posee la facultad de ser descompuesta: “El significado (saussureano) de las raíces podría entonces ser una entidad de segundo orden, extraída por abstracción de los pares sonido-significado” (37). De esta manera, habría un mismo signo lingüístico pero este contendría ciertas variaciones de tipo gramatical.

Así, por tanto, desde el área de la morfología, las palabras pueden segmentarse en unidades más pequeñas, las cuales son denominadas morfemas (unidad gramatical mínima) (Torner, 2008). La clasificación de los morfemas se basa en criterios. El criterio distribucional, establece que existe una posición o un orden para el morfema (relacionándose esto con la combinación de raíces, afijos, sufijos, prefijos, interfijos).Otro criterio es el semántico, en él se encuentran los llamados morfemas léxicos y gramaticales. Y el criterio sintáctico, está encargado de los morfemas libres y ligados, los cuales presentan funciones propias de la sintaxis (29).

El criterio distribucional manifiesta la presencia de raíz o lexema. Un ejemplo de esto es *cant-* en las palabras *cantar, cantaremos, cantas,* etc. Los afijos son aquellos que se adjuntan a la raíz (*cant-*) y que reciben distintas denominaciones dependiendo de la posición en la que se encuentren: prefijos, sufijos, interfijos y circunfijos. Un ejemplo de prefijo en la palabra *antepasado* corresponde a *ante-,* en cuanto a los sufijos un ejemplo sería *–azo* en la palabra *codazo,* en el caso de los interfijos el ejemplo de *te* en la palabra *tetera*, y los circunfijos como *a / do* en la palabra *anaranjado.* El criterio semántico presenta los morfemas léxicos (poseen significado) y los morfemas gramaticales (no poseen significado y aportan sentido gramatical como el tiempo, género, persona, etc.). El criterio sintáctico, aparte de estudiar los morfemas ligados y libres, se encarga de las funciones sintácticas que pueden realizar los morfemas. Este punto se puede relacionar con la clasificación de las palabras según lo que establece la *Nueva gramática de la lengua española. Manual* (2010), ya que posee una visión de carácter sintáctico en cuanto a la palabra. Además, en ella se menciona y aborda uno de los puntos centrales de esta investigación, es decir, el tipo de palabra sustantivo. El cual será abordado en profundidad más adelante.

En línea con lo anterior, Escandell (2007) menciona la idea de Lyons “…la noción de cohesión interna como rasgo definidor de *palabra*: una palabra puede tener otros componentes menores, pero estos no

pueden reordenarse, ni admiten la interpolación de otras palabras" (21). Claro que otros autores han intentado unir esta idea con lo que se denomina movilidad sintáctica, pero la noción de palabra mencionada antes es clara y se refiere a que la palabra puede ser reordenada morfológicamente, pero en base a criterios y no al azar como por ejemplo en el caso de "mano", ya que esta palabra no podría ordenarse como "omna",

Se han expuesto algunas consideraciones sobre la concepción de palabra en general, y a continuación trataremos el concepto de unidad léxica en particular.

El significado de la unidad léxica se puede relacionar directamente desde el punto de contenido descriptivo, con lo que comenta Renau (2012) según lo expuesto por Sinclair del significado de la unidad léxica "Según la teoría del lexical ítem de este lingüista (Sinclair 1999), la unidad léxica está desprovista de significado por sí misma, de modo que necesita de un contexto para activarlo. De este modo, por ejemplo, levantar tiene un significado intrínseco o «intensión»12 reducido, que podría describirse como 'mover hacia arriba', pero es el contexto el que activa las diferentes «extensiones» o matices de significado"(59).

Un ejemplo a lo referido anteriormente sería el caso de la unidad léxica *Perro,* puesto que en la oración "el *perro* se comió la hamburguesa de la mesa", está unidad léxica acompañada con este contexto (y con la concurrencia al verbo comer) hace que se interprete la referencia al animal (mamífero, cuadrúpedo, etc.), por el contrario, la misma unidad léxica acompañada de otro contexto cambia el significado adquirido, por ejemplo "Hay que comprar más *perros* para la ropa porque no hay" en esta oración la unidad léxica es acompañada con la concurrencia de ropa, por lo tanto hace referencia a las pinzas de ropa (pinzas de madera o plástico para afirmar la ropa colgada luego del lavado). En relación a la unidad léxica, Escandell (2007) expone que el concepto de palabra se puede entender dependiendo a la categoría a la que se adhiera, en cuanto a esto la autora presenta categorías acorde al tipo de significado ya sea de carácter gramatical o léxico. Este último se relaciona con las unidades léxicas porque considera "… que las palabras que tienen significado léxico remiten a conceptos, a partir de los cuales es posible identificar entidades (reales o imaginarias) …las expresiones con significado gramatical indican de manera abstracta el modo en que hay que combinar entre sí los conceptos"(Escandell, 2007, p. 29).

Entonces, se puede decir que según Escandell (2007) las unidades léxicas, son entendidas como aquellas palabras que poseen una estructura abierta (se pueden incorporar elementos), un contenido descriptivo (la relación que se establece con el conocimiento del mundo) y se ligan a representaciones conceptuales accesibles a la introspección (al saber usar la palabra se puede interpretar sus

significados) (Escandell, 2007, p. 30), de esta manera la el tipo de palabra sustantivo sería considerado una unidad léxica, el cual posee una carga significativa para esta investigación y se abordará más adelante.

Es así, como a pesar de que la unidad léxica posea un significado independiente tomando en cuenta esa percepción para esta investigación de las unidades léxicas, es necesario el contexto de la palabra para llevar a cabo la comprensión de lo que se está mencionando. Sin embargo, en esta investigación no será necesario centrarse en el significado, ya que es de carácter léxico, por lo tanto se tomarán las unidades léxicas desde la concepción como diría Saussure del significante, y por supuesto enfocándose en lo mencionado por Sinclair (1999).

2.2 Morfología del sustantivo

En el apartado anterior se describió la concepción de la palabra desde la mirada del signo lingüístico y se explicaron las particularidades de la unidad léxica. En este tipo de palabras se encuentran los sustantivos, que son objeto de estudio de esta tesis y que se describirán de manera general en este apartado.

Según la NGLE, el sustantivo corresponde a una clase de palabra, y desde el punto de vista de la morfología, "se caracteriza por admitir género y número, así como participar en varios procesos de derivación y composición. Desde el punto de sintáctico el sustantivo forma parte de grupos nominales" (2010, p. 209). Estos grupos nominales pueden ser sujeto, complemento directo, término de preposición, entre otros. Además, los sustantivos pueden denominar entidades materiales o inmateriales, por lo que existen varias clases gramaticales para ellos.

Los sustantivos pueden clasificarse según diversos criterios. En primer lugar, se encuentra la clasificación tradicional de los sustantivos en comunes y propios. Esta categoría es conocida como aquella que denomina animales, objetos, personas, etc. Un ejemplo de sustantivo propio son los nombres de ciudades, como *Valparaíso, Santiago, Calama,* etc. Los sustantivos comunes son propios de la nominación de objetos, cosas, entre otras, como sería el caso de *televisión, mochila, árbol.* Con respecto a esta clasificación, en esta investigación solo serán analizados los sustantivos de tipo común. Estos sustantivos se pueden clasificar en diferentes agrupaciones como los contables - no contables, individuales - colectivos, abstractos - concretos (NGLE, p. 210).

Los sustantivos contables son aquellos que pueden ser enumerados. Un ejemplo de este tipo sería *(tres) flores, (una) guitarra, (dos) perros*. En cambio, los sustantivos no contables son aquellos que no

poseen la facultad de enumerarse, como por ejemplo *(mucha) lluvia, (poco) tránsito, (un poco de) vino*. Por otro lado, "los sustantivos individuales denotan personas, animales o cosas que concebimos como entidades únicas *(profesor, oveja, barco)*; los nombres colectivos pueden designar, construidos en singular, conjunto de personas o cosas similares (*profesorado, rebaño, flota)*" (NGLE, p. 210). De esta manera, se puede ver que los individuales y colectivos pueden entenderse como uno parte del otro, es decir, en cómo se apreció en el ejemplo de oveja (individual) pero que es encontrada en el rebaño (colectivo).

La tercera agrupación corresponde a los sustantivos abstractos y concretos. En el caso del primer tipo, son sustantivos abstractos los que designan aquellas entidades que no se pueden materializar o que son acciones que son llevadas a cabo por otras entidades; un ejemplo de esto es el sustantivo *amor*. El amor, si bien es un sentimiento y la mayoría de los seres humanos son capaces de comprender lo que significa, no es algo tangible, más bien es algo abstracto. En cambio, los sustantivos concretos son objetos, personas o entidades que son tangibles; ejemplos de estos son *lápiz, león*, *té*, etc. (NGLE, 2010, p.210).

Las clasificaciones antes mencionadas son las más tradicionales, como menciona el NGLE (2010) en cuanto a la clasificación de los sustantivos. Sin embargo, cabe destacar otra categoría conocida como sustantivos argumentales: estos son "los que se construyen con modificadores o complementos o complementos que designan participantes a pedido de su propio significado" (210). Un ejemplo de esta categoría es *novio*, pues para que este exista es necesaria la participación y relación de dos entidades (un *novio* no puede existir sin otro *novio*). Luego, se encuentran los sustantivos eventivos, que designan eventos o situaciones que pueden estar sujetas al predicado *tener lugar* o términos de la preposición *durante*, además están ubicados temporalmente y pueden ser sujetos del verbo *ser: El almuerzo es a las doce; durante el almuerzo; El almuerzo tuvo lugar en el comedor* (211). De igual manera, se pueden encontrar otras clases como los cuantificadores y los clasificativos, llamados también nombres de clase.

De esta manera, se puede apreciar el sustantivo como una clase de unidad léxica que, como postula Peronard y Gómez (2005), "se caracteriza formalmente porque posee categoría gramatical de género (masculino y femenino) y lleva categoría de número (singular y plural). La categoría de género del sustantivo se reconoce incluso en aquellos casos en que el vocablo no manifiesta el morfema característico" (204).De esta manera, y como ya se ha mencionado anteriormente es que se conforma la concepción de sustantivo que se ocupara como el objeto de análisis de esta investigación.

2.3 Variación en el español

2.3.1 Variación en español, lengua, lengua estándar

La lengua española es hablada "Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, España, Guatemala, Honduras, México, Nicaragua, Panamá, Paraguay, Perú, Portugal, Puerto Rico, República Dominicana, Uruguay y Venezuela"(Moreno, 2007, p. 32). En cada uno de estos países, el español tiene sus propias variantes. Las variaciones pueden ser entre países o dentro de un mismo país. Por ejemplo, la palabra *guagua* en Chile se refiere al 'bebé', sin embargo, en República Dominicana y otros países del Caribe se refiere al 'bus'. Otro ejemplo de variación entre países es el caso de *palta,* que se utiliza en Chile para denominar al 'fruto de carne verde', mientras que en México, España y otros países se llama *aguacate* al mismo fruto. A estas variantes de la lengua se las llama dialectos. Los dialectos son "un término para toda variante de una lengua ligada a una zona geográfica o a un grupo social determinados."(Bernárdez, 1999, p. 59). En relación con esta concepción, Silva Corvalán (2001) menciona este término pero abarca desde una mirada más amplia: "Es un término que se refiere simplemente a una variedad de la lengua compartida por una comunidad. Las lenguas, conceptos abstractos, se realizan en dialectos. Hablar una lengua es hablar un dialecto de una lengua y la forma estándar o de prestigio de una lengua es simplemente otra realización dialectal más" (14). Entonces, se podría afirmar que todas las personas hablan un dialecto, por lo tanto no hay una única lengua española, sino que las lenguas están formadas, en su uso, por un conjunto de dialectos.

La lengua estándar según Bernárdez (1999) es entendida como: "El estándar no es otra cosa que la forma de la lengua socialmente aceptada como la más adecuada para los contextos formales de uso…" (38). Esto quiere decir que finalmente lo que se está enseñando a los extranjeros, como se mencionó antes no es más que un versión de la lengua, en este caso la más aceptada. Según esta concepción, todas las lenguas poseen un estándar.

Para establecer una definición aún más completa de la lengua estándar se debe considerar a Demonte (2003), el cual explica esta concepción desde la mirada de Lewandowski (1982), según el cual

> La lengua de intercambio de una comunidad lingüística, legitimada e institucionalizada históricamente, con carácter suprarregional, que está por encima de la(s) lengua(s) coloquial(es) y los dialectos y es normalizada y transmitida de acuerdo con las normas del uso oral y escrito correcto. Al ser el medio de

> intercomprensión más amplio y extendido, la LE [lengua estándar] se transmite en las escuelas y favorece el ascenso social; frente a los dialectos y sociolectos, [es] el medio de comunicación más abstracto y de mayor extensión social (p.4).

2.3.2 Dialectos y tipos de variaciones

Entonces, se afirma en relación a lo ya mencionado que la lengua posee variación, esta variación para comprenderla se tomará como "Hechos o productos lingüísticos concebidos como realidades cambiantes es decir que suponen o experimentan cambios (…)Toda lengua, concebida como estructura, varia o experimenta variaciones en los distintos niveles de organización" (Peronard, 2005:206).En relación a la variación de la lengua, existe una disciplina que se enfoca en algunos cambios de esta lengua llamada sociolingüística, según Moreno Fernández (1998) la investigación en esta área permite identificar las variables sociales que influyen sobre la variación, existen ciertos niveles de la lengua en los que es más probable la incidencia de factores extralingüísticos (fonética-fonología, morfología), y a pesar de que ocurren hechos lingüísticos sociales recurrentes, no es factible suponer o poder adivinar que variables sociales actuaran sobre los elementos lingüísticos en la comunidad establecida (33).

Con respecto a esto hay que aclarar que la sociolingüística presenta ciertos rasgos similares a otra disciplina, la sociología, a diferencia de esta, la sociología tiene como tarea fundamental la "identificación de las características según las cuales se pueden agrupar o clasificar las situaciones sociales en conjuntos, que tengan correlativos únicos y específicos de conducta lingüística". (Silva-Corvalán, 2001:7). Sin embargo a pesar de ser diferentes poseen ciertos puntos en común, así mismo sucede con la dialectología y la etnografía o la etnolingüística.

Las variables dentro de la lengua son entendidas como un elemento o rasgo que puede manifestarse de modos diversos y cada una de las manifestaciones de una variable que presenta un hablante, se conoce como variaciones. Es un término que suele modificarse según el autor o el punto de vista. Los casos más comunes de variación, son en primer lugar la difásica, esta variación se centra en situaciones específicas, y en como los hablantes utilizan su lengua en determinadas ocasiones, como por ejemplo la variación de un hablante en una cita médica y en una reunión con amigos. En segundo lugar la diacrónica, en donde se pueden aprecias variaciones de la lengua en el tiempo, o incluso verlas reflejadas por la edad de los hablantes.

En tercer lugar está la variación diastrática, esta se relaciona con el estatus o clase social de los hablantes y finalmente en cuarto lugar está la variación diatópica, esta variación está presente en los

cambios de la lengua acorde a la geografía, ya sea en diferentes países o dentro de un mismo país se podría encontrar por zonas específicas del territorio. Cabe destacar que esta última variación será relevante para la investigación y por lo tanto se desarrollara más adelante tomando en consideración cifras y ejemplos de esta.

La primera variación (diafásica) se relaciona con lo mencionado por Moreno Fernández (1998) "educación, nivel o grado de instrucción, estudios o escolaridad son algunas denominaciones que ha recibido la variable que se refiere al tipo de formación académica o de titulación conseguido por los hablantes, lo que está íntimamente relacionado con la cantidad de años que se ha estado estudiando" (55). Cabe destacar, que hay categorías que son utilizadas por un concepto general el grado de escolaridad o estudios, algunos de estos son: analfabeto, educación primaria, educación segundaria, universitaria completa, post-grados y más. Esta variante está relacionada con los estudios y enfocada en cómo cambia el uso de la lengua según el grado de instrucción del hablante, en este caso se puede establecer una apreciación en el ejemplo de un hablante que ha terminado su enseñanza solo básica (hablante A), es decir acomodándose en Chile el hablante habría cumplido hasta octavo básico con una edad de catorce años, a diferencia del uso de la lengua que posee un hablante que ha recibido educación universitaria completa (hablante B), en este ejemplo es inferirle que podría existir una relación entre menos variación o apego más a la norma de parte del hablante A, y una variación mayor de aquel hablante B.

La segunda Variante que se mencionó antes es la diacrónica, en ella se encuentra variaciones en la lengua que se relacionan con el tiempo, o también se pueden apreciar en los rangos de edades que hay en una sociedad. Y con respecto a esto, es percibirle por la comunidad por ejemplo que los jóvenes no utilizan la lengua del mismo modo que lo hace un adulto mayor, o un niño no se expresa de la misma manera que lo hace un adulto. Refiriéndose a esto Moreno Fernández (1998) explica que la edad condiciona la variación lingüística con más intensidad que otros factores, el individuo cambia de edad de forma continua y sin remisión, esta va cambiando y determinando los caracteres, los hábitos sociales de los individuos, incluidos los comunicativos y los puramente lingüísticos (40).

Además, ese autor comenta lo propuesto por J.K Chambers respecto a la evolución de la lengua acompañada de la edad la progresión en el habla se desarrolla en conjunto con la fonología y la sintaxis, y se lleva a cabo en tres periodos formativos en la adquisición de los sociolectos:

> En primer lugar la infancia, aquí se desarrolla lengua bajo la influencia de amigos y familia, luego está la adolescencia, en donde se llevan a mas allá los límites establecidos por generaciones anteriores y se utilizan por ejemplo jergas

> marcando diferencias con generaciones adultas y finalmente esta la edad adulta joven, aquí se suele hacer un mayor uso de la variedad normativa, y tiene influencia las aspiraciones sociales y profesionales en el uso de la lengua. Y se supone que a partir de la última etapa los hablantes estabilizan sus sociolectos (43).

Estos grupos establecidos por Chambers poseen características en el uso de la lengua acorde a la visión de mundo y al nivel de madurez que presentan los hablantes, hay que destacar que esta variante también se relaciona con las demás ya sea por el grupo social al que pertenecen o el nivel de educación.

La tercera variante es la diastrática, en donde se enfocan las diferencias según clases sociales en cuanto a esto Silva-Corvalán (2001) hace referencia a que "la pertenencia a un grupo social u otro influye tanto sobre la manera de hablar como las actitudes hacia estas diferentes maneras de hablar. El termino estratificación social se emplea para referirse al orden jerarquizado de grupos de individuos dentro de una sociedad" (104). Esta autora menciona la estratificación social, en donde se establece relación a niveles superiores e inferiores dentro de lo que es el estatus social, en base a esto se puede vincular la clasificación que presenta Moreno Fernández (1998), sin antes mencionar la importancia que tiene Karl Marx como pionero en establecer el concepto de clases sociales, pues para Marx las clases se dividen en función de la propiedad del capital, y de los medios de producción. De modo que la población se divide en los que tienen capital (clase capitalista) y los que no tienen (proletariado); y los grupos que se adjuntan a esta división (agricultores, pequeño comerciantes y propietarios) son considerados como residuos de la economía pre capitalista destinados a desaparecer (45).

2.3.3 Variación diatópica en el español

La última variación es la diatópica, en esta variación se presentan grados de la variable de la lengua según espacios geográficos, en países, regiones, barrios, incluso influyen las etnias o la raza. En cuanto a lo mencionado Moreno Fernández (1998) apela a que "la procedencia geográfica del hablantes y el barrio de residencia son variables pertinente para la correcta interpretación de algunos fenómenos sociolingüísticos" (62). Este autor presenta la diferencia de las variantes en base a la relación campo ciudad, y a la vez hace alusión a los barrios en donde crecen los hablantes, en base a ellos hace una comparación con lo que serían barrios tradicionales y nuevos en Madrid "En los barrios nuevos, la norma mayoritaria en cuanto a la distinción de *s* y *z,* el seseo o el ceseo, es una distinción; en los barrios tradicionales la norma casi única es el seseo".(63).En cuanto a este caso y en relación a lo que es la realidad Chilena se puede hacer una comparación, por ejemplo en Santiago de Chile,

entre los barrios de la comuna de la *Dehesa (barrio 1*) en comparación con los barrios de la comuna de la *Pintana (barrio 2*), en primer lugar hay un cambio de carácter visual en el entono y los hablantes, en cuanto a lo fonológico se puede considerar el uso del fonema /ch/ y /tch/, para aclarar este ejemplo se verificara en se expondrá la palabra seguido de su forma fonológica (como se pronuncia la palabra acorde al lugar), por ejemplo : dieciocho; /diesiotcho/ correspondiente al *barrio 1*, y /diesiosho/ correspondiente al *barrio 2.*

El ejemplo mencionado anteriormente, está ligado a lo que se menciona por Moreno Fernández (1998), y aquí se presenta el *barrio 1* como una variable que se apega más a la norma que a diferencia del *barrio 2*. Sin embargo, en ambos se presenta una variación de lo que sería la lengua estándar. En esta variante también se encuentra la variación por zonas dentro de un mismo territorio en este caso Chile, en este país al ser extremadamente largo y angosto presenta una variedad de dialectos como variantes del español Chileno, uno de ellos es la denominación al pan llamado *marraqueta,* este tipo de pan suele ser llamado de tres manera según la zona geográfica en donde se encuentre el hablante. En la zona norte del país se llama *pan batido*, en las zonas centro es denominado *marraqueta* y en la zona sur como *pan francés*, hay que destacar que los hablantes de este país son conscientes de esta variación pero no en su totalidad.

En cuanto al español, según el Instituto de Cervantes (2018), “en 2018, más de 480 millones de personas tienen el español como lengua materna. A su vez, el grupo de usuarios potenciales de español en el mundo (cifra que aglutina al Grupo de Dominio Nativo, el Grupo de Competencia Limitada y el Grupo de Aprendices de Lengua Extranjera) supera los 577 millones” (5). Entonces se podría decir que la lengua española es la una de las lenguas más habladas del mundo: “es la segunda lengua materna del mundo por número de hablantes, tras el chino mandarín, y también la segunda lengua en un cómputo global de hablantes (dominio nativo + competencia limitada + estudiantes de español)” (5). El Instituto de Cervantes (2018) realiza una tabla comparativa que da cuenta del número de hablantes del español, ya sea por ser nativos de la zona o con alguna competencia de esta (Figura 2).

Figura 2. Población de los países hispanohablantes según el Instituto de Cervantes (2018).

País	*Población*[1]	*Hablantes nativos (%)*[2]	*Grupo de Dominio Nativo (GDN)*[3]	*Grupo de Competencia Limitada (GCL)*[4]
México	124.737.788[5]	96,80	120.746.179	3.991.609[6]
Colombia	49.608.366[7]	99,20	49.211.499	396.867
España	46.572.132[8]	92,09[9]	42.890.437[10]	3.681.695[11]
Argentina	44.494.502[12]	98,10	43.649.106	845.396
Perú	32.162.184[13]	86,60	27.852.451	4.309.733
Venezuela	31.828.110[14]	97,30	30.968.751	859.359
Chile	18.552.218[15]	95,90	17.791.577	760.641
Guatemala	16.838.489[16]	78,30	13.184.537	3.653.952
Ecuador	15.924.465[17]	95,70	15.239.713	684.752
Cuba	11.417.398[18]	99,70	11.383.146	34.252
Bolivia	11.307.314[19]	83,00	9.385.071	1.922.243
República Dominicana	10.266.149[20]	97,60	10.019.761	246.388
Honduras	9.012.229[21]	98,70	8.895.070	117.159
Paraguay	7.052.983[22]	67,90	4.788.975	2.264.008
El Salvador	6.375.467	99,70	6.356.341	19.126
Nicaragua	6.283.437	97,10	6.101.217	182.220
Costa Rica	5.003.402[23]	99,30	4.968.378	35.024
Panamá	4.158.783[24]	91,90	3.821.922	336.861
Uruguay	3.468.879	98,40	3.413.377	55.502
Puerto Rico	3.337.177[25]	99,00	3.303.805	33.372
Guinea Ecuatorial	1.222.442[26]	74,00	904.607	317.835
Total	459.623.914		434.875.921	24.747.993

En la figura 3 se puede observar la población actual que habla español y se subdividen en aquellos que la poseen como la lengua materna, segunda lengua y los hablantes estudiantes del español.

Tomando en consideración las figura 2 y 3, se establece una relación con lo mencionado por Moreno (2007), puesto que este autor expone las lenguas internacionalmente más usadas y en concordancia con el Instituto de Cervantes se puede observar al español dentro de los tres primero lugares, “Algunas de ellas cuentan con una población nativa muy extensa, como es el caso de las cuatro lenguas de mayor peso demográfico: el chino mandarín, el español, el hindi/urdu y el inglés” (24).

Figura 3. Población hablantes del español. Fuente: Instituto de Cervantes, 2018. Elaboración: I. Renau.

“El Grupo de Competencia Limitada incluye a los hablantes de español de segunda y tercera generación en comunidades bilingües, a los usuarios de variedades de mezcla bilingües y a las personas extranjeras de lengua materna diferente del español residentes en un país hispanohablante” (7). A través de estas cifras, el Instituto hizo una progresión en el tiempo e indica que, “en 2060,

Estados Unidos será el segundo país hispanohablante del mundo, después de México. Las estimaciones realizadas por la Oficina del Censo de los Estados Unidos hablan de que los hispanos serán 119 millones en 2060" (13). De esta manera, se prevé que el uso de la lengua en español irá en crecimiento y por ende hay que tomar en consideración que existirán nuevas variaciones en la lengua española de carácter diatópico.

Figura 4. Superficie de las lenguas oficiales de la ONU. Moreno (2007).

Figura 5. Grandes Áreas Dialectales del Español de América, Moreno (2007).

Como se puede apreciar en la figura 3 el Español está en cuarto lugar, seguido del inglés que tiene el primero, luego el Francés y después el Ruso. Es destacable que si bien la ONU ha expuesto estos datos, no es una de las fuentes más representativas cuando de valores del Español se trata, no obstante uno de los datos que es semejante tanto con Moreno (2007) y el Instituto de Cervantes (2018), es la posición de la lengua Española dentro de los primero 4 primeros lugares a nivel mundial, considerando en esta categoría como una lengua frecuente y con gran cantidad de hablantes.

A la vez, como se encuentra dentro de los primeros lugares se establece una progresión mencionada anteriormente de esta posición a nivel mundial, en cuanto a esta situación el autor Moreno (2007) afirma que “El español se mueve entre las posiciones tercera y cuarta, aunque se considera la posibilidad de que en la primera década del siglo XXI supere al inglés en número de hablantes nativos”(26). De esta manera es esperable que en unos años más el Español poseerá tantas variaciones como lo hace en el presente o incluso más, en cuanto a estas variaciones más lo presentado en cuanto a los dialectos y estudio específico del español según regiones Moreno (2007) presenta que “ es posible distinguir ocho importantes variedades dialectales o geolectales del español en el mundo: en España, la castellana, la andaluza y la canaria; en América, la caribeña, la mexicano-

centroamericana, la andina, la rioplatense y la chilena"(33). Las zonas dialectales que el menciona y que son relevantes para esta investigación están visualizadas en la figura 4 de forma dinámica en el mapa que está en el costado derecho de esta hoja, en él se pueden ver reflejado el dialecto Chileno (morado), Rioplatense (amarillo), Anadino (café), Caribeño (celeste) y Mexicano –Centroamérica (rojo).

Si bien los países que hablan la lengua Española poseen similitudes también tienen rasgos característicos al emplear la lengua, estos rasgos se pueden ver en los dialectos, los cuales se pueden agrupar como se ha mencionado antes de la mano de Moreno (2007) por zonas dialectales. En las zonas dialectales se encuentran en el español y con las cuales se trabajará en esta investigación son las siguientes:

Figura 6. Zonas dialectales de América.

Zonas dialectales	**Países que las conforman**
Caribe	-Puerto Rico -República Dominicana -Cuba
México y Centroamérica	-México -Guatemala -Honduras -Costa Rica -Panamá -Nicaragua -El Salvador
Área Andina	-Venezuela -Bolivia -Colombia -Ecuador - Perú
Rio de la Plata	-Argentina -Paraguay -Uruguay
Chile	-Chile

Estas zonas dialectales fueron extraídas y comparadas por lo postulado por Moreno (2007) y Alba (1992). Cabe destacar de igual forma que, además de los países mencionados se trabajará con España. Además, Puerto Rico, en esta investigación, ha quedado fuera del corpus porque, al no constituir un país sino un estado de Estados Unidos, el EsTenTen no contenía un subcorpus de esta zona.

2.4 El español en la Red

El corpus utiliza en esta investigación, como ya se ha explicado, es el EsTenTen, formado por textos extraídos de páginas web y, por ende, el vocabulario analizado es el que aparece en ellas. En este apartado trataremos brevemente acerca del uso del español en Internet.

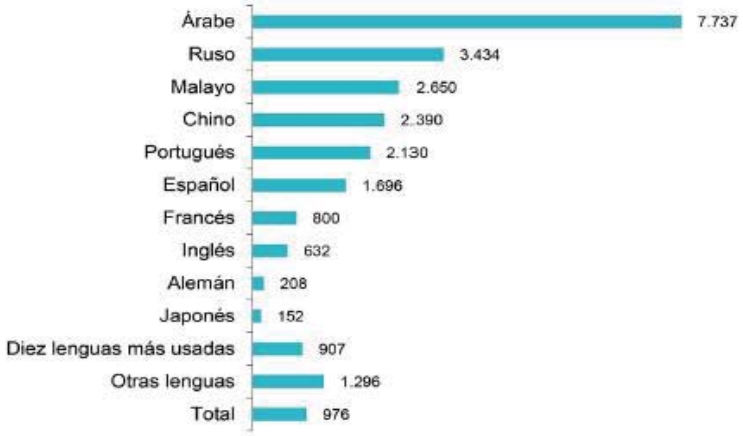
En cuanto a web y a el uso de las diferentes aplicaciones o plataformas digitales para la comunicación, difusión de información, etc., según el Instituto de Cervantes (2018),

> "en la actualidad, el español es la tercera lengua más empleada en Internet por número de internautas. De los casi 3.885 millones de usuarios que tenía Internet en todo el mundo en diciembre de 2017, el 8,1% se comunicaba en español. Los dos idiomas que están por delante del español son el inglés y el chino. Si se tiene en cuenta que el chino es una lengua que, en general, solo la hablan sus nativos, el español se situaría como la segunda lengua de comunicación en Internet tras el inglés" (41).

Entonces, se puede interpretar que el español es una de las lenguas más usadas dentro de la web, y de esta manera la presente investigación da cuenta de un vocabulario bastante representativo del español. A continuación se presenta una comparación de las lenguas más utilizadas en el internet en 2018.

Figura 7. Las lenguas más usadas en la red. Instituto de Cervantes (2018: 41)

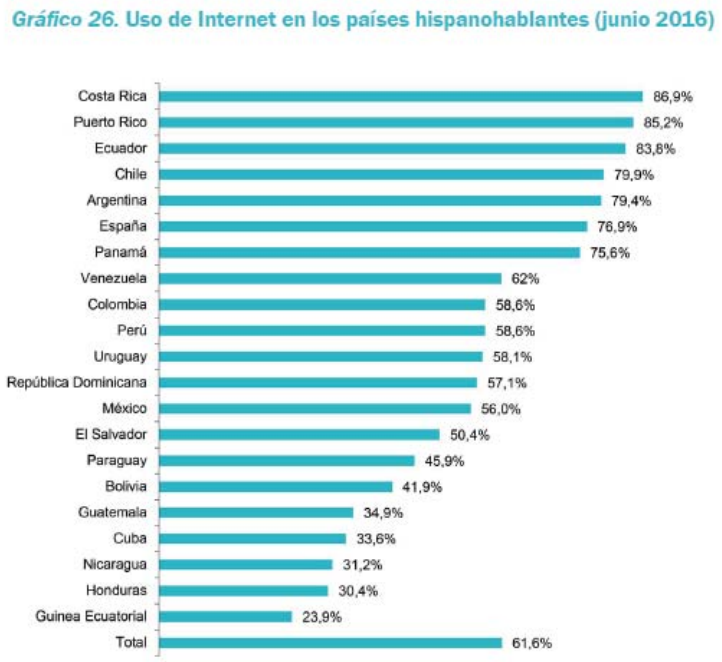
Gráfico 25. **Porcentaje de crecimiento de las lenguas más usadas en la Red (2000-2017)**



Fuente: Internet World Stats, consultado el 3 de marzo de 2018.

No obstante, a pesar de que existe una gran demanda del español en el internet, el uso que se mantiene por parte de los usuarios hispanohablantes es relativamente bajo: “La penetración media de Internet en los países hispanohablantes, o el porcentaje de población que usa Internet, es del 63,3%, lejos de la media europea, que alcanza el 73,5%, y del 76,9% de España” (Instituto Cervantes, 2018: 42) (Figura 8):

Figura 8. Uso del internet en los países hispanohablantes. Instituto de Cervantes (2018) p. 42.



Esto nos indica que, pese a la gran cantidad de textos que hay en Internet, el acceso a la red puede no ser frecuente en algunos países hispanohablantes pobres, lo que constituye una limitación para la generalización de los datos de este estudio.

3. Marco metodológico

3.1 Tipo de investigación

Esta investigación es de tipo exploratorio y descriptivo. En primer lugar, es de carácter exploratorio debido a que no existen muchas investigaciones acerca del estudio de los sustantivos de las páginas web en los países hispanohablantes. Respecto al carácter exploratorio de una investigación, Hernández Sampieri (2014) menciona qué los estudios exploratorios:

> sirven para familiarizarnos con fenómenos relativamente desconocidos, obtener información sobre la posibilidad de llevar a cabo una investigación más completa respecto de un contexto particular, indagar nuevos problemas, identificar conceptos o variables promisorias, establecer prioridades para investigaciones futuras, o sugerir afirmaciones y postulados (91)

Por lo tanto, es un estudio relativamente nuevo que aportará dentro del área de la lexicología en relación con el cálculo de la frecuencia de los sustantivos en varios países, y también en relación con la comparación y registro de estos sustantivos en esta lengua.

Además, el estudio es descriptivo, pues como explica como Hernández Sampieri (2014) en estos tipos de estudios “se busca especificar las propiedades, las características y los perfiles de personas, grupos, comunidades, procesos, objetos o cualquier otro fenómeno que se someta a un análisis” (92). Esta investigación se considera descriptiva debido a que su objetivo principal es detectar los sustantivos utilizados por país, por lo tanto, se estaría caracterizando a estos sustantivos.

En cuanto a los enfoques de este proyecto se puede evidenciar que esta investigación posee un enfoque mixto, es decir, por una parte es cuantitativa porque se extraerán los sustantivos y analizarán por medio de cálculos estadísticos, número de frecuencia, cantidad de países, etc. Pero a la vez, es cualitativa ya que al extraer los datos de la frecuencia, presencia o inexistencia de los sustantivos por países, y se caracteriza semánticamente a los sustantivos encontrados

3.2 Preguntas de investigación

Las preguntas que guían esta investigación son las siguientes:

1. ¿Cuáles son los sustantivos más frecuentes utilizados en español en la web?
2. ¿Cuáles son los sustantivos específicos de Chile?

3.3 Objetivos

Los objetivos generales y específicos que guían esta investigación son los siguientes:

3.3.1 Objetivo general

- Detectar los sustantivos en las páginas web de los países hispanohablantes en el corpus esTenTen.

3.3.2 Objetivos específicos

1. Establecer un listado de sustantivos del español en las páginas web, en general y por países.
2. Identificar la frecuencia y uso activo de los sustantivos generales del español.
3. Extraer los sustantivos específicos de Chile

3.4 Materiales y métodos

3.4.1 Materiales

Se emplearon los siguientes materiales para esta investigación:

1. Corpus esTenTen de Sketch Engine, constituido por 8.038.000.000 de palabras y dividido en subcorpus por países (Kilgarriff & Renau, 2013). Se utilizó el corpus como fuente de los datos porque se pretendía estudiar los sustantivos en uso y no aquellos posiblemente en desuso como los que recogen a menudo los diccionarios.

2. Listado de sustantivos por países extraído del esTenTen. Para obtener estos datos fue necesario extraer todos los sustantivos que en el esTenTen estuviesen etiquetados como

"sustantivos", divididos por países y con la frecuencia de uso. El listado extraído efectivamente se conformó por los etiquetados sustantivos, y dado que al ser un corpus extraído de páginas web contiene múltiples fallas o errores, algunos de los cuales se revisaron a mano.
*(El listado de sustantivos por países fue extraído por el profesor Rogelio Nazar)

3.4.2 Métodos

A continuación se describirán los métodos realizados para el desarrollo de esta investigación.

El listado de sustantivos indicado en el apartado anterior se volcó en un documento Excel para el análisis. Del total de sustantivos, se excluyeron (arbitrariamente) los que tuvieran una frecuencia menor o igual a 1.000. Este método se llevó a cabo con el objetivo de reducir los múltiples errores de etiquetado del corpus.
Luego, se dio paso a realizar una nueva limpieza en este caso manual de los datos, fueron revisados los 12.000 sustantivos más frecuentes (cifra que fue escogida arbitrariamente y considerando la gran cantidad de datos que posee el corpus). Los datos que se fueron eliminando eran todos aquellos que no correspondían a sustantivos (como caso de artículos, pronombres, palabras con símbolos, errores de expresión, etc.). En la figura 9 se muestran algunos ejemplos de errores que se eliminaron del listado.

Figura 9.Tabla Ejemplo errores corpus inicial.

Vandálicos	sistems	Estratégico
transformando	pancakes	Hyogo
infraestrucura	inculpaciones	Trasgresores
derehos	dispensadoras	Primeramente
auriverde	harapientos	Interista
honrradez	aseguradas	Solitud
mirlande	paquetazos	Condecorados
ecepción	foguearse	Autosostenibles
telenoticiero	porquerizas	Hidrólogo
exdelantero	ostente	Oea
drechos	abnegadas	Missouri
priorizados	injustica	Billonaria

ameritas	giz	Nosotros
germoplasmas	fitosanitarias	Jajaja

Una vez realizada la limpieza de datos se procedió a categorizar los 200 sustantivos más frecuentes según su tipo semántico. La categorización se estableció según los tipos semánticos que aparecen en la CPA Ontology del *Pattern Dictionary of English Verbs* (Hanks, en proceso), un diccionario de patrones ubicado en línea para el cual se requiere como en la CPA Ontology " la creación de una muestra aleatoria de concordancias extraídas de corpus, en la que se advierten los patrones verbales a partir del análisis de la estructura sintáctica y argumental y del análisis de los tipos semánticos de los argumentos" (Renau et al. 2019. p. 882)

Posteriormente, se creó una tabla, también en Excel, con los mismos datos pero agrupados por intervalos de frecuencia de 500.000, con el fin de tener mayor claridad acerca de la gran cantidad de datos manejados. Estos intervalos, fueron filtrados y en ellos dependiendo de la cantidad de palabras, es decir, los intervalos se presentó un ejemplo de 50 palabras en aquellos que el número de estas era grande, y en los que presentaban menor cantidad de palabras se expuso el total de la muestra.

A continuación, se dio paso a la utilización de los datos "Especificidad de Chile", cabe destacar que si bien los sustantivos específicos de Chile son expuestos en esta investigación, este procedimiento solo se pudo realizar con este país, el cual fue escogido por preferencia personal. Sin embargo se encuentran el coeficiente de especificidad por país, estos valores representados a través de decimales representan un número decimal el cual entre más cercano este al 1 más específico será de ese país el sustantivo. Con la tabla de sustantivos específicos de Chile se realizó una categorización

según su categoría semántica, utilizando también la CPA Ontology como en el caso de los 200 sustantivos más frecuentes. En este caso también se clasificaron los primeros 200 sustantivos considerados más propios de Chile.

Finalmente, se expusieron los resultados de los métodos expuestos anteriormente y se dio paso a las conclusiones y trabajos futuros.

4. Resultados

4.1 Sustantivos compartidos y clasificación semántica

Una vez eliminados los sustantivos que aparecían 1.000 veces o menos en el total de países, y una vez realizada la limpieza de los datos de los primeros 12.000 sustantivos, los resultados arrojan un total de 148.991 sustantivos compartidos por los 19 países según lo extraído en el corpus de las páginas web. Esto significa que aparecen como mínimo una vez en cada país, aunque lo normal es que aparezcan muchas más veces.

De este total de sustantivos se escogieron los 200 más frecuentes para realizar una clasificación semántica.

Tabla de sustantivos clasificados según tipo semántico.

Sustantivos	Tipo Semántico
año	tiempo
parte	parte
día	tiempo
vez	tiempo
país	lugar
estado	estado
persona	persona
trabajo	evento
caso	evento
tiempo	tiempo
forma	propiedad
servicio	evento
vida	evento
lugar	lugar
empresa	institución
proyecto	evento
ciudad	lugar
sistema	ambiguo
grupo	persona

presidente	persona
momento	tiempo
equipo	ambiguo
hora	tiempo
programa	ambiguo
desarrollo	evento
derecho	ambiguo
cuenta	ambiguo
medio	ambiguo
actividad	actividad
mundo	lugar
punto	ambiguo
problema	eventualidad
centro	lugar
manera	propiedad
tema	concepto
cosa	entidad
tipo	entidad abstracta
gobierno	institución
partido	ambiguo

mes	tiempo
proceso	proceso
información	información
nivel	estado
fin	tiempo
obra	evento
acuerdo	acción
niño	persona
estudio	evento
casa	lugar
través	ambiguo
gente	persona
artículo	documento
sector	lugar
zona	lugar
relación	relación
producto	objeto físico
universidad	institución
cambio	evento
situación	estado
semana	tiempo

acción	acción
salud	propiedad
mujer	persona
resultado	estado
hombre	persona
hecho	evento
seguridad	ambiguo
comunidad	grupo
recurso	concepto
ley	norma
familia	personas
hijo	persona
gracia	ambiguo
agua	agua
historia	evento
sociedad	entidad abstracta
ejemplo	proposición
educación	eventualidad
valor	propiedad
calidad	propiedad
embargo	actividad
condición	norma
frente	ambiguo
medida	eventualidad
número	número
objetivo	meta
nombre	propiedad
investigación	investigación
uso	uso
mercado	lugar
dato	información
organización	actividad
pueblo	lugar
política	actividad
juego	actividad

mano	parte del cuerpo
peso	peso
fecha	tiempo
espacio	lugar
experiencia	evento
autoridad	entidad abstracta
necesidad	eventualidad
plan	acción
lado	lugar
sentido	entidad abstracta
base	parte
institución	institución
idea	proposición
provincia	lugar
director	persona
precio	valor monetario
efecto	evento
comisión	institución
paso	ambiguo
red	artefacto
control	privilegio
escuela	ambiguo
interés	actitud
calle	lugar
libro	objeto fisico
región	lugar
producción	actividad
cargo	peso
padre	persona
posibilidad	eventualidad
municipio	lugar
realidad	eventualidad

palabra	palabra
apoyo	eventualidad
final	eventualidad
comentario	preposición
conocimiento	entidad
comunicación	actividad
trabajador	persona
oportunidad	oportunidad
personal	ambiguo
verdad	ambiguo
función	evento
gestión	actividad
campo	lugar
atención	propiedad
modelo	entidad
población	grupo
unidad	entidad
fuerza	propiedad
estudiante	persona
imagen	imagen
línea	ambiguo
ser	entidad abstracta
administración	institución
respuesta	acto de habla
principio	tiempo
alumno	persona
cuerpo	cuerpo
joven	persona
consejo	eventualidad
capacidad	propiedad
fondo	parte
dirección	propiedad
ciudadano	persona
decisión	eventualidad

acto	acto
usuario	persona
participación	eventualidad
mañana	tiempo
materia	entidad
modo	entidad abstracta
clase	entidad abstracta
noche	tiempo
tierra	lugar
curso	actividad
encuentro	actividad
falta	evento
propuesta	eventualidad

miembro	ambiguo
dios	entidad abstracta
razón	entidad abstracta
tecnología	entidad
entidad	institución
riesgo	entidad abstracta
aplicación	ambiguo
cantidad	entidad abstracta
ministerio	institución
mayoría	parte
cultura	entidad

	abstracta
serie	grupo
camino	lugar
carrera	actividad
elemento	entidad
formación	eventualidad
prueba	eventualidad
profesor	persona
ministro	persona
edad	propiedad
jugador	persona
venta	eventualidad
evento	eventualidad

Figura 10.Clasificación Semántica 200 sustantivos más frecuentes.

Como se puede observar en la figura 10 y 11, los tipos semánticos que poseen mayor repetición son: con repetición de 19 veces son el tipo semántico de *lugar* (por ejemplo: país, lugar, mundo centro, etc.) Y persona (por ejemplo: grupo, presidente, niño), además se encuentra la categoría de ambiguo en esta categoría se encasillan todos aquellos sustantivos que no pueden ser establecidos como un solo tipo semántico o que son de carácter ambiguo valga la redundancia (por ejemplo algunos de los considerados ambiguos son: punto, medio, través). Este tipo semántico refleja que el vocabulario que se comparte entré los países se relaciona directamente a los lugares dentro de los países, ya sea cuando se mencionan lugares específicos o cuando son de carácter universal. Esto último da cuentas de que el vocabulario compartido es el que posee menor variación.

Eventualidades cuenta con 16 sustantivos (más amplio que evento como por ejemplo: educación, medida, necesidad) y evento con 15 sustantivos (ejemplos de esta categoría son: experiencia, historia, hecho, etc.). Continuando, con frecuencia de repetición 13 tiempo (año, día, mes, momento, etc.). La entidad abstracta 12 (tipo, sociedad, autoridad, etc.). Y con frecuencia de repetición 11 están propiedad (nombre, valor, calidad, etc.) y actividad (política, juego, producción). Los tipos semánticos que se poseen menor repetición de 10 son: Institución 8 (universidad, gobierno, empresa, etc.), entidad 7 (unidad, materia, tecnología, etc.). Los otros tipos de categoría semántica inferiores a 5 en repetición se pueden observar en la tabla anterior.

En general, con los tipos semánticos expuestos reflejan que el vocabulario que es compartido se acerca más al uso general de la lengua, independientemente de si es la española, ya que por ejemplo sustantivos como lo son *año y mes,* se pueden observar transversalmente en las lenguas, a diferencia de los grupos semánticos menos ocupados o encontrados en estos 200 sustantivos clasificados que corresponden a instituciones y entidades.

En el grafico expuesto en la figura 11, se puede apreciar los tipos semánticos en un orden decreciente, al igual que han sido organizados en la tabla (figura 12).

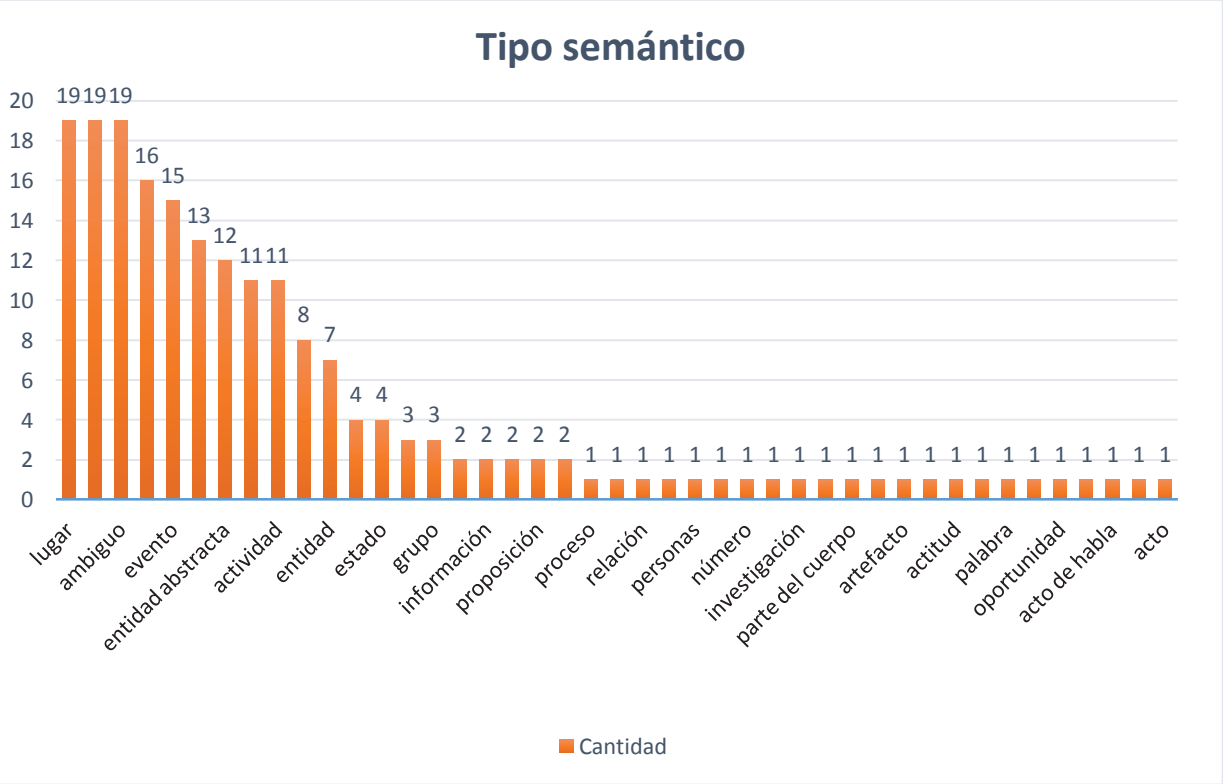


Figura 11. Gráfico clasificación 200 sustantivos tipo semántico.

Grupo Semántico	Cantidad
lugar	19
persona	19
ambiguo	19
eventualidad	16
evento	15

tiempo	13
entidad abstracta	12
propiedad	11
actividad	11
institución	8
entidad	7
parte	4
estado	4
acción	3
grupo	3
concepto	2
información	2
norma	2
proposición	2
peso	2
proceso	1
documento	1
relación	1
objeto físico	1
personas	1
agua	1
número	1
meta	1
investigación	1
uso	1
parte del cuerpo	1
valor monetario	1
artefacto	1
privilegio	1
actitud	1
objeto fisico	1
palabra	1
preposición	1
oportunidad	1
imagen	1
acto de habla	1

cuerpo	1
acto	1

Figura 12. Tabla orden decreciente de tipos semánticos.

4.2 Frecuencia por intervalos y sumas

En la siguiente tabla (figura 13), se presentan los intervalos de 500.000 según el total de datos. En ella quedó como una muestra de los tipos de palabras que se encuentran entre los intervalos, pero no en su totalidad, es decir en la tabla no se muestran la totalidad de palabras sino solo una muestra de las primeras 50 palabras en el caso de aquellos intervalos con bastantes sustantivos y en los que son menor a esta suma el total de la muestra de palabras que se encontraba entre los intervalos, el motivo de esta decisión era la cantidad de palabras que se encontraban por intervalos, puesto que eran cantidades excesivas para realizar una muestra:

Numero	Suma total intervalos	Frecuencia de uso	Sustantivos
1	766.959.107	1.000 a 500.000	empleado, puesta ,emisión ,cocina ,cuento, expectativa, bloque, hoja ,gana, obrero ,aceite ,medicamento, cobertura, argentino, difusión ,documentación, viento, piedra, espera, célula, legislación, enlace ,adolescente, personalidad, entrenamiento, reacción, fiscal, brazo, municipalidad, transformación, local, ingeniero, clasificación, novedad ,clima, agricultura, vuelo, guía, bosque, expediente, reto, idioma, excepción, entrenador, aparato, alegría, proveedor, sorpresa.
2	290.562.351	500.001 a 1.000.000	enseñanza, dificultad, gas, consumidor, integración, aporte, abogado ,estación, investigador, discusión, sujeto, puesto, modificación, regla, piel ,contrario, pieza, miedo, circunstancia, sentencia, pantalla, perspectiva, cancha, coordinador, campeón, campeonato, gusto, lengua, señora ,adulto, ganador, tarjeta, fuego, oposición, nación, toma,

			golpe, fenómeno, temperatura, largo, capítulo, convocatoria, conclusión ,dueño, esposo, estadio, don.
3	193.172.418	1.000.001 a 1.500.000	criterio, martes, agente, plaza, exposición, gol, parque, costa, impacto ,convenio, teatro, público, código, periodista, juicio, titular, disco ,destino, firma, metro, opción, transporte, marca, consumo, creación ,lista, cabeza, herramienta, central, alcalde, comercio, secretario ,mensaje, registro, español, república, declaración, bien, voz, especie ,ojo, voto, versión, líder, jefe, éxito, puerta, reforma, unión, actor.
4	184.536.856	5.500.001 o más	parte, día, vez, país, estado, persona, trabajo, caso, tiempo, forma, servicio, vida, lugar, empresa, proyecto, ciudad, sistema, grupo, presidente, momento, equipo, hora, programa, desarrollo, derecho, cuenta.
5	175.260.518	1.500.001 a 2.000.000	septiembre, madre, esfuerzo, siglo, club, congreso, música, beneficio ,cuestión, consecuencia, ingreso, amor, presentación, diseño, ayuda ,texto, tarea, concepto, compromiso, chico, julio, premio, vivienda ,luz, viernes, entrada, futuro, elemento, formación, prueba, profesor ,ministro, edad, jugador, venta, evento, autor, capital, cliente, orden ,duda, movimiento, construcción, funcionario, diferencia, reunión.
6	130.389.334	2.000.001 a 2.500.000	comentario, conocimiento, comunicación, trabajador, oportunidad ,personal, verdad, función, gestión, campo, atención, modelo, población ,unidad, fuerza, estudiante, imagen, línea, ser, materia, modo, clase ,noche, tierra, curso, encuentro, falta, propuesta, miembro, dios, razón

			,tecnología, entidad, riesgo, aplicación, cantidad, ministerio, mayoría, cultura, serie, camino.
7	83.969.113	2.500.001 a 3.000.000	espacio, experiencia, autoridad, necesidad, plan, lado, sentido, base ,institución, idea, provincia, director, precio, efecto, comisión, paso ,red, control, escuela, interés, calle, libro, región, producción, cargo padre, posibilidad, municipio, realidad, palabra, apoyo, final.
8	83.947.191	3.000.001 a 3.500.000	gracia, agua, historia, sociedad, ejemplo, educación, valor, calidad, embargo, condición, frente, medida, número, objetivo, nombre, investigación, uso, mercado, dato, organización, pueblo, política ,juego, mano, peso, fecha.
9	73.130.229	5.000.001 a 5.500.000	medio, actividad, mundo, punto, problema, centro, manera, tema ,cosa, tipo, gobierno, partido, mes, proceso.
10	52.591.284	3.500.001 a 4.000.000	situación, semana, acción, salud, mujer, resultado, hombre, hecho, seguridad, comunidad, recurso, ley, familia, hijo.
11	51.246.324	4.000.001 a 4.500.000	niño, estudio, casa, través, gente, artículo, sector, zona, relación ,producto, universidad, cambio.
12	24.238.575	4.500.001 a 5.000.000	información, nivel, fin, obra, acuerdo.

La tabla (figura 13) esta ordenada de forma decreciente tomando en consideración la suma total entre los intervalos (columna 2).En esta no se observa ninguna relación directa entre la suma total de las repeticiones por intervalos y su frecuencia de uso. Sin embargo, se puede observar una similitud en la suma total de repeticiones entre algunos intervalos con una marcada diferencia en la cantidad de los sustantivos pertenecientes a cada uno de estos, para ejemplificar lo anteriormente expuesto se debe comparar los intervalos numerados 3, 4 5. Si bien existe una diferencia en la suma entre los intervalos

esta es de carácter despreciable en comparación con las diferencias valóricas presentes en otros intervalos.

Dentro de los intervalos 3 y 5 hay una gran cantidad de sustantivos, por su parte el intervalo 4 solo posee 26, esto sugiere que en este intervalo se encuentran los sustantivos con mayor incidencia dentro del corpus estudiado en esta tesis.

Figura 13. Tabla suma de frecuencia por intervalos.

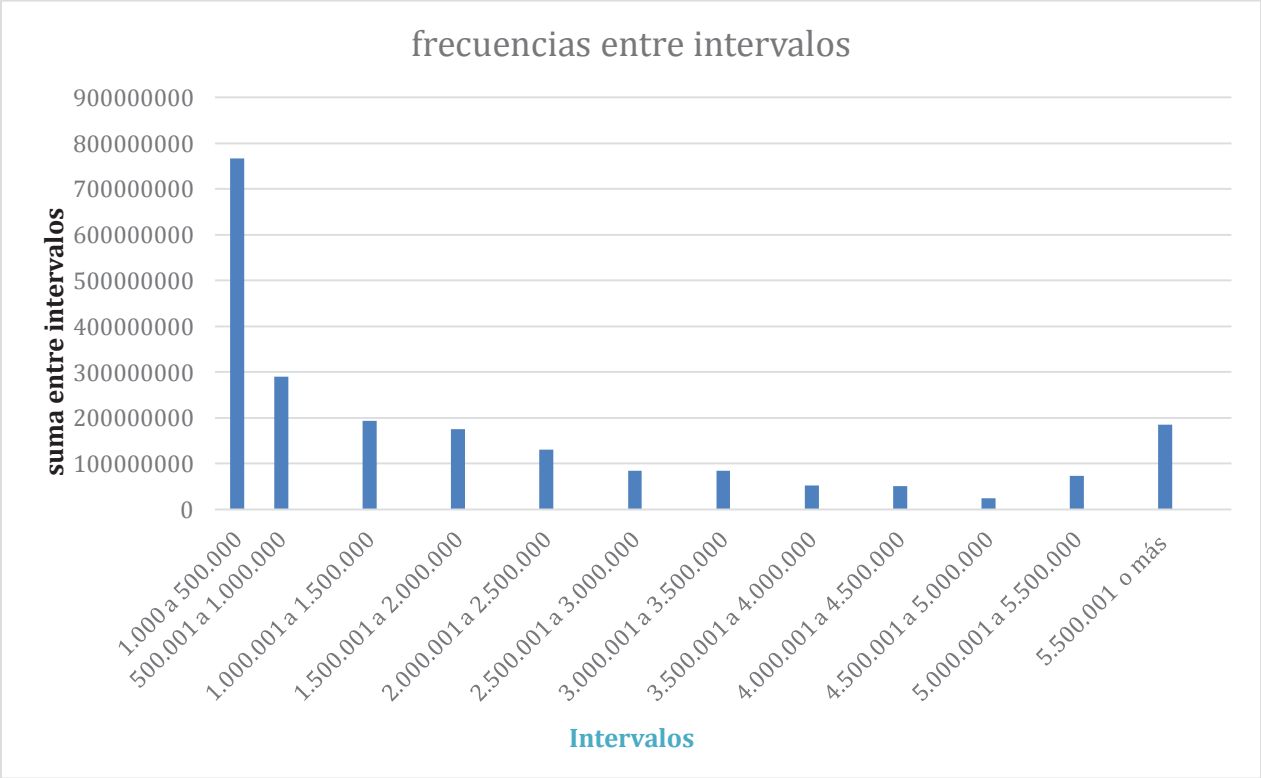


Figura 14. Grafico suma entre intervalos por frecuencia.

4.3 Especificidad de Chile y tipos semánticos

Tomando el coeficiente de especificidad de los sustantivos de Chile, se realizó la búsqueda de identificación de los sustantivos utilizados solo en este país, llevando a cabo además una la tabla por categorización de los primeros 200 sustantivos en cuanto al tipo semántico, y para establecer una visión multimodal se realizó un gráfico representativo de estos datos. Esta clasificación semántica se realizó nuevamente utilizando la CPA Ontology del *Pattern Dictionary of English Verbs* (Hanks, en proceso). De esta manera, los resultados en cuanto a la tabla por clasificación se expuso de la siguiente manera:

Como se puede apreciar en la figura 15 los primeros 200 sustantivos de la hoja Excel correspondiente a la “Espec. Chile”. En negro se encuentran los sustantivos extraídos de los datos y en azul se pueden observar la clasificación de los tipos semánticos encontrados. En la figura 16 basándose en la tabla clasificatoria anterior, en cuanto al tipo semántico se tuvo que marcar como error a aquellas palabras que son nombres propios debido a que no cumplen con el formato establecido de sustantivos, aun asi no se decidió eliminar de los datos pero se expusieron como error y no se clasificaron por este motivo, se puede observar la mayor repetición al tipo semántico *lugar* con 15 sustantivos como por ejemplo *paradero y alameda*. Posteriormente, se encuentra evento con13 sustantivos como por ejemplo *maremoto y clasificatorios*. Estos tipo semánticos dan cuenta de que existe una relación en cuanto a la clasificación semántica de los sustantivos compartidos, pues como en este caso de especificidad de Chile y en el vocabulario compartido el tipo semántico que más se repite es el de *Lugar,* significando que así que el vocabulario de sustantivos específico y compartido se relaciona en la ideología de una semánticidad universal. Esto se interpreta ya que independientemente de la lengua en uso los sustantivos utilizados frecuentemente son de carácter universal, asimismo podría interpretarse la diferencia entre la especificidad de Chile con la otra clasificación, porque en segundo lugar se encuentran los eventos en cambio en el vocabulario compartido corresponde a personas. Destacando de esta manera, que la representatividad de Chile al menos en estos 200 sustantivos se aleja bastante del corpus compartido.

Sustantivo	**Tipo semántico**
talca	Error
temuco	Error
udi	Institución
entretención	Estado
carabinero	Persona
mantención	Actividad
talcahuano	Error
ovalle	Error
serena	Error
atacama	Error
guata	Parte del
	cuerpo
roja	Ambiguo
chucha	Palabra
condes	Error
ppd	Institución
fono	Objeto físico
ubilla	Error
rangers	Error
chileno	Persona
tontera	Palabra
utm	Institución
auspiciadores	Grupo
quintiles	Valor
	monetario
mapuche	Persona
esperados	ambiguo
mapuches	Persona
arriendo	Actividad
cesantes	Estado
malls	Lugar
machas	Entidad
albos	Grupo
imperial	Evento
parque	Lugar
antiterrorista	Estado
polar	Estado

magíster	Estado	remezón	Evento	personeros	Grupo
liceo	Institución	ostiones	Entidad	parlamentario	Persona
concertación	Institución	cocina	Lugar	terremoto	Evento
farandula	Entidad	volcamiento	Actividad	viña	Error
comuna	Lugar	modelamiento	Actividad	fim	Institución
ensilaje	Actividad	cato	Institución	salmón	Entidad
fiscalizador	Persona	licitados	Evento	camila	Error
locomoción	Entidad	unitaria	Eventualidad	habitacionales	Ambiguo
innova	Institución	clasificatorias	Actividad	mba	Estado
quintil	Valor monetario	fiscaliza	Actividad	pedagogia	Actividad
		sotomayor	Entidad	erazo	Error
accidentabilidad	Evento	neruda	Error	magister	Estado
salmones	Entidad	hidroelectricidad	Eventualidad	octanos	Valor monetario
ancha	Estado	dga	Institución		
laurence	Persona	subcontratados	Estado	maricon	Palabra
endesa	Institución	asociatividad	Estado	hurtado	Error
microempresarias	Institución	probidad	Actividad	paraderos	Objeto fisico
fiscalizaciones	Actividad	pucha	Palabra	reajuste	Evento
adenda	Entidad abstracta	postulación	Actividad	cobre	Objeto fisico
		microempresarios	Grupo	nana	Persona
angelica	Persona	voluntario	Persona	competitivo	Estado
suazo	Propiedad	casablanca	Lugar	puntajes	Valor monetario
chile	Error	femicidio	Evento		
amigotes	Grupo	panoramas	Actividad	regiòn	Estado
condorito	Entidad	manipuladoras	Estado	freire	Entidad
termoeléctricas	Institución	bus	Entidad	alameda	Lugar
michelle	Error	cesantía	Estado	peo	Entidad Abstracta
maremoto	Evento	litigación	Actividad		
teletón	Institución	computador	Objeto físico	pisco	Objeto fisico
católica	Institución	dina	Institución	copete	Objeto fisico
colusión	Actividad	educadoras	Grupo	católica	Institución
providencia	Lugar	graneros	Lugar	matrona	Persona
fiscalizadores	Grupo	florida	Lugar	salitre	Objeto fisico

glicemia	Información	vibras	Entidad abstracta	acuicultura	Actividad
timonel	Error			personero	Estado
mejillón	Error	flamengo	Institución	crs	Lugar
cifuentes	Entidad	subgerente	Persona	cotizantes	Persona
parricidio	Evento	turn	Actividad	descontaminación	Evento
aprensión	Estado	srs	Palabra	secundarios	Grupo
sinverguenzas	Grupo	siniestro	Estado	conviviente	Persona
holding	Entidad	paralizaciones	Evento	clasificatorios	Evento
denisse	Error	institucionalidad	Evento	encontre	Actividad
pasquín	Entidad	metodista	Persona	empleabilidad	Estado
papayas	Entidad	bio	Lugar	chiquilla	Persona
uss	Institución	cesante	Estado	prc	Institución
graderías	Objeto fisico	relajo	Estado	clínico	Estado
anglo	Ambiguo	ena	Institución	plebiscito	Entidad
frutícola	Objeto fisico	cofinanciamiento	Actividad	subcontratistas	Grupo
karen	Error	clasificadora	Entidad	connotado	Estado
core	Institución	insulza	Persona	oregón	Lugar
botaderos	Lugar	orellana	Error	priscilla	Error
animadora	Persona	alexis	Error	adiestrador	Persona
pomada	Objeto fisico	chancho	Entidad	lixiviación	Actividad
alamedas	Lugar	maricón	Palabra	allende	Persona
estanque	Objeto fisico	folclorista	Persona	mercurio	Objeto fisico
estanques	Objeto fisico	ascencio	Persona	municipalidades	Institución
asistenciales	Estado	insto	Evento		

Figura 15 . Tabla clasificación de sustantivos por tipo semántico.

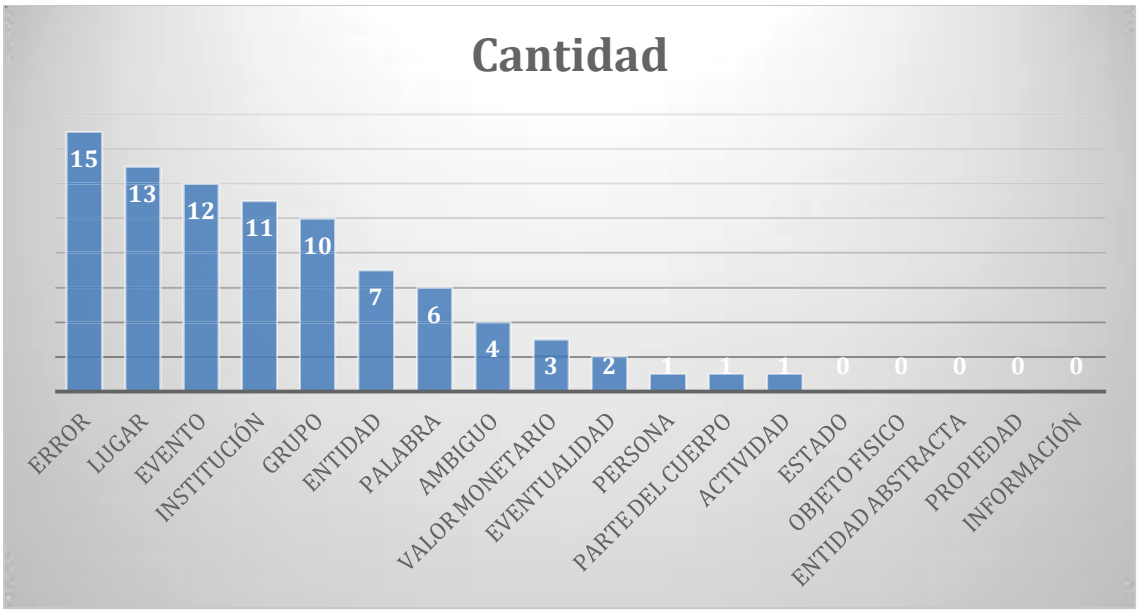


Figura 16. Grafico tipo semántico Chile

5. Conclusiones

5.1 Sustantivos compartidos y la clasificación semántica

En cuanto al total de los sustantivos compartidos por los 19 países del corpus, la cifra oficial es de 148.991, esta cifra, si bien se corresponde solamente al vocabulario del español de Internet, refleja de forma bastante fiable el uso de los sustantivos que son compartidos por los países hispanohablantes, sobre todo en relación con los de frecuencia más alta. Este conjunto de sustantivos puede tenerse en cuenta para la enseñanza del español como lengua materna o extranjera.

En cuanto a la clasificación semántica, es interesante observar que aquellos tipos más frecuentes fueron *persona, lugar y eventualidad*, ya que demuestra en cierto grado la utilización más frecuente de estos tipos de palabras en la web. A modo de especulación se podría comprender que la mayoría de los hablantes de español conversan, buscan o se comunican en torno a lo que serían hechos de la vida en concreto u actividades y sobre lugares específicos de los mismos países o incluso partes de estos mismos espacios geográficos. Por contra, lo que menos es utilizado en la web son *partes del cuerpo, actos de habla y artefactos.*

5.2 Especificidad de Chile y tipo semántico de sustantivos

Se ubicaron los sustantivos más utilizados en este país, y se realizó mediante la aproximación de cuanto más cerca estuviese este cálculo (coeficiente de especificidad) al número 1 significaba que era más específico de este país. En los resultados de este apartado, se pudo observar que al igual que en el vocabulario compartido el tipo semántico más utilizado era el de lugar, reflejando que los sustantivos más frecuente en Chile se apegan a un vocabulario más universal. Es así, como este mismo procedimiento que se ha realizado para detectar los sustantivos específicos de Chile podría ser realizado como un trabajo futuro en el resto de países hispanohablantes, dado que es un trabajo interesante no solo saber cómo se ha demostrado en este estudio la especificidad de Chile y el vocabulario compartido, sino también del resto de países. Claramente este proceso daría cuenta de una parte de la identidad de cada cultura de estos países y sus intereses.

En cuanto a la clasificación semántica realizada, es interesante observar las diferencias hechas con los tipos semánticos más frecuentes del vocabulario compartido, ya que estos varían sutilmente, pues en el compartido era: persona, lugar y eventualidad. En Chile destacan por sobre el resto son: lugar, evento e institución. Por lo tanto, se pudo observar que al igual que en el vocabulario compartido el tipo semántico más utilizado era el de lugar, reflejando que los sustantivos más frecuente en Chile se apegan a un vocabulario más universal, al igual que en los 19 países.

5.4 Trabajos futuros

Como se puede deducir, el principal trabajo futuro sería el caso de la especificidad por país, debido a que en este proyecto solo se pudo obtener la información acerca de Chile. Además, un trabajo o investigación que puede apoyar esta idea sería el caso del vocabulario compartido por zonas dialectales, establecer que sustantivos comparten Zona Andina (Bolivia, Colombia, Ecuador y Perú), México y Centroamérica (México, Guatemala, Honduras, Costa Rica, Panamá, Nicaragua y El

Salvador), Caribe (República Dominicana y Cuba), y Rio de la plata (Argentina, Uruguay y Paraguay).Incluso se puede incluir la zona dialectal a la cual pertenece España. Finalmente, se pueden hacer investigaciones tomando la frecuencia absoluta de esta investigación como base para realizar la frecuencia relativa, indicando el porcentaje de repetición respecto del total de los datos. De esta forma se completaría la idea del vocabulario compartido por los países.

6. Bibliografía

Alba, O. (1992). Zonificación dialectal del español en América. República Dominicana, Santiago. Brigham Young University: BYU ScholarsArchive.

Bernárdez, E. (1999). *¿Qué son las lenguas?* Madrid: Alianza Editorial.

Davies, M. (2006). A frequency dictionary of Spanish: core vocabulary for learners.United Estates, New York: Routledge.

Demonte, V. (2003). Lengua estándar, norma y normas en la difusión actual de la lengua española. España, Madrid: Fundación José Ortega y Gasset. Disponible en: https://digital.csic.es/handle/10261/13074

Díaz-Campos, M. (2014). Introducción a la sociolingüística hispánica. Washington, D. C: Wiley Blackwell.

Escandell, M.A. (2007).Apuntes de la Semántica Léxica. Madrid, España: UNED.

Gómez, L y Peronard, M. (2005). *El lenguaje humano: Léxico fundamental para la iniciación Lingüística*. Valparaíso, Chile: Ediciones universitarias de Valparaíso.

Hernández, R. (2014). Metodología de la Investigación. México, Ciudad de México: McGraw-Hill / INTERAMERICANA EDITORES, S.A. DE C.V.

Instituto de Cervantes. (2018). El español: una lengua viva. España, Madrid: Departamento de Comunicación Digital del Instituto Cervantes. Disponible en: https://www.cervantes.es/imagenes/File/espanol_lengua_viva_2018.pdf.

Kilgarriff, A., & Renau, I. (2013). EsTenTen, a vast web corpus of Peninsular and American Spanish. Procedia-Social and Behavioral Sciences, 95, 11-14.

Moreno Fernández, F. (2007). Atlas de la lengua Española en el Mundo. España, Barcelona: Ariel.

Moreno Fernández, F. (1998). *Principios de sociolingüística y sociología del lenguaje*. Barcelona: Ariel.

Nazar, R.; Vivaldi, J.; Cabré, MT. (2008). A Suite to Compile and Analyze an LSP Corpus. Proceedings of LREC 2008 (The 6th edition of the Language Resources and Evaluation Conference) Marrakech (Morocco), May 28-30, 2008.

Nazar, R.; Vidal, V. (2008). Aproximación cuantitativa a la neología. Proceedings of CINEO 2008, (I Congreso Internacional de Neología en las lenguas románicas), Barcelona, May 07-10, 2008.Piera, C. (2009). “Una idea de palabra”. En De Miguel (ed.), Panorama de lexicología. Barcelona: Ariel.

Renau, I. (2012). *Las construcciones con se en las entradas verbales del diccionario de español como lengua extranjera* (Tesis doctoral). Universidad Pompeo Fabra, s/d.

Renau, I., Nazar, R., Castro, A., López, B. y Obreque, J. (2019). Verbo y contexto de uso: Un análisis basado en corpus con métodos cualitativos y cuantitativos. *Revista Signos. Estudios de Lingüística,* 52(101), 879-882.

Real Academia Española (2014). Diccionario de la lengua española. (23. a ed.). Disponible en http://www.dle.rae.es. [Última consulta: 16 de junio del 2019].

Real Academia Española. (2010). Nueva gramática de la lengua española. Manual. Buenos Aires, Argentina: Espasa Libros, S.L.

Saussure, F. (1916/2007). *Curso de lingüística general.* Buenos Aires: Losada.

Silva-Corvalán, C. (2001). *Sociolingüística y pragmática del español*. Washington, D. C.: Georgetown University.

Torner, S. (2008). Gramática: lengua española. Barcelona: Larousse