

# DETECCIÓN DE LAS FRECUENCIAS DEL LÉXICO DE UN TEXTO CIENTÍFICO CON LA APLICACIÓN WEB LEX-A PARA APOYAR EL APRENDIZAJE DE LENGUAS EXTRANJERAS

Yadira A. Hernández-Pérez <sup>1\*</sup>, Susana A. Alaniz-Álvarez <sup>2</sup>, Julio C. Martínez-Troncoso <sup>3</sup>, Michelangelo Martini <sup>4</sup> y Ana D. Esparza-Herrera <sup>5</sup>

<sup>1</sup> Escuela Nacional Preparatoria, Plantel 5, Universidad Nacional Autónoma de México, Calz. Del Hueso 729, Coapa, Ex Hacienda Coapa, Tlalpan, C.P. 14300, Ciudad de México, México.

<sup>2</sup> Centro de Geociencias, Campus Juriquilla Universidad Nacional Autónoma de México, Blvd. Juriquilla 3001, C.P. 76230, Querétaro, Querétaro.

<sup>3</sup> Facultad de Ingeniería, Universidad Nacional Autónoma de México, Av. Universidad 3000, Ciudad Universitaria, Coyoacán, C.P. 04510, Ciudad de México, México.

<sup>4</sup> Instituto de Geología, Universidad Nacional Autónoma de México, Av. Universidad 3000, Ciudad Universitaria, Coyoacán, C.P. 04510 Ciudad de México, México.

<sup>5</sup> Escuela Nacional Preparatoria, Plantel 3, Universidad Nacional Autónoma de México, Av. Ing. Eduardo Molina No. 1577, Delegación Gustavo A. Madero, C.P. 07469, Ciudad de México, México.

\*[yadira.hernandez@enp.unam.mx](mailto:yadira.hernandez@enp.unam.mx) (autora para correspondencia)

## RESUMEN

Se presenta una lista de vocabulario italiano-español con las palabras más recurrentes en un texto de divulgación científica de Ciencias de la Tierra escrito en la lengua extranjera. La lista se creó a partir de los resultados arrojados por el analizador léxico LEX-A (Martínez y Hernández, 2020), que es una aplicación web en construcción. Se aborda la creación del listado en tres etapas, así como el tratamiento de la aplicación piloto con un grupo de estudiantes de italiano como lengua extranjera con nivel A1. La versión final de la lista de vocabulario fue validada por un geocientífico nativo hablante del italiano. La lista fue creada como herramienta para facilitar la comprensión de lectura del texto base usado en un taller de Ciencias de la Tierra dirigido a docentes y estudiantes de diferentes áreas de conocimiento con nivel A1 de italiano.

**Palabras clave:** Ciencias de la Tierra, italiano, divulgación científica, comprensión de lectura, vocabulario, aplicación web.

## ABSTRACT

*A list of Italian-Spanish vocabulary is presented with the most recurrent lexical items (words) in a science dissemination text written on Earth Sciences in a foreign language. The list was created on the basis of the results of the LEX-A lexical analyzer (Martínez and Hernández, 2020), which is the web app under construction. The creation of the list is addressed in three stages, as well as the pilot application treatment with a group of students of Italian as a foreign language with A1 level. The final version of the vocabulary list was validated by a native Italian-speaking geoscientist. The list was created as a tool to facilitate the reading comprehension of the basic text used in an elementary Earth Sciences course aimed at teachers and students from different areas of knowledge, with A1 level of Italian.*

**Keywords:** Earth Sciences, Italian, science dissemination, reading comprehension, vocabulary, web app

## INTRODUCCIÓN

La lengua siempre ha acompañado al conocimiento científico, ya sea para plantearlo, para cuestionarlo, para comunicarlo o para preservarlo. Lengua y ciencia son inseparables.

En el caso de un texto de ciencia escrito en una lengua extranjera, conocer su léxico (entendido el léxico como el acervo de palabras que lo conforman) agiliza los procesos de comprensión de lectura. En principio, se asume que el lector ya conoce el tema de ciencia por instrucción formal en su lengua nativa, de modo que el texto plantea la posibilidad de que el lector recicle<sup>1</sup> conocimientos científicos mientras se enfoca en el aprendizaje de la ciencia en una lengua distinta a la suya.

La construcción de listas de palabras ha sido cuestionada en múltiples ocasiones dentro del campo de la enseñanza de lenguas extranjeras debido a que, en múltiples ocasiones, su uso se ha limitado a resolver superficial y momentáneamente problemas de desconocimiento léxico. Tal es el caso de las listas de vocabulario que acompañan a un texto, o la inserción de ligas intratextuales que conducen al significado de una palabra.

Por el contrario, la introducción de listas de vocabulario para apoyar la comprensión de textos dentro de un curso de enseñanza de lengua extranjera cobra relevancia cuando se acompaña de actividades complementarias para el aprendizaje de ese léxico en contexto (Schmidt, 1990; Laufer y Hustijn, 2001; Hustijn y Laufer, 2001; Kim, 2008; Rott, 2012). En este marco, una lista de palabras puede apoyar los procesos de inferencia (deducción razonada) que se promueven durante la lectura (Hughes, 2023). Este es el caso del curso de introducción al italiano para estudiantes mexicanos por medio del estudio de experimentos de Galileo. La experiencia se reporta en este mismo número de la Revista ECG.

La construcción de listas de vocabulario que serán objeto de aprendizaje se lleva a cabo mediante el análisis léxico del texto base, dicho análisis solía hacerse manualmente. En la actualidad existen programas de cómputo que lo realizan en segundos y arrojan resultados precisos. Uno de los ejemplos más completos de analizador léxico digital es el *Compleat Lexical Tutor* de Tom Cobb (1997).

El *Compleat Lexical Tutor* es un software de procesamiento de datos que ofrece una serie de opciones al estudiante de vocabulario de inglés y francés. Las herramientas permiten evaluar el conocimiento léxico, leer y escuchar vocabulario en contexto, determinar el tamaño y profundidad del léxico mental en estado pasivo con el fin de convertirlo en activo, así como medir la presencia y corrección de la información gramatical de las piezas léxicas y de grupos de palabras, entre otros. La asunción teórica sobre la que se diseñó el *Lex Tutor*, como se le conoce comúnmente, es que el lector que sabe al menos las 200 palabras más frecuentes de una lengua, sumadas a las 570 piezas léxicas más comunes del discurso académico (según la *Academic Word List*, AWL), puede comprender cualquier texto académico (cf. Morris y Cobb, 2004).

<sup>1</sup> Debe hacerse referencia al modelo pedagógico de la Teoría del Cambio Conceptual (Chi, 2008; Carey, 2011) que propone la existencia de interacción entre los conocimientos previos y los nuevos siempre que inicia el aprendizaje. Tal interacción sucede a través de dos procesos de construcción: asimilación (que es una adaptación del proceso piagetiano del mismo nombre), y acomodación o reestructuración.

<sup>2</sup> *Backend* es el código de programación que opera del lado del servidor incluyendo las funciones CRUD, la base de datos y la lógica del servidor.

<sup>3</sup> *Frontend* se refiere a la parte de un sitio web que interactúa directamente con el usuario.

## LEX-A

En el presente estudio se implementó el analizador léxico LEX-A (Martínez y Hernández, 2020). Esta aplicación web usa *Python* como lenguaje primario de programación, los módulos científicos *Pandas* y *Matplotlib* para el *backend*<sup>2</sup> y *Flask* para el *frontend*<sup>3</sup>, todos son de código abierto. A fin de que LEX-A sea capaz de manejar grandes volúmenes de texto en tiempo real, se eligió un marco ligero como lo es *Flask*, para asegurar rapidez y eficacia. El código se almacena en un repositorio privado, lo que permite rastrear los cambios con el transcurso del tiempo (Figura 1). La combinación de estas tecnologías dió como resultado una poderosa aplicación web para el análisis léxico a partir de textos auténticos. Nuestra *app* es accesible, confiable, eficiente y escalable, lo que hace posible obtener datos a partir de un texto a fin de apoyar una amplia variedad de propósitos académicos y de investigación.

LEX-A permite obtener resultados de análisis léxico de textos de hasta 50 cuartillas de longitud en formatos de archivos como: .docx, .pdf y .txt. A la fecha, LEX-A realiza el procesamiento detallado del texto y arroja los resultados en cuatro formas de organización distintas: número de frecuencia, conteo de todas las palabras del texto, conteo de palabras que excluye las palabras repetidas, lista de frecuencia de ocurrencias (ascendente o descendente), análisis léxico descartando todas las palabras de función (artículos, preposiciones, pronombres, verbos auxiliares) y gráficas con las ocho palabras más frecuentes en el texto.

LEX-A realiza las mismas tareas en alemán, español, francés, inglés, italiano y purépecha. Los resultados pueden servir de punto de partida para la construcción de corpus léxicos y glosarios y para realizar el análisis de extensión y profundidad léxicos de un texto. Estos recursos han sido reconocidos por su pertinencia para promover el aprendizaje de vocabulario así como el diseño de materiales y de instrumentos para la evaluación del conocimiento léxico. Para mostrar el uso de LEX-A como auxiliar en la enseñanza de una lengua extranjera, se hizo el análisis léxico del texto *La pressione atmosferica e la caduta dei corpi* (Alaniz-Álvarez y Nieto-Samaniego, 2017) de la colección *Esperimenti semplici per capire una Terra complessa*.

En primera instancia, se construyó un listado con todas las palabras del texto. El proceso comprendió los siguientes pasos:

- Análisis léxico del texto por medio de LEX-A.
- Exportación de la lista a una base de datos tipo Excel.
- Integración de la información semántica y categoría gramatical.
- Clasificación de las piezas léxicas de acuerdo con su categoría gramatical conservando el orden de frecuencia.



Figura 1. Página de acceso de la aplicación web LEX-A en: <http://cuatromiradas.herokuapp.com>

## ANÁLISIS LÉXICO DE UN TEXTO MEDIANTE LEX-A

Una vez seleccionado el texto a analizar, el primer paso consistió en pegarlo en la ventana que aparece en la primera pantalla. Para este estudio se usó el listado por orden de frecuencia descendente, esto es, de la palabra que aparece más frecuentemente en el texto a la que menos. Este orden de frecuencia permitió:

- Enlistar las piezas léxicas de función (artículos, pronombres, preposiciones, conjunciones, verbos copulativos y verbos auxiliares) que normalmente constituyen el mayor porcentaje de palabras en un texto. La jerarquización por frecuencia de todas estas piezas del corpus obedece a que se concentrará el aprendizaje en los vocablos más comunes de cada categoría.
- Recabar en los últimos lugares las piezas léxicas cuyos significados tengan la menor aportación a la comprensión del texto.
- Asegurar que todas las piezas léxicas que formen parte del texto estén incluidas en el corpus.
- Conocer la frecuencia con que cada palabra aparece en el texto, ya que ésta se asoció con su aportación a nivel conceptual para la comprensión.

### Selección de las piezas léxicas

Se contabilizaron 4704 palabras que corresponden a 1272 piezas léxicas diferentes (Tabla 1) del libro *La pressione atmosferica e la caduta dei corpi* (Alaniz-Álvarez y Nieto-Samaniego, 2017).

El siguiente paso fue agruparlas por su categoría gramatical conservando el orden de frecuencia. La lista consta de 103 sustantivos de los cuales 9 corresponden a unidades de medida, 3 nombres propios, 57 a verbos, 43 a adjetivos, 25 a preposiciones, 18 a adverbios, 10 a conjunciones y 4 a pronombres. A cada sustantivo se le agregó la forma plural entre paréntesis (Tabla 2). Se insertó una columna con todos los significados en español, otra con significados alternativos posibles (cuando los hubiera) y, una más con el género de cada sustantivo (Tabla 3).

Este listado se sometió al escrutinio de un especialista en el área, nativo hablante del italiano.

## IMPLEMENTACIÓN DE LEX-A

### Prueba piloto

Esta primera versión se puso a prueba con un grupo de 6 estudiantes del sexto grado de bachillerato con nivel A1+ en italiano, según el Marco Común Europeo de Referencia para las lenguas. El procedimiento llevó a cabo las siguientes acciones:

- Presentación del texto base al grupo. Fase de prelectura: los estudiantes revisaron el texto a nivel de percepción visual para identificar elementos gráficos y tipográficos que orientaran sus inferencias acerca del tema general.
- Reconocimiento de la idea principal. Fase de prelectura: los estudiantes leyeron superficialmente el texto para ubicar las palabras de contenido (sustantivos, verbos, adjetivos y adverbios) que aparecían con mayor frecuencia. A partir de esa información y la obtenida de las imágenes, títulos y subtítulos, se realizó una lluvia de ideas hasta establecer cuál era la idea principal.
- Detección de problemas en la comprensión del texto por desconocimiento léxico. Fase de lectura: 1ª lectura. Los estudiantes leyeron el texto resaltando con color las oraciones que no pudieron entender ya fuera parcial o totalmente. Siempre que fue posible,

usaron un color distinto para señalar las palabras que impedían o nublaban la comprensión.

- Presentación del listado (1ª versión). Los estudiantes recibieron el listado con las siguientes indicaciones: observar su estructura, leerlo completo con atención, subrayar las palabras que habían encontrado en el texto y que impedían la comprensión de algún fragmento.
- Aplicación a la lectura. Los estudiantes volvieron a leer el texto completo, en esta ocasión consultando el listado cada vez que encontraban una oración o palabra de las que habían marcado anteriormente. A fin de promover una mayor profundidad en el procesamiento, se les indicó no escribir los significados en el texto.
- Evaluación de la comprensión del texto. Se solicitó a los estudiantes que elaboraran 10 preguntas sobre el texto que involucraran las palabras que habían marcado en el segundo paso, sin tener la lista a la mano. Al terminar, se intercambiaron los cuestionarios para que los resolvieran. Concluido el trabajo, se requirió a los estudiantes que compartieran la mayor información que les fuera posible recordar sobre el texto en equipos de tres personas. Se observó que los estudiantes habían comprendido el texto completo.
- Retroalimentación. Los estudiantes señalaron aquellas palabras que les resultaron más útiles para la comprensión del texto y escribieron un párrafo corto con sus sugerencias.

A partir de los resultados obtenidos de la aplicación piloto, se elaboró la versión actual del listado, la cual consta de 126 piezas léxicas. Las acciones incluyeron: eliminar 51 sustantivos, 3 nombres propios, 35 verbos, 7 unidades de medida, 37 adjetivos y 10 adverbios. Se excluyeron los datos relativos a la frecuencia. Se conservó la clasificación por categorías gramaticales a fin de promover la atención sobre la sintaxis de la lengua italiana y se ordenaron alfabéticamente las piezas de cada categoría. Asimismo, se mantuvieron las formas del plural y los significados de todas las piezas léxicas (Tabla 3). El listado actual se conforma de los contenidos de la Tabla 1 y Figura 2, el formato se modificó para facilitar la búsqueda.

### LESSICO - VOCABULARIO

## La pressione atmosferica e la caduta dei corpi



### SOSTANTIVI - SUSTANTIVOS

tempo (pl. tempi)	tiempo (pl. tiempos), clima (pl. climas)
pressione (pl. pressioni)	presión (pl. presiones)
aria (pl. arie)	aire (pl. aires)
palloncino (pl. palloncini)	globo (pl. globos)
terra (pl. terre)	tierra (pl. tierras)
bicchiere (pl. bicchieri)	vaso (pl. vasos)
oggetto (pl. oggetti)	objeto (pl. objetos)
bottiglia (pl. bottiglie)	botella (pl. botellas)
contagocce (pl. contagocce)	gotero (pl. goteros)
gravità (pl. gravità)	gravedad (pl. gravedades)
corpo (pl. corpi)	cuerpo (pl. cuerpos)
massa (pl. masse)	masa (pl. masas)
successo (pl. successi)	éxito (pl. éxitos)
candela (pl. candele)	vela (pl. velas)
legge (pl. leggi)	ley (pl. leyes)
svolgimento (pl. svolgimenti)	ejecución (pl. ejecuciones)
velocità (pl. velocità)	velocidad (pl. velocidades)
vita (pl. vite)	vida (pl. vidas)
spazio (pl. spazi)	espacio (pl. espacios)
barattolo (pl. barattoli)	frasco (pl. frascos)
carta (pl. carte)	papel (pl. papeles)
densità (pl. densità)	densidad (pl. densidades)
foglio (pl. fogli)	hoja (pl. hojas)
livello (pl. livelli)	nivel (pl. niveles)
moto (pl. moti)	moto (pl. motos), movimiento (pl. movimientos)
natura (pl. nature)	naturaleza (pl. naturalezas)
pavimento (pl. pavimenti)	piso (pl. pisos)
estrato (pl. estrati)	capa (pl. capas), estrato (pl. estratos)
volta (pl. volte)	vez (pl. veces)
cannuccia (pl. cannucce)	popote (pl. popotes)
gravitazione (pl. gravitazioni)	gravitación (pl. gravitaciones)
misura (pl. misure)	medida (pl. medidas)
biglia (pl. biglie)	canica (pl. canicas)

Figura 2. Fragmento de la Tabla 3. Consulta la [Tabla completa](#).

## APLICACIÓN DE LEX-A EN EL TALLER DE ITALIANO CON EXPERIMENTOS DE GALILEO

Se presenta aquí el caso de aplicación en el "Taller de introducción al italiano a través de los experimentos de Galileo" ofrecido a estudiantes mexicanos que están aprendiendo italiano, en donde se usa como texto base *La pressione atmosferica e la caduta dei corpi* (Alaniz-Álvarez y Nieto-Samaniego, 2017) de la colección *Esperimenti semplici per capire una Terra complessa*. El artículo que reporta esa experiencia se publica en este mismo número con el título 'Acercamiento a la lengua italiana con experimentos de Galileo' (Alaniz-Álvarez, et al. 2023).

## CONCLUSIONES

En este trabajo se presenta el analizador léxico LEX-A, una aplicación web que facilita los procesos de análisis léxico de textos en alemán, francés, inglés, italiano y purépecha.

Se propone el uso de LEX-A como apoyo para la elaboración de material didáctico en un curso de lenguas extranjeras con propósitos académicos. En este texto se presentó un ejemplo de su implementación en un taller de italiano dirigido a estudiantes de bachillerato que aspiran a una licenciatura relacionada con las Ciencias de la Tierra: Geociencias, Geomática, Geología y Ciencias de la Tierra.

Las diferentes experiencias en su implementación aportarán nuevos usos de la aplicación web LEX-A, a fin de tener una herramienta cada vez más útil para el diseño de materiales didácticos para la comprensión de textos, así como para mejorar el aprendizaje de vocabulario en cualquier lengua extranjera.

## AGRADECIMIENTOS

El listado de palabras que se presenta en este artículo es uno de los productos del proyecto 'A las lenguas por la Ciencia', (Centro de Geociencias, CGEO y Escuela Nacional Preparatoria, ENP de la Universidad Nacional Autónoma de México). El proyecto busca impulsar el aprendizaje de las Geociencias desde el nivel de bachillerato mediante cursos de Ciencias de la Tierra en alemán, francés, inglés e italiano. La planeación y diseño de estos cursos demanda la creación de glosarios terminológicos que favorezcan la comprensión de textos del área.

## REFERENCIAS

- Alaniz-Álvarez, S. A., Nieto-Samaniego, A. F., 2017, *Esperimenti Semplici per Capire una Terra Complessa: 1. La Pressione Atmosferica e la Caduta dei Corpi*, Verónica Nájera Martínez, María Luisa Pedraglio, traductores. Universidad Nacional Autónoma de México, Centro de Geociencias, 26 p., ISBN: 978-607-02-9408-2.
- Alaniz-Álvarez S.A., Martini M. , Esparza A.D., Gutiérrez-Navarro R., Hernández-Pérez Y.A., Mori L. y Nieto-Samaniego A.F. (2023). Acercamiento a la lengua italiana con experimentos de Galileo. *Enseñanza y Comunicación de las Geociencias*, vol. 2, núm. 1.
- Carey, S. (2011). Précis of the origin of concepts. *Behavioral and Brain Sciences*, 34(3), 113-124..
- Chi, M. T. (2009). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In

- International handbook of research on conceptual change (pp. 89-110). Routledge.
- Cobb, T. (1997). Version (8.5). Compleat Lexical Tutor. Retrieved April 9, 2023, from <https://www.lex tutor.ca/>.
- COLMEX. (2023). Diccionario del español de México. DEM. Retrieved April 12, 2023, from <https://dem.colmex.mx/>
- Hughes, L. S. (2019). Contextualized Versus Decontextualized Vocabulary Learning as a Pre-reading Task. Research Gate. Seoul; Korea.
- Hulstijn, J., y Laufer, B. (2001). Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. *Language Learning*, 51(3), 539–558. <https://doi.org/10.1111/0023-8333.00164>
- Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, 58, 285–325. <https://doi.org/10.1111/j.1467-9922.2008.00442.x>
- Laufer, B., y Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1–26. <https://doi.org/10.1093/applin/22.1.1>
- Martínez, J.C. y Hernández, Y. A. (2020). LEX-A Version (1.3). LEX-A App. Retrieved April 12, 2023. <http://cuatromiradas.herokuapp.com/>.
- Morris, L., y Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System*, 32(1), 75-87.
- (N.d.). Version (Understanding complex systems series). In *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*. Retrieved May 8, 2023, from [https://link.springer.com/chapter/10.1007/978-3-662-47238-5\\_3](https://link.springer.com/chapter/10.1007/978-3-662-47238-5_3).
- Rott, S. (2012). The effect of task-induced involvement on vocabulary acquisition: An approximate replication of Hulstijn and Laufer (2001). In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 228–267). Cambridge University Press.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Trautwein, J. (2019). *The Mental Lexicon. Assessment, size and structure* (dissertation). Retrieved May 8, 2023 NRC-1. *Saline systems*, 2, 1-13.

Manuscrito recibido: 17 de marzo de 2023

Manuscrito corregido recibido: 16 de mayo de 2023

Manuscrito aceptado: 24 de mayo de 2023