



UNIVERSIDAD
AUTÓNOMA
METROPOLITANA
Unidad Cuajimalpa

UNIVERSIDAD AUTÓNOMA METROPOLITANA

DIVISIÓN DE CIENCIAS DE LA COMUNICACIÓN Y DISEÑO
LICENCIATURA EN TECNOLOGÍAS Y SISTEMAS DE INFORMACIÓN

Herramienta para la visualización y contraste de vocabulario relevante para diferentes poblaciones

Por:
Gerardo Uriel Ruiz Santiago

Asesores:
Mtra. Adriana Gabriela Ramírez de la Rosa
Dr. Esaú Villatoro Tello

Octubre 2021

Índice

1. Introducción	2
1.1. Planteamiento del problema	3
1.2. Solución propuesta	3
1.3. Objetivos	3
1.3.1. Objetivo general	3
1.3.2. Objetivos Específicos	4
1.4. Organización del documento	4
2. Marco Teórico	5
2.1. Información general sobre el lenguaje	5
2.1.1. ¿Qué es el lenguaje?	5
2.1.2. Funciones del lenguaje	5
2.2. Disponibilidad léxica	6
2.2.1. ¿Qué es el léxico?	6
2.2.2. ¿Qué es léxico mental?	6
2.2.3. ¿Qué es la disponibilidad léxica?	6
2.2.4. Prueba de la disponibilidad léxica	8
2.2.5. Índice de la Disponibilidad Léxica (IDL)	9
3. Trabajo relacionado	12
3.1. Herramientas pioneras en disponibilidad léxica	12
3.1.1. Dispogen y Dispografo	12
3.1.2. LexMath	13
3.1.3. Tabla comparativa entre aplicaciones	14
4. Arquitectura del sistema	15
4.1. Esquema general de la aplicación web	15
4.1.1. Módulo de gestión de archivos	16
4.1.2. Módulo de procesamiento	16
4.1.3. Módulo de visualización	19
4.2. Implementación del sistema web	20
4.3. Visualización de la aplicación web	21
4.3.1. Gráfica de Pastel	21
4.3.2. Gráfica de barras	22
4.3.3. Gráfica de radar	23
4.3.4. Diagrama de Sankey	24
5. Ejemplo de uso de la aplicación web	25
6. Conclusiones y trabajo a futuro	34
7. Referencias	35
7. Anexos	37

1. Introducción

El lenguaje es uno de los elementos que distinguen a los seres humanos del resto de los seres vivos, vista como una función lingüística, podemos definir el lenguaje como una función que permite expresar ideas, pensamientos y la comunicación entre dos o más personas. Esto se puede realizar mediante signos (escritura) o mediante sonidos (voz)[1].

Asimismo podemos distinguir el lenguaje en dos clases: el lenguaje natural (español, inglés, alemán, etc.) y el lenguaje formal (matemáticas, lógica, programación, etc.). El lenguaje natural es el medio que utilizamos para comunicarnos de manera cotidiana con las demás personas, por esta razón es que el análisis del lenguaje natural ha tenido una gran relevancia en la historia del hombre, ya que éste puede revelar aspectos y/o características importantes entre personas, desde cómo lo usan, qué es lo que dicen y cuándo lo dicen. [2]. Esto también se puede ver en las Ciencias Sociales y Humanidades que se encargan del estudio del ser humano, dando gran énfasis en cómo utilizan el lenguaje natural en aspectos como el comportamiento, la interacción humana y la cultura [3].

El lenguaje es un aspecto importante para varias disciplinas, por ejemplo, la Antropología lingüística es una disciplina de la Antropología que estudia el lenguaje de las personas en su contexto social y cultural. Esto ayuda a entender su comportamiento y mostrar cómo el habla refleja diferencias sociales [4]. La psicología es otra ciencia que al igual que la antropología, estudia el comportamiento humano a través de sus pensamientos, las emociones, los instintos, los sueños, el lenguaje, la inteligencia, la percepción, tanto en su desarrollo normal como en los diferentes trastornos o problemas del pensamiento, de las emociones y del comportamiento [5]. Otro ejemplo donde encontramos presente el lenguaje es en la sociología, ciencia que estudia el comportamiento de las personas en relación con los demás y las situaciones en las que viven. En especial, presta atención al impacto que el comportamiento de la gente tiene en la sociedad a través de sus acciones y su participación en los procesos de cambio [6].

En la computación y más específicamente en la inteligencia artificial, una meta fundamental es el procesamiento de lenguajes naturales usando herramientas de computo. Podemos decir que el procesamiento de lenguaje natural (PLN) es la función de componentes en un sistema de cómputo que busca comunicar a las personas con las máquinas mediante el uso de lenguajes naturales. [7]

La Disponibilidad Léxica (DL) es el campo de investigación que dentro de la lingüística tiene como objetivo la extracción del vocabulario espontáneo que viene a la mente de un determinado grupo de personas en respuesta a un tema en concreto, los temas pueden estar relacionados con la vida diaria, tales como comida, salud, el campo, animales, partes del cuerpo, música, y muchas cosas más[8].

En lo anterior podemos observar que muchas disciplinas utilizan el lenguaje para el estudio de las personas, pero a la hora de hacer el estudio a nivel de poblaciones existen limitaciones por la escala tan grande que esto significa. Es aquí donde se propone colaborar con ayuda del PLN y la DL para hacer posible el análisis y generar un contraste de vocabulario a nivel de poblaciones.

1.1. Planteamiento del problema

Como podemos observar, el estudio del comportamiento humano a través del lenguaje es algo que tienen en común las ciencias sociales y diferentes disciplinas. Aunque siempre se han hecho análisis de comunidades, existe una limitación para los especialistas de estas disciplinas al momento de analizar comunidades y su comportamiento utilizando el lenguaje natural como base. Esto se debe a que haciendo el análisis de forma tradicional, tendrían que realizar un análisis por cada individuo. El tiempo que tomaría hacer el análisis de toda una comunidad dependería del número de personas. Quizá si el grupo de personas es pequeño no le tome mucho tiempo hacer el análisis a un especialista, en cambio, si suponemos que se tratara de un grupo donde hay miles de personas, es muy probable que realizar el análisis de forma tradicional sea casi imposible para el especialista.

1.2. Solución propuesta

La propuesta de este proyecto es diseñar y desarrollar una herramienta web que ayude al análisis del vocabulario de dos o más poblaciones. Esta herramienta estará diseñada para que a través del procesamiento de lenguaje natural y la disponibilidad léxica, muestre de forma visual contrastes o similitudes entre el vocabulario de distintas poblaciones. Con esto también se busca que el especialista pueda realizar el análisis de conjuntos de textos grandes en poco tiempo y que los datos obtenidos por la herramienta le sirvan para realizar más estudios.

Se piensa en utilizar la DL en la herramienta, debido a que ésta nos ayuda a obtener el lenguaje más significativo y no necesariamente el lenguaje con mayor frecuencia del léxico de una persona o en general de una población.

Un ejemplo del uso de la DL en la educación es un estudio que se realizó en estudiantes de bachillerato de la ciudad de Matanzas Cuba y Morelia México para comparar y enfocarse en la importancia del vocabulario para la didáctica de la lengua y la literatura a nivel medio superior [9]. Con este estudio se logró observar la similitud de vocabulario entre los dos grupos de personas aunque eran de dos países diferentes y también se pudo ver las palabras que no tienen ninguna similitud entre estos grupos.

Con el ejemplo anterior nos damos cuenta que utilizando la DL podemos obtener el léxico más relevante de una comunidad y con esto, realizar distintos análisis del vocabulario como hacer contrastes y encontrar similitudes en el vocabulario de diferentes poblaciones.

1.3. Objetivos

1.3.1. Objetivo general

- Desarrollar una herramienta web para visualizar y contrastar el vocabulario extraído mediante la disponibilidad léxica de dos o más poblaciones a través del análisis de sus textos.

1.3.2. Objetivos Específicos

- Implementar el cálculo de la disponibilidad léxica para describir el léxico mental de una población de interés a través del análisis de los textos que ésta población produce.
- Incorporar herramientas de visualización de información que permitan describir y contrastar el léxico mental de las poblaciones de interés.
- Validar la utilidad de la herramienta desarrollada en un caso de estudio específico.

1.4. Organización del documento

Este documento está organizado por los siguientes capítulos:

- Marco Teórico: En este capítulo, se explican los conceptos y elementos teóricos más importantes para entender mejor el contexto del proyecto.
- Trabajo Relacionado: En este capítulo, se muestran los trabajos previos ya realizados sobre la disponibilidad léxica y se hace una pequeña comparación entre las diferentes herramientas que ya existen y la propuesta de este proyecto.
- Arquitectura del sistema: En este apartado se describe a grandes rasgos y de manera detallada las partes (módulos) que conforman la aplicación para el contraste de vocabulario de distintas poblaciones. Esto ayuda a comprender como funciona la aplicación y las distintas tareas que se realizan para llegar al resultado final.
- Ejemplo del uso de la aplicación: En este capítulo se muestra un ejemplo usando la aplicación web con un conjunto de poblaciones y sus textos. La finalidad es ver de principio a fin cómo se utiliza la aplicación y las gráficas generadas para el contraste del vocabulario.
- Conclusiones y trabajo a futuro: En esta sección, se habla de los resultados obtenidos en el proyecto, los objetivos logrados y además se menciona que trabajo puede haber a futuro para este proyecto.
- Anexos: Por último, en esta parte se muestran los enlaces utilizados para la recolección de los textos en el ejemplo del uso de la aplicación web.

2. Marco Teórico

El siguiente apartado tiene como objetivo dar a conocer conceptos necesarios e información que ayude al lector entender de mejor manera el proyecto. Como se mencionó anteriormente, el proyecto consiste en diseñar e implementar una herramienta web que ayude a especialistas a contrastar el vocabulario de poblaciones a través de su léxico y los textos que éstas producen.

2.1. Información general sobre el lenguaje

2.1.1. ¿Qué es el lenguaje?

El lenguaje es un proceso de comunicación que permite la emisión y recepción de información al ser humano, asimismo, es un proceso fisiológico y psíquico, que pertenece tanto al dominio individual como al social [10]. Entendido así, el lenguaje es la capacidad propia de la especie humana y una de las mayores riquezas para la interacción de las personas con su entorno, dado que es el principal medio de comunicación entre las mismas [11].

Otros autores como Kristeva [10] nos dice que el lenguaje es un proceso de comunicación basado en la emisión de un mensaje entre dos sujetos hablantes (al menos), siendo uno el emisor, y otro el receptor.



Figura 1: Proceso de intercambio de mensajes

Ahora bien, cuando se lleva a cabo este proceso de comunicación, cada sujeto hablante puede tener tanto el rol de emisor como receptor, puesto que es capaz de emitir y recibir mensajes como se observa en la Figura 1.

2.1.2. Funciones del lenguaje

Es importante tener en cuenta las funciones que tiene el lenguaje en nuestro entorno, ya que cuando lo utilizamos, lo hacemos con una intención o función determinada, puede ser para comunicar ideas, pensamientos, sentimientos o simplemente para persuadir, ordenar, conmover o interactuar. Esto depende de cuál sea nuestra intención en el momento [12].

Dicho lo anterior, encontramos que las principales funciones que nos otorga el lenguaje, son las siguientes.

- **Representativa:** Transmite contenidos de forma objetiva.
- **Expresiva:** Expresa sentimientos, opiniones, deseos, es decir, la subjetividad del hablante.
- **Conativa:** Influye en el comportamiento del receptor y provoca en él una reacción.
- **Fática:** Prolonga, establece o interrumpe la comunicación.
- **Poética:** Atrae la atención sobre la forma de la expresión lingüística.
- **Metalíngüística:** Utiliza la lengua para hablar de la propia lengua.

El lenguaje es una función psicológica que se manifiesta desde que somos pequeños y luego aumenta la organización gramatical, mediante reacciones sociales. Comunica significados por medio de sonidos articulados y es el principal instrumento que establece una cultura. Además el lenguaje es un sistema compuesto por signos lingüísticos, cuya adquisición y uso posibilita formas peculiares y específicas de relación y acción sobre el medio, especialmente el medio social [13].

2.2. Disponibilidad léxica

2.2.1. ¿Qué es el léxico?

La palabra, como tal, proviene del griego *λεξικός* (lexicós) y podemos denominar como léxico al conjunto de palabras que constituyen una lengua. En este sentido, también se conoce como léxico al conjunto de palabras que son de uso particular en una región, en una actividad o pueden ser modismos que caracterizan el lenguaje que usa una persona para expresarse. Es por esto que podemos decir que léxico y vocabulario son palabras sinónimas ya que ambas se refieren al conjunto de palabras o modismos que utiliza una comunidad para comunicarse[14].

2.2.2. ¿Qué es léxico mental?

El léxico mental o lexicón mental podemos definir que es el conocimiento que una persona tiene interiorizado sobre el vocabulario de un lenguaje. Con esto se nos referimos a la capacidad de entender y utilizar las unidades léxicas del vocabulario[15].

El léxico mental también lo podemos ver como un diccionario mental donde se encuentran todas las palabras que una persona tiene sobre un lenguaje. La diferencia entre el léxico mental y un diccionario habitual es la forma en que están organizados, mientras que en cualquier diccionario habitual, las palabras se encuentran organizadas de manera alfabética, en el léxico mental se encuentran ordenadas de tal manera que facilita el acceso y la recuperación de palabras a la persona[16].

2.2.3. ¿Qué es la disponibilidad léxica?

Según un artículo de Chris Payne [8] nos dice que la DL se puede definir como el conjunto de palabras que vienen a la mente de una persona en respuesta a un tema en

concreto (estímulo), los temas pueden estar relacionados con la vida diaria, tales como comida y bebida, medicina y salud, el campo, la ciudad, animales, partes del cuerpo, jardín, ropa, y muchas cosas más. La disponibilidad léxica se mide a través de una prueba que refleja la producción espontánea del vocabulario.

Al principio los investigadores recurrieron al criterio de la frecuencia porque pensaban que las palabras más frecuentes en el vocabulario eran las más útiles y basándose en dicho criterio se formula el término de léxico básico, que incluye las palabras más usuales o frecuentes en el lenguaje de una comunidad [17].

Bartol Hernández [17] afirmó que la frecuencia de las palabras no puede ser el único criterio en el cual se debe basar para la selección del léxico ya que palabras muy normales y conocidas pueden presentar una frecuencia baja debido a que se producen sólo cuando se da un estímulo concreto al hablante. Por este motivo, surge el concepto de la disponibilidad léxica, como un criterio en el cual se basa la formulación del léxico disponible de una lengua. El léxico disponible se compone por las palabras más comunes de una lengua que no son necesariamente las más frecuentes ya que aparecen en casos determinados cuando el tema específico requiere su uso. En concreto, en el léxico disponible se engloban principalmente sustantivos, algunos verbos y adjetivos [18].

En la Figura 2 se observa un diagrama sobre el léxico disponible en una persona. Podemos ver que dado un estímulo (tema en concreto) a la persona, se obtienen las palabras más relevantes de dicho tema para la persona.

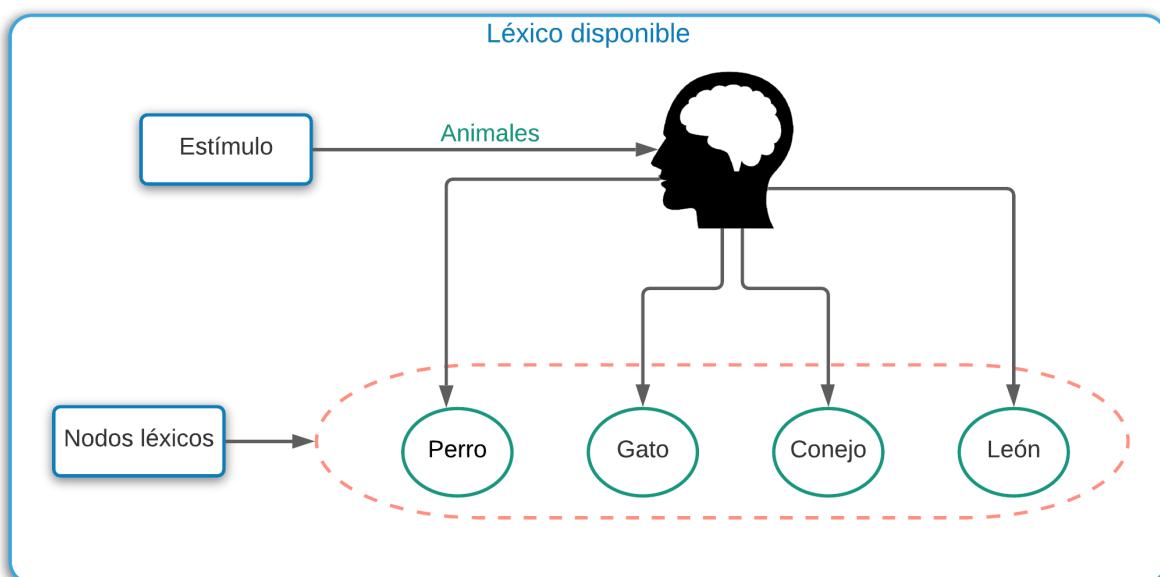


Figura 2: Ejemplo del léxico disponible. El estímulo es el tema en concreto en el que la persona va a pensar, mientras que los nodos léxicos son las palabras obtenidas de la persona después de pensar en el tema dado.

2.2.4. Prueba de la disponibilidad léxica

La prueba de disponibilidad léxica se le aplica a un grupo de participantes hablantes. Se utilizan temas específicos sin ningún orden en particular (por ejemplo, salud, medicina, jardín, etcétera). Cada tema se muestra en la parte superior de una página diferente, después se pide a los participantes escribir tantas palabras como sea posible en un tiempo determinado. Se proporciona espacio para las palabras o elementos léxicos que pueden constar de más de una palabra, por ejemplo, cortar el césped. A los participantes no se les permite pasar al siguiente tema hasta que el límite de tiempo termine.

En la prueba de disponibilidad léxica no existe un modelo estándar único, por lo que hay una gran libertad en lo que respecta al diseño de la prueba. Por ejemplo, los temas utilizados para provocar respuestas pueden ser diferentes, el número de temas se puede cambiar, esto puede alargar o acortar la prueba. Los participantes tienen un límite de tiempo, pero no un límite en el número de palabras que producen [8].

Es importante tener en cuenta que la disponibilidad léxica no es una propiedad fija de las palabras, ya que la disponibilidad de una palabra puede variar de acuerdo con el tema. Por lo tanto, la palabra coche tiene un alto valor de DL para el tema de transporte, un valor medio de DL para el tema de la ciudad y una baja para el tema del campo.

En la Tabla 1 se muestra un pequeño ejemplo del test de disponibilidad léxica donde el tema específico es 'Animales' y se observan las palabras producidas por cuatro personas.

Posición	Persona 1	Persona 2	Persona 3	Persona 4
01	Perro	Gato	Perro	Vaca
02	Vaca	Perro	Pez	Gato
03	Gato	Ratón	Ave	Perro
04	Pez	Pez	Ratón	Pez
05	Ave	Ave	Gato	Ave
06	Perro			Ratón
07	Gato			

Tabla 1: Ejemplo del test de Disponibilidad Léxica a cuatro personas donde el centro de interés es Animales

2.2.5. Índice de la Disponibilidad Léxica (IDL)

Las primeras investigaciones sobre disponibilidad léxica, fueron desarrolladas en Francia y consideraban la “frecuencia” de las palabras dentro de los listados como el único factor ponderador de la disponibilidad léxica. Desde entonces, muchos investigadores se interesaron por encontrar una fórmula matemática que permitiera valorar no sólo la frecuencia, sino también la posición que ocupaban las palabras en las listas a la hora de calcular tal índice [19].

Con el paso del tiempo se propusieron varias formular para realizar el cálculo de la disponibilidad léxica pero fue hasta la presentación de la fórmula de López Chávez y Strassburger Frías que se superaron muchas dificultades de las fórmulas anteriores, esto permitió medir más adecuadamente el índice de disponibilidad de una unidad léxica desde el punto de vista lingüístico [19]. La formulación de este nuevo Índice de Disponibilidad Léxica (IDL) es la siguiente:

$$D_{w,k,m}^{st}(V_j) = \sum_{i=1}^n w^{(\frac{i-1}{k-i})^m} \left(\frac{f_{ji}}{I} \right)$$

Figura 3: Ecuación utilizada para calcular el valor estandarizo de IDLV

V_j : Es la palabra cuya disponibilidad léxica se va a medir, siendo j el índice que identifica a V_j en la lista de palabras del centro de interés.

i = Posición de las posiciones en que puede mencionarse la palabra V_j

f_{ji} = Número de menciones de la palabra j en la posición i para una categoría determinada.

n = Máxima posición alcanzada por las palabras en el centro de interés.

I = Número total de participantes en la prueba.

k = Indicador de la posición en la que se desea que el peso valga " w ".

w = Nivel del peso, comprendido entre 0 y 1, deseado en la posición " k ".

m = Constante que modula la intensidad a la que se aplican los pesos. ($m \leq 1$ decaimiento de la primera posición rápido; $m > 1$ decaimiento de la primera posición muy lento).

Para estas implementaciones de IDL normalizamos (factor $\frac{f_{ji}}{I}$) por la proporción de veces que la palabra j se escribe en la posición i de todas las veces que una palabra se puede escribir en todos los textos.

Esta fórmula proporciona una medición más fiable de la disponibilidad de una palabra, a partir de la suma de diferentes términos como:

- a) La frecuencia absoluta con que fue dicha la palabra en cada posición.
- b) El número de informantes que participan en la prueba.
- c) El número n , o máxima posición alcanzada por las palabras en la prueba en el centro de interés analizado.

- d) Las posiciones en que fue dicha la palabra por los participantes que la mencionaron.

En la Figura 3 podemos observar la ecuación para obtener el valor de el índice de disponibilidad léxica por palabra.

Como vemos, m nos ayuda a controlar el valor de los pesos de las palabras. Para obtener el valor de m que mejor nos ayude a controlar estos pesos, existe la fórmula que vemos en la Figura 4, ésta nos ayuda a controlar el peso de las palabras dependiendo su posición.

$$m = \frac{\ln\left(\frac{z}{w}\right)}{\ln\left(\frac{j-1}{k-1}\right)}$$

Figura 4: Ecuación utilizada para calcular el valor de m

m = Constante que modula la intensidad a la que se aplican los pesos.

z = Peso que se desea que no decaiga antes de la posición j .

w = Peso que se desea que decaiga hasta la posición k .

j = Posición a la que se desea que no decaiga el el valor de z .

k = Posición a la que se desea que decaiga hasta el valor de w .

Como ejemplo, si queremos que el valor tome 0.1 en la posición 35 ($w = 0.1$ y $k = 35$) y que decaiga de tal forma que en la quinta posición el peso no haya bajado de 0.99 ($z = 0.99$ y $j = 5$, para que las cinco posiciones iniciales tengan pesos parecidos), el parámetro de modulación ' m ' debería ser como se muestra en la Figura 5:

$$m = \frac{\ln\left(\frac{0.99}{0.1}\right)}{\ln\left(\frac{5-1}{35-1}\right)} = 2,5392587 = 2,54$$

Figura 5: Ecuación utilizada para calcular el valor de m donde se busca que hasta la quinta posición no decaiga de 0.99 y hasta la posición 35 decaiga a 0.1

En la Tabla 2 tenemos el cálculo de IDL para la palabra 'perro' del test de disponibilidad léxica hecha a cuatro personas que vimos en el ejemplo de la Tabla 1. En la Tabla 2 podemos observar el valor que toma la palabra en cada posición de la lista de palabras de las personas y al final se suman estos valores para obtener el valor final de IDL de la palabra. Este procedimiento se hace con cada palabra escrita por los participantes en el test de DL.

Valor de i	Fórmula	Resultado
1	$\sum_{i=1}^7 0,50^{(\frac{1-i}{3-i})^2} (\frac{2}{3})$	0.5
2	$\sum_{i=2}^7 0,50^{(\frac{2-i}{3-i})^2} (\frac{1}{3})$	0.210224104
3	$\sum_{i=3}^7 0,50^{(\frac{3-i}{3-i})^2} (\frac{1}{3})$	0.125
4	$\sum_{i=4}^7 0,50^{(\frac{4-i}{3-i})^2} (\frac{0}{3})$	0
5	$\sum_{i=5}^7 0,50^{(\frac{5-i}{3-i})^2} (\frac{0}{3})$	0
6	$\sum_{i=6}^7 0,50^{(\frac{6-i}{3-i})^2} (\frac{1}{3})$	0.003284752
7	$\sum_{i=7}^7 0,50^{(\frac{7-i}{3-i})^2} (\frac{0}{3})$	0

Tabla 2: Cálculo de IDL para la palabra perro en cada posición de la lista de palabras de la Tabla 1

Después de hacer el cálculo de la palabra 'perro' de cada posición de la lista de palabras, se suman esos valores que vemos en la Tabla 2 y obtenemos como resultado **0.838508855**, esto nos indica el valor de IDL de esa palabra. Como ya se mencionaba anteriormente, lo mismo se tiene que hacer para cada palabra y al final nos quedaría una lista de las palabras con su valor de IDL como se muestra en la Tabla 3.

Palabra	IDL
Perro	0.838508855
Gato	0.7017447599531434
Pez	0.5977973202207825
Ave	0.4681792830507429
Vaca	0.39152065613971476
Ratón	0.3713176150912768

Tabla 3: Palabras de la Tabla 1 con su valor de índice de disponibilidad léxica

En el ejemplo anterior podemos observar que con la disponibilidad léxica obtenemos las palabras más significativas de una población, vemos que la posición de las palabras puede cambiar en la lista final dependiendo su valor de disponibilidad léxica, entre mayor sea el índice de disponibilidad léxica de la palabra, se coloca en una posición más alta y esto quiere decir que es una palabra más significativa para la población que otras que tengan un índice de disponibilidad léxica más baja.

Con el cálculo de disponibilidad léxica, los especialistas pueden obtener el vocabulario más relevante de una persona y si se lleva a grandes escalas, se puede obtener el vocabulario más relevante de toda una comunidad.

3. Trabajo relacionado

El desarrollo de la herramienta propuesta mediante una interfaz web, permite que el usuario interactúe de forma dinámica. En el siguiente apartado se muestran herramientas existentes que incorporan el cálculo de la disponibilidad léxica.

3.1. Herramientas pioneras en disponibilidad léxica

3.1.1. Dispogen y Dispografo

Dispogen y Dispografo¹ son programas computacionales que ayudan al análisis de la DL, estos programas fueron creados en la Universidad de Concepción de Chile por el Dr. Máx Echeverría. Con estas herramientas es posible obtener el IDL de cada palabra y las redes semánticas que se forman a partir de estas utilizando grafos. Además se puede obtener el IDL personal de cada sujeto y las listas que se forman con las palabras [20]. En la Figura 6 se muestra un ejemplo de una red semántica hecha con la aplicación de Dispogen y Dispografo utilizando grafos.

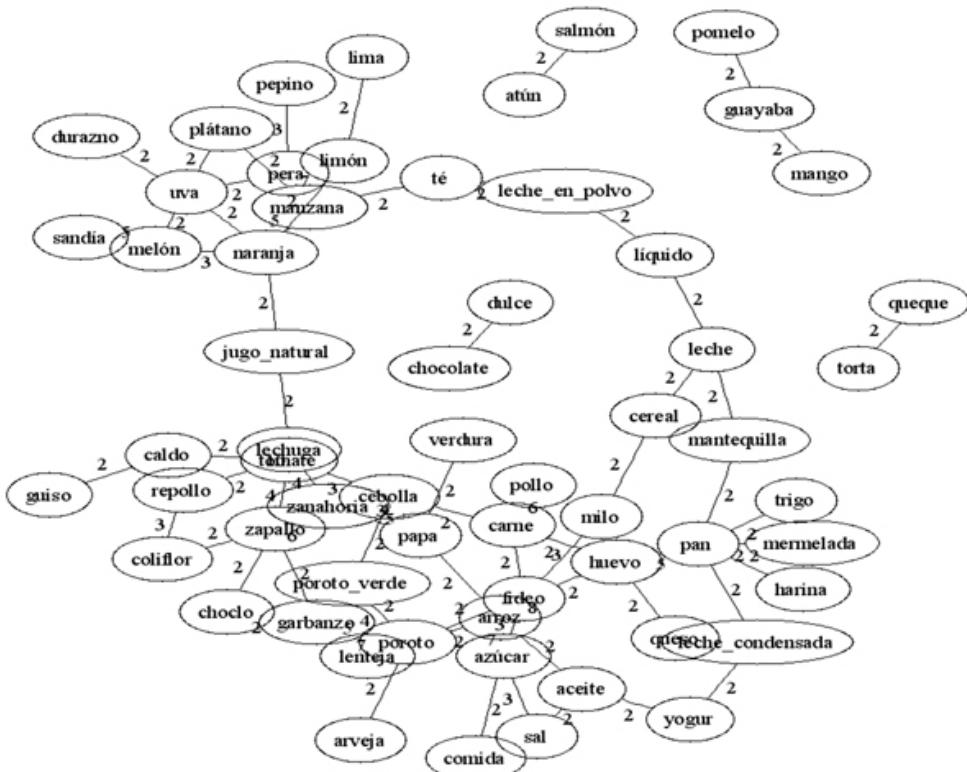


Figura 6: Red semántica creada con Dispogen y Dispografo

¹https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-48832008000100005

3.1.2. LexMath

La plataforma de LexMath² ha sido el producto de dos proyectos de investigación de la comisión nacional de investigación científica y tecnología de Chile (CONICYT). Se tomó en cuenta que los estudiantes manejan cierto léxico que les permite comunicarse, pero no es suficiente para comprender de forma adecuada textos con temáticas específicas como los del ámbito escolar (por ejemplo un libro de matemáticas). Esto se debe al deficiente vocabulario de este tipo de textos y a su incidencia en los procesos de lectura y escritura [21].

LexMath, es una plataforma web que entrega el lexicón mental con el IDL de cada palabra de forma ordenada de mayor a menor, permite además descargar el lexicón a EXCEL o a formato GEPHI, lo que permite estudiar los grafos con otras herramientas especializadas en grafos. Además de eso, tiene la posibilidad de generar los Grafos online, con opción de presentar las estructuras más fuertes y las métricas del grafo. Por último, se tiene la posibilidad de que cualquier usuario pueda procesar un archivo en formato .CSV que contenga un test de disponibilidad léxica y así obtener el lexicón mental y los grafos correspondientes [21].

En diciembre de 2020, la plataforma web estaba en funcionamiento y cualquier persona podía acceder a ella. Actualmente la plataforma se encuentra fuera de servicio, no se sabe si es debido a mantenimiento o si se debe a otras causas.

²<http://www.lexmath.com>

3.1.3. Tabla comparativa entre aplicaciones

En la Tabla 4 se muestra una comparación entre las aplicaciones que hay en cuanto a la disponibilidad léxica y la propuesta hecha en este proyecto. La comparación refleja aspectos que cubre la aplicación web y aspectos que no cubre pero que se pueden trabajar a futuro. Es importante mencionar que aunque hay un gran aporte en trabajos de investigación sobre la disponibilidad léxica, lo que vemos en la tabla son las únicas aplicaciones que están disponibles para el uso de las personas.

Características	Dispogen Dispografo	LexMath	Herramienta propuesta
1. Cuenta con plataforma web		X	X
2. Se puede usar por cualquier persona	X		X
3. Se puede cargar el test desde archivos	X	X	X
4. Se puede hacer el test desde la plataforma		X	
5. Se puede analizar una cantidad grande de textos			X
6. Utiliza grafos para visualizar los datos	X	X	
7. Se puede extraer los resultados obtenidos	X		X
8. Se puede analizar y contrastar la disponibilidad léxica de dos grupos	X		X
9. Se ve de manera gráfica el contraste de vocabulario			X

Tabla 4: Tabla comparativa entre herramientas existentes y propuesta

Como vemos en la tabla anterior, con la aplicación web propuesta en este proyecto se busca cubrir la mayoría de aspectos que tienen otras aplicaciones, aunque también se busca brindar nuevos aspectos que proporcionen algo diferente a las demás aplicaciones.

Las funciones más relevantes de la aplicación propuesta ante las demás aplicaciones que ya existen, es poder analizar un gran número de textos planos y poder realizar y visualizar un contraste de vocabulario entre dos o más poblaciones distintas mediante la disponibilidad léxica de los textos que estas producen. También la aplicación permite que a futuro se pueda incorporar mejoras y que cada vez sea más útil para las personas que la utilicen.

4. Arquitectura del sistema

En esta sección se describe el funcionamiento de la aplicación web desarrollada para el contraste de vocabulario entre distintas poblaciones, así como también, se describirá con más detalle la estructura de los módulos que la componen. La aplicación para el contraste de vocabulario se compone por tres módulos principales, los cuales son: Módulo de gestión de datos, Módulo de procesamiento de datos y Módulo de visualización de datos. A continuación, se muestra el esquema general de la aplicación, asimismo se describen cada uno de los módulos que lo integran.

4.1. Esquema general de la aplicación web

En la Figura 7 podemos observar el esquema general de la aplicación web, también se observan los módulos que la componen y más adelante se explica con más detalle cada módulo.

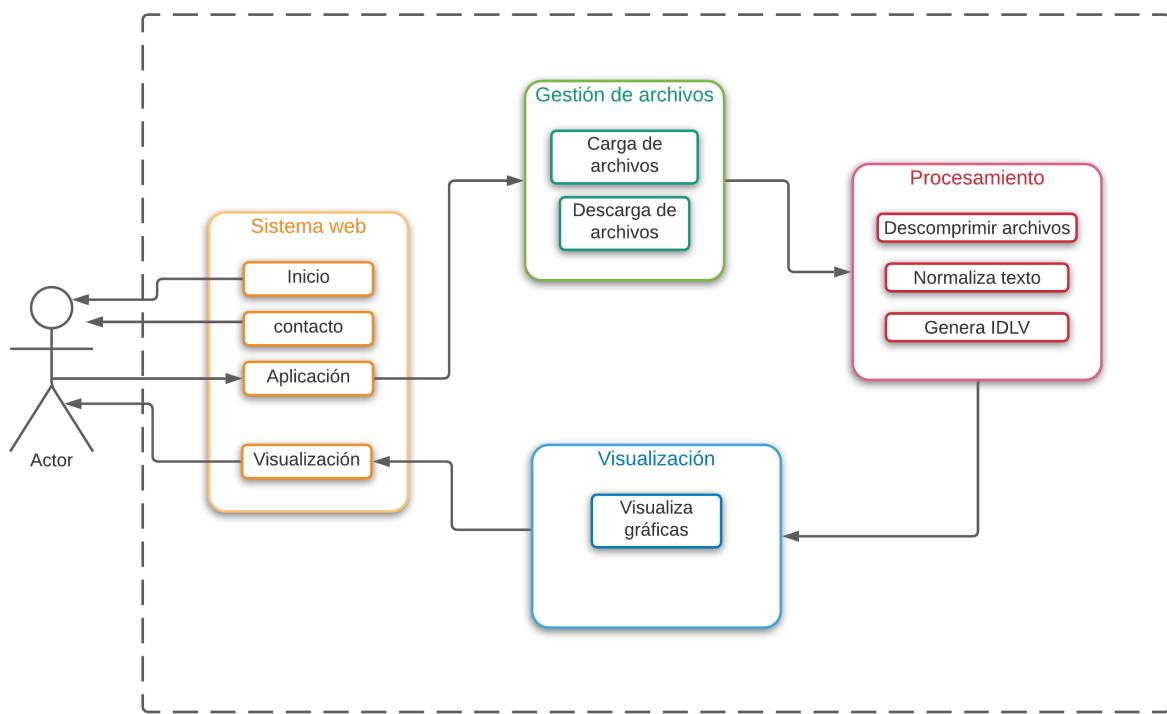


Figura 7: Esquema general de la aplicación web

Como se observa en la Figura 7, en el recuadro 'Sistema web' (recuadro naranja) se muestran las funciones principales de la aplicación web con las que el usuario puede interactuar, las funciones son: inicio, contacto, aplicación y visualización. Es en el subproceso de aplicación donde se desencadenan los demás procesos que realizan el contraste de vocabulario entre distintas poblaciones, esto hace que se utilicen todos los módulos del sistema y finalmente se muestre una visualización de los resultados del contraste de vocabulario mediante gráficas. En las siguientes subsecciones se describen cada uno de los módulos que contiene la aplicación web.

4.1.1. Módulo de gestión de archivos

Este módulo recibe como entrada un archivo comprimido en formato .zip el cual debe contener una carpeta por cada población que se desea contrastar. Dentro de cada carpeta se deben encontrar los archivos de texto en formato .txt y cada archivo de texto representa a una persona de la población. Una vez que el usuario elige un archivo, la aplicación web se encarga de validar que sea correcto, es decir, que el archivo esté en el formato solicitado, además de que el archivo contenga al menos un par de carpetas con archivos de texto en formato txt y que no contenga archivos en un formato diferente.

Finalmente, una vez que se valida el archivo seleccionado se procede a almacenar el archivo .zip en el servidor y se guarda con un nombre basado en la hora, la fecha de subida y el nombre original para ser utilizado posteriormente por el módulo de procesamiento. En la Figura 8 podemos observar el proceso que realiza el módulo de la gestión de datos.

Por otro lado, en este módulo se encuentra la función de bajar archivos, esto facilita a que el usuario pueda guardar archivos en su dispositivo, tales como las listas de palabras y su valor de disponibilidad léxica de las distintas poblaciones.

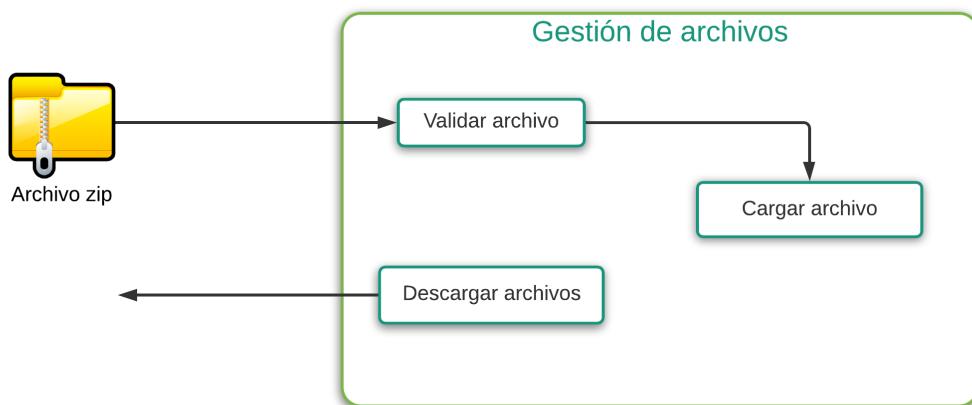


Figura 8: Módulo de gestión de archivos

4.1.2. Módulo de procesamiento

En este módulo es donde se procesa el archivo elegido por el usuario. La finalidad del procesamiento es generar por cada población un archivo con las palabras más importantes y el índice de disponibilidad léxica de cada palabra. En la Figura 9 se muestra el esquema de los subprocesos que realiza el modulo de procesamiento.

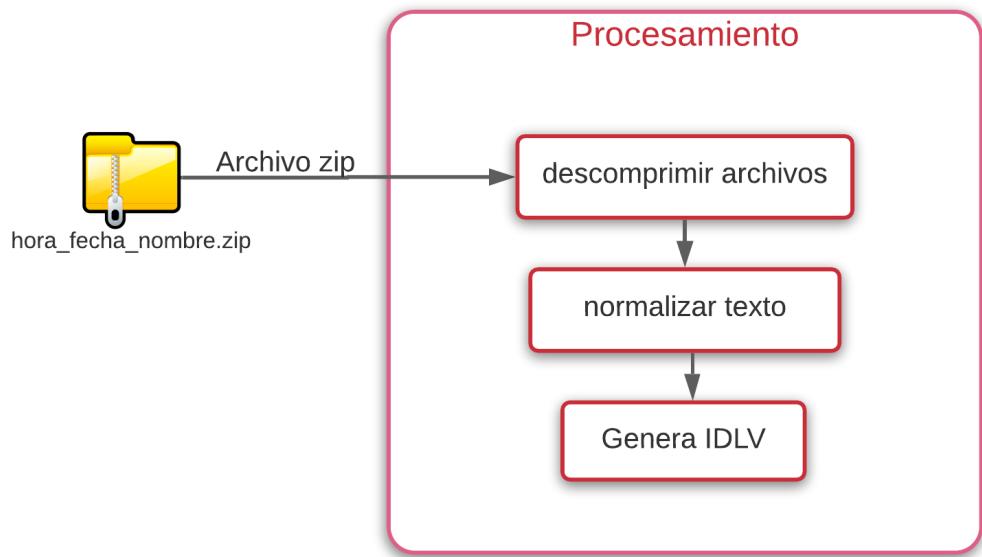


Figura 9: Módulo de procesamiento

A continuación se muestran las distintas tareas que se realizan durante el procesamiento del archivo:

- **Descompresión del archivo:** Teniendo en cuenta que el archivo elegido por el usuario ya fue validado y se encuentra almacenado en el servidor, se procede a descomprimir el archivo para poder acceder a todas las carpetas y archivos de texto que este contiene. En la Figura 10 se muestra un esquema de la descompresión de un archivo y en la Figura 11 podemos observar como es la estructura de un archivo descomprimido, podemos ver que la carpeta principal contiene una carpeta por cada población y dentro de cada carpeta hay archivos de texto.

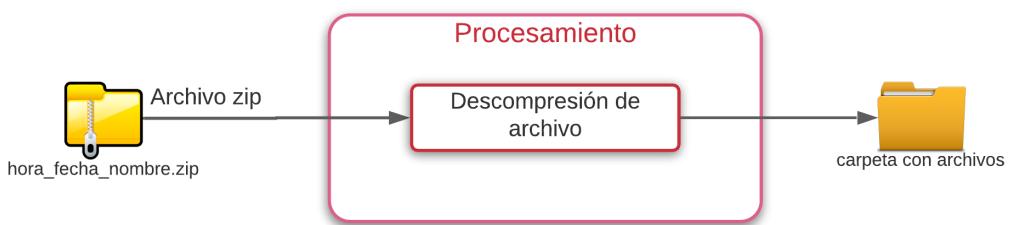


Figura 10: Esquema de la descompresión de un archivo

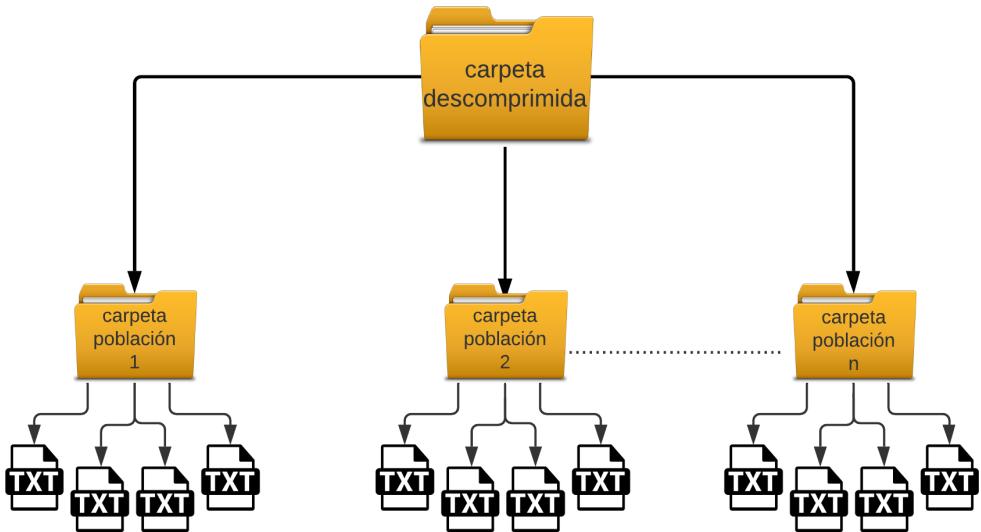


Figura 11: Estructura de un archivo descomprimido

- **Normalización de texto:** Una vez que se tienen todos los archivos de texto por carpetas (poblaciones), se procede a extraer el texto de los archivos para darle un formato en letras minúsculas y eliminar los caracteres especiales, los signos de puntuación y las palabras vacías.³
- **Creación de archivos con texto normalizado:** Despues del proceso anterior se crea un archivo por cada población con todo el texto ya normalizado. En la Figura 12 se muestra un esquema del subproceso para la normalización de textos donde la entrada es un conjunto de archivos de texto y la salida es un archivo de texto ya normalizado.

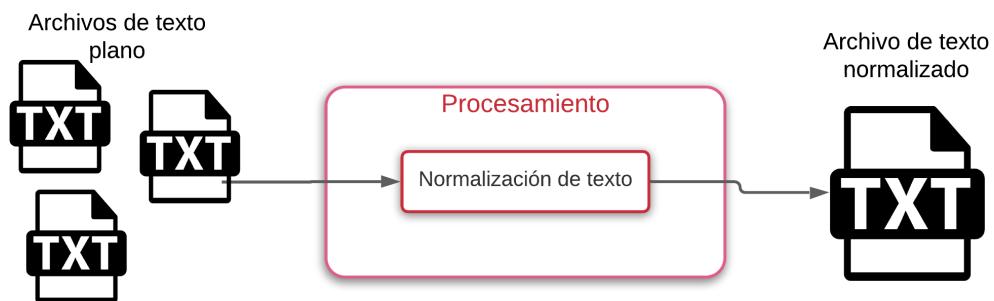


Figura 12: Esquema de la normalización de textos

- **Cálculo del Índice de Disponibilidad Léxica del Vocabulario (IDLV):** En este subproceso se utilizan los archivos de texto normalizado para poder crear una lista de palabras más relevantes por población y su índice de disponibilidad léxica por cada palabra.
- **Creación de archivos con IDLV:** Por último se crea un archivo por población

³palabras que no aportan significado por sí solas como artículos, pronombres, preposiciones, etc.

con la lista de palabras más relevantes y su IDLV. En la Figura 13 se muestra el esquema del cálculo de IDLV.

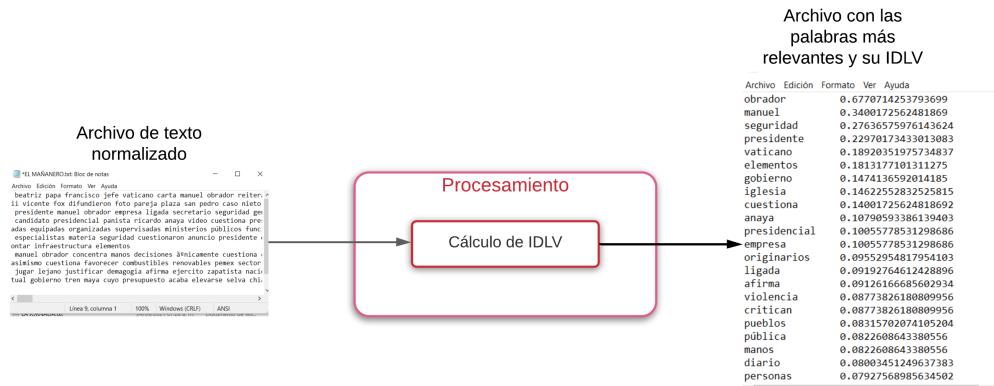


Figura 13: Esquema del cálculo de IDLV. La entrada es un archivo de texto normalizado y la salida es una lista de palabras más relevantes con su valor de disponibilidad léxica.

Todo este procesamiento es necesario para poder hacer el contraste de vocabulario entre las distintas poblaciones y poder generar las gráficas para la visualización de dicho contraste.

4.1.3. Módulo de visualización

En este módulo es donde se generan las gráficas para que el usuario pueda visualizar el contraste de vocabulario entre las distintas poblaciones. Para generar un contraste de vocabulario se utilizan las listas de palabras más relevantes de las distintas poblaciones, además de esto se toman en cuenta cosas como:

- poblaciones que se desea contrastar
- número de palabras a visualizar. Depende del número de poblaciones que se esté manejando pero se recomienda que no sean más de 20 palabras las que se visualicen para una mejor apreciación del contraste de vocabulario.
- Se debe elegir entre visualizar las palabras con mayor o menor disponibilidad léxica.

Una vez que el usuario ha configurado esta parte de la visualización, se continua a elegir el tipo de gráfica para visualizar el contraste de vocabulario. en la Figura 14 se observa el esquema del módulo de la visualización.

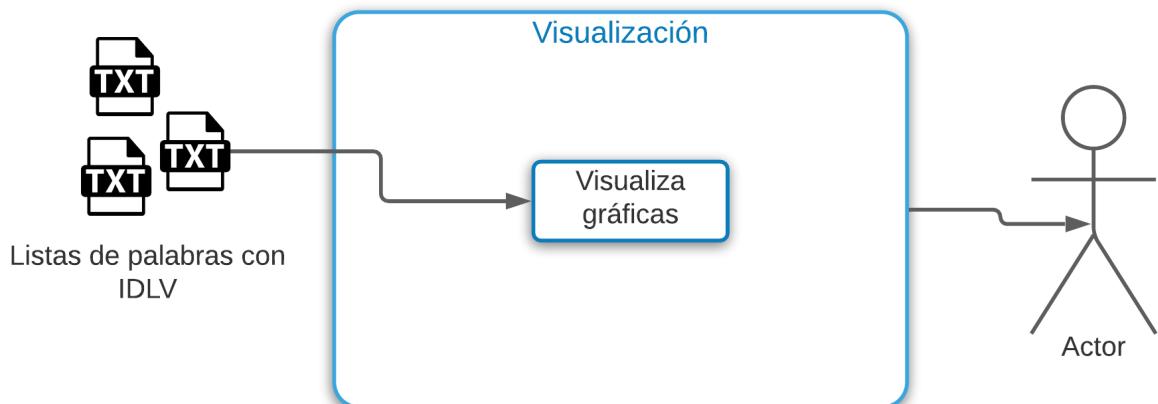


Figura 14: Módulo de Visualización

4.2. Implementación del sistema web

En esta sección se menciona como fue el desarrollo del sistema, también encontramos cuales fueron las distintas tecnologías como lenguajes de programación que se utilizaron e incorporaron para poder crear dicha aplicación web. A continuación se muestra y se explica sobre las tecnologías utilizadas para el sistema web:

- **Python y Flask:** Flask es un framework sencillo que permite crear aplicaciones web utilizando el lenguaje de programación python. Se eligió crear la aplicación web mediante flask porque es fácil de utilizar, cuenta con documentación extensa para crear aplicaciones y como está basado en python, es posible utilizar código y funciones hechas en python. Toda la estructura de la aplicación web esta coordinada por flask, desde el frontend que es la parte con la que el usuario interactúa con la aplicación, hasta el backend que son todos los procesos que hacen posible llegar a una visualización del contraste de vocabulario.
- **Bootstrap:** Al igual que flask, bootstrap es un framework facil de utilizar, está basado en lenguajes como html y css que nos permiten crear un entorno más estético para el usuario y mejorar su experiencia en la aplicación. Debido a que es fácil de utilizar y no se necesitan muchos conocimientos específicos, además de que cuenta con una documentación amplia en línea, fue que se decidió incorporarlo en la aplicación web.
- **JSON:** Este nombre corresponde a las siglas JavaScript Object Notation y es un formato de intercambio de datos. Dentro del sistema web es utilizado para almacenar información resultante de algunos procesos y así mismo, se utiliza para brindar información necesaria a otros procesos para hacer posible sus tareas. Aunque hay otras opciones que hacen lo mismo, se eligió utilizar JSON por ser muy sencillo para el uso y manejo de información.
- **Chart.js y D3.js:** Para la visualización del contraste de vocabulario se utilizaron distintas gráficas que ayudan a ver similitudes y contrastes en el vocabulario de las diferentes poblaciones. Chart.js y D3.js son bibliotecas de JavaScript que nos

permite visualizar los datos en forma de gráficas, Algunas gráficas que se muestran como resultado en la aplicación web están hechas mediante estas bibliotecas de JavaScript. Se utilizaron estas bibliotecas porque brindan una visualización clara al contraste de vocabulario y además el usuario puede manipular dichas gráficas para facilitar aún más su comprensión.

4.3. Visualización de la aplicación web

La visualización de información es una parte muy importante dentro del sistema web, esta sección es la parte final de las tareas de la aplicación web y es donde se muestra información de manera visual y el contraste de vocabulario de las distintas poblaciones al usuario. Para esto, se eligió incorporar una gráfica de pastel que muestra información sobre el archivo cargado por el usuario, de igual manera se incorporaron tres gráficas que ayuden al usuario a visualizar dicho contraste y que sea fácil la interpretación de estas gráficas. A continuación se muestran las gráficas incorporadas de bibliotecas de JavaScript para visualizar el contraste de vocabulario de distintas poblaciones.

4.3.1. Gráfica de Pastel

La gráfica de pastel nos ayuda a mostrar de manera visual y fácil, la información sobre el archivo elegido por el usuario. En esta gráfica se muestra cuantas poblaciones hay dentro de un archivo y el numero de archivos de texto que contiene cada población. Se muestra de un color distinto cada población para que el usuario pueda identificarlos. En la Figura 15 se muestra la visualización de la gráfica de pastel utilizada en la aplicación web.

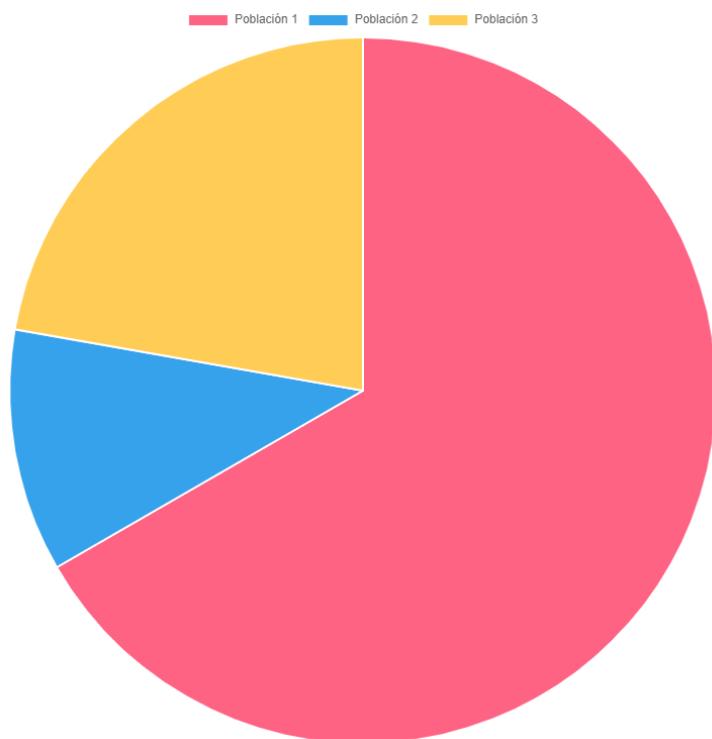


Figura 15: Ejemplo de la gráfica de pastel que muestra información sobre las diferentes poblaciones elegidas por el usuario

4.3.2. Gráfica de barras

En esta gráfica los datos se representan visualmente en forma de barras verticales. Se utiliza para mostrar comparaciones entre las palabras más comunes o menos comunes de las distintas poblaciones utilizando su valor de disponibilidad léxica de la palabra. Con esta gráfica podemos observar como una palabra puede ser muy importante para una población mientras que para otras poblaciones no.

En la Figura 16 se muestra un ejemplo de como se ve la gráfica de barras de la aplicación web. Podemos observar que para cada palabra el valor de disponibilidad léxica cambia dependiendo el grupo de interés al que pertenece y nos indicaría que tan importante es esta palabra para el léxico de ese grupo.

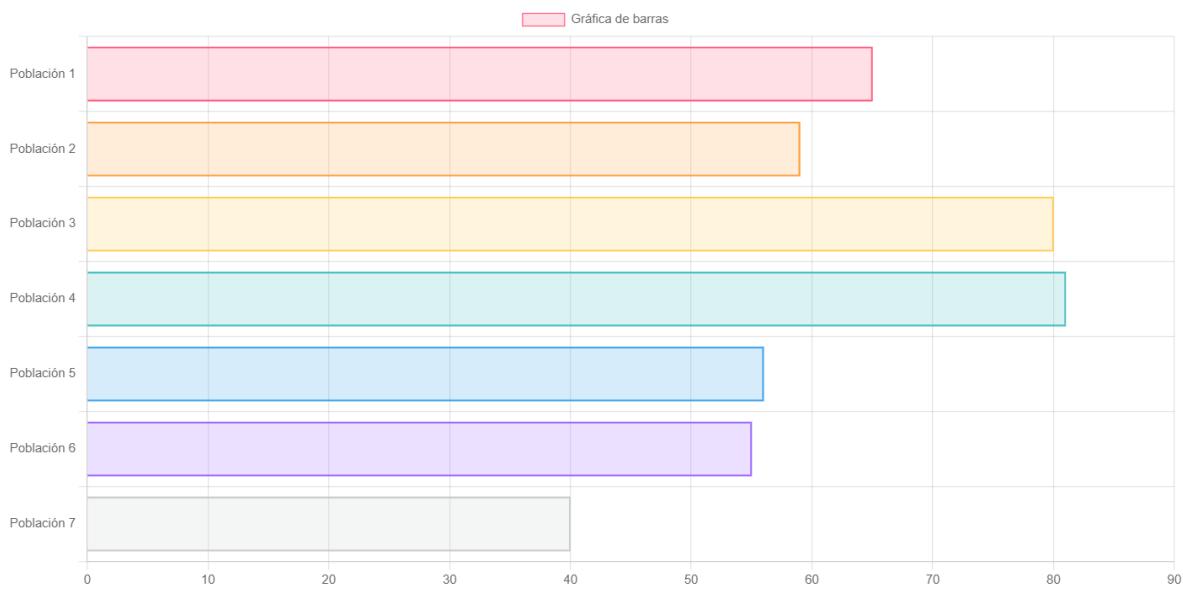


Figura 16: Ejemplo de gráfica de barras vertical

4.3.3. Gráfica de radar

La gráfica de radar es utilizada para mostrar múltiples puntos de datos y la variación entre ellos. Tiene una semejanza con la gráfica de barras pero mientras que en la gráfica de barras vemos por separado cada palabra y su relación entre las poblaciones, con la gráfica de radar podemos observar al mismo tiempo todas las palabras y su relación entre las poblaciones. Es otra manera de ver como las poblaciones se contrastan entre si o como pueden tener similitudes entre ellas.

En la Figura 17 se muestra un ejemplo de como se ve la gráfica de radar de la aplicación web. Podemos observar que las palabras se encuentran alrededor de la gráfica y se observa la relación que hay entre los distintos grupos de interés y sus palabras con mayor o menor disponibilidad léxica.

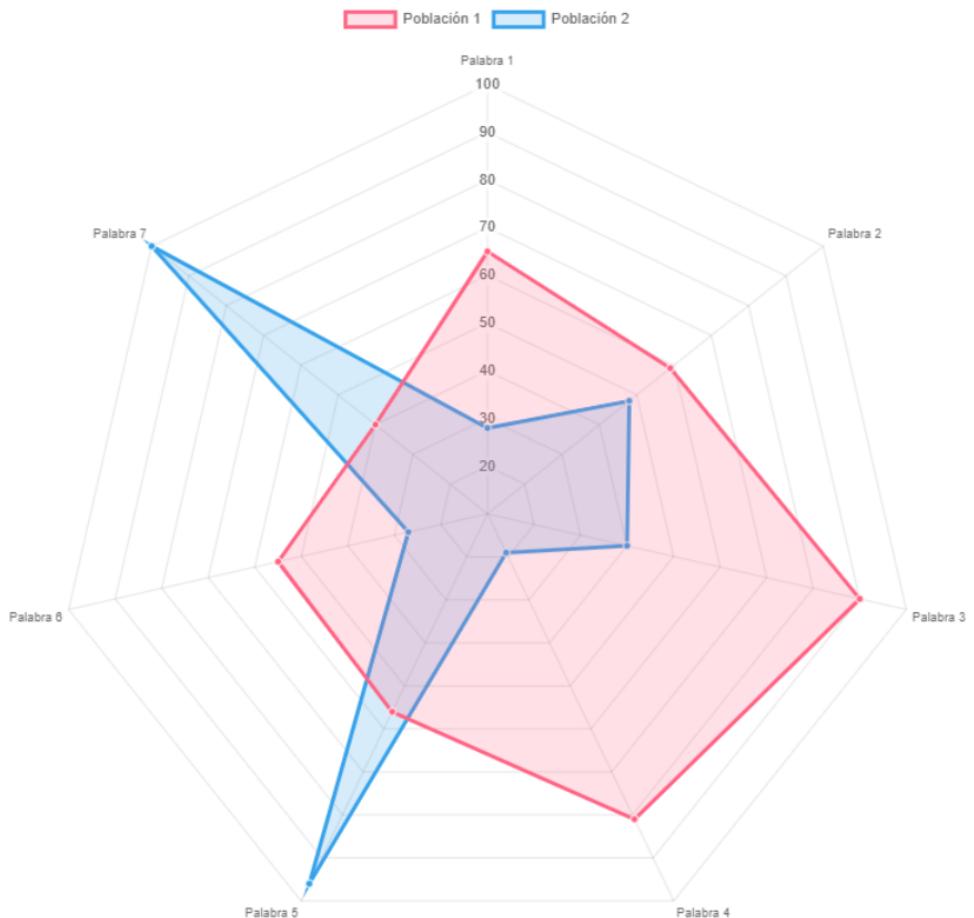


Figura 17: Ejemplo de gráfica de radar

4.3.4. Diagrama de Sankey

El diagrama de Sankey nos permite visualizar el flujo de las palabras que hay entre las distintas poblaciones. Este diagrama nos permite visualizar de una manera diferente a las otras gráficas cómo se relacionan las palabras más relevantes y menos relevantes de dos poblaciones.

En la Figura 18 se muestra un ejemplo de como se ve el diagrama de Sankey de la aplicación web. Podemos observar como las palabras más importantes en el léxico de una población tienen relación con otras poblaciones y su léxico.

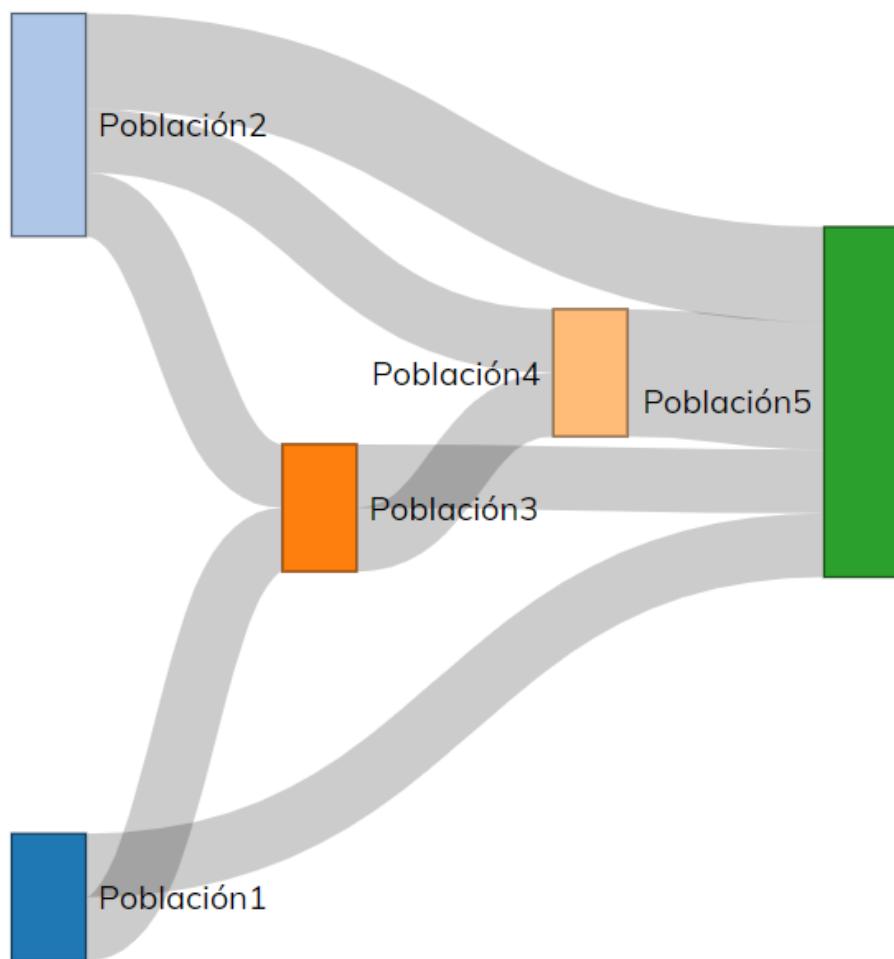


Figura 18: Ejemplo del diagrama de Sankey

5. Ejemplo de uso de la aplicación web

En el siguiente apartado se muestra cómo se ve la aplicación web cuando es utilizada por una persona. Se muestra desde la página de inicio hasta llegar a la visualización del contraste de vocabulario de las distintas poblaciones, se explica con un ejemplo paso a paso cómo se obtiene la visualización de las distintas gráficas de la aplicación web.

En la Figura 19 se observa la página de inicio de la aplicación web. En esta, se muestra información general, tanto de la página como del proyecto y de la disponibilidad léxica. Desde aquí el usuario puede dirigirse a las distintas páginas que contiene el sistema web como el contacto, páginas de información y lo más importante, la aplicación para el contraste de vocabulario.



Figura 19: Página de inicio de la aplicación web

Para nuestro ejemplo de uso de la aplicación se tomó como poblaciones al presidente mexicano AMLO (Andrés Manuel López Obrador) y distintos medios de comunicación como periódicos y facebook. Aunque uno de estos grupos es sólo una persona y los demás son medios de información, se busca analizar los textos que cada grupo ha producido en relación al presidente. Las poblaciones utilizadas quedaron de la siguiente manera.

- AMLO
- El mañanero
- El universal
- La jornada
- Milenio
- Proceso

- Facebook

Los textos de cada población fueron sobre temas tratados en las mañaneras⁴ presentadas del 5 de octubre del 2020 al 16 de octubre del 2020. Los textos fueron creados copiando de la página oficial de cada periódico notas hechas y que hablaran sobre la mañaneras de las fechas señaladas anteriormente, para el caso de facebook, se copiaron comentarios hechos por personas en las páginas de los distintos medios y para el caso de AMLO, los textos fueron extraídos de las mañaneras mediante un crawler⁵.

En la Tabla 5 se muestra el número de textos utilizados por población para este ejemplo de la aplicación web.

Población	Textos
AMLO	10
El mañanero	6
El universal	9
La jornada	17
Milenio	12
Proceso	15
Facebook	8

Tabla 5: Poblaciones y número de textos obtenidos de cada población

Una vez que se obtuvieron los textos de las poblaciones, se creó una carpeta por cada población con sus archivos de texto correspondientes. En la Figura 20 se muestran las carpetas de las distintas poblaciones antes de ser comprimidas en zip.

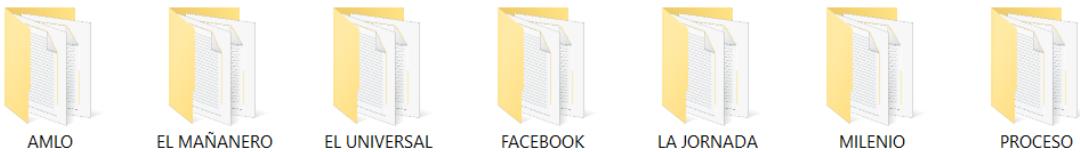


Figura 20: Carpetas de las poblaciones a contrastar

En la Figura 21 se muestra la página de la aplicación para el contraste de vocabulario. Aquí el usuario inicia la aplicación del sistema web para realizar el contraste de vocabulario entre las distintas poblaciones.

⁴Las mañaneras son conferencias hechas a las 7:00am donde el presidente Andrés Manuel López Obrador contesta preguntas a los medios de prensa, explica políticas y habla de temas relacionados con el país

⁵Un crawler es un programa que rastrea páginas de internet para extraer información.

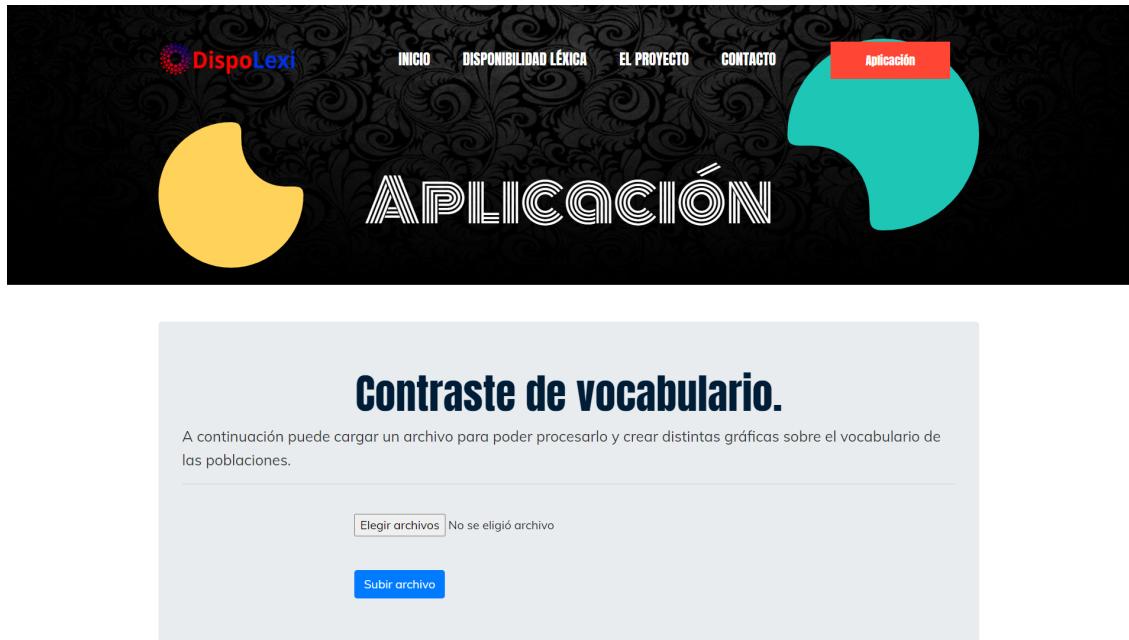


Figura 21: Página para cargar un archivo zip

Lo primero que tiene que hacer el usuario es dar clic en el botón de “Elegir archivo” para que se despliegue una ventana emergente. Una vez que se despliegue la ventana, el usuario podrá elegir un archivo zip que cumpla con las características requeridas. En el caso de este ejemplo de uso, el usuario elige el archivo llamado “mañaneras.zip”. En la Figura 22 se muestra la ventana emergente donde se elige el archivo zip.

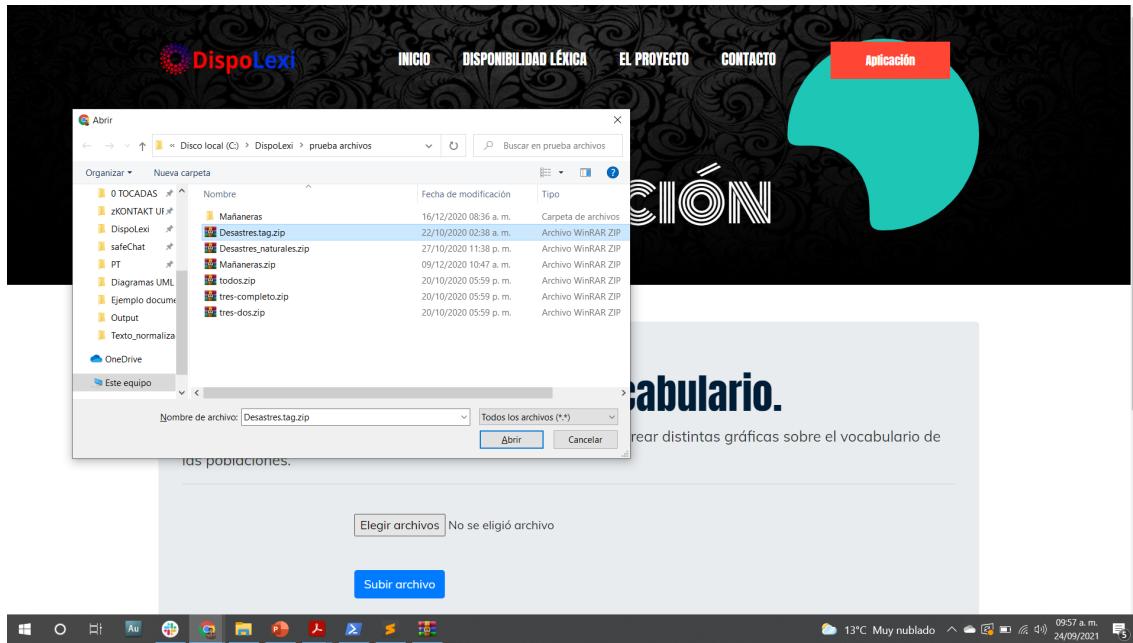


Figura 22: Ventana emergente para seleccionar un archivo .zip

Una vez que el archivo se ha validado y cargado, se muestra la primer gráfica de la aplicación web. Como vemos en la Figura 23, se muestra la gráfica de pastel que contiene la información de las poblaciones subidas por el usuario. Esta parte da la opción al usuario de elegir las poblaciones para el contraste de vocabulario.



Figura 23: Visualización de las poblaciones contenidas en el archivo cargado

Después de que el usuario elige las poblaciones que desea contrastar, se muestra la página donde elige un tipo de gráfica para visualizar el contraste de vocabulario. La aplicación web cuenta con tres gráficas distintas para visualizar el contraste de vocabulario de distintas poblaciones.

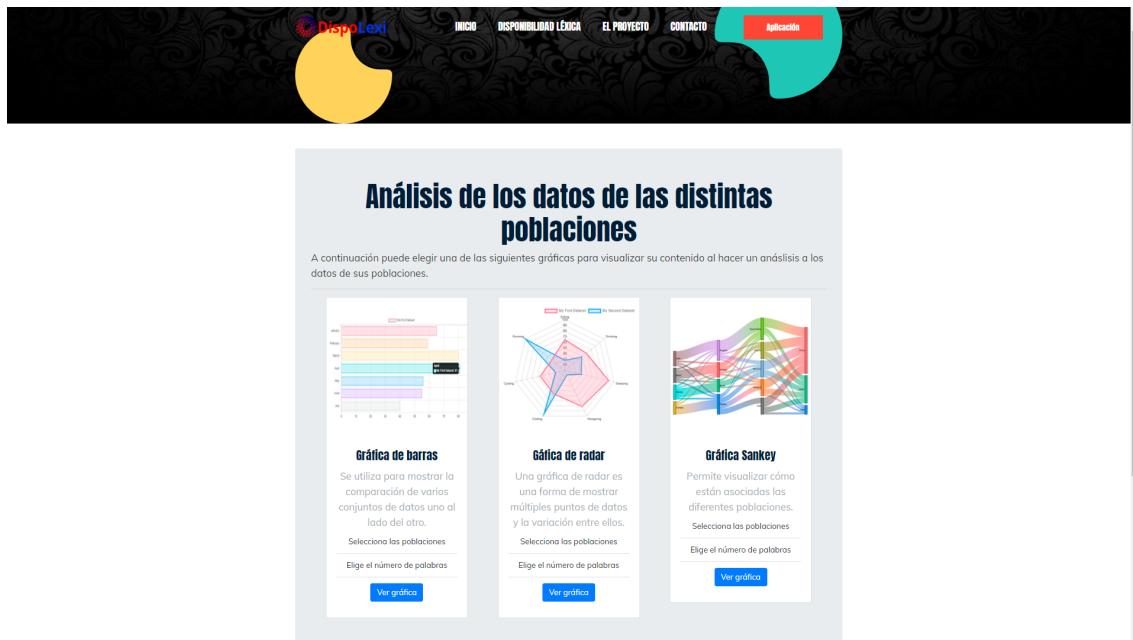


Figura 24: Página donde se elige el tipo de gráfica para visualizar un contraste de vocabulario

Después de que el usuario eligió una gráfica para visualizar el contraste de vocabulario, aparece una página donde el usuario debe seleccionar que poblaciones quiere visualizar (por defecto están todas las poblaciones seleccionadas), el número de palabras que desea visualizar y elegir entre las palabras más relevantes para las poblaciones o las palabras menos relevantes para las poblaciones (por defecto están seleccionadas las palabras más relevantes). Una vez que el usuario haya configurado esta parte le dará clic en el botón azul que dice “Generar gráfica” para poder generar y visualizar la gráfica.

En la Figura 25 se muestra la página donde el usuario configura las partes ya explicadas anteriormente.

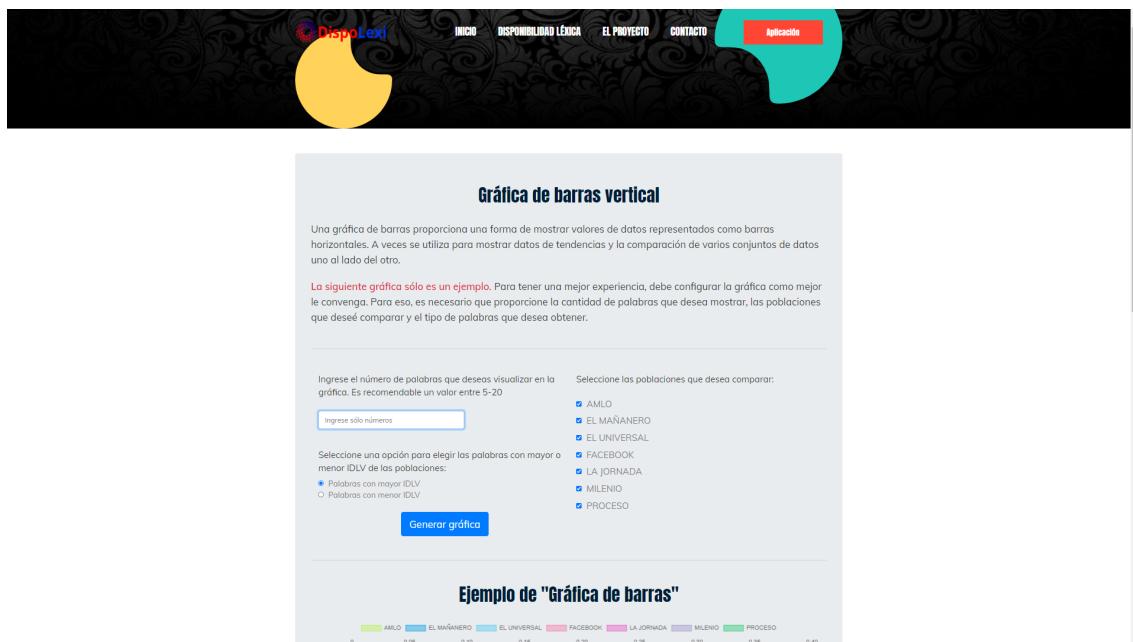


Figura 25: Página de configuración para visualizar gráfica

Después de que el usuario configura la página anterior y da clic en el botón azul, se genera la gráfica seleccionada y el usuario puede visualizar mediante la gráfica contrastes o similitudes de vocabulario entre las distintas poblaciones. En la Figura 26 se muestra la página donde el usuario visualiza el contraste de vocabulario mediante la gráfica.

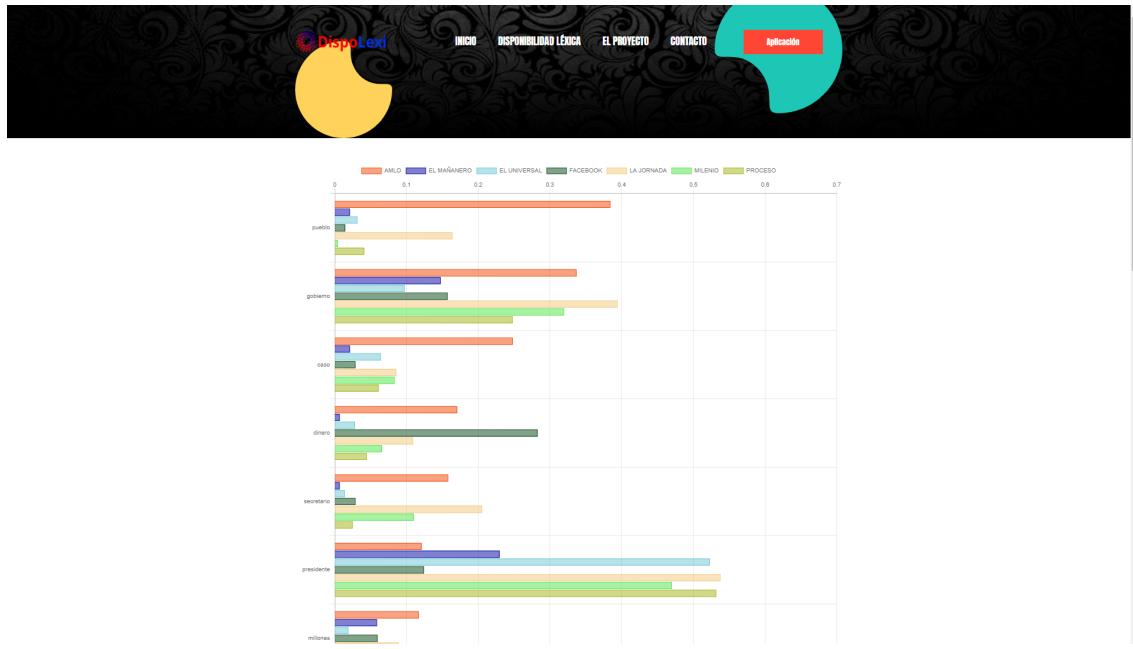


Figura 26: Visualización de la gráfica de barras generada con las especificaciones hechas por el usuario

En la Figura 27 podemos ver más de cerca la gráfica generada y observamos el contraste de vocabulario que hay entre las distintas poblaciones. Hay aspectos que podemos observar en este análisis de vocabulario, por ejemplo vemos que la palabra ‘pueblo’ es más relevante para la población ‘AMLO’ a diferencia de poblaciones como milenio. Otro ejemplo es la similitud que encontramos entre poblaciones como ‘proceso’ y ‘la jornada’ cuando se trata de la palabra presidente ya que es una palabra mas relevante en su vocabulario a diferencia de la población ‘AMLO’ o ‘facebook’.

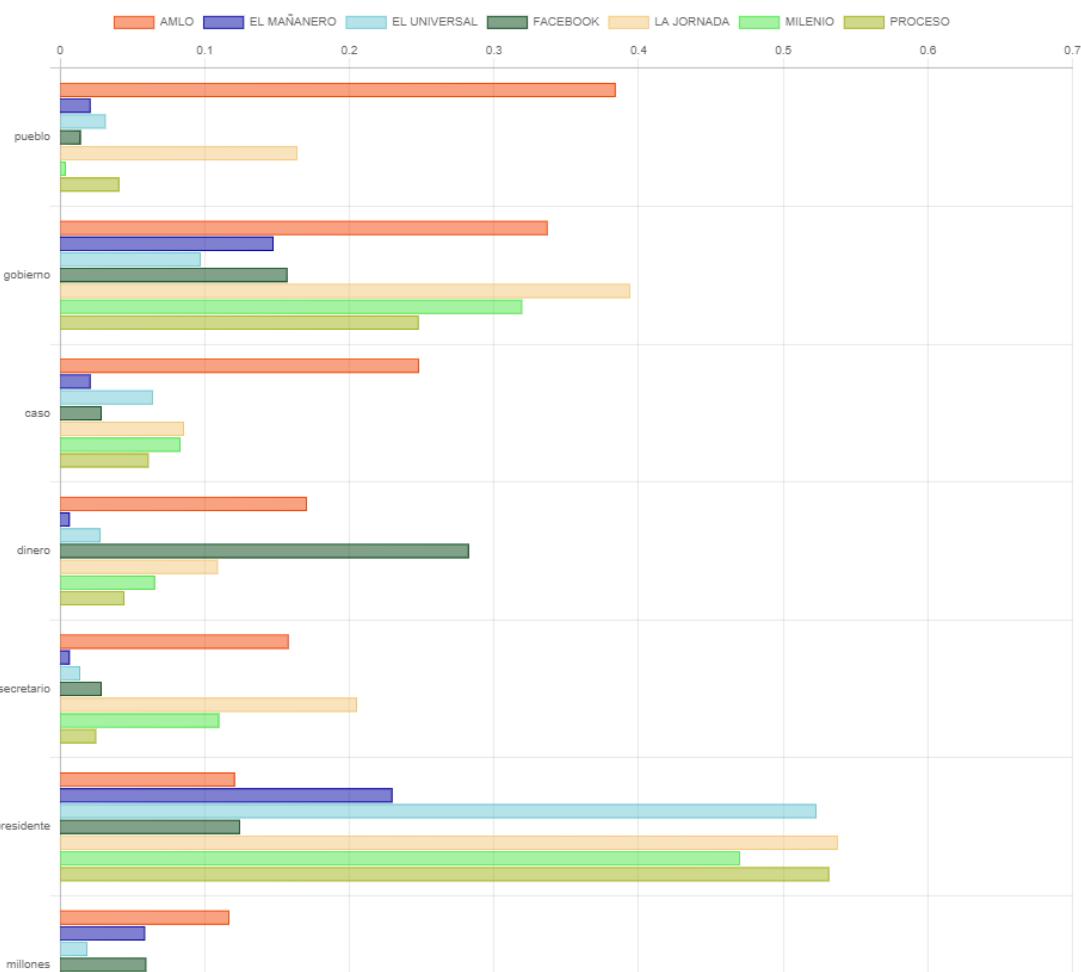


Figura 27: Gráfica de barras generada con las 7 primeras palabras en común entre las poblaciones

En la Figura 28 se muestra la gráfica de radar generada con las 7 primeras palabras que hay en común entre todas las poblaciones, no significa que esas palabras sean las más significativas para todas las poblaciones pero se puede observar que para la mayoría de poblaciones la palabra 'presidente' es más significativa que para la población 'AMLO'.

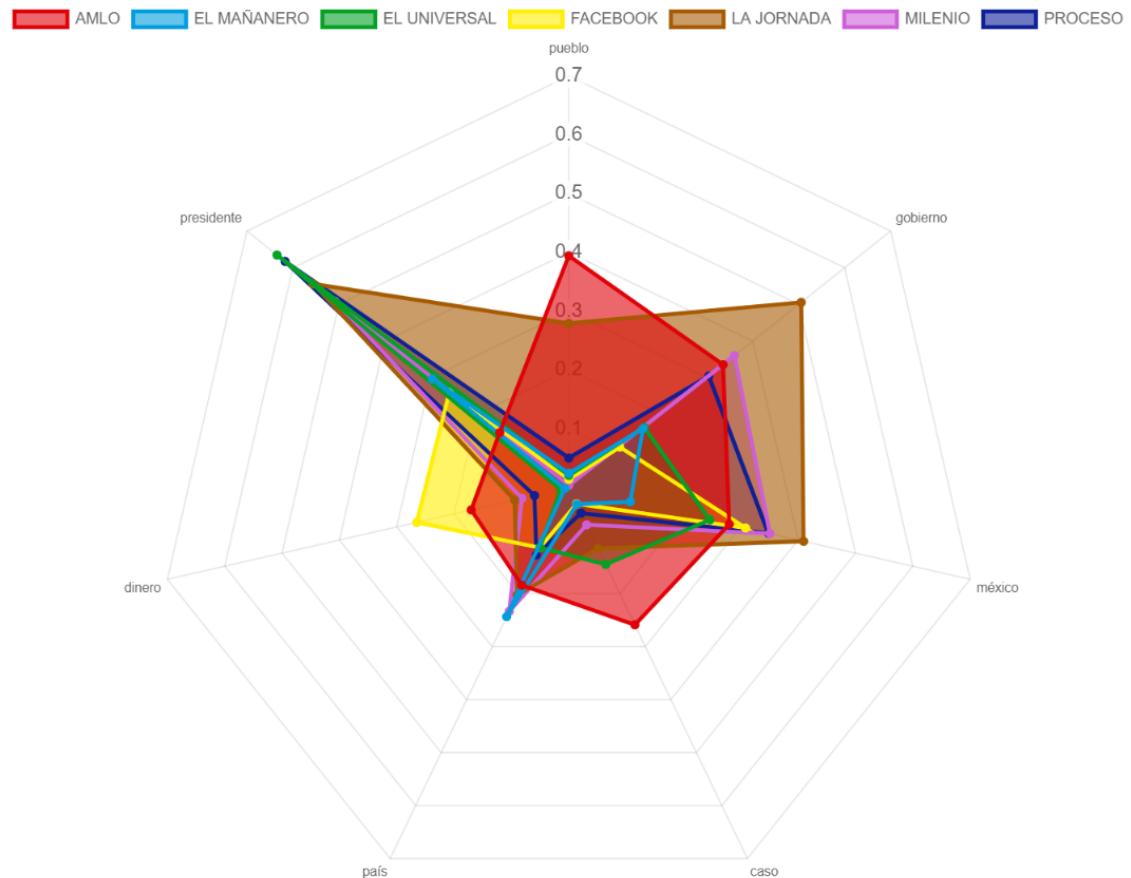


Figura 28: Gráfica de radar generada con las primeras 7 palabras en común entre las poblaciones

En la Figura 29 se muestra el diagrama de sankey donde podemos observar que se comparan dos poblaciones (AMLO y LA JORNADA), estas poblaciones se dividen en cuatro grupos cada una que son: palabras muy relevantes, palabras relevantes, palabras poco relevantes y palabras muy poco relevantes. Este diagrama nos ayuda a ver la relación del vocabulario que tienen dos poblaciones entre las palabras más relevantes y menos relevantes de las poblaciones.

En el diagrama se puede observar que las palabras más relevantes para la población 'AMLO' coinciden en su mayoría con las palabras más relevantes de la población 'LA JORNADA'. También se observa que hay palabras muy relevantes para la población 'AMLO' que están dentro del grupo de palabras muy poco relevantes para la población 'LA JORNADA'.

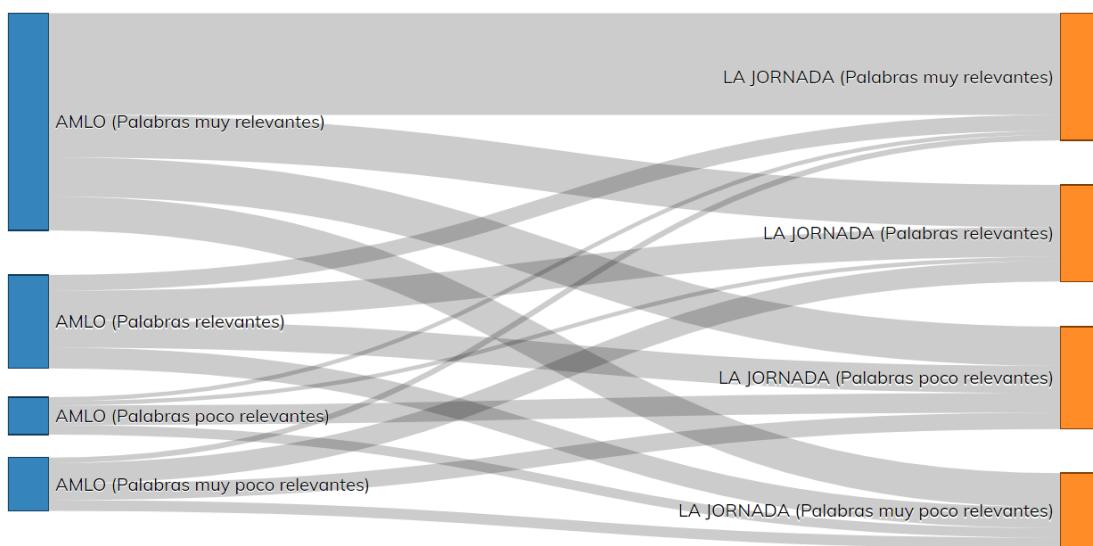


Figura 29: Diagrama de Sankey que muestra la relación del vocabulario que hay entre dos poblaciones

6. Conclusiones y trabajo a futuro

En esta sección se hablará acerca de los objetivos que se lograron cumplir durante el desarrollo de este proyecto, así como las conclusiones que se obtuvieron tras terminar este proyecto terminal.

Al principio de este proyecto se planteo como objetivo general, desarrollar una herramienta web para visualizar y contrastar el vocabulario extraído mediante la disponibilidad léxica de dos o más poblaciones a través del análisis de sus textos. Una vez desarrollado el sistema web, con algunas complicaciones que se lograron resolver, se logró cumplir el objetivo general planteado al inicio.

Al igual que el objetivo general, se lograron cumplir los objetivos específicos que se plantearon, como implementar el cálculo de la disponibilidad léxica para obtener el léxico mental de una población a través del análisis de los textos que ésta produce. También se logró incorporar herramientas de visualización de información que nos ayudan a presentar un contraste de vocabulario entre las distintas poblaciones.

Cabe mencionar que otro objetivo específico que se tenía para este proyecto, era poder validar la utilidad de la herramienta desarrollada en un caso de estudio específico y aunque no se contó con la ayuda de un especialista, se logró validar la utilidad de la aplicación haciendo un contraste de vocabulario entre lo que dijo el presidente mexicano en las mañaneras sobre temas específicos y lo que dijeron diversos medios de comunicación sobre esos mismos temas. Lo anterior mencionado nos ayudó a ver la importancia que tiene la manera en que muestran los datos en un contraste de vocabulario a través de gráficas. Aunque hay múltiples formas de mostrar la información y muchas gráficas que ayudan a esto, se incorporaron las gráficas que más ayudan a entender los contrastes o similitudes que hay entre las distintas poblaciones analizadas.

Es por todo lo anterior que la motivación principal para desarrollar este proyecto terminal fue ayudar a que con la aplicación web sea más fácil para los especialistas poder analizar y contrastar el vocabulario de distintas poblaciones usando la disponibilidad léxica de los textos que estas producen. Sin embargo, cabe resaltar que este proyecto y la aplicación web creada no pretende sustituir el trabajo ya existente de otras personas, más bien busca ser una opción más que facilite el trabajo de los especialistas.

Aunque este proyecto logró cumplir los objetivos propuestos, también se puede mejorar la aplicación web con trabajo a futuro, para esto, se propone incorporando nuevas gráficas que permitan visualizar un contraste de vocabulario de una manera diferente a las que ya se muestra. De igual manera se puede mejorar la aplicación web incorporando cosas como poder subir archivos en diferentes formatos y configuraciones que permitan cambiar la manera en que se maneja la información o se visualizan las gráficas.

La aplicación web puede ser aún más completa si se incorpora una base de datos y un sistema de usuarios, donde se puedan registrar los especialistas y crear una cuenta para poder guardar los contrastes de vocabularios y listas de disponibilidad léxica en distintos momentos. Con esto los usuarios podrían utilizar la aplicación en cualquier momento, ya que toda su información estaría almacenada en la base de datos.

7. Referencias

- [1] Ravi Sethi. *Lenguajes de programación, Conceptos y Constructores*. Addison-Wesley, oct. de 1992.
- [2] Importancia. *Importancia del Lenguaje*. URL: <https://www.importancia.org/lenguaje.php>. (accessed: 10.10.2019).
- [3] Importancia. UNIVERSIDAD DE GUADALAJARA. URL: <http://www.udg.mx/es/oferta-academica/ciencias-sociales-humanidades>. (accessed: 31.10.2019).
- [4] EcuRed. *Antropología Lingüística*. URL: [https://www.ecured.cu/Antropolog%C3%ADA_Ling%C3%BDstica](https://www.ecured.cu/Antropolog%C3%ADA_Ling%C3%ADstica). (accessed: 10.10.2019).
- [5] Aaron Beck. *¿Qué es realmente un psicólogo?* URL: <https://www.cpaaronbeck.com/psicologos-granada/psicologo.html>. (accessed: 10.10.2019).
- [6] FEDERACIÓN ESPAÑOLA DE SOCIOLOGÍA. QUÉ ES LA SOCIOLOGÍA. URL: <http://www.fes-sociologia.com/que-es-la-sociologia/pages/27/>. (accessed: 10.10.2019).
- [7] Isabelle Moulinier Peter Jackson. *Natural Language Processing for Online Applications*. Thomson Corporation, 2002.
- [8] Chris Payne. «Lexical availability». En: *English Teaching professional Pavilion Publishing and Media Ltd, Rayford House, School Road, Hove BN3 5HX, UK* (2016), pág. 18.
- [9] Ángela María García Caballero Manuel Martín Oramas Díaz. «Estudio de Disponibilidad Léxica en estudiantes de bachillerato de las ciudades de Matanzas y Morelia». En: *Universidad de Matanzas Camilo Cienfuegos* (oct. de 2017). URL: <https://www.redalyc.org/journal/4780/478055152002/html/>.
- [10] Kristeve Julie. *Le langage cet inconnu. Fundamentos*. Caracas, 15. 28010 Madrid, 1988.
- [11] Lewandowski Theodor. *Diccionario de lingüística*. Madrid, Cátedra, 2000.
- [12] Gobierno de España Ministerio de educación. *Funciones del lenguaje*. 2007. URL: http://recursos.cnice.mec.es/lengua/profesores/eso2/t1/teoria_1.htm#arriba.
- [13] Wikiversidad. *Lenguaje. Una introducción al estudio psicológico de las habilidades humanas para significar*. Ene. de 2019. URL: https://es.wikiversity.org/wiki/Lenguaje._Una_introducci%C3%B3n_al_estudio_psicol%C3%B3gico_de_las_habilidades_humanas_para_significar. (accessed: 17.10.2019).
- [14] Casado Velarde Manuel. *La innovación léxica en el español actual*. Síntesis, 2015.
- [15] Mercé Pujol Vila Javier Lahuerta Galán. *El lexicón mental y la enseñanza del vocabulario*. 1996.
- [16] Ricardo Mairal Usón Pamela B. Faber. *Construyendo un léxico de verbos en inglés*. Walter de Gruyter, 1999.
- [17] J. Bartol Hernández. *Disponibilidad léxica y selección del vocabulario*. En R. Castañer Martín y V. Lagüéns Gracia, 2010, págs. 85-107.
- [18] J. López Chávez. *omportamiento sintáctico de algunos verbos ordenados según su grado de disponibilidad léxica*. REALE. vol. 1, 1994, pág. 67.

- [19] Francisco Javier Callealta Barroso y Diego Javier Gallego Gallego. «Medidas de disponibilidad lÃ¡: comparabilidad y normalizaciÃ». es. En: *BoletÃn de filologÃa* 51 (jun. de 2016), pÃ¡gs. 39-92. ISSN: 0718-9303. URL: https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0718-93032016000100002&nrm=iso.
- [20] Pedro L. *Tutorial para aprender a investigar el lÃ©xico disponible y las redes semÃ¡nticas a travÃ©s de las herramientas Dispogen y Dispografo*. Oct. de 2018. doi: 10.13140/RG.2.2.30569.34402.
- [21] LexMath. *Bienvenidos a LexMath*. Ene. de 2013. URL: <http://www.lexmath.com/sistema/index.php>. (accessed: 13.10.2019).

7. Anexos

A continuación se muestran las listas de urls de los distintos medios de comunicación (periódicos y facebook) donde se obtuvieron los textos para el ejemplo del uso la aplicación web.

■ AMLO

- <https://lopezobrador.org.mx/temas/amlo-mananera/>

■ El mañanero

1. <https://www.elmananerodiario.com/lo-mas-sonado-la-semana-4-ezln-refuta\amlo-no-nada-perdonar-espana-la-conquista/>
2. <https://www.elmananerodiario.com/lo-mas-sonado-la-semana-2-advierte\ricardo-anaya-riesgos-la-democracia-amlo/>
3. <https://www.elmananerodiario.com/lo-mas-sonado-la-semana-1-amlo-nueva\figura-del-autoritarismo-latinoamerica-financial-times/>
4. <https://www.elmananerodiario.com/amlo-solicita-al-vaticano-disculpa\pide-codices-prestados/>
5. <https://www.elmananerodiario.com/critica-ricardo-anaya-ocurrencias\amlo-materia-seguridad/>
6. <https://www.elmananerodiario.com/acusa-amlo-empresa-ligada-garcia\luna-esta-cobrando-6-mdd-al-gobierno/>

■ El universal

1. <https://www.eluniversal.com.mx/nacion/emilio-alvarez-icaza-absurdo-que\amlo-este-entrengado-las-fuerzas-armadas>
2. <https://www.eluniversal.com.mx/cartera/hay-que-estar-preparados-para\cualquier-reves-del-covid-19-advierte-directora-del-fmi>
3. <https://www.eluniversal.com.mx/estados/con-reclamos-por-fraudes-en\programa-de-reconstruccion-reciben-amlo-en-el-istmo>
4. <https://www.eluniversal.com.mx/nacion/critica-amlo-intromision-de\agencias-extranjeras>
5. <https://www.eluniversal.com.mx/opinion/luis-estrada/es-amlo-el-mas\atacado-en-100-anos>
6. <https://www.eluniversal.com.mx/cultura/no-hay-asedio-ni-ataques-la\libertad-de-expresion-julio-astillero>
7. <https://www.eluniversal.com.mx/nacion/que-se-haga-justicia-con-toda\la-rectitud-pide-cuahtemoc-cardenas-en-caso-de-cienfuegos>
8. <https://www.eluniversal.com.mx/nacion/amlo-pide-fgr-informe-sobre\caso-lozoya-fue-gravissimo-lo-que-denuncio-dice>
9. <https://www.eluniversal.com.mx/cultura/diario-hay-una-buena-dosis\de-veneno-social-en-la-mananera>

■ La jornada

1. [https://www.jornada.com.mx/ultimas/politica/2020/10/05/amlo-llama-a-gabinete-a-redoblar-esfuerzos-para-cumplir-metas\fijadas-1573.html](https://www.jornada.com.mx/ultimas/politica/2020/10/05/amlo-llama-a-gabinete-a-redoblar-esfuerzos-para-cumplir-metas-fijadas-1573.html)
2. <https://www.jornada.com.mx/ultimas/politica/2020/10/05/se-pidio-intervencion-a-honduras-y-guatemala-para-detener-caravana-amlo-1717.html>
3. <https://www.jornada.com.mx/ultimas/cultura/2020/10/08/201clos-periodistas-deben-luchar-en-aras-de-la-democracia201d-jon-lee-anderson-4521.html>
4. <https://www.jornada.com.mx/ultimas/capital/2020/10/07/ciudadanos-anuncian-la-marcha-del-millon-en-apoyo-a-amlo-3512.html>
5. <https://www.jornada.com.mx/ultimas/politica/2020/10/08/mas-tiempo-y-mas-victimas-que-las-previstas-por-pandemia-amlo-2213.html>
6. <https://www.jornada.com.mx/ultimas/politica/2020/10/09/damnificados-por-lluvias-en-tabasco-piden-ayuda-a-amlo-7136.html>
7. <https://www.jornada.com.mx/ultimas/politica/2020/10/09/oposicion-a-eliminacion-de-fideicomisos-tiene-fines-electorales-amlo-7438.html>
8. <https://www.jornada.com.mx/ultimas/politica/2020/10/10/reitera-ejecutivo-federal-que-hay-finanzas-publicas-sanas-6363.html>
9. <https://www.jornada.com.mx/ultimas/politica/2020/10/12/amlo-buena-parte-de-la-ruta-del-tren-maya-sera-electrificada-9859.html>
10. <https://www.jornada.com.mx/ultimas/politica/2020/10/12/importante-que-se-ofrezcan-disculpas-a-pueblos-originarios-amlo-1280.html>
11. <https://www.jornada.com.mx/ultimas/politica/2020/10/12/evita-amlo-opinar-sobre-proceso-de-eleccion-en-morena-2818.html>
12. <https://www.jornada.com.mx/ultimas/politica/2020/10/14/quienes-defienden-los-fideicomisos-defienden-la-corrupcion-amlo-5872.html>
13. <https://www.jornada.com.mx/ultimas/politica/2020/10/14/gobierno-de-la-cdmx-investiga-robo-de-medicinas-contra-cancer-amlo-5548.html>
14. <https://www.jornada.com.mx/ultimas/politica/2020/10/14/aun-sin-acuerdo-pero-avanza-iniciativa-de-reforma-a-ley-del-infonavit-amlo-3016.html>

15. <https://www.jornada.com.mx/ultimas/politica/2020/10/15/no-hace-falta-usar-cubrebocas-si-se-mantiene-sana-distancia-amlo-9339.html>
16. <https://www.jornada.com.mx/ultimas/politica/2020/10/15/anuncia-amlo-nueva-etapa-de-atencion-medica-para-covid-19-6665.html>
17. <https://www.jornada.com.mx/ultimas/politica/2020/10/16/caso-cienfuegos-muestra-descomposicion-del-regimen-neoliberal-amlo-9987.html>

■ Milenio

1. <https://www.milenio.com/politica/amlo-temas-conferencia-mananera-8-octubre-2020>
2. <https://www.milenio.com/politica/mario-delgado-el-movimiento-esta-mas-vivo-que-nunca>
3. <https://www.milenio.com/politica/cabeza-vaca-4t-desaparece-fideicomisos-sustento-manejo>
4. <https://www.milenio.com/politica/amlo-temas-conferencia-mananera-13-octubre-2020>
5. <https://www.milenio.com/politica/fideicomisos-si-hay-corrupcion-que-se-denuncie-oposicion-a-amlo>
6. <https://www.milenio.com/negocios/financial-times/restringira-mexico-permisos-ip-sector-energetico>
7. <https://www.milenio.com/politica/amlo-temas-conferencia-mananera-16-octubre-2020>
8. <https://www.milenio.com/politica/amlo-marzo-2021-recuperaremos-empleos-perdidos-pandemia>
9. <https://www.milenio.com/negocios/profeco-cierra-10-gasolineras-litros-completos>
10. <https://www.milenio.com/negocios/tren-mexico-queretaro-amlo-revive-planes-construirlo>
11. <https://www.milenio.com/politica/frena-marcha-zocalo-cdmx-saldo-positivo>

■ Proceso

1. <https://www.proceso.com.mx/nacional/2020/10/16/onu-pide-mexico-no-socavar-presupuesto-destinado-combatir-la-violencia-contra-mujeres-251070.html>
2. <https://www.proceso.com.mx/nacional/2020/10/15/lopez-obrador-entre-mas-me-golpean-mas-digno-me-siento-estoy-orgulloso-de-lo-que-hago-250969.html>

3. <https://www.proceso.com.mx/nacional/2020/10/14/la-prensa-conservadora-se-esta-agrupando-en-contra-del-gobierno\ -dice-amlo-250904.html>
4. <https://www.proceso.com.mx/internacional/2020/10/14/amlo-el-muro-la-guardia-nacional-donald-trump-250939.html>
5. <https://www.proceso.com.mx/nacional/cdmx/2020/10/12/robo-de-medicamentos-ocurrio-la-madrugada-del-domingo-vigilante\ -aviso-el-lunes-fgjcdmx-250777.html>
6. <https://www.proceso.com.mx/economia/2020/10/12/coparmex-advierte-amlo-la-extincion-de-fideicomisos-es-\ inconstitucional-250780.html>
7. <https://www.proceso.com.mx/opinion/2020/10/13/amlo-las-feministas-cronica-de-una-confrontacion-reiterada-250833.html>
8. <https://www.proceso.com.mx/nacional/2020/10/9/amlo-ante-criticas-de-la-prensa-mi-escudo-es-la-mananera-mi-\ angel-guardian-el-pueblo-250656.html>
9. <https://www.proceso.com.mx/cultura/2020/10/9/vistos-buenos-del-inah-para-el-proyecto-tren-maya-250708.html>
10. <https://www.proceso.com.mx/nacional/2020/10/8/no-hay-que-hacer-conjeturas-dice-amlo-tras-declaracion-de-\ inocencia-de-garcia-luna-250595.html>
11. <https://www.proceso.com.mx/nacional/2020/10/8/analisis-de-carstens-sobre-bancarrota-no-aplica-mexico-amlo-\ 250596.html>
12. <https://www.proceso.com.mx/nacional/2020/10/8/atencion-por-pandemia-nos-llevo-mas-tiempo-de-lo Esperado-amlo-\ 250602.html>
13. <https://www.proceso.com.mx/nacional/2020/10/7/el-fonden-era-caja-chica-de-funcionarios-dice-amlo-\ 250539.html>
14. <https://www.proceso.com.mx/nacional/2020/10/7/rodriguez-castaneda-amlo-es-hombre-de-pelea-no-se-bajara-del-\ ring-250570.html>
15. <https://www.proceso.com.mx/nacional/2020/10/6/poner-orden-transparencia-alega-amlo-en-extincion-de-\ fideicomisos-250460.html>

■ Facebook

- Links de El universal, Milenio y Proceso.