



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
PROGRAMA DE MAESTRÍAS Y DOCTORADO EN  
LINGÜÍSTICA

TÍTULO

ALONIMIA DE TOPÓNIMOS DE MÉXICO EN ALEMÁN EN  
UN CORPUS DE MINERÍA Y CIENCIAS DE LA TIERRA Y  
SU LOCALIZACIÓN CON PYTHON

TESIS

QUE PARA OPTAR POR EL GRADO DE MAESTRO EN LINGÜÍSTICA  
APLICADA

PRESENTA

Erick Brian Villanueva Villaseñor

TUTORA:

Dra. Marisela Colín Rodea  
ENALLT, UNAM

MÉXICO, CDMX. a 25 de abril de 2022



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Esta DECLARACIÓN PERSONAL DE NO PLAGIO debe ser insertada en primera página de todos los trabajos terminales para la obtención de grado.

### DECLARACIÓN PERSONAL DE NO PLAGIO

Yo, Erick Brian Villanueva Villaseñor  
[NOMBRE COMPLETO DEL SUSTENTANTE]

estudiante de Lingüística Aplicada  
(NOMBRE DEL PLAN DE ESTUDIOS CURSADO)

con número de cuenta 313009929, declaro que soy autor/a de este documento académico titulado:

Alonimia de topónimos de México en alemán en un corpus de  
minería y ciencias de la tierra y su localización con Python,

el cual presento como trabajo terminal para la obtención del grado correspondiente.

Certifico que el mismo es fruto de mi trabajo personal, y que no he copiado, utilizado ideas, formulaciones, citas integrales o ilustraciones extraídas de cualquier obra, artículo, memoria, etc., en cualquier formato y soporte, sin mencionar de forma clara y estricta su origen, tanto en el cuerpo del texto como en la bibliografía. Asiento también que no he hecho uso de información no autorizada, proveniente de cualquier fuente.

Soy plenamente consciente de que el hecho de no respetar estos lineamientos de autenticidad y originalidad es objeto de sanciones universitarias y/o de otro orden legal.

Cd. Universitaria, CdMx, 25 de abril de 2023

Erick Brian Villanueva Villaseñor

(NOMBRE Y FIRMA DEL SUSTENTANTE)



## Reconocimientos:

Quiero agradecer a mi familia, Paola, Ricardo, Leticia, Alejandra y Gabriela. A mi tutora y a los miembros de mi sínodo, Dra. Marisela Colín, Dra. Gemma Bel, Dra. Ioana Cornea, Dr. Erik Franco y Dr. Daniel Rodríguez-Vergara. A Luis Guillermo Bustillo Estrada y a la Dra. Luz Fernanda Azuela, a todos mis amigos y al Consejo Nacional de Ciencia y Tecnología por la beca otorgada para hacer de mis estudios de maestría.

## RESUMEN

La toponimia mexicana ha sufrido cambios al incorporarse a otros sistemas lingüísticos e incluso dentro del español es posible encontrar múltiples variantes nominales para un mismo objeto geográfico. El fenómeno de la multiplicidad de nombres propios para un mismo referente, llamado alonimia, representa un problema durante la búsqueda documental, ya que el investigador puede no conocer todas las variantes nominales que puede tener el objeto de su interés. En esta investigación se analizará este fenómeno en la combinación español-alemán y se propondrá una metodología para la localización automática de las diversas variedades nominales de un topónimo en un corpus usando el lenguaje de programación Python.

## ABSTRACT

Mexican toponymy has undergone changes during its incorporation into other linguistic systems and even within Spanish it is possible to find multiple nominal variants for the same geographical object. The phenomenon of the multiplicity of proper names for the same referent, called allonym, represents a problem during the documentary search, as the researcher may not know all the nominal variants that the object of his interest may have. This research will analyze this phenomenon in the Spanish-German combination and propose a methodology for the automatic localization of the various nominal varieties of a toponym in a corpus using the programming language Python.

## PALABRAS CLAVE

Topónimo – México – Traducción – Alemán – Python

## ÍNDICE

<b>INTRODUCCIÓN .....</b>	<b>9</b>
<b>MARCO TEÓRICO .....</b>	<b>13</b>
1.1 LA ONOMÁSTICA.....	13
1.2 EL NOMBRE PROPIO.....	14
1.3 EL TOPÓNIMO.....	16
1.3.1 EL SEMANTISMO DE LOS TOPÓNIMOS.....	16
1.3.2 EL ESTUDIO DE LOS TOPÓNIMOS.....	19
1.3.2.1 ETIMOLOGÍA TOPONÍMICA.....	19
1.3.2.3 TOPONOMÁSTICA APLICADA.....	22
1.3.2.4 SUBCLASES DEL TOPÓNIMO.....	23
1.3.2.1.1 OICÓNIMOS.....	24
1.3.2.1.2 ORÓNIMOS.....	25
1.3.2.1.3. MICROTOPÓNIMOS.....	26
1.4 ALONIMIA .....	29
1.4.1 DEFINICIÓN.....	29
1.4.2 SUBCLASIFICACIONES DE LA ALONIMIA.....	30
1.4.2.1 ALONIMIA SUPERFICIAL O FONÉTICO-GRÁFICA.....	30
1.4.2.2 ALONIMIA DE VARIACIÓN DE RAÍZ.....	31
1.4.2.3 ALONIMIA DE CONCORDANCIA SEMÁNTICA.....	31
1.4.2.4 ALONIMIA DE DIVERSIDAD DE PALABRAS.....	32
1.5 LA ALONIMIA Y SU RELACIÓN CON LA TRADUCCIÓN.....	33
1.5.1 TRANSCRIPCIÓN Y TRANSLITERACIÓN.....	33
1.5.2 MÉTODO DOMESTICANTE Y MÉTODO EXTRANJERIZANTE.....	34
1.5.3 TÉCNICAS Y MÉTODOS TRADUCTORES.....	36
1.5.4 FUNCIÓN Y TIPOLOGÍA TEXTUAL.....	39
1.5.4.1 TEXTOS NARRATIVOS.....	40
1.5.4.2 TEXTOS DESCRIPTIVOS.....	40
1.5.4.3 TEXTOS ARGUMENTATIVOS.....	41
1.5.4.4 TEXTOS EXPLICATIVOS.....	41
1.5.4.5 TEXTOS DIALÓGICOS.....	41
1.5.4.6 APROXIMACIONES PRELIMINARES A LOS TIPOS DE TEXTOS EN EL CORPUS COMPILADO.....	42
1.6 GRAFEMÁTICA Y GRAFOTAXIS.....	43
1.7 CAMBIOS EN LOS TOPÓNIMOS.....	44
1.7.1 TOPÓNIMOS DE MÉXICO.....	45
1.7.2 ETAPAS TOPONÍMICAS DE MÉXICO.....	45
1.7.3 ASPECTOS PARTICULARES DEL PERIODO COLONIAL Y POSINDEPENDENTISTA.....	47
1.8 LINGÜÍSTICA DE CORPUS Y LINGÜÍSTICA COMPUTACIONAL.....	49
1.8.1 LINGÜÍSTICA DE CORPUS.....	49
1.8.1.1 COMPILACIÓN DEL CORPUS.....	53
1.8.2 LINGÜÍSTICA COMPUTACIONAL.....	54
1.8.3 RELACIÓN ENTRE LA LINGÜÍSTICA DE CORPUS Y LA LINGÜÍSTICA COMPUTACIONAL.....	55
1.8.3.1 ANTECEDENTES DE ESTUDIOS DE NOMBRES BASADOS EN CORPORA ASISTIDOS POR COMPUTADORA.....	55
<b>METODOLOGÍA .....</b>	<b>58</b>
2.1 CORPUS.....	60
2.1.1 CARACTERÍSTICAS GENERALES DEL CORPUS.....	60
2.1.2 COMPILACIÓN DEL CORPUS.....	62
2.1.3 REVISIÓN Y EDICIÓN DEL CORPUS.....	64
2.1.3.1 PROBLEMAS CON LOS DOCUMENTOS ORIGINALES.....	64
2.1.3.2 PROBLEMAS CON LOS DOCUMENTOS Y EL PROCESO DE DIGITALIZACIÓN.....	69
2.1.3.3 CORRECCIÓN DE LOS TEXTOS DEL CORPUS.....	71
2.1.4 RECABACIÓN DE DATOS DEL CORPUS.....	74
2.1.5 REGISTRO DE TOPÓNIMOS DEL CORPUS.....	74

2.2 REVISIÓN Y SUAVIZADO DE DATOS.....	76
2.2.1 MEDIDAS DE AGRUPACIÓN Y DISPERSIÓN DEL CORPUS .....	77
2.2.2 CORPUS SIN VALORES ATÍPICOS.....	80
2.3 SELECCIÓN Y APLICACIÓN DE PRUEBAS ESTADÍSTICAS .....	81
2.3.1 PRUEBA KOLMOGOROV-SMIRNOV.....	81
2.3.2 PRUEBA $\chi^2$ DE BONDAD DE AJUSTE.....	81
2.4 PROCESAMIENTO DE TEXTOS PARA LA EXTRACCIÓN DE DATOS DE ARCHIVOS CSV Y LOCALIZACIÓN DE TOPÓNIMOS DE ARCHIVOS PDF CON PYTHON.....	83
2.4.1 AMBIENTE .....	83
2.4.2 IMPORTACIÓN DE MÓDULOS.....	84
2.3.3 PROCESAMIENTO DE TEXTOS PARA LA LOCALIZACIÓN DE TOPÓNIMOS EN PDFS CON PYTHON.....	85
2.3.3.1 EVALUACIÓN DE RESULTADOS DE LA LOCALIZACIÓN CON PYTHON .....	86
<b>RESULTADOS.....</b>	<b>88</b>
3.1 DESCRIPCIÓN DEL CORPUS.....	88
3.2 EVALUACIÓN DE RESULTADOS DE LA LOCALIZACIÓN DE TOPÓNIMOS.....	88
3.2.1 CONCLUSIONES DE LA EVALUACIÓN DE RESULTADOS DE LA LOCALIZACIÓN DE TOPÓNIMOS.....	90
3.3 RESULTADOS DE LAS PRUEBAS ESTADÍSTICAS .....	90
3.3.1 Kolmogorov Smirnov - Año.....	91
3.3.1.1 CONCLUSIONES GENERALES DE LA APLICACIÓN DE LA PRUEBA KOLMOGOROV- SMIRNOV PARA LA VARIABLE INDEPENDIENTE AÑO .....	98
3.3.2 $\chi^2$ DE BONDAD DE AJUSTE - FUENTE DE PUBLICACIÓN.....	99
3.3.2.1 CONCLUSIONES GENERALES DE LA APLICACIÓN DE LA PRUEBA $\chi^2$ DE BONDAD DE AJUSTE PARA LA VARIABLE FUENTE .....	105
3.3.3 $\chi^2$ DE BONDAD DE AJUSTE - TIPOLOGÍA TEXTUAL .....	106
3.3.3.1 CONCLUSIONES GENERALES DE LA APLICACIÓN DE LA PRUEBA $\chi^2$ DE BONDAD DE AJUSTE PARA LA VARIABLE TIPOLOGÍA TEXTUAL .....	112
<b>DISCUSIÓN Y CONCLUSIONES FINALES .....</b>	<b>115</b>
<b>BIBLIOGRAFÍA .....</b>	<b>124</b>

## Índice de figuras

Figura 1. Monoreferencia directa del nombre propio ejemplificado con Hidalgo (basado en el diagrama de Nübling (p. 18)).....	15
Figura 2. Texto con tipografía Fraktur (Gerolt y Berghes, 1828) .....	65
Figura 3. Dispersión de tinta (Burkart, 1839) .....	66
Figura 4. Difuminación de la tinta en el párrafo superior (Gerolt y de Berghes, 1828).....	67
Figura 5. Texto con hojas oscuras (Gerolt, 1827).....	68



Figura 6. Texto con orientación incorrecta (Mühlepfordt, 1847) .....	70
Figura 7. Texto con defecto de escaneo (Berghes, 1847) .....	71
Figura 8. Texto original (Redacción, 1848).....	72
Figura 9. Texto reconocido (Redacción, 1848).....	72
Figura 10. Texto reconocido con correcciones para los topónimos encontrados .....	73
Figura 11. De izquierda a derecha, marcación en texto, registro en archivo CSV y su visualización con Python.....	76
Figura 12. Núm. de topónimos por texto.....	77
Figura 13. Diagrama de cajas y bigotes del núm. de topónimos por texto .....	79
Figura 14. Núm. de topónimos por texto sin valores atípicos.....	80
Figura 15. Serie de tiempo de la alonimia de México .....	91
Figura 16. Variedades nominales de <i>México</i> por fuente .....	100
Figura 17. Variedades nominales de <i>México</i> por tipología textual .....	106

## Índice de tablas

Tabla 1: Tabla de transliteración y transcripción de Otto Back.....	34
Tabla 2. Preguntas de investigación.....	59
Tabla 3. Resultados de la localización de topónimos .....	90
Tabla 4. Cálculo de frecuencias acumulativas para las distribuciones observadas y teóricas de <i>Mexico</i> por año .....	92
Tabla 5. Cálculo de frecuencias acumulativas para las distribuciones observadas y teóricas de <i>Mexiko</i> por año .....	94
Tabla 6. Cálculo de frecuencias acumulativas para las distribuciones observadas y teóricas de <i>Mejico</i> por año .....	95
Tabla 7. Cálculo de frecuencias acumulativas para las distribuciones observadas y teóricas de <i>Mejiko</i> por año .....	96
Tabla 8. Cálculo de frecuencias acumulativas para las distribuciones observadas y teóricas de <i>Neuspanien</i> por año. ....	98
Tabla 9. Cálculo de frecuencias esperadas y frecuencias observadas por revista para <i>Mexico</i> .....	101
Tabla 10. Cálculo de frecuencias esperadas y frecuencias observadas por revista para <i>Mexiko</i> .....	102
Tabla 11. Cálculo de frecuencias esperadas y frecuencias observadas por revista para <i>Mejico</i> .....	103
Tabla 12. Cálculo de frecuencias esperadas y frecuencias observadas por revista para <i>Mejiko</i> .....	104
Tabla 13. Cálculo de frecuencias esperadas y frecuencias observadas por revista para <i>Neuspanien</i> .....	105
Tabla 14. Cálculo de frecuencias esperadas y frecuencias observadas por tipología textual para <i>Mexico</i> .....	107
Tabla 15. Cálculo de frecuencias esperadas y frecuencias observadas por tipología textual para <i>Mexiko</i> .....	108
Tabla 16. Cálculo de frecuencias esperadas y frecuencias observadas por tipología textual para <i>Mejico</i> .....	109

Tabla 17. Cálculo de frecuencias esperadas y frecuencias observadas por tipología textual para <i>Mejiko</i> .....	110
Tabla 18. Cálculo de frecuencias esperadas y frecuencias observadas por tipología textual para <i>Neuspanien</i> .....	111

## INTRODUCCIÓN

El estudio de la toponimia ha sido de gran importancia desde la antigüedad. La necesidad de un país o grupo poblacional de nombrar los espacios responde a necesidades de habitación, comercio, tránsito y regulación social. Esto resulta relevante desde una perspectiva histórica y antropológica, ya que permite saber cómo se han resignificado y transformado ciertos lugares. Esta necesidad de nombrar los espacios está presente hasta nuestros días en los que la cartografía juega un papel fundamental en actividades informativas, económicas, políticas y de investigación.

Hay numerosos factores que pueden complicar el estudio y uso de los conocimientos de la toponimia, por ejemplo, la falta de datos o la coexistencia de datos aparentemente incongruentes entre sí. Este último punto es bastante frecuente cuando el estudio de la toponimia de cierta región se realiza desde distintos lugares y, consecuentemente, distintas lenguas, resultando en un estudio de toponimia comparada.

Es sabido que los nombres de lugares no son únicos ni universales. El nombre de un país o región puede tener variantes nominales en cada lengua distinta a la del grupo que habita y nombra su espacio. Esto puede representar un problema porque muchas veces no se tiene conocimiento de las variantes nominales que tiene un mismo lugar, lo que resulta en la inaccesibilidad a la información disponible acerca de ese lugar en otras lenguas o incluso dentro de una misma.

En México se puede encontrar esta problemática. El topónimo que en español mexicano denominamos *México* se encuentra representado

gráficamente como “Méjico” en otros países de habla hispana, *Mexiko* en países de habla alemana, *Le Mexique* en países francófonos y de muchas otras maneras en países de hablas y sistemas de escritura distintos.

El hecho de que un nombre tenga distintas escrituras y variantes nominales no es sorprendente, se trata de procesos de adaptación e incorporación de palabras a sistemas lingüísticos con características y reglas distintas a las de la lengua de la que provino; sin embargo, sí puede ser más sorprendente, aunque también reflejo de la variación que caracteriza a cualquier lengua, que un nombre tenga distintas variantes nominales dentro de una misma lengua, región y periodo. Este es el caso de la toponimia mexicana incorporada al sistema lingüístico del alemán. Podemos encontrar en diversos textos con características similares al topónimo que designamos en español mexicano como “México” de distintas maneras: *Mexico*, *Mexiko*, *Mejico*, *Mejiko*, entre otras.

Esto ha probado ser un problema durante la búsqueda bibliográfica y es un problema que, en su acepción más general, hace inaccesible la información de lugares para hablantes de lenguas distintas. Frecuentemente un país cuenta con información acerca del desarrollo industrial de otro país, mientras que un tercero cuenta, más bien, con información económica y otro cuarto con información relacionada con los recursos minerales, todas informaciones de utilidad para fines en conjunto, pero que son inútiles separadas. Estamos frente a un problema típico de la traducción, por una parte, y de las tecnologías de la información, por otra.

Son necesarios los conocimientos traductológicos relacionados con la adaptación de nombres propios, tales como los topónimos, de un sistema lingüístico a otro y de las condiciones que provocan o favorecen ciertos procesos

traductores sobre otros; también son necesarios los conocimientos de las tecnologías de la información para poder recabar, ordenar y analizar los datos para llegar a conclusiones generalizables acerca de estos procesos traductores.

Para poder hacer aproximaciones generales al fenómeno de la multiplicidad de nombres propios para un objeto geográfico es necesario hacer una investigación que se aboque de manera particular a esta problemática. Para esto se ha hecho una investigación basada en corpora que analice la multiplicidad de nombres propios de la toponimia mexicana en lengua alemana durante el periodo de 1804-1848. Se ha compilado un corpus con características homogéneas para demostrar que hay múltiples factores que intervienen en la coexistencia de las variantes nominales de la toponimia mexicana en alemán y para poder analizar estas variables en un espacio controlado.

Se ha utilizado el lenguaje de programación Python, que hace más rápido y eficiente el análisis de los datos del corpus, y se ha propuesto una metodología de investigación basada en corpora usando esta herramienta para demostrar la utilidad de herramientas computacionales para la investigación lingüística, ya que frecuentemente se recurre en esta área al uso de conteos manuales, que conllevan una gran cantidad de tiempo, y al uso de otros lenguajes de programación como R que, pese a su utilidad estadística, obliga a los investigadores a recurrir a herramientas adicionales para la recabación de datos y su graficación.

Las preguntas de investigación, por lo tanto, son: ¿cuáles son los topónimos y variantes nominales de ellos a los que hacen referencia los autores alemanes en los textos coleccionados del corpus constituido?, ¿cuáles son las motivaciones de las variaciones internas y externas de los topónimos? y ¿cómo

pueden localizarse todas las entradas de un topónimo y sus respectivas variantes?

Relacionado con las preguntas, el primer objetivo de esta investigación es explicar el fenómeno de la multiplicidad de variantes nominales de la toponimia mexicana en alemán, por una parte, y demostrar la utilidad para estos objetivos y para la investigación lingüística basada en corpora del uso del lenguaje de programación Python, como segundo objetivo.

Con ello se espera hacer una aportación al acceso a la información en varias lenguas usando herramientas computacionales para ampliar el alcance de datos y reducir el tiempo necesario para su obtención y análisis.

## MARCO TEÓRICO

En este capítulo se presentan las propuestas teórico-metodológicas que sustentan esta investigación y que han permitido contar con un marco conceptual para tratar los datos del corpus de nuestro estudio. Primero se expone la onomástica como área específica de conocimiento. En seguida las definiciones del *nombre propio* y del *topónimo* en este mismo marco. Inmediatamente se expondrá el concepto de *alonomia* y se le pondrá en relación con el topónimo. Posteriormente se relacionarán los conceptos de *alonomia* y de *traducción*. Se presentará una tipología de los posibles cambios que pueden sufrir los topónimos al pasar de una lengua a otra. Después se expondrán los posibles cambios que pueden sufrir los topónimos. Se definirá la relación de esta investigación con la lingüística de corpus y la lingüística computacional. Finalmente se expondrán los fundamentos del procesamiento de lenguaje natural con el lenguaje de programación Python que permiten el procesamiento de textos para reconocer en ellos los topónimos y sus variedades nominales.

### 1.1 LA ONOMÁSTICA

De acuerdo con Hough (2016, p. 22-23), la onomástica es el estudio de los nombres. Los nombres son un universal lingüístico, presente en todas las lenguas. En el mundo occidental, desde la antigüedad, el estudio de los nombres ha estado ligado al estudio de sus referentes, pero recientemente estos estudios se han orientado a la investigación del origen de los nombres con énfasis etimológico (como Mentz, 2017, y Pocklington, 2022). Para esto, se han hecho

investigaciones desde diferentes perspectivas: etimológicas, sociolingüísticas, semánticas, etcétera.

Estas investigaciones suelen priorizar el estudio de un tipo de nombre en particular: ciencias como la arqueología y geografía se enfocan principalmente en el estudio de nombres de lugares (topónimos), mientras que la antropología y genealogía en los nombres de personas y los estudios literarios por nombres de personas o personajes.

## 1.2 EL NOMBRE PROPIO

Dentro de la onomástica, así como desde la descripción y análisis gramatical, se clasifican las distintas clases de palabras en adjetivos, nombres (también llamado sustantivos), verbos, adverbios, entre otros, siguiendo las reglas de la gramática de la lengua. A su vez, los sustantivos pueden clasificarse en abstractos y concretos; un ejemplo de un sustantivo abstracto es la palabra *geografía* y de uno concreto, *montaña*. El nombre propio entra dentro de esta última categoría y se define como una expresión lingüística que se refiere, idealmente, a un objeto preciso en el mundo.

Es por esto que los nombres propios se consideran monoreferentes e inherentemente definidos y, en general, cumplen la función de identificar (Nübling, 2015, pp. 17-20). Un nombre propio como *Hidalgo* no tiene descripción, a diferencia de su contraparte *hidalgo* como nombre común, que sí la tiene, y está ligado directamente a un referente, un estado de México; este nombre permite identificar este objeto geográfico e identificarlo entre los otros 31 estados del país (Fig. 1).



Los nombres propios se consideran nombres con una referencia directa, ya que puede reconocerse en ellos la función de representar a un objeto rápidamente, directamente y unívocamente. Aunque puedan conservar estructuras léxicas, los nombres propios no sirven para la descripción de un objeto porque, aun cuando parezca ser el caso, como, por ejemplo, en Río Grande, los nombres propios generalmente están desemenatizados y no tienen relación con su referente (pp. 18 y 27-28), como se puede observar en la Figura 1.

En los casos de nombres con estructuras léxicas identificables es posible encontrar una motivación en el mundo real para el nombramiento del objeto, aunque frecuentemente estas partes apelativas en la toponimia mexicana, particularmente, se encuentran ya obscurecidas de tal manera que un nombre como *Coyoacán*, que en náhuatl es “lugar de dueños de coyotes”, en realidad, ya no corresponde con la motivación que le dio lugar, empero el objeto geográfico conserva el nombre.

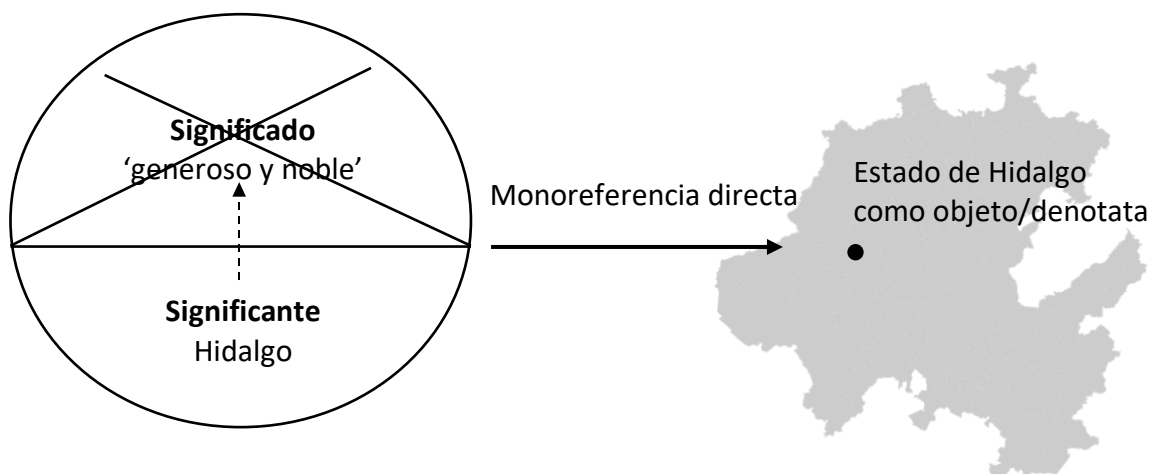


Figura 1. Monoreferencia directa del nombre propio ejemplificado con Hidalgo (basado en el diagrama de Nübling (p. 18))

### 1.3 EL TOPÓNIMO

La clase de nombre propio que más interesa en esta investigación es la del topónimo. El topónimo se refiere a objetos de la superficie de la tierra o del universo y son generados por los usuarios de la lengua en función del ambiente en el que se encuentran; a medida que la relación de los hablantes con el ambiente cambia, puede cambiar también la función y el nombre mismo del objeto, por ejemplo en renombramientos del periodo independentista, anterior al cual se encontraban nombres como *Paso de la Fábrica*, que después fue cambiado por *Minatitlán* para reflejar los cambios políticos resultados de la independencia. Es por esto que los topónimos pueden ser considerados como artefactos culturales que surgen de la relación entre la lengua y el ambiente (Hough, 2016, 86).

#### 1.3.1 EL SEMANTISMO DE LOS TOPÓNIMOS

Se ha postulado que existen dos clases de topónimos: los topónimos específicos o nominativos, que solo funcionan como topónimos, como *México*, *Alemania*, *Oaxaca*, y los topónimos genéricos o descriptivos, que han pasado a la toponimia desde una originaria naturaleza como apelativos, es decir, nombres comunes, p.e. *Villanueva*, *Mineral del Chico*, *Real del Monte* (García Sánchez, 2011, p. 178-9).

Los topónimos genéricos o descriptivos, o las partes del topónimo que lo sean, se consideran formalmente léxico común y funcional, en el que pueden operar procedimientos morfológicos y sintácticos como cambios en morfemas de género y número, procesos derivativos y compositivos y formación de perífrasis léxicas (p. 179), como en *Villanovense*.

Sin embargo, el topónimo, en tanto nombre propio, no cuenta con un significado, y no debe confundirse la evocación del significado del topónimo, o del componente apelativo de este, con la referencia, ya que el significado es algo propiamente lingüístico y que solo se encuentra en el plano del contenido, mientras que la referencia, el designado o el portador del nombre, corresponde con la realidad extralingüística. El nombre puede tener un valor connotativo, pero no necesariamente satisfacerlo, como en el nombre del estado de *Hidalgo*, en el cual no necesariamente hay más hidalgos; puede tratarse, más bien, de un valor simbólico que, en este caso, alude a un personaje de la historia de México.

Conforme a lo anterior, un topónimo transparente como *Aguascalientes*, aunque esté compuesto por elementos claramente identificables, está al mismo nivel que un topónimo opaco como *Zacatecas* (p. 180).

Si bien hay muchos topónimos, especialmente microtopónimos (lugares geográficos deshabitados y menos prominentes en cuanto a su extensión o tránsito), que cumplen su función designativa como nombres propios sin diferenciarse de los apelativos de los que proceden –lo que crea la apariencia de que significan– son, en realidad, los apelativos los que significan y meramente coinciden con los topónimos que han constituido y de los que se han separado, resultando en un caso de homonimia entre ambos nombres, como *La estancia*, localidad en Sonora, y *la estancia*, acción de estar o permanecer cierto tiempo en un lugar.

Los topónimos solamente significan en la medida en la que “se contagian” por los apelativos, mismos que evocan o connotan su significado. Los topónimos también pueden significar cuando los hablantes están interesados en poner en

relación los topónimos con los apelativos, pero esta es una cuestión más pragmática que de norma lingüística (p. 181-2).

El apelativo, en realidad, deja de significar en el momento en el que se constituye como topónimo. Al hacerlo se llevan a cabo los procesos de dessemantización y de duplicación onomasiológica, ya que el apelativo mantiene su función y categoría léxica tras crearse el topónimo. Lo mismo sucede cuando el topónimo procede de un antropónimo o de otro topónimo (p. 181), como en *Guanajuatillo*, que no es *Guanajuato* pequeño, sino una localidad en Michoacán de Ocampo.

Volviendo al topónimo nominativo, la opacidad de este tipo de nombres debe ser reconocida como el resultado de la distancia que puede haber entre el nombre y su proceso denominativo. Los nombradores nombran los espacios basándose, frecuentemente, en las características geográficas, orográficas, hidrográficas, en los animales, plantas o grupos de personas que los habitan, la actividad económica o cultural que se realiza ahí, entre otras (Mentz, 2017, pp. 9-10), por lo que podemos afirmar que hay una motivación en el acto de nombrar un objeto geográfico.

La motivación para el nombre del lugar puede verse opacada por las transformaciones sufridas por el espacio nombrado, en el topónimo mismo, o en los habitantes del espacio geográfico. Estas transformaciones son de interés para los investigadores de los nombres de lugares, ya que el reconocimiento de las causas de estos cambios permite reconstruir la historia de los espacios nombrados y de sus habitantes, lo que es de gran utilidad para las investigaciones de corte histórico, antropológico, geográfico, entre otras (p. 52).

### 1.3.2 EL ESTUDIO DE LOS TOPÓNIMOS

Pocklington distingue tres áreas en la toponomástica (el estudio de los topónimos): la toponomástica pura, la investigación acerca de los aspectos universales de la toponimia como el origen, la naturaleza y su comportamiento; la etimología toponímica, la reconstrucción de la historia de cada topónimo; y la toponomástica aplicada, el estudio de la utilidad y validez de la aplicación de los resultados en otros campos (Pockington, 2022, p. 2).

En los apartados anteriores se han dado algunas de las características generales del topónimo que se enmarcan en la primer área de la toponomástica; a continuación se muestran las características de los estudios de etimología toponímica.

#### 1.3.2.1 ETIMOLOGÍA TOPONÍMICA

Los topónimos no tienen significado, pero sí motivación. La motivación es la conexión entre el o los apelativos que dieron lugar al topónimo y el referente. Puede decirse que los topónimos significan diacrónicamente, pero que el significado apelativo de las palabras que lo conforman puede haberse opacado o, bien, permanecer transparente (García Sánchez, 2011, p. 182), por ejemplo, *Puebla* y *Aguascalientes* respectivamente.

Como se ha mencionado, los topónimos también pueden provenir de antropónimos y de otros topónimos. Esto quiere decir que la motivación para designar un objeto con un topónimo no se encuentra exclusivamente en los apelativos y su significado con respecto al referente, sino que también en las relaciones, más de carácter histórico, entre esos nombres propios y los lugares que se denominan con ellos (p. 182-3).

Aunque desde una perspectiva sincrónica parece que los topónimos, como las palabras en general, son arbitrarios y opacos, desde una perspectiva asincrónica puede considerarse que todos tienen una motivación, aunque esta ya esté oculta. Para entender un topónimo (su motivación, cuándo fue acuñado y por quién) es necesario recolectar información tanto de la lengua como del ambiente del objeto designado con el topónimo.

El ambiente se estudia al visitar el sitio, conversar con los locales y al examinar mapas y registros; la lengua se estudia al recolectar formas primitivas y sus contextos, aunque también la conversación con los locales puede ser de importancia.

Muchos topónimos comienzan siendo parte del lexicón, la parte ordinaria del vocabulario, para después ser parte del onomasticón, el nombre técnico para la parte de la lengua que consiste solo en nombres propios. Estos son elementos de la lengua que son frecuentemente intraducibles, cuyo significado como palabras no es relevante para la comunicación cotidiana, aunque esto no significa necesariamente que estén completamente vaciadas de significado semántico (Hough, p. 86) o, como se ha mencionado, que no tengan un valor connotativo o simbólico.

El estudio de las formas primitivas es un punto de partida en el estudio de la toponimia. Las dos preguntas más importantes son: ¿qué significa el topónimo? y ¿de qué lengua proviene? El responder estas preguntas permite revelar la historia del lugar y del área en la que se encuentra, lo que frecuentemente lleva a periodos pre-documentales (p. 86-7).

Aunque las formas primitivas son una parte fundamental del estudio de la toponimia, el estudio de un topónimo no concluye con la respuesta de estas dos

preguntas. Los topónimos no son unidades lingüísticas inertes producto de cierto lapso, sino que están en constante cambio, reflejo de nuestra relación cambiante con el ambiente.

Mientras que todos los topónimos cuentan con detalles espaciales que pueden ser expresados en términos de latitud, longitud y altitud, las fuentes de un nombre particular pueden variar entre fuentes orales (registradas con un sistema de grabación de sonido o a través de notas del investigador) o fuentes escritas (como textos publicados o manuscritos antiguos). En este sentido, el estudio de la toponimia implica una investigación de carácter histórico que involucra la revisión de fuentes antiguas y modernas, publicadas y no publicadas, escritas y orales (p. 88).

Una fuente escrita, aunque puede ser una fuente valiosa que da testimonio de los nombres propios, no puede ser la única fuente para considerar en el estudio de la toponimia. Esto debido a que, en la transcripción, sea de un escriba copiando el trabajo de otro o para fines editoriales y de publicación, los nombres propios, y especialmente los topónimos, son las unidades lingüísticas más proclives a los errores y la variación. Por ello cada fuente debe ser tratada individualmente, pero también puesta en relación con otras fuentes.

Una fuente oral, pese a la dificultad de su registro, cuenta con propiedades valiosas que no se encuentran generalmente en las fuentes escritas, como la pronunciación. La pronunciación puede ser un elemento fundamental para el análisis correcto de un topónimo, ya que permite desambiguar la lengua o lenguas que le dieron lugar a partir del reconocimiento de las propiedades de acentuación o de las propiedades fonotácticas de las lenguas. Cuando hay más de una pronunciación para un nombre, es importante caracterizar los grupos de

hablantes que se distinguen a partir de esto. Idealmente, se reportan las pronunciaciones usando el Alfabeto Fonético Internacional (p. 96-7).

Por otra parte, las fuentes orales, principalmente los locales,, los habitantes del sitio nombrado, pueden proporcionar información acerca de microtopónimos poco conocidos y no registrados en fuentes escritas (p. 89-90), tales como nombres de campos, rocas y, de especial interés para esta investigación, nombres de minas, montañas y demás lugares importantes para la minería y ciencias de la tierra.

En ambos casos, es importante considerar el contexto, ya que puede ser en ocasiones tan importante como el topónimo mismo. El contexto puede proporcionar evidencia que permita identificar y desambiguar lugares con el mismo nombre. Es especialmente importante en este proceso contar con la información de la tenencia, es decir, la información acerca de quién poseía el objeto nombrado, quién lo concedió y a quién (p. 91).

En esta investigación la importancia de la tenencia se refleja en el hecho de que una mina o algún otro lugar relacionado con los procesos de extracción, refinamiento y distribución de materiales pudo haber tenido varios dueños, mismos que ocuparon un nombre, solo para que el siguiente lo cambiara o modificara, voluntaria o involuntariamente, este último siendo el caso más común cuando con el cambio de tenencia también hubo un cambio de la lengua con la que se nombraba el objeto.

### 1.3.2.3 TOPONOMÁSTICA APLICADA

Como se ha mencionado en apartados anteriores, los estudios de la toponimia pueden ser de utilidad para otros campos. El conocimiento de los cambios en los



topónimos sirve como herramienta para investigaciones históricas, geográficas, antropológicas, arqueológicas, entre otras, que buscan reconstruir el pasado de un lugar o un grupo de personas.

Son los estudios con énfasis etimológico los que más sirven para estos fines, sin embargo, los objetivos de esta investigación no son ni hacer una reconstrucción de la historia de México a partir de su toponimia, ni aclarar el significado de los topónimos que la comprenden.

El estudio de la toponimia mexicana en esta investigación tiene como propósito el reconocimiento de los nombres de lugares en circulación durante el periodo del corpus en ambas lenguas para identificar los casos de alonimia (la multiplicidad de nombres para un mismo objeto), los tipos de alonimia y sus causas.

Para esto es necesario primero ahondar en algunas de las subclases del topónimo para poder pasar después a las características particulares de la toponimia mexicana.

#### 1.3.2.4 SUBCLASES DEL TOPÓNIMO

El topónimo puede clasificarse dicotómicamente en cuerpos localizados en la tierra (geónimos) y cuerpos localizados en el cosmos (cosmónimos). Nombres relacionados a objetos terrestres (orónimos) y nombres relacionados a cuerpos de agua (hidrónimos). En macro- y microtopónimos y en nombres relacionados a objetos habitados (oicónimo) y no habitados (anoicónimo) (Nübling, 2015, p. 206).

Las subcategorías del topónimo son numerosas. En este trabajo solo se contemplan para su localización automática los topónimos que hacen referencia

a zonas terrestres de México, tales como ciudades, regiones y montañas, es decir, geónimos relacionados con objetos terrestres como los orónimos, oicónimos y anoicónimos, ya que estas fueron las más significativas para los estudios de minería y ciencias de la tierra.

#### 1.3.2.1.1 OICÓNIMOS

Los oicónimos (Nübling, 2015), polinóminos (Iglesias, 1991) o nombres de lugares habitados son el nombre de lugar prototípico en el que se designa un área habitada por una comunidad. Ya que los nombres pueden usarse sin reparo a su posible significado semántico, un grupo de hablantes que entra al espacio nombrado puede retomar el nombre del lugar o crear un nombre nuevo para él (Hough, p. 103).

Los oicónimos pueden dividirse estructuralmente en nombres descriptivos y no descriptivos. Los oicónimos descriptivos pueden ser frases nominales compuestas por un solo núcleo “genérico”, cuyos tres tipos de referentes más comunes son el ambiente natural, el ambiente construido, o los habitantes de estos, por ejemplo, *Las piedras*, *Las Torres* y *Los Mártires*. Sin embargo, la mayoría de los oicónimos son nombres compuestos que combinan un elemento genérico con un modificador específico, como *Tlalpan*, “sobre tierra firme”, *Iztapalapa*, “adoquinado sobre el agua” y *Mexicalzingo*, “lugar donde habitan los mexicas distinguidos”. Además, tanto los oicónimos simples como los compuestos pueden contener un afijo que sirve para diferenciar entre nombres de lugares homónimos o para indicar posesión o hacer una descripción del lugar, como *Pueblo viejo de Tampico*. El orden de los elementos es determinado por las reglas del sistema lingüístico de la lengua (Hough, p. 104).

Los oicónimos no descriptivos son aquellos cuyo nombramiento evoca una asociación con un evento histórico o legendario, o con una persona, caso abundante en la toponimia mexicana (*Minatitlán, Allendetitlán, Morelos, Hidalgo, Quintana Roo*, entre otros). Este último caso es el más frecuente durante el periodo colonial europeo desde comienzos del s. XV. Estos nombres conmemorativos abundan también en México y pueden encontrarse en la forma de nombres de lugares con el nombre de exploradores y colonizadores europeos, nombres de lugares existentes ya en Europa que fueron transferidos a los territorios conquistados (frecuentemente diferenciados por el afijo *Nuevo/New*), y nombres de lugares que aluden a eventos relacionados con la era de exploraciones europeas (p. 108), como *Nuevo Santander* y *Veracruz*.

#### 1.3.2.1.2 ORÓNIMOS

Los orónimos o nombres de montañas son nombres raramente registrados o registrados tardíamente debido a su poca productividad económica en lo general, aunque en el caso de México no ha sido de esta manera. El interés en nombrar los cuerpos montañosos parece ser resultado del montañismo como actividad recreativa en registros europeos a partir del s. XVI; en el caso de México, en el corpus compilado para esta investigación, el interés parece recaer completamente en la investigación científica y la explotación minera.

El registro de los orónimos tampoco ha sido uniforme, siendo los cuerpos montañosos bajos y cercanos a poblados habitados los primeros en ser nombrados por su potencial de tránsito humano o de alguna otra actividad económica, mientras que las cúspides y puntos más altos permanecieron sin

nombre por su poca relevancia para las actividades económicas (Hough, p. 131-2).

Una característica importante de los orónimos, frente a los hidrónimos y oicónimos, es la multiplicidad de nombres para un mismo cuerpo montañoso sin tratarse de traducciones de un nombre original o de adaptaciones para incorporarlo a otro sistema lingüístico, por ejemplo Nauhcampatépet y *Cofre de Perote*. Este tipo de polimorfismo, llamado *alnimia* por Otto Back y que en el apartado de Alonimia será abordado a profundidad, se debe a que un mismo cuerpo montañoso puede tener diferentes grupos de habitantes en sus distintos lados, cada uno de estos grupos con un nombre distinto para el cuerpo montañoso, en distinta lengua o dialecto, o que muestre una diferente percepción del cuerpo desde su respectivo ángulo. Estos nombres pueden coexistir por mucho tiempo, hasta que se llegue a un consenso sistemático de cartógrafos que eligen qué nombre poner en el mapa (pp. 132-3).

De la misma manera que los oicónimos, los orónimos pueden componerse por partes genéricas o específicas, siendo las primeras las que indican el tipo de cuerpo montañoso, y las específicas las que lo describen, como *Chapultepec*. En este último grupo se encuentran como elementos más comunes el color, la forma y el tamaño. Los nombres conmemorativos también son muy comunes en los cuerpos montañosos y están relacionados con las prácticas de montañismo (pp. 137-9).

#### 1.3.2.1.3. MICROTOPÓNIMOS

Un tercer tipo de topónimo que es importante definir para esta investigación es el del *microtopónimo* o *nombre rural*. Estos nombres hacen referencia a lugares

inhabitados, en zonas rurales o adyacentes a ellas. Los microtopónimos hacen referencia frecuentemente a nombres de acres, campos, viñedos, campos, bosques, entre otros (Hough, p. 151). A esta lista podríamos agregar, para el caso particular de México, nombres de minas, haciendas de amalgamación, entre otras zonas no habitadas y solo destinadas para la minería, frecuentemente lugares con nombres de santos o elementos religiosos como *Sta. Rosa*, *St. Pablo* y *Mina del Christo*.

Otra definición de microtopónimos, basada en el tamaño y no en calidad de habitado o no, puede incluir nombres de calles, cuerpos de agua y cuerpos montañosos (p. 152). En esta investigación se considera la definición de microtopónimo que atiende a la calidad de habitado o no.

Los microtopónimos tienen poco valor comunicativo, ya que el conocimiento del nombre de estos lugares o de los lugares mismos puede ser de pocas personas. Los microtopónimos son en lo general lugares deshabitados, pero pueden tener una conexión con lugares que sí estén habitados (p. 152).

Una de las principales motivaciones de los microtopónimos es la orientación e identificación de áreas dentro de un área mayor que puede ser subdividida. La base de este tipo de nombres son frecuentemente apelativos o genéricos que designan áreas que son usadas en contextos agrónomos o, en el caso de esta investigación, mineros, como *Veta grande*. A partir de esto puede hacerse una distinción entre nombres naturales y nombres culturales. Los nombres naturales hacen referencia a las características del lugar, mientras que los culturales a la influencia o actividad humana del lugar nombrado, *Peñón Blanco* y *Estanzia de las Vacas*. Los nombres naturales, además, pueden reflejar

condiciones climáticas, la existencia de agua, animales, plantas o árboles, así como las condiciones del suelo.

Es difícil hacer un corte definido entre nombres culturales y naturales, ya que no es raro encontrar nombres compuestos que contengan elementos que aludan a ambos tipos de referentes. Otra característica de los microtopónimos, y una que está también ampliamente presente en el corpus de esta investigación, es el uso de nombres indirectos. Los nombres indirectos son aquellos que designan un lugar a través de la referencia de otro, frecuentemente a través de preposiciones de lugar, como *Mineral del Chico*.

Morfológicamente, los microtopónimos están formados principalmente por apelativos y morfemas derivativos que pueden indicar colectivos, diminutivos o aumentativos y la ubicación del lugar, como *Estanquillo*. Otro patrón morfológico común es el de la composición, que puede resultar de la operación de nombre más nombre, nombre más adjetivo, nombre más numeral, o la combinación de más de dos elementos. Los elementos del nombre pueden o no escribirse como una sola palabra, es decir, sin espacios entre los elementos. Además, pueden contar con preposiciones o artículos (pp. 155-7).

Los microtopónimos pueden ser problemáticos durante la investigación de la toponimia de una zona debido a su poca productividad económica, ya que provoca que no sean considerados como lugares relevantes para las actividades económicas humanas y, por lo tanto, son raramente registrados. La mayoría de los documentos históricos impresos hacen referencia solo a oicónimos, y los registros de microtopónimos deben ser extraídos de fuentes escritas, como se ha hecho en este trabajo (p. 159).

## 1.4 ALONIMIA

La *alonymia* es el fenómeno central de esta investigación. Este término es un calco del alemán *Allonymie* (Back, 2002) y se profundiza en él en los apartados siguientes.

### 1.4.1 DEFINICIÓN

Otto Back (2002, p. 3-4) define este fenómeno como “la diversidad de nombres propios para un objeto”. Back precisa que la alonimia puede ser “interlingual” cuando se contemplan los nombres de un mismo objeto desde más de una lengua. Un ejemplo de esto es el nombre del topónimo que hoy denominamos en español *México*: *México* es el nombre propio del objeto en español, el endónimo (nombre de un objeto en la lengua donde está localizado geográfica y políticamente), pero *Mexiko* es el nombre propio convencionalizado del objeto en alemán, el exónimo (el nombre extranjero correspondiente).

El anterior no es un caso aislado. Los textos del corpus constituido para esta investigación hacen referencia a diversos topónimos de México. Este corpus está acotado por tiempo (s. XIX), lugar (México), tema (minería y ciencias de la tierra) y lengua (alemán), sin embargo, pueden encontrarse distintos nombres que hacen referencia al mismo objeto. Las diferencias entre los distintos nombres son, principalmente, de carácter ortográfico. Un ejemplo muy conocido y documentado por su relevancia para la minería en México es *Guanajuato/Guanaxuato*, aunque no es el único topónimo con un cambio mínimo donde solo un grafema puede cambiar. Otros ejemplos son *Tlalpujahuá/Tlalpujaua*, *Popocatepetl/Popokatepetl* y *Real de Catorze/Real de Catorce*.

Otros autores han tratado superficialmente el tema. Uno de ellos es García Sánchez, quien lo llama sinonimia en toponimia. Se aclara en las exposiciones acerca de este fenómeno que la sinonimia no debe entenderse como una relación entre significados, sino entre los significantes. Son, entonces, sinónimos de topónimos los significantes que pueden aplicarse al mismo referente, también las perífrasis explicativas o expresiones descriptivas que sirven para identificar un lugar, como *la perla tapatía*, así como los endónimos y sus correspondientes exónimos (2011, pp. 185-6).

#### 1.4.2 SUBCLASIFICACIONES DE LA ALONIMIA

##### 1.4.2.1 ALONIMIA SUPERFICIAL O FONÉTICO-GRÁFICA

En los ejemplos dados al principio de este apartado puede observarse el fenómeno de la alonimia superficial o fonético-gráfica. Este tipo de alonimia surge de la necesidad de hacer una “adaptación mínima” de los rasgos fonéticos o gráficos del endónimo al sistema lingüístico de la lengua meta (Back, 2002, p. 23, 49).

En el caso del par *México/Mexiko*, este último el exónimo más empleado en el corpus, es difícil hacer una aproximación a las propiedades fonéticas del exónimo alemán en este periodo histórico (aunque en la actualidad es indudable que hay una adaptación del sonido fricativo velar /x/ a /ks/, ya que el alemán cuenta con los recursos fonéticos para representar /x/ como ‘ch’); sin embargo, la <c> no es un grafema que pueda emplearse por sí solo en el alemán (Wermke *et al.*, 2009, p. 67), lo que lleva al uso del grafema <k> y al cumplimiento de la adaptación mínima comprobable en el plano gráfico.



#### 1.4.2.2 ALONIMIA DE VARIACIÓN DE RAÍZ

Después de la alonimia superficial o fonético-gráfica, Back describe la alonimia de variación de raíz (Back, 2002, p. 23). En este tipo de alonimia la raíz de la palabra se conserva, pero con modificaciones que permiten la operación morfológica de adición de sufijos con el valor semántico de lugar. Entre estos sufijos o terminaciones típicas con valor de lugar destacan en alemán *-land* (Deutschland, Finnland, England), *-reich* (Österreich, Frankreich, Griechenland), “-en”/“-ien” (Spanien, Ägypten, Syrien), y *-ei* (Mongolei, die Türkei, Slowakei). En el corpus solo se encuentra este tipo de alonimia en *Californien* y sus variedades con prefijación; otros ejemplos son *Bolivien*, *Argentinien* y *Kolumbien*, que están fuera del foco de esta investigación, y para microtopónimos, notablemente los que comienzan con *Hacienda*, en donde pueden observarse procesos morfológicos de declinación como en *Hazienden* (*Haciendas* en dativo plural).

#### 1.4.2.3 ALONIMIA DE CONCORDANCIA SEMÁNTICA

Un tercer tipo de alonimia es el de concordancia semántica (p. 23). En este tipo de alonimia las partes apelativas del nombre son transparentes tanto en la lengua origen como en la meta. Esto permite hablar de un proceso de traducción en el que se considera una traducción literal entre ambas lenguas, como el que sucede en *Messerberg* (*Messer*: cuchillo o navaja y *Berg*: montaña o cerro) y *Cerro de las Navajas*. En esta categoría se pueden incluir los casos de formas largas y cortas de un topónimo, mismos que son comunes tanto en los endónimos como en exónimos (Nübling, p. 209), por ejemplo, *Mexico* y *Estados Unidos Mexicanos* o *Vereinigten Staaten von Mexico*.

La variación de formas cortas y largas se caracteriza por la exclusión (forma corta) o inclusión (forma larga) de una parte apelativa que conlleva el estatus político o institucional del nombre geográfico. La parte apelativa es, naturalmente, la que se traduce frecuentemente en la lengua meta. La estructura sintáctica de los nombres largos puede ser aposicional, como en *República Mexicana/Republik Mexiko* o atributiva con un sustantivo o adjetivo subordinado, como en *Estados Unidos Mexicanos /Vereinigten Staaten von Mexiko*.

Ya que los topónimos tienen una función de referencia directa, cuando el conocimiento del objeto geográfico esté dado en el contexto se usa la forma corta del nombre con mayor frecuencia, aunque hay registros en los que es norma o es más común usar la forma larga (Nübling, p. 18). Ejemplos de esta categoría encontrados en el corpus son *Veta Grande/Hauptgang*, *Sierra Madre/Muttergebirge* y *Ciudad de México/México-Stadt*.

#### 1.4.2.4 ALONIMIA DE DIVERSIDAD DE PALABRAS

El cuarto tipo de alonimia descrito por Back, y también el que considera el autor como alonimia plena, es el de diversidad de palabras (2002, p. 23). En este tipo de alonimia no hay concordancia semántica, un equivalente natural entre ambas lenguas (*Montaña* y *Berg*, por ejemplo), ni fonético-gráfica. Ejemplos encontrados en el corpus son *México/Nueva España* o *Mexiko/Neuspanien*. Queda pendiente considerar si también puede considerarse como un mismo objeto geográfico *Tenochtitlán*, que, aunque comparte el mismo espacio geográfico, no el mismo tiempo.

## 1.5 LA ALONIMIA Y SU RELACIÓN CON LA TRADUCCIÓN

### 1.5.1 TRANSCRIPCIÓN Y TRANSLITERACIÓN

Un problema importante por considerar es que, mientras que en la actualidad los endónimos y exónimos cuentan con una escritura registrada en fuentes escritas (Back, 2006, p. 84), *México* y *Mexiko* respectivamente, en el periodo histórico de nuestro corpus, la primera mitad del s. XIX, no necesariamente es de esta manera.

Esto ha provocado una falta de certeza acerca de las técnicas de traducción de los autores alemanes para pasar del español (o de alguna lengua indígena) al alemán los topónimos. Pueden suponerse dos métodos: el de transliteración y el de transcripción.

Ambos métodos buscan representar texto de una lengua con el sistema gráfico de otra. Mientras que la transcripción parte de la forma hablada y considera la pronunciación de la forma escrita de la lengua de destino, la transliteración intenta reflejar la forma escrita (Bezoz, 2006, p. 149).

Tanto la transcripción como la transliteración pueden considerarse procesos traductores, ya que cumplen con la función de incorporar un elemento a un sistema lingüístico diferente, adhiriéndose principalmente o a las normas fonéticas o a las gráficas de este sistema, según las intenciones del proceso traductor.

En la siguiente tabla de Back (2006, p. 92) (Tabla 1) se distinguen ambos métodos, aunque es importante recalcar que en el s. XIX estas no contaban con una aplicación rigurosa:

	Transliteración	Transcripción	Ejemplo/explicación
<b>Reproducción de...</b>	Letras o signos silábicos	Fonemas	<i>Mexico y Mejico</i>
<b>Información acerca de...</b>	La escritura original	La pronunciación original	<i>Kwawnawak y Cuernavaca</i>
<b>Las reglas para la escritura son...</b>	Relativamente sencillas	Relativamente complejas	Reproducción de grafemas y reglas basadas en el API
<b>La creación de expresiones y aclaraciones para la pronunciación son...</b>	Frecuente- y relativamente complicadas	Relativamente sencillas	Soportadas en información externa al nombre y en el nombre mismo
<b>Propias de...</b>	Documentación científica	Publicidad, literatura	Literatura especializada y textos dirigidos a un público general
<b>Las fuentes son...</b>	Filológicas, normas bibliotecológicas, normas internacionales	Diccionarios e indicaciones acerca de las oraciones de la lengua meta	Textos científicos y textos informativos

Tabla 1: Tabla de transliteración y transcripción de Otto Back

### 1.5.2 MÉTODO DOMESTICANTE Y MÉTODO EXTRANJERIZANTE

En los dos procedimientos presentados en el apartado anterior puede reconocerse la dicotomía que está presente a lo largo de numerosas teorías de la traducción y que fue planteada por Cicerón, Friedrich Schleiermacher y otros autores como la oposición entre el método extranjerizante y el domesticante, que ha sido nombrada de distintas maneras a lo largo de los estudios de traducción desde el s. XX.

Antes de poder hablar del método traductor, es importante definir la equivalencia como noción central de cualquier proceso traductor. La equivalencia es la supuesta relación de igualdad semántica entre palabras o conceptos entre dos lenguas que permite la “sustitución” o “reemplazo” de los elementos de un texto en una lengua por los elementos correspondientes en otra (Pym, pp. 27-28).

Otto Kade (1968) ha hecho una clasificación de la equivalencia que considera cuatro tipos: equivalencia uno a uno (la palabra en la lengua de salida tiene un equivalente exacto en la lengua meta), equivalencia uno a varios o varios a uno (la palabra en la lengua de salida tiene varios posibles equivalentes en la lengua meta o al revés, hay varias posibles equivalencias en la lengua de salida para un solo equivalente en la lengua meta), uno a parte (una palabra en la lengua meta tiene solo un equivalente parcial en la lengua meta) y uno a ninguno (no hay equivalencia) (Pym, p. 29).

En el caso de los nombres propios podemos reconocer que hay dos posibles perspectivas frente al concepto de la equivalencia: el nombre propio cuenta, o puede contar, con un equivalente (como en *Hans* y *Juan*) y el nombre propio no cuenta ni puede contar con equivalentes en otras lenguas (*Guanajuato* deberá ser *Guanajuato* en los otros sistemas lingüísticos).

Alguna de estas dos perspectivas habría sido utilizada por los investigadores alemanes, lo que los habría llevado al uso metódico de procesos traductores en sus textos particulares o en su producción científica en general. A partir de este concepto de equivalencia es, por lo tanto, que se desarrolla la distinción entre los métodos traductores extranjerizantes y domesticantes.

El método traductor es la manera en que el traductor se enfrenta al conjunto del texto original y desarrolla el proceso traductor según determinados principios (Hurtado, 2001, p. 241). Se trata de un procedimiento global, cuyas manifestaciones se encuentran presentes a lo largo de todo el texto.

La distinción entre ambos métodos recae en la función que tomará el texto a traducir. El método extranjerizante se caracteriza por acercar al lector a la cultura del texto origen, mientras que el domesticante sigue el camino inverso,

acerca el texto a la cultura del lector. Venuti, quien retoma esta dicotomía, distingue entre el método apropiante (domesticante) y extranjerizante y los define, respectivamente, como la reducción etnocéntrica del texto extranjero a los valores culturales de la lengua de llegada, y como una desviación de los valores culturales locales para establecer las diferencias lingüísticas y culturales del texto extranjero (Hurtado, p. 243-4).

### 1.5.3 TÉCNICAS Y MÉTODOS TRADUCTORES

Ya que se trata, en su mayoría, de textos originales de autores alemanes, y no de traducciones, es importante hacer una distinción entre técnicas y métodos de traducción.

Mientras que el método traductor es ejecutado en todo el texto, la técnica traductora es el procedimiento verbal concreto, visible en el resultado de la traducción para conseguir equivalencias traductoras. A diferencia del método traductor, que es una opción global que recorre todo el texto y que afecta al proceso traductor y al producto, la técnica solo afecta al resultado y a unidades menores del texto (pp. 256-7).

Se han propuesto numerosas clasificaciones de las técnicas de traducción, pero orientadas principalmente al estudio de la traducción de textos y no de palabras aisladas. Una propuesta pionera que es de suma utilidad para reconocer dos grandes grupos de técnicas es la de Vinay y Darbelnet (1958).

En esta categorización de los “procedimiento técnicos de traducción” (Hurtado, p. 257), los autores distinguen siete procedimientos, mismos que están agrupados en procedimientos directos y oblicuos. Los procedimientos literales solo son posibles entre lenguas y culturas muy cercanas porque hay una

correspondencia tanto léxica como sintáctica. Los procedimientos oblicuos se dan entre lenguas en las que no hay esta correspondencia y, por lo tanto, no puede hacerse una traducción palabra por palabra, sino que se requieren procedimientos más complejos. Amparo Hurtado ha descrito estos procedimientos de la siguiente manera (pp. 257-8):

- Procedimientos de traducción literal:
  1. Préstamo: palabra incorporada a otra lengua sin modificaciones.
  2. Calco: préstamo de un sintagma extranjero con traducción literal de sus elementos.
  3. Traducción literal: la traducción palabra por palabra.
- Procedimientos de traducción oblicua:
  1. Transposición: el cambio de categoría gramatical.
  2. Modulación: el cambio de punto de vista, de enfoque o de categoría de pensamiento.
  3. Equivalencia: da cuenta de una misma situación empleando una redacción completamente distinta.
  4. Adaptación: cuando se utiliza una equivalencia reconocida entre dos situaciones.

En esta clasificación, aunque está concebida para textos completos, se encuentra también la distinción entre procedimientos literales que se caracterizan por la incorporación de palabras extranjeras tal como fueron encontradas en el original, y que tienen gran relación con el método extranjerizante, y los procedimientos oblicuos que se caracterizan por la adaptación del material extranjero a las condiciones del sistema lingüístico de la lengua meta, y que tiene gran relación con el método domesticante.

Es por lo anterior que propongo una nueva clasificación específica para las técnicas de traducción, que siga atendiendo la distinción entre método domesticante y extranjerizante, pero que también considere las técnicas particulares de traducción de nombres propios de la clasificación de alonimia de Back.

Las técnicas de traducción pueden ser, en esta propuesta de clasificación, extranjerizantes, usando el endónimo del objeto geográfico, o domesticantes, usando una de las posibles formas del exónimo: la técnica de adaptación fonético-gráfica, la de variación de raíz, la de concordancia semántica y la de diversidad de palabras.

Esta clasificación permitiría agrupar discretamente dos tipos de topónimos, los traducidos por técnica domesticante (*Mexiko*) y los de técnica extranjerizante (*México*), y contrastar su frecuencia de aparición en textos con características particulares según su año, autor, fuente de publicación y tipología textual.

Otro factor que habría de ser influyente en la categoría del topónimo es si la lengua del endónimo es español o una lengua indígena. Se espera que los topónimos de lenguas indígenas, por su mayor opacidad frente al español, – lengua europea como el alemán– se encuentren principalmente como endónimos o como topónimos traducidos con la técnica de adaptación fonético-gráfica, como puede verse en los topónimos *Mexiko*, *Zakatecas*, *Yukatan*, entre otros.



#### 1.5.4 FUNCIÓN Y TIPOLOGÍA TEXTUAL

Determinar las posibles funciones que las técnicas de traducción de topónimos habrían de cumplir es un reto. Esto porque es necesario observar más allá de la unidad de análisis y hacia su cotexto y contexto también, lo que nos aleja de la unidad más acotada de la técnica de traducción empleada en un topónimo concreto.

Es necesario identificar el público al que está dirigido, lo que generalmente se explicita en las introducciones, cuando se trata de un libro o en los comentarios del editor, cuando se trata de un artículo en una revista. También es necesario identificar la tipología a la que se adscribe el texto, porque pueden ser las características mismas de este las que ayuden a determinar la función general que habrían de cumplir los textos y, consecuentemente, la posible intención que habrían tenido los autores al utilizar determinada técnica de traducción.

Los textos del corpus son homogéneos en su tema, pero heterogéneos en su género textual, e incluso un mismo texto puede contar con secuencias textuales de diferentes tipos textuales, como ya lo ha propuesto J.M. Adam. Es importante determinar la secuencia textual dominante de cada texto para poder establecer una relación entre el género, y por lo tanto función del texto, y el proceso de traducción de los topónimos del corpus. Para esto se ha considerado la tipología textual propuesta por Adam, que comprende textos narrativos, descriptivos, argumentativos, explicativos y dialógicos (Adam, 1992), esto en razón de que su clasificación, en comparación con otras propuestas como la de Sandig, Werlich y Grosse, presenta categorías mejor diferenciadas entre sí, además de considerar las aportaciones de los teóricos de la lingüística textual

mencionados (Simón, 2006, pp. 156-161). Las características de los tipos de textos se resumen en los siguientes apartados.

#### 1.5.4.1 TEXTOS NARRATIVOS

Los textos narrativos contienen una sucesión de eventos en el tiempo. La información se organiza siguiendo una cronología y utilizan frecuentemente oraciones temporales, adverbios, conjunciones de tiempo, verbos en pretérito o el uso del presente histórico. Los cuentos, películas, chistes, novelas, fábulas y noticias periodísticas son ejemplos de textos con una secuencia textual predominantemente narrativa.

#### 1.5.4.2 TEXTOS DESCRIPTIVOS

Los textos descriptivos informan acerca del estado, presente o en otro tiempo, de una persona, objeto o evento. Adam distingue una serie de operaciones básicas para la descripción: la operación de anclaje, que señala lo que será descrito, generalmente a través del uso de su nombre; la operación de aspectualización, que presenta los elementos más importantes del objeto descrito; la operación de puesta en relación, que puede hacer una comparación por contigüidad o por analogía; y la operación facultativa de subtematización, que toma un aspecto del objeto como subtema. La descripción es una de las formas más elementales y antiguas de otras formas del discurso, y está presente en textos literarios, históricos, jurídicos, publicitarios, entre otros.

#### 1.5.4.3 TEXTOS ARGUMENTATIVOS

Los textos argumentativos tienen como fin persuadir al lector de cierta idea o actitud a través de la presentación de razones o premisas. Adam (1992) distingue tres proposiciones básicas de los textos argumentativos: las premisas o datos que sostienen la opinión presentada en el texto, las inferencias hechas por la opinión defendida, y la conclusión o tesis del texto. Los textos con secuencias predominantemente argumentativas pueden encontrarse en artículos de opinión y científicos, textos publicitarios, jurídicos, entre otros. Estos textos se caracterizan por el uso de conectores contrastivos, causales, consecutivos y distributivos, por el uso de adverbios modales, y por la modalidad oracional.

#### 1.5.4.4 TEXTOS EXPLICATIVOS

Los textos explicativos hacen comprender al lector un fenómeno o evento. El texto explicativo resuelve dudas para una mejor comprensión de la realidad. Para Adam los textos de este tipo están compuestos por una esquematización inicial, en la que se plantea la pregunta o problema, seguida por una explicación o respuesta, que aclara esta cuestión, y finalmente una conclusión. Los textos con una secuencia textual predominantemente explicativa son principalmente del ámbito científico y académico, tales como entradas enciclopédicas, tratados científicos o de divulgación científica. Este tipo de secuencia textual se encuentra frecuentemente en relación con secuencias textuales argumentativas.

#### 1.5.4.5 TEXTOS DIALÓGICOS

Los textos dialógicos se refieren a las conversaciones entre dos o más personas. Pueden considerarse dialógicos tanto textos orales como escritos. Este tipo de

secuencia textual no es necesariamente el intercambio explícito entre participantes, sino que puede ser un texto en el que el autor “dialoga” con otros textos o autores (discute ideas, posturas, etc.).

#### 1.5.4.6 APROXIMACIONES PRELIMINARES A LOS TIPOS DE TEXTOS EN EL CORPUS COMPILADO

A partir de una revisión preliminar manual del corpus, un texto con secuencias textuales predominantemente descriptivas, tal como el reporte de estado de minas o reporte financiero, se caracterizaría por acercar la cultura, o su realidad material, al lector. Esto es generalmente en la dirección cultura/realidad local mexicana hacia el lector alemán, aunque también hay textos que están dirigidos a locales, sean mexicanos o trabajadores de minas extranjeros en México. Por lo anterior, pueden esperarse en estos textos más adaptaciones que permitan incorporar de manera natural las palabras extranjeras al sistema lingüístico local, el alemán.

Un texto con secuencias textuales predominantemente narrativas, propias de la literatura de viajes, género muy popular en este periodo debido a la influencia de los viajes y escritos de Humboldt, se caracterizaría por acercar al lector a la cultura expuesta en el texto, por lo que podemos esperar que haya menos adaptaciones de grafemas y menos traducciones de términos genéricos considerados equivalentes, o, bien, el nombre original acompañado de su traducción y/o explicación etimológica. Se hablará más acerca de la importancia de los grafemas y sus adaptaciones y de casos en los que se encuentran estas aportaciones a las apariciones de topónimos en textos en los siguientes apartados.

## 1.6 GRAFEMÁTICA Y GRAFOTAXIS

Otras consideraciones importantes relacionadas con la alonimia, especialmente la alonimia fonético-gráfica, están relacionadas con la grafemática, el estudio de los sistemas de escritura o grafemas. Ya que los nombres propios no cuentan, en lo general, con significado semántico, es pertinente atender su forma, su parte “perceptible”: su medio gráfico (Nübling, p. 86).

Es el medio gráfico a través del cual se puede disociar un nombre de otros con la misma forma fonética. Esto es posible en tanto que los nombres propios, junto con las interjecciones, son las únicas unidades lingüísticas que no están normadas ortográficamente, lo que permite la generación de numerosas variantes para un nombre con las que el objeto portador del nombre puede identificarse fuertemente, como en *Erik, Eric, Erick, Erich*, entre otras (p. 87).

Esta falta de normatividad aplica también para el aspecto morfológico y grafotáctico. En nombres morfológicamente complejos, no es raro encontrar alografías para partes apelativas de un nombre propio (p.e. el sufijo que tiene valor semántico de familia *-sohn* ‘hijo’ en alemán cuenta con las alografías *-sen* y *-Ben*), cambios vocálicos, duplicaciones y otros cambios que no corresponden con las reglas grafotácticas de la lengua (p. 88).

Estas violaciones a la grafotaxis o sintaxis de grafemas son especialmente frecuentes en la incorporación de palabras extranjeras. En los nombres propios aparecen conexiones de grafemas que en la ortografía del sistema lingüístico local están prohibidas o son de poca frecuencia. Así mismo, en la incorporación de palabras extranjeras es frecuente la adopción de grafemas periféricos. En el caso del alemán, p.e., los grafemas <c>, <y>, <x>, <v> y <q> aparecen en

nombres propios con mucha mayor frecuencia que en el resto del léxico y en palabras de origen extranjero (p. 89).

La grafotaxis libre de los nombres propios, y por lo tanto de los topónimos, representa grandes retos para el estudio de la toponimia de México en alemán, ya que la falta de un requerimiento de sistematicidad en los procesos de traducción y adaptación habría permitido la proliferación de numerosas variedades nominales para un mismo objeto.

Sin embargo, el reconocimiento de la libertad en el uso de grafemas y de ciertas combinaciones grafo-fonéticas nos permite distinguir los dos grandes métodos de traducción que fueron discutidos en el apartado anterior: el método domesticante, caracterizado por el apego al uso de los grafemas y combinaciones grafo-fonéticas permitidos o más frecuentes del sistema lingüístico local, el alemán en este caso; por otra parte, un método extranjerizante, caracterizado por el apego a las convenciones grafo-fonéticas de la lengua de origen del topónimo, que puede ser el mismo español o una lengua indígena.

## 1.7 CAMBIOS EN LOS TOPÓNIMOS

En los apartados anteriores se han expuesto algunos factores importantes a considerar con respecto a la traducción de los topónimos del español o de otra lengua indígena al alemán, pero en este apartado se pretende dar cuenta de algunos factores importantes dentro del propio territorio mexicano que afectaron los topónimos. Estos apartados, que pueden ser considerados más como marco contextual que como marco teórico, deben ponerse en relación con las teorías

mencionadas anteriormente para contextualizar el caso particular de la toponimia mexicana.

### 1.7.1 TOPÓNIMOS DE MÉXICO

Los topónimos de México, en su origen, son esencialmente apelativos y motivados, principalmente, por las propiedades y configuraciones del terreno, las características climáticas, así como la flora y la fauna del lugar, pero también por los dioses, templos, ceremonias religiosas y eventos históricos importantes (Lefebvre y Paredes, 2017: p. 88), como *Xochimilco* y *Reforma*.

### 1.7.2 ETAPAS TOPONÍMICAS DE MÉXICO

Han existido diversas etapas toponímicas diferenciables:

1. Periodo prehispánico: representado por topónimos provenientes principalmente de las culturas maya, tarasca, mixteca, zapoteca y totonaca, por ejemplo, *Tulum*, *Michoacán* y *Oaxaca*.
2. Periodo del Imperio Azteca: durante la expansión de este imperio, este periodo está representado por los topónimos de origen náhuatl, los cuales fueron transformados durante el siguiente periodo, por ejemplo, *Acapulco*, *Tlalnepantla* y *Teotihuacán*.
3. Periodo Colonial: durante la conquista española, los españoles trajeron consigo nombres de localidades relacionadas con las actividades económicas más importantes, normalmente relacionadas con la minería y la agricultura como *Nopalera*, *Carboneros*, *Calera*, etc; nombres de localidades homólogas en España, como *Valladolid* y *Mérida*; y nombres

relacionados con la religión católica y sus labores misioneras (Bahr, 2009, p. 27)

Durante este periodo hubo muchas modificaciones de nombres indígenas. Las tres más frecuentes fueron:

- a. Adaptaciones fonéticas: ya que los españoles desconocían las lenguas indígenas, la transcripción y comunicación oral de los topónimos locales, compuestos principalmente por descripciones apelativas, se caracterizó por la modificación de estos nombres para adaptarlos a las reglas fonotácticas del español, opacando el significado original como en el topónimo *Cuauhnáhuac* y su adaptación española *Cuernavaca*, donde se hace uso de palabras ya existentes en español, aunque sin relación semántica con el nombre original, o en *Teotihuacan* y *Teotihuacán*, donde solo hay un cambio en la posición del acento (p. 30).
  - b. Incorporación de elementos religiosos: con la llegada de los misioneros evangelizadores se extendió la costumbre de anteponer el nombre de santos o el de algún elemento propio de la iglesia católica a nombres indígenas. Por ejemplo, *San Bernardino Citlaltépec* y *Santa María Nenetzintla* (p. 27).
  - c. Hibridaciones hispano-indígenas: este es el caso de los nombres que sufrieron transformaciones como la incorporación de palabras del español. Como *Cruzitla* y *Tepunte* (p. 30).
4. Periodo independentista y posterior: este periodo se caracterizó por los homenajes a héroes nacionales a través de la incorporación de su nombre a un topónimo ya existente, o el renombramiento completo, y



también por el surgimiento de nombres que reflejaban expectativas del futuro o valores nacionales, como *en Hidalgotitlán, Tuxtla de Guerrero y Minatitlán*, y *El Porvenir, La Libertad*. Durante la Reforma (Lefebvre y Paredes, pp. 88-94), por otra parte, también hubo una revitalización de los nombres y las culturas prehispánicas, resultando en el renombramiento de nombres coloniales con topónimos indígenas como *Nuevo Santander a Tamaulipas y Nueva Galicia a Jalisco*, y en el renombramiento con nombres de figuras importantes indígenas como en *Tlaxcala de Xicoténcatl y Tepeyahualco Cuauhtémoc* (Bahr, p. 33).

### 1.7.3 ASPECTOS PARTICULARES DEL PERIODO COLONIAL Y POSINDEPENDENTISTA

El largo proceso de colonización española en América fue acompañado por un proceso de contacto lingüístico entre el español y las lenguas indígenas. Durante el periodo colonial se alternaron las políticas lingüísticas de fomento al español y al náhuatl (Wright, 2007, pp. 6-11), lo que provocó una falta de homogenización en el panorama lingüístico de Nueva España, principalmente por la disparidad en el acceso a la educación y por la numerosa variedad de lenguas al interior del territorio de la Nueva España (p. 16).

Las dificultades en materia de políticas lingüísticas no cesaron con la independencia de México, especialmente bajo la idea, compartida tanto por conservadores como liberales, de que una lengua era un instrumento necesario para la unificación del país (Heath, 1972, p. 58). El fomento de la educación tuvo la misma eficacia que tuvieron los esfuerzos análogos en la Nueva España: solo una parte de la población, los españoles y los criollos de clase alta, pudieron

tener acceso a los servicios educativos mientras que los indios, que constituían tres quintos de la población, permanecieron en condiciones precarias (p. 61).

Esto es importante por las siguientes razones: es lícito pensar que los trabajadores de minas, guías y pobladores de las zonas en las que se encontraron los investigadores alemanes eran ágrafos, por lo que los investigadores obtuvieron el conocimiento de los nombres de los lugares, por lo menos en contacto con ellos, solo por la oralidad. Por otro lado, sabemos que los investigadores tuvieron en ocasiones contacto con científicos, empresarios, traductores e intérpretes locales, aunque este ha sido poco documentado, de quienes es más probable que hayan podido recibir información escrita (Azuela y Villanueva, 2021, pp. 143-146).

Otras fuentes escritas de las que pudieron haber obtenido estos conocimientos fueron textos canónicos en materia de minería en América como el mencionado *Essai politique...* de Humboldt (1808) o *Mineralogische Beschreibung der vorzüglichsten Bergwerks-Reviere von Mexico* de Friedrich Sonneschmidt (1804) o información proporcionada por las compañías mineras inglesas o alemanas por las que fueron contratados. La correcta documentación de estos intercambios permitirá elucidar de manera más precisa si los investigadores siguieron métodos propios de la transcripción o de la transliteración.

Para lo anterior, es necesario compilar y analizar un corpus. Para poder pasar a la parte metodológica relacionada con la compilación y análisis de corpus es necesario presentar los fundamentos teóricos, la relación y la diferencia entre la lingüística de corpus y la lingüística computacional.

## 1.8 LINGÜÍSTICA DE CORPUS Y LINGÜÍSTICA COMPUTACIONAL

La lingüística de corpus y la lingüística computacional son áreas de la lingüística, relativamente modernas, que tienen una gran relación entre sí. En los siguientes apartados se describen ambas áreas, para después ponerlas en relación una con otra, y finalmente con esta investigación.

### 1.8.1 LINGÜÍSTICA DE CORPUS

Un corpus es, según Lemnitzer y Zinsmeister (2010, p. 40), una compilación de textos orales o escritos en una o más lenguas. Los textos del corpus están digitalizados, es decir, almacenados y legibles por computadora y cuentan con metadatos, anotaciones lingüísticas.

La lingüística de corpus es un área de la lingüística con un enfoque empírico, en oposición al enfoque racionalista que había dominado la lingüística durante la mayor parte del s. XX. Esta oposición entre empirismo y racionalismo se da en el marco de las llamadas “Guerras lingüísticas”, como la ha nombrado Randy Allen (1993).

Desde finales de los años 90, Jan Svartvik sostuvo que “los estudios de corpora ayudarán a promover la descripción más adecuada de las gramáticas”. Entre otros lingüistas, Michael Halliday consideró muy notable este posicionamiento porque va en contra de las ideas corrientes de la época acerca de lo que es o debe hacer la lingüística. Los estudios en lingüística de finales del s. XX se encontraban dominados por las perspectivas generativistas transformacionales de Noam Chomsky y no se reconocían en ellos las aportaciones que podía hacer la lingüística de corpus o cualquier análisis cuantitativo para el entendimiento del lenguaje.

Las teorías de Chomsky y su división entre la competencia y la actuación, entre el sistema lingüístico y su respectiva instanciación, postulaban que era imposible que el análisis de textos reales pudiera llevar a la explicación de la gramática de una lengua y aún menos podía formularse una teoría lingüística general a partir de estas muestras reales de lenguaje. El enfoque desde la teoría chomskiana debía recaer siempre en la competencia, en el sistema lingüístico mismo, mas no en una lengua particular, incluso a costa de la poca accesibilidad a este sistema dentro de la mente de los hablantes.

Svartvik (1966) mismo reconoció que los estudios cuantitativos no pueden ser trivializados al encontrar diferencias sistemáticas y predecibles de algunas propiedades gramaticales del inglés a partir del estudio de corpora. Svartvik elaboró una taxonomía de las clases de cláusulas pasivas en inglés a partir de los datos de un corpus para demostrar la existencia de una proporción de cláusulas pasivas con respecto a las activas que variaba a lo largo de diferentes registros (Halliday, 2005, p. 63).

Halliday considera que para principios de los noventa no es cuestionable que los estudios basados en corpora sean una fuente de información de la gramática de una lengua particular e incluso de la gramática en general. Svartvik y el centro de investigación Survey of English Usage han demostrado el valor de los estudios de corpora, pero Halliday considera que el estatus teórico de la frecuencia en corpora todavía es una cuestión abierta (p. 64).

El lingüista británico considera que el sistema lingüístico mismo es inherentemente probabilístico y que la frecuencia en un texto se debe a la instanciación de la probabilidad de su gramática. Esto parece ser algo sumamente evidente, aunque es una idea que todavía tiene oposición por su

carácter, hasta cierto punto, determinista; sin embargo, es una idea que es ampliamente aceptada y reconocida como algo válido en algunas áreas de la lingüística, mientras que en otras no recibe la misma aprobación.

Halliday ejemplifica la aceptación de los estudios probabilísticos con el caso del estudio del vocabulario. Zipf, desde 1935, usó estrategias cuantitativas en su estudio del vocabulario para establecer principios generales como la relación entre la frecuencia relativa de un ítem léxico y su lugar en el orden de rango. Aunque las generalizaciones realizadas surgidas de este estudio pueden ser cuestionadas, es un hecho que los patrones de frecuencia de las palabras y su sistematización fueron aceptados como propiedades de la lengua. La pregunta de Halliday es, entonces, ¿por qué hay tanta resistencia al reconocimiento de los patrones cuantitativos de la gramática? (p. 64).

Para Halliday no es lógico hacer tal distinción entre el vocabulario o léxico y la gramática. Él considera que ambos son parte del mismo continuo: la léxico-gramática; pero también reconoce que la gramática, el extremo más “profundo” del continuo, es menos accesible a la conciencia, por lo que el tratamiento de la gramática, en cualquier forma, siempre resulta en mayor resistencia. Empero es el estudio de la gramática desde esta perspectiva lo que permitiría hacer interpretaciones directas y significativas acerca del sistema de la lengua, ya que en el sistema de la gramática el número de opciones es cerrado y más reducido, en comparación con el de la lexis.

Halliday enfatiza que estas probabilidades son propias tanto del sistema lingüístico como de su instanciación porque se está tratando aquí con un mismo fenómeno. Para ilustrar mejor este punto, Halliday propone que la relación entre sistema lingüístico y su instanciación es similar a la relación entre el clima y el

tiempo: las lecturas o mediciones de la instanciación y del tiempo se hacen siempre bajo condiciones particulares, pero estas mediciones siguen siendo parte de un conjunto de características comunes, las del sistema lingüístico o las del sistema climático. Las mediciones de las instanciaciones particulares son lo que permite hablar de generalidades y particularidades en la totalidad del sistema lingüístico, así como se le hace con toda naturalidad en el climático (p. 66).

Los estudios basados en corpora permiten llegar a reconocer estas propiedades en términos de selecciones no marcadas y marcadas o de probabilidades en la selección de opciones. Para poder determinar las probabilidades de algún subsistema de la gramática de una lengua es necesario contar con un corpus que esté organizado por el registro. Esto permite estudiar las condiciones externas que determinan estas probabilidades.

Nesbitt y Plum (1988), por ejemplo, constituyeron un corpus de conversaciones orales en inglés que podían ser agrupadas bajo el género narrativo para medir la distribución del sistema de taxis (parataxis e hipotaxis) con relación al sistema de relaciones lógico-semánticas (proyección, a su vez subdividida en locución e idea, y expansión, subdividida en elaboración, extensión y realce).

En su estudio pudieron comprobar que hay una relación probabilística clara entre ambos sistemas: hay una clara asociación entre parataxis y locución, y entre hipotaxis e idea. Esto demostró que mientras que la elección de taxis y el tipo de relación lógico-semántica son gramaticalmente independientes, son estadísticamente dependientes; hay combinaciones más probables que otras (pp. 71-2).

De lo anterior podemos concluir que los estudios basados en corpora tienen un papel central en las investigaciones teóricas del lenguaje. Un corpus puede utilizarse de numerosas maneras y para distintos fines, entre ellos para la elaboración o el robustecimiento de teorías acerca del desarrollo del lenguaje en niños, la variación diatópica o el registro, la sistematicidad de la gramática en términos de las opciones paradigmáticas de sus subsistemas y el reconocimiento de procesos de semogénesis a lo largo de un periodo histórico (p. 73). En el caso de la toponomástica, los estudios basados en corpora han servido como registros a estudiar desde distintas perspectivas o disciplinas estos nombres.

#### 1.8.1.1 COMPILACIÓN DEL CORPUS

La compilación del corpus debe ser un proceso metódico en el que se consideren tanto con el objeto de investigación mismo como las fuentes de datos a través de las cuales se estudiará (Lemnitzer y Zinsmeister, 2010, p. 57).

En la actualidad, la mayoría de los textos se encuentran también digitalizados, lo que facilita su recopilación y estandarización de la codificación de los datos (pp. 57-58). Sin embargo, en el caso de esta investigación, se trata de textos impresos digitalizados por escaneo, lo que también conlleva una serie de procesos de edición y preparación que permita su correcta legibilidad por computadora.

Lo que se busca los textos de un corpus es contar con muestras representativas del fenómeno u objeto lingüístico en cuestión. Para lograr compilar una muestra representativa, deberán aplicarse diferentes criterios que dependerán de la naturaleza del estudio (pp. 50-52); en esta investigación, se

aplicarán criterios que permitan llegar a textos en los cuales se encuentre el fenómeno de la alonimia de topónimos de México en alemán.

### 1.8.2 LINGÜÍSTICA COMPUTACIONAL

La Asociación de Lingüística Computacional (ACL, por sus siglas en inglés) define la lingüística computacional como el estudio científico del lenguaje desde una perspectiva computacional. Esta área de la lingüística busca proveer modelos computacionales para diversos fenómenos lingüísticos. Estos modelos pueden ser basados en teorías lingüísticas o en datos lingüísticos particulares.

Los trabajos en lingüística computacional pueden ser motivados por la búsqueda de una explicación computacional para un fenómeno lingüístico; en otros casos, como los de esta investigación, la motivación puede ser meramente técnica: se busca proveer un componente funcional que pueda realizar las funciones de un sistema lingüístico. Las aplicaciones más comunes de la lingüística computacional son sistemas de reconocimiento de voz, respuestas de voz automáticas, buscadores web, editores de texto, entre otras.

Pese a las numerosas utilidades de esta área, hasta donde tengo conocimiento, son relativamente pocos y muy recientes los estudios en lingüística que toman conocimientos o aplicaciones de esta área. Por otra parte, la lingüística computacional es un área con rápido crecimiento en las ciencias de la computación, esto debido a las necesidades informativas de la sociedad actual.



### 1.8.3 RELACIÓN ENTRE LA LINGÜÍSTICA DE CORPUS Y LA LINGÜÍSTICA COMPUTACIONAL

La lingüística computacional es un campo en el que el uso de corpora toma un papel fundamental. Esta corpora sirve, en primer lugar, como fuente de datos para la investigación empírica. El lingüista computacional observa estos datos para desarrollar y comprobar hipótesis (Lemnitzer y Zinsmeister, 2006, pp. 137-8). En esta investigación se usa una herramienta computacional para observar y cuantificar datos que permitan llegar a conclusiones.

#### 1.8.3.1 ANTECEDENTES DE ESTUDIOS DE NOMBRES BASADOS EN CORPORA ASISTIDOS POR COMPUTADORA

Los estudios de nombres propios basados en corpora son relativamente nuevos y su desarrollo se ha hecho a la par del de las tecnologías de la información. El primer estudio de este tipo del que tengo conocimiento fue la tesis de doctorado de Martina Schwanke en 1992, titulada *Name und Namengebung bei Goethe. Computergestützte Studien zu epischen Werken* (Los nombres y el nombramiento en Goethe. Estudios asistidos por computadora de textos épicos).

En esta tesis, Schwanke pilotea la herramienta de la computadora para convencer a sus lectores de que esta es una nueva herramienta que permite ampliar el alcance de la investigación y procesar más datos para que los investigadores puedan hacer una interpretación de estos (Hough, p. 390).

En la tesis de Schwanke se muestra el proceso de compilación del corpus de obras de Goethe, los procesos asistidos por computadora de lematización, registro y concordancia, y las interpretaciones de la autora acerca de la relación entre los nombres usados en las obras de Goethe, su frecuencia de aparición,

su relación con los otros personajes de las obras, y sus alusiones a personajes extratextuales (p. 391).

Schwanke concluye afirmando que el beneficio del uso de corpora procesado por computadora consiste, principalmente, en que el investigador puede visualizar y analizar más datos, así creando una base más firme para sus interpretaciones. Schwanke considera que su tesis es solo un estudio piloto en la onomástica literaria asistida por computadora (p. 391).

Después de Schwanke, otros investigadores han seguido metodologías similares basadas en corpus para el análisis de nombres. Algunos de estos autores son Ines Sobanski (estudio acerca de los nombres propios en textos de G.K. Chesterton en el 2000), Jan-Christian Schwarz (estudio acerca de los nombres propios y sus funciones en los cantos de Neidhart en el 2004) y Karina van Dalen-Oskam (estudio acerca de los apellidos en obras de Países Bajos) (pp. 391-4).

A medida que se han desarrollado las herramientas tecnológicas, se han refinado las investigaciones basadas en corpora y se han orientado a enfoques cuantitativos basados también en la estadística. También se han desarrollado tecnologías como el Reconocimiento de Entidades Nombradas (NER por sus siglas en inglés) que han probado ser de mucha utilidad, aunque también con muchos desafíos por delante, para el estudio de nombres propios, aunque la efectividad de este reconocimiento de entidades depende mucho de la lengua, el género y el campo de aplicación (p. 395). En este estudio no se hace una aproximación a la localización de topónimos basada en el reconocimiento de entidades nombradas, aunque se ha reconocido como una manera de localizar entidades como lugares en textos.

Estos antecedentes muestran que los estudios basados en corpora se ven beneficiados por herramientas computacionales que permitan procesar grandes volúmenes de datos. En el caso de esta investigación se pilotea la herramienta de Python para los mismos fines: procesar grandes volúmenes de textos que permitan al investigador llegar a conclusiones más generales y sólidas.

## METODOLOGÍA

En el capítulo anterior nos referimos a la literatura especializada más relevante en el área de la onomástica, para una conceptualización del topónimo, para el reconocimiento de los procesos de cambio lingüístico con motivación interna, y de la sociolingüística, para el reconocimiento de los procesos de cambio lingüístico con motivación externa que pudieron haber sufrido los topónimos durante el contacto entre el español, las lenguas amerindias y el alemán. De este análisis se tomaron los elementos más importantes para este trabajo: a) la identificación de los topónimos y sus variantes nominales en un corpus y b) la identificación de las motivaciones externas e internas de los cambios en los topónimos.

Se empleó un enfoque mixto con instrumentos cuantitativos para la contabilización y gestión de los topónimos y sus variedades nominales en el corpus, su clasificación y la representación gráfica de estos valores numéricos, así como instrumentos cualitativos, como la búsqueda en fuentes especializadas y la consulta de documentos históricos, para clasificar correctamente los topónimos e identificar por qué pueden tener variantes nominales.

En la siguiente tabla (Tabla 2) se especifica la relación entre las preguntas de investigación planteadas en la introducción y los procedimientos necesarios para responderlas.

Preguntas de investigación	Fuentes de datos y métodos	Justificación	Recursos del investigador para tener acceso a los datos
1) ¿Cuáles son los topónimos a los que hacen referencia los autores alemanes en los textos coleccionados del corpus constituido?	Revisión manual del corpus en alemán	Es necesario contabilizar con exactitud el núm. de topónimos y sus variedades nominales en el corpus	Conocimientos de la onomástica, lexicología, lengua alemana y la geografía de México
2) ¿Cuáles son las variaciones alonimicas de topónimos en alemán encontradas en el corpus?	Revisión manual del corpus en alemán y de algún otro corpus o alguna otra fuente en español	Es necesario partir de un supuesto acerca de cuál fue el topónimo corriente en español del cual se desprendieron las variantes	Conocimientos de geografía de México y sus topónimos
3) ¿Cuáles son las motivaciones de las variaciones internas y externas de los topónimos?	Investigación de los cambios lingüísticos que posiblemente sufrieron los topónimos	Es necesario identificar los procesos de cambio lingüístico para llegar a conclusiones generales a partir de una muestra (el corpus en alemán)	Conocimientos en lingüística histórica, sociolingüística, fonética y morfología
d) ¿Cómo pueden localizarse todas las entradas de un topónimo y sus respectivas variantes?	Elaboración de un programa con Python que localice automáticamente los topónimos y sus variantes	Contar con una herramienta que pueda localizar diversas palabras en un corpus a partir de una serie de reglas con sustento lingüístico y no solo a partir de una cadena de caracteres ingresada puede facilitar otras investigaciones en cuestión de tiempo y eficacia.	Conocimientos de programación con Python

Tabla 2. Preguntas de investigación

## 2.1 CORPUS

El propósito del corpus es contar con una muestra representativa de textos con características en común en la que se pueda observar la alonimia de topónimos de México para analizarla a partir de sus motivaciones de cambio, es decir, a partir de las causas por las que un topónimo puede tener múltiples variantes nominales.

Este corpus ha de servir como corpus testigo para poder identificar los cambios que se dieron en la traducción al alemán a partir de las fuentes originales en español, cambios que son producto del contacto entre el español y lenguas amerindias, entre otros. La unidad de análisis en este corpus es el topónimo, que se entiende desde la onomástica como el nombre de países, regiones, pueblos y lugares (Lara, 2006, p. 243) y que tiene subclases como el orónimo, hidrónimo, macrotopónimo, microtopónimo, polinónimo, entre otras (Iglesias, 1991, p. 240).

### 2.1.1 CARACTERÍSTICAS GENERALES DEL CORPUS

Para la constitución del corpus se han recuperado textos con las siguientes características:

- **Tema:** minería y ciencias de la tierra de México, esto porque, desde el periodo colonial, la minería ha sido una de las actividades económicas más importantes, por lo que se cuenta con numerosos textos al respecto y, consecuentemente, también de ciencias de la tierra que permitan explicar mejor el mantenimiento, usos y manipulación de los productos de las minas.

- **Periodo:** 1804-1848, esto porque en 1808 se publica *Mineralogische Beschreibung der vorzüglichsten Bergwerks Revieren von Mexico oder Neuspanien* (Descripción Mineralógica de los principales distritos mineros de México o Nueva España) de Friedrich Sonneschmidt, texto en el que por primera vez, hasta donde tengo conocimiento, se encuentra el fenómeno de la alonimia plena (*Mexico y Neuspanien*) en un texto con las características temáticas descritas arriba. Además, este texto será ampliamente citado durante toda la primera mitad del s. XIX, por lo que se espera reconocer su influencia directa en otros textos del corpus. Por otro lado, se cierra el periodo en 1848, año en que termina la intervención estadounidense en México y se firma el Tratado de Guadalupe Hidalgo, en el que México pierde más de la mitad de su territorio, y en que los frecuentes problemas económicos de los inversores y trabajadores extranjeros de minas en México se vieron seriamente agravados durante este conflicto armado, lo que llevó al quiebre de las principales compañías mineras extranjeras y a la necesidad de dirigir las investigaciones científicas al campo de la cartografía y la delimitación territorial.
- **Lengua:** alemana, esto en razón de que fueron los ingenieros y técnicos de estados miembros de la Confederación Germánica los que mayor participación tuvieron en compañías mineras en México, aunque las compañías hayan sido, en su mayoría, inglesas, y, consecuentemente, los que mayor producción científica realizaron.
- **Género textual:** textos informativos con formato de artículo, informe de estado de minas, correspondencia y narrativos, esto debido a que los géneros textuales todavía no estaban completamente estandarizados, por

lo que con frecuencia se encuentran artículos científicos escritos a manera de literatura, y literatura con contenido y temáticas científicas, esto en gran medida bajo la influencia de la literatura de viajes científica de Humboldt (Louise Pratt, 2007).

- **Fuente:** anuarios, boletines, periódicos, libros y revistas científicas con sede en territorios germanoparlantes, esto en razón de que fueron estos territorios los que mayor producción científica tuvieron en alemán.

### 2.1.2 COMPILACIÓN DEL CORPUS

El universo de estos textos es vasto y no cuento con su número. Se utiliza como guía *Bibliografía geológica y minera de la República Mexicana completada hasta el año de 1904* de Rafael Aguilar y Santillán. En este texto Aguilar y Santillán reporta una gran cantidad de textos acerca de la minería y ciencias de la tierra de México en varias lenguas, incluida la lengua alemana. Aguilar y Santillán da cuenta de muchos textos, aunque no todos, pero esto debe atribuirse a las limitaciones informativas de su época.

Pese a que solo un 30% de los textos del corpus son textos reportados en la *Bibliografía* de Aguilar y Santillán, este texto ha sido de gran ayuda en la identificación de las numerosas revistas científicas, anuarios y boletines en las que se publicó constantemente acerca de la minería y geología mexicanas. Al identificar estos lugares de publicaciones, se procedió a revisar cada uno de sus números o volúmenes en búsqueda de textos que cumplieran con las características arriba descritas del corpus. Las revistas, boletines, periódicos y anuarios identificados son:

- *Taschenbuch für die gesamte Mineralogie*



- *Hertha, Zeitschrift für Erd-, Völker- und Staatenkunde*
- *Annalen der Erd-, Völker- und Staatenkunde*
- *Journal für Chemie und Physik*
- *Annalen der Physik und Chemie*
- *Archiv für Bergbau und Hüttenwesen, después renombrado Archiv für Mineralogie, Geognosie, Bergbau und Hüttenkunde*
- *Neues Jahrbuch für Mineralogie, Geognosie, Geologie und Petrefaktenkunde*
- *Berg- und Hüttenmännischen Zeitung*
- *Fortschritte der Geographie und Naturgeschichte*

Estas revistas, anuarios, periódicos y boletines han sido consultados principalmente en el sitio de la biblioteca digital de acceso libre [hathitrust.org](http://hathitrust.org). Esta página cuenta con un gran acervo digitalizado y permite la descarga de sus documentos en formato PDF, JPEG o TXT. Una dificultad para la descarga de los documentos en este sitio es que solo se permite la descarga de una página del documento a la vez.

Para realizar este paso con mayor velocidad se ha utilizado el programa de uso libre Hathi Downloader Helper (v.1.2.2) que permite descargar el documento completo o el rango deseado al ingresar la URL del documento de [hathitrust.org](http://hathitrust.org). Este programa descarga las páginas del documento y las unifica en un solo documento formato PDF.

La mayoría de los textos han sido descargados siguiendo este procedimiento. Solo 8 textos han sido consultados y descargados de otros sitios: 7 de [digitale-sammlungen.de](http://digitale-sammlungen.de), el sitio web del Centro de digitalización de Múnich

y uno de Google Books. El acervo del corpus siguiendo estos criterios comprende 113 textos (Anexo A).

### 2.1.3 REVISIÓN Y EDICIÓN DEL CORPUS

Los textos del corpus en formato PDF cuentan con numerosas dificultades para su correcta lectura por computadora, algunas de ellas relacionadas con la impresión del texto original y otras relacionadas con su proceso de digitalización.

#### 2.1.3.1 PROBLEMAS CON LOS DOCUMENTOS ORIGINALES

Los textos originales, ya que son textos impresos de la primera mitad del s. XIX, cuentan con una calidad de impresión que dificulta el reconocimiento óptico de caracteres asistido por computadora, es decir, el reconocimiento de texto en los documentos del corpus. Los problemas más frecuentes para el correcto OCR (*Optical Character Recognition* por sus siglas en inglés) debidos a la impresión misma del texto original son la tipografía utilizada, la dispersión de tinta y la claridad del papel de impresión.

La tipografía de los textos del corpus varía, principalmente, según la casa editorial de publicación. Dos tipografías caracterizan el corpus: la *Fraktur* y la *Antiqua*. La primera es subgrupo de la escritura gótica que fue ampliamente utilizada en regiones germanoparlantes desde el s. XVI y hasta mediados del s. XX; la segunda es una tipografía proveniente del periodo renacentista italiano que rápidamente ganó popularidad hasta convertirse en la tipografía más utilizada en toda Europa, en detrimento del uso de la *Fraktur*.

El periodo de transición de la *Fraktur* a la *Antiqua* duró aproximadamente 200 años (de mediados del s. XVIII a mediados del XX) y es conocido como la

disputa *Antiqua-Fraktur* (*Antiqua-Fraktur-Streit*) por sus implicaciones políticas. En este periodo se encuentran también mezclas entre ambos tipos de escrituras, pero en los textos del corpus pueden reconocerse ambas tipografías claramente diferenciadas entre sí.

La tipografía *Antiqua* no representa, por sí misma, dificultades para el OCR. Esto debido a que esta tipografía es, aún en la actualidad, extensamente utilizada y, por lo tanto, los softwares de OCR están preparados para reconocerla. La tipografía *Fraktur*, por su tiempo y espacio de uso limitado, no ha sido considerada para el entrenamiento de softwares de OCR, por lo que es ilegible para estos (Fig. 2).

Die Formation des Schiefergebirges, obgleich sie auf der Oberfläche größtentheils mit Porphyren und vulkanischen Gebilden bedeckt ist und sich nicht zu den höchsten Bergen erhebt, scheint das Grundgebirge auf dem Hochlande von Mexiko zu bilden, und verdient wegen ihrer großen Mannigfaltigkeit von Erzlagerstätten die größte Aufmerksamkeit der Gebirgsforscher. Das Schiefergebirge, bildet in seinen verschiedenen Gliedern die durch frühen Bergbau berühmten Bergwerksdistrikte von Tlalpujahua, Temascaltepec, Sultepec, Christo, alle in Thonschiefer; Zacualpan, abwechselnd im Thon-, Grauwacken- und Grünsteinschiefer; Zimapan und Peshuga im Kalkschiefer mit Kieselschiefer; Cardonal im Kalkschiefer mit Kalkstein; Istapa del oro im Thonschiefer und Kalkstein gelegen. Das Einfallen des Schiefergebirges ist größtentheils zwischen Süd und West, obgleich bei Temascaltepec, Zacualpan und Tasco östliches und nordöstliches Einfallen stattfindet.

Die Gänge der Schieferformation sind im Allgemeinen weit schmaler als im Porphyry, auch sind die Erzmittel nicht so aushaltend, aber sie führen die reichsten Geschicke, vorzüglich gediegenes Silber, Rothgültigerz, Glaserz und mehre derselben haben einen bedeutenden Goldgehalt, wie in Tlalpujahua. Außer der unzähligen Menge von Erzgängen nimmt das Schiefergebirge auch viele

Figura 2. Texto con tipografía Fraktur (Gerolt y Berghes, 1828)

Otros factores como la dispersión de tinta (Fig. 3), su difuminación (Fig. 4) y la claridad del papel de impresión de los textos originales (Fig. 5) son hechos en los que poco podemos ahondar. Se trata de complicaciones propias de las limitaciones materiales y tecnológicas de la época de impresión de los textos, pero que tienen repercusiones en las capacidades de los softwares de OCR.

Eben diese Schwierigkeit der Auffindung der Grenze zwischen diesen beiden Porphyren des Gebietes von *Real del monte* fand auch schon A. v. Humboldt \*) Ueber diesen Porphyr führt der Weg fort bis in die Nähe des *Cerro de Guajalote*, das durch ein weites Thal von dem *Cerro de las Nabajas* getrennt wird. Auch das *Cerro de Guajalote* besteht aus Porphyr, der hier zwar noch nicht auffallend verschieden von jenem am *Real del monte* ist, aber doch davon abweicht. Im Allgemeinen besteht der Porphyr des *Cerro de Guajalote* aus einem dichten, rötlichgrauen Feldspath-Teige, Krystalle von Feldspath, Hornblende und Quarz und einzelne runde Ausscheidungen eines heller gefärbten Porphyrs mit vielem Quarz umschließend. Der in diesem Porphyr auftretende Quarz ist sehr fettglänzend (Rauchtopas), die Hornblende ist häufig an der Oberfläche von einer dünnen,artigen, grünen Substanz umschlossen, und der Feldspath scheint bisweilen glasiger Feldspath zu seyn. Häufig treten hier Trümmer-Porphyre auf, welche unförmliche Massen in dem gewöhnlichen Porphyr zu bilden scheinen; in einer perlgrauen Grundmasse und schliesst dieser Trümmerporphyr, ausser Krystallen von gemeinem Feldspath, sehr feine, kleine Stücke eines schwarzen thonigen Gesteins und runde, dunkler gefärbte Ausscheidungen seiner Grundmasse.

So wie man von der Höhe nach dem *Rancho Guajalote* in das Thal hinunter steigt, findet man eine Menge von Obsidianstücken auf einem gelblichgrauen Trachyt-Porphyr umherliegen.

Der gänzliche Mangel an Quarz, der glatte Feldspath, die langen dünnen Hornblende-Krystalle liessent mich nicht länger zweifeln, dass ich mich hier nicht mehr in dem erstführenden Porphyr von *Real del monte*, sondern auf Trachyt-Porphyr befand.

Auch weiter südöstlich von diesem Punkte zeigt sich noch einmal Trümmer-Porphyr in einem rauchgrauen dichten Feldspath-Teige, Krystalle von Feldspath und Hornblende und verschiedenartig gefärbte runde Porphyr-

\*) Vergl. dessen *Essai politique* Tome IV, S. 8.

Figura 3. Dispersión de tinta (Burkart, 1839)

sondere mit der heutigen Geognosie völlig vertraut, so daß sie auf der von Alexander v. Humboldt gebrochenen und vorgezeichneten Bahn leicht verschreiten können, um die von ihrem erhabenen Vorbilde gegebenen Aufschlüsse zu vervollständigen und zu vermehren, zum Nutzen und Frommen nicht allein ihrer Kommittenten, sondern der ganzen wissenschaftlichen Welt. Dank der Gesellschaft, die in der Wahl ihrer Beamten so glücklich war!

Bei der Entwerfung dieser Karte lag hauptsächlich die Absicht zu Grunde, die verschiedenen Formationen der erzführenden Gebirge heranzuheben, in deren Schoos ein so unermeßlicher Reichtum von edlen Metallen niedergelegt ist. Die geographische Grundlage dieser Karte sind die astronomisch-geodätischen Operationen des Herrn A. v. Humboldt und insbesondere die ganze Karte desselben des Thales von Mejiko, so wie denn auch der Maßstab dieser Karte beibehalten worden ist. Von den geographischen Resultaten Alexanders v. Humboldt ausgehend bestimmten die Verfasser die Lage mehrerer anderer Punkte, theils durch astronomische Beobachtungen, theils durch trigonometrische Messungen, mit dem Theodoliten und mit dem Dosen-Sextanten, wobei auch die Hauptgebirgszüge und der Lauf der Flüsse bemerkt wurden; da, wo die Gestalt der ersteren bisher unbekannt geblieben, ist der Raum frei gelassen und bloß mit dem Namen *Cerranía* (Gebirgslandschaft) auf der Karte bezeichnet worden. Die Wege und ihre nächsten Umgebungen wurden mit dem Kompaß in der Hand aufgenommen, und die vertikalen Dimensionen mit englischen *Mountain-Barometern* gemessen, so wie denn auch auf diesen wiederholten Reisen die Notizen über die geognostischen Verhältnisse gesammelt wurden. Als Beispiel der großen Genauigkeit und Uebereinstimmung der Barometermessungen unter jenem Himmelsstriche führt Herr v. Gerolt an, daß die Höhe von Mejiko über der Meeresfläche nach seinen Beobachtungen 7450 engl. Fuß betrage, was von der Bestimmung des Herrn v. Humboldt nur etwa eine Loise abweicht.

Herr v. Gerolt hat der Karte eine kurze Notiz über die geognostischen Verhältnisse der Hochebenen von Mejiko beigefügt, die er als den Vorläufer einer ausführlicheren Darstellung dieser Verhältnisse und insbesondere der Formationen der Erzgebirge und der darin aufstehenden Lagerstätten von edlen Metallen ankündigt.

Figura 4. Difuminación de la tinta en el párrafo superior (Gerolt y de Berghes, 1828)

larifizirt haben, und zwischen diesen Spalten in der Quere stehend, fest anhängen.

Dieser Eisenstein ist theils dichter, theils körniger oder poröser Magnet-Eisenstein. Der letztere ist häufig krystallisirt, in Dodekaedern, und an den Drusen desselben findet man zuweilen silberweißen Glimmer. Auf den Klüften bemerkt man hier und dort Spuren von Kupfergrün, jedoch scheint die Erzmasse keinen Gehalt an Kupfer zu haben, indem die salzsaure Auflösung, mit Ammoniak im Ueberschuss versetzt, keine Spur von blauer Farbe zeigt. Das spezifische Gewicht des körnigen Magnet-Eisensteins von Real del Dro ist 4,44; das des dichten 4,96. Die Härte des ersteren = 5; die des letzteren = 5,5. Beide haben einen schwarzen Strich, und eine bräunlich-schwarze Farbe.

Auf demselben Gebirgsrücken,  $\frac{1}{2}$  Stunde südwestlich von diesem Punkte, setzt eine zweite mächtige Niederlage von Magnet-Eisenstein, ebenfalls in Kalkstein auf, welcher ganz dieselben Erscheinungen giebt. Auch diese scheint ein Gangkörper zu seyn, welcher, nahe der obersten Hütte, im Thale Encarnacion durchsetzt; jedoch ist der Gang an letzterem Punkt mit viel Jaspis und Kupfergrün vermengt.

Die Umstände unter welchen dieser Eisenstein vorkommt, geben die hoffnungsvollsten Ausichten zum Betrieb einer Eisenhütte im Thal Encarnacion. Die umliegenden Gebirge sind mit schönen Eichen- und Nadel-Hölzern bedeckt. Der Centner Eichenkohlen kostet 2 Realen, und der Centner Nadelholz-Kohlen  $2\frac{1}{2}$  Realen. Die Gewinnung und der Transport der gewonnenen Eisenerze zur Hütte,  $\frac{1}{2}$  Legua weit, kann man auf höchstens 2 Realen für die Carga annehmen. Auch ist die Lage von Real del Dro vortheilhaft zum Absatz des Eisens.

Zuletzt muß ich noch einer Lagerstätte erwähnen, welche  $\frac{1}{2}$  Legua westlich der Grube Flajonales, in demselben Thale in Kalkstein aufsetzt, und über 6 Vara's zu Tage heraussieht. — Die Masse dieses Gangkörpers, welcher ebenfalls

Figura 5. Texto con hojas oscuras (Gerolt, 1827)

### 2.1.3.2 PROBLEMAS CON LOS DOCUMENTOS Y EL PROCESO DE DIGITALIZACIÓN

Además de los problemas relacionados a las características de los documentos originales, el proceso mismo de su digitalización acarrea también otras complicaciones.

Los textos originales han sido escaneados, presumiblemente, aunque no se muestra esa información en las bibliotecas digitales de las que se descargaron los documentos, en grandes volúmenes y siguiendo métodos homogéneos. Esto implica que un texto con tipografía *Fraktur* fue escaneado y digitalizado de la misma manera que otra con tipografía *Antiqua*, o que un texto con hojas oscurecidas fue escaneado y digitalizado igual que un texto con hojas claras y legibles.

La falta de un proceso de escaneo y digitalización particular para cada texto, o para un grupo de textos con características en común, ha llevado a que la calidad del reconocimiento de texto varíe mucho de texto a texto del corpus. Aunado a lo anterior, errores humanos en este proceso tales como escaneo con orientación incorrecta (Fig. 6) y defectos en el proceso de digitalización (Fig. 7) representan también retos para la preparación de los documentos.

Linie gelegene Punkt der die Landenge aus O. in W. durch-  
ziehenden Cordillerenkette, der *Cerro Pelado*, erhebt sich  
kaum mehr als 2,000 Fuss. Alle übrigen Höhen sind bedeu-  
tend niedriger, und die Thäler der Flüsse Guasacualco und  
Chimalapa steigen auf dieser Linie an keinem Punkte zu  
mehr als 525 Fuss absoluter Höhe auf. Die Lagunen von  
Tehuantepec vertiefen sich 6 Leguas weit in das Land, und  
machen dadurch die Landenge hier noch schmaler.  
Eine grosse Menge von Flüssen bewässern die Landenge,  
unter welchen der *Rio de Guasacualco* für eine Wasser-  
verbindung zwischen beiden Meeren besonders wichtig, zu-  
gleich aber auch der bedeutendste von allen ist, und die  
meisten der übrigen aufnimmt. Dieser Fluss entspringt in  
etwa 16° 58' n. Br., 96° 19' w. L. v. Par. in den Gebirgen  
von Tarifa, an der Gränze zwischen den Staaten Oajaca und

— 101 —

— 102 —

*Rio Guélaguésza* oder *Almolóyan* zum *Rio Alamán* oder  
*Malpáso*. Weit im W. entspringt in den Gebirgen der Mijes  
der *Rio de Chichicóvi*. Er strömt gegen O., empfängt aus  
S. W. die zuvor vereinten Wasser der Flösschen *Malatengo*  
und *Citini*, aus N. W. den Bach *Mugána*, und fällt in den  
Malpáso, welcher seinerseits wenige Leguas unterhalb der  
Mündung des Iscuilápa in den Guasacualco fällt. Unmittelbar  
unterhalb der Vereinigung der Flüsse Guélaguésza und Chi-  
chihua sollte nach den bisherigen Plänen am Rio Alamán

Figura 6. Texto con orientación incorrecta (Mühlepfordt, 1847)



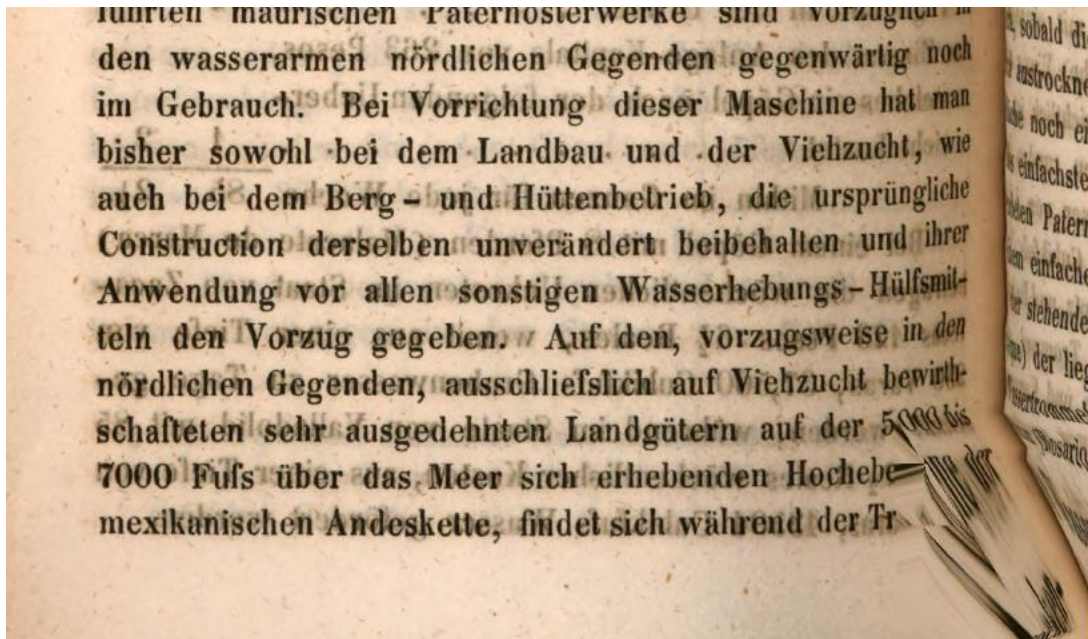


Figura 7. Texto con defecto de escaneo (Berghes, 1847)

### 2.1.3.3 CORRECCIÓN DE LOS TEXTOS DEL CORPUS

Los textos del corpus tienen imperfecciones que complican el correcto reconocimiento de texto por computadora. Es necesario identificar estos errores de reconocimiento de caracteres y corregir el texto para que este pueda ser correctamente leído. Para esto se ha usado el programa Adobe Acrobat Pro DC (v. 2021.011.20039). Con este programa es posible realizar el OCR, revisar el texto reconocido y corregir los fragmentos que fueron reconocidos incorrectamente.

Adobe Acrobat Pro DC cuenta con la herramienta Digitalización y OCR, dentro de la cual está la función de Reconocer texto y Corregir texto reconocido. Dentro de esta última función está la opción Revisar texto reconocido, que permite ver el texto que ha reconocido el programa (Fig. 8 y 9).

## Die Silbergruben in Mexiko.

Mexiko hat, seitdem Cortez es eroberte, bereits 30 Milliarden Gold und Silber in den Handel geliefert, und doch sieht man im Lande selbst keine Spur dieser ungeheuren Reichthümer und auch jetzt verbreitet die Menge Silbers keinen eigentlichen Wohlstand. Beides ist Folge des fehlerhaften Regierungssystems, das Wohlstand durch Verhinderung der Geldausfuhr erzwingen will, während er sich durch Beförderung des Ackerbaues und des Gewerbefleißes von selbst machen würde, und dann würden auch die Bergwerke einträglich werden. Jetzt sind

Figura 8. Texto original (Redacción, 1848)

~fe efiW?B?UWU fa fltgffo.  
:Wt-nifo tat, fritbtm fürtrA el rro&trtr. l,mitl 30 !Dm-  
liarbtn @ol'oun'o Silber in bm ,t;cinl>glfürrt, unb 'ood fie~t  
mattim fanbe felbfl feine ~i,ur 'oicff ungrl)rurm !Jhicl/ttü  
atcrunb audl iett Nttbrtitrt blc !Vltngi Silbrri feintn figmt•  
litben! Botiflanb. Beiotl ifl S'olgrbd frlrirtaiten Olrglrrungl•  
f~tml, bal !Doblflcinburdi /Urrtinbrung ber @tlbauöfutr  
crgmingem>iU,n,ifmnb rr ftd: »urd) !Btfürbrungbt) Sfuftr  
baue&nb 'oel @rtucrbcjeleifitton flbjl ma<fltnh>Urbt, unb  
'Dann>Übtnciud) oic~ttgn>trfc rinrliglld)rottrbn. 3e~t finl  
fitcl n,mig, ba ~r&enmittl !l)uhm, Srilt, 111itlnem !IDor  
it alle! Bebürniffaunt !fütrieb iijjerjl tl)rutt flnb, unl>oCl  
ISil&nn ijolt lebnerf<flitlerth) futr nur fitlt<Ut umvrr•  
t~tt mmen fann. 3n btt !l)tat flnb aucti für brn !Uiafterbr  
in ~anfrtid) 5 ijranfen 30 a:entimrnfoht, in btr !Vlinmge

Figura 9. Texto reconocido (Redacción, 1848)

En la mayoría de los casos el reconocimiento del texto es deficiente por los problemas mencionados arriba acerca de los procesos de impresión de los textos originales y de sus procesos de digitalización. La función *de Reconocer texto* permite identificar y corregir el texto reconocido incorrectamente

seleccionando el recuadro correspondiente a la palabra a corregir y editar el texto en la casilla *reconocido como* (Fig. 10).



Figura 10. Texto reconocido con correcciones para los topónimos encontrados

Ya que es más el texto reconocido incorrectamente que el que sí es reconocido correctamente, el uso de la función *Revisar texto reconocido* es poco práctica. Son pocos los casos en los que se ha reconocido el texto correctamente, por lo que lo más rápido es usar directamente la herramienta *Editar PDF*, que reconoce el texto del documento PDF y permite seleccionar, editar, eliminar o insertar texto. Solo se han hecho correcciones para los topónimos de México en el cuerpo del texto y el resto del texto se ha dejado tal cual como ha sido reconocido por el OCR.

Otra herramienta que ha sido útil en el proceso de edición y corrección de los textos es *Organizar* páginas. Con esta herramienta se han eliminado las páginas en blanco, páginas repetidas por el proceso de digitalización, los escaneos de las pastas duras y los mapas al final de los textos, y se le ha dado la orientación correcta a las páginas que por errores en el proceso de digitalización se encontraban al revés.

Siguiendo los procedimientos descritos en este apartado se ha llegado a una versión del corpus que cuente con textos que sean legibles por computadora y por Python. Aunque los textos en general cuenten con muchos fragmentos incorrectamente reconocidos por el OCR, la corrección manual de los topónimos de México garantiza que estos puedan ser buscados y recuperados usando Python o algún otro procesador de textos.

#### 2.1.4 RECABACIÓN DE DATOS DEL CORPUS

Del corpus compilado se han recabado los datos de interés para esta investigación: los topónimos. Estos se han registrado manualmente para describir su distribución en el corpus y para contar con una lista testigo que sirva de punto de comparación con los resultados obtenidos del proceso de búsqueda de topónimos en el corpus con el lenguaje de programación Python.

#### 2.1.5 REGISTRO DE TOPÓNIMOS DEL CORPUS

La preparación del corpus descrita en los apartados anteriores permite contar con archivos en formato PDF que sean legibles por procesadores de texto y por Python para realizar la búsqueda de topónimos, pero es necesario garantizar

que los resultados obtenidos de estas herramientas sean los mismos que se obtendrían de la revisión “manual” o no asistida por computadora.

Para esto se han contado y registrado los topónimos y el número de páginas de cada texto del corpus de manera manual. Se ha hecho el registro de los topónimos por texto del corpus en un archivo formato CSV (Comma-separated-values por sus siglas en inglés).

Este registro cuenta con cuatro columnas: núm. de texto del corpus, página del texto, fuente de publicación, tipología textual del texto, tipo de topónimo y topónimo.

Este tipo de archivo permite representar datos en forma de tabla y es legible por numerosos procesadores de texto y también por Python. En este registro se han declarado el número de texto del corpus, la página (siguiendo la numeración del texto original), la categoría del topónimo y el topónimo. Además, se ha usado la función *Marcatextos* de Adobe Acrobat para la rápida consulta de los topónimos encontrados en el archivo PDF (Fig. 11).

130 Geographische Zeitung.

bestimmte. In so fern sie sich auf den Ausgangspunkt Malparaiso beziehen, vermehrt werden müssen, um sie auf nähergeheißig wahre Länge zu bringen.

Hr. Prof. Dittmanns kündigt neuere Untersuchungen über die Geographie von Südamerika an, bei denen er, wie wir hören, die Beobachtungen von Matefina und andern spanischen Entdeckern einer neuen Distansion unterwerfen wird; wir sehen ihnen mit Freuden entgegen.

137. Höhenmessungen auf dem Plateau von Mejico. Von Hrn. J. Burkart, Chef des Bergwesens der Bergwerks-Kompagnie von Tlalpujahua.

Hr. Prof. und Oberbergrath Wöggerath hat in Leonhard's Zeitschrift für Mineralogie, Jahrgang 1826, Nr. 7, Julius, die geognostischen Beobachtungen bekannt gemacht, welche Hr. Burkart auf Reisen in Mejico zu sammeln Gelegenheit hatte. Diesen gehaltreichen Beobachtungen ist ein Höhenprofil beigelegt, welches die Hochebene zwischen Tlalpujahua und Atotonilco el Chico darstellt. Hr. Burkart sagt darüber a. a. D. S. 24: „Das angelegene Höhenprofil wird Ihnen (dem Hrn. W. Wöggerath) einigen Aufschluß geben über die relativen Höhen des Vorpöbir- und Tragt: Gebirges, auf dem Wege von Chico nach Tlalpujahua; die Messung geschah mittelst eines engl. Mountain-Barometers, jedoch nur aus einem Standpunkte; sehr unangenehm war es mir, diese Messung nicht auf dem ganzen Wege von Tampico hierher (nach Tlalpujahua) anstellen zu können; ich würde gewiß ein sehr interessantes Profil erhalten haben; doch die Mountain-Barometer eignen sich für die diesigen großen Höhen nicht, da das Quecksilber in ihnen nicht tief genug fallen kann; die- sem Mangel mußte ich erst selbst abhelfen, da man hier Alles ausleichen muß.“

Die gemessenen Höhen sind auf der Linie von Tlalpujahua S. N. h. 4 nach Atotonilco el Chico, N. O. h. 4, über dem Niveau des Meeres im meißtlichen Meerbusen folgende:

Tlalpujahua	8805'
Serro de Sa. Rosa	9643
La Jordana	8774
Rio de Lerma	8575
Changia de las Vacas	9239
Höhe zwischen diesem und dem folgenden Punkte	9839
San Andreas	8986
Höhe zwischen diesem und dem folgenden Punkte	9072
Gilotepec	8345
Hazienda de Tandaja	7949
Serro de San Antonio	9250
Tula	6116
Tetepango	7232
Tulan calco	7530
Rio Jubero	8300

1826, Burkart a.csv

```

texto;página;categoria;topónimo
27;130;país;Mejiko
27;130;municipio;Tlalpujahua
27;130;país;Mejiko
27;130;municipio;Tlalpujahua
27;130;localidad;Atotonilco el Chico
27;130;municipio;Tlalpujahua
27;130;localidad;Chico
27;130;municipio;Tlalpujahua
27;130;localidad;Tampico
27;130;municipio;Tlalpujahua
27;130;municipio;Tlalpujahua
27;130;localidad;Atotonilco el Chico
27;130;municipio;Tlalpujahua
27;130;montaña;Serro de la Sa. Rosa
27;130;localidad;La Jordana
27;130;localidad;Estanzia de las Vacas
27;130;localidad;San Andreas
27;130;municipio;Gilotepec
27;130;localidad;Hazienda de Tandaja
27;130;montaña;Serro de San Antonio
27;130;municipio;Tula
27;130;municipio;Tetepango
27;130;localidad;Tulan calco
27;131;localidad;Ilguautila
27;131;localidad;Istanzuelo
27;131;mina;Purissima
27;131;localidad;Atotonilco el Chico

```

1826, Burkart a

Txt Id	página	categoria	topónimo
27	130	país	Mejiko
27	130	municipio	Tlalpujahua
27	130	país	Mejiko
27	130	municipio	Tlalpujahua
27	130	localidad	Atotonilco el Chico
27	130	municipio	Tlalpujahua
27	130	localidad	Chico
27	130	municipio	Tlalpujahua
27	130	localidad	Tampico
27	130	municipio	Tlalpujahua
27	130	municipio	Tlalpujahua
27	130	localidad	Atotonilco el Chico
27	130	municipio	Tlalpujahua
27	130	montaña	Serro de la Sa. Rosa
27	130	localidad	La Jordana
27	130	localidad	Estanzia de las Vacas
27	130	localidad	San Andreas
27	130	municipio	Gilotepec
27	130	localidad	Hazienda de Tandaja
27	130	montaña	Serro de San Antonio
27	130	municipio	Tula
27	130	municipio	Tetepango
27	130	localidad	Tulan calco
27	131	localidad	Ilguautila
27	131	localidad	Istanzuelo
27	131	mina	Purissima
27	131	localidad	Atotonilco el Chico

Figura 11. De izquierda a derecha, marcación en texto, registro en archivo CSV y su visualización con Python

## 2.2 REVISIÓN Y SUAVIZADO DE DATOS

El registro de topónimos y sus características permite contar con datos que pueden ser analizados con pruebas de significación estadística para comprobar si las relaciones entre distintas variables del corpus son producto del azar o si sí hay una causa subyacente a ellas.

Para saber qué prueba estadística aplicar es necesario conocer la distribución de los datos del corpus y para esto se calculan las medidas de agrupación y dispersión del número de topónimos por texto.

Durante la compilación de los textos del corpus noté que hay algunos textos que tienen considerablemente más topónimos que el resto. Estos textos también

tienen un número de páginas mayor al de la media del corpus y son principalmente publicaciones unitarias (libros).

Esto es problemático para el análisis porque estos pocos textos con un gran número de topónimos distorsionan los datos del corpus completo afectando el valor de la media, lo que afecta los resultados de las pruebas estadísticas. Por lo anterior es necesario descartar los textos que tengan valores atípicos de número de topónimos y garantizar la homogeneidad de la muestra en este aspecto.

Para conocer la manera en la que estos datos del corpus se agrupan y dispersan se grafican los datos y se calcula la media, mediana y moda del número de topónimos por texto (medidas de agrupación), y el rango, rango intercuartil y la desviación estándar (medidas de dispersión). Esto nos permite saber si la distribución de los textos es normal, cuando las media, mediana y moda son iguales o parecidas, o anormal, cuando no.

## 2.2.1 MEDIDAS DE AGRUPACIÓN Y DISPERSIÓN DEL CORPUS

a) Agrupación: media = 200; mediana = 10; moda = 2 (Fig. 12).

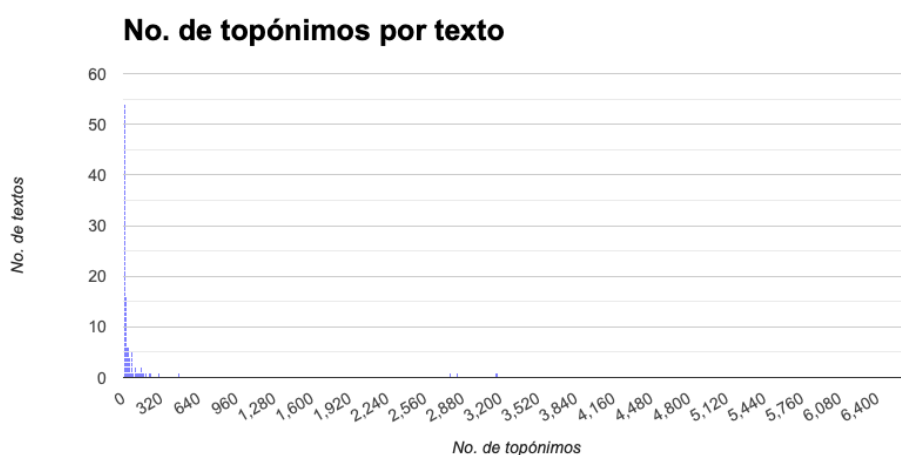


Figura 12. Núm.de topónimos por texto

Ya que los valores de media, mediana y moda no son parecidos, podemos saber que los datos no están distribuidos de manera normal. El histograma muestra también que hay numerosos datos que se alejan considerablemente de estos valores, lo que explica la distancia del valor de la media de la mediana y moda.

b) Dispersión: rango = 6656; rango intercuartil = 38; desviación estándar = 825.1134.

Ya que los datos tienen una distribución asimétrica (no normal) porque la media se ve afectada por los valores extremos, debe considerarse el rango intercuartil como medida de dispersión y la mediana como medida de agrupación.

Para mostrar la distribución (anormal) de los datos, junto con sus valores atípicos, se grafica un diagrama de cajas y bigotes basado en la mediana y el rango intercuartil.

Núm. de topónimos:

- i. Mediana: 10
- ii. Cuartil bajo: 3
- iii. Cuartil alto: 46
- iv. Rango intercuartil: 43
- v. Límite inferior: -61
- vi. Límite superior: 110
- vii. Primera observación por arriba del límite inferior: 1
- viii. Primera observación por abajo del límite superior: 102
- ix. Variaciones atípicas: 113, 126, 133, 142, 144, 155, 162, 181, 218, 222, 294, 468, 2771, 2836, 3169, 3171, 6657



x. Rango: 101

xi. Gráfico (Fig. 13):

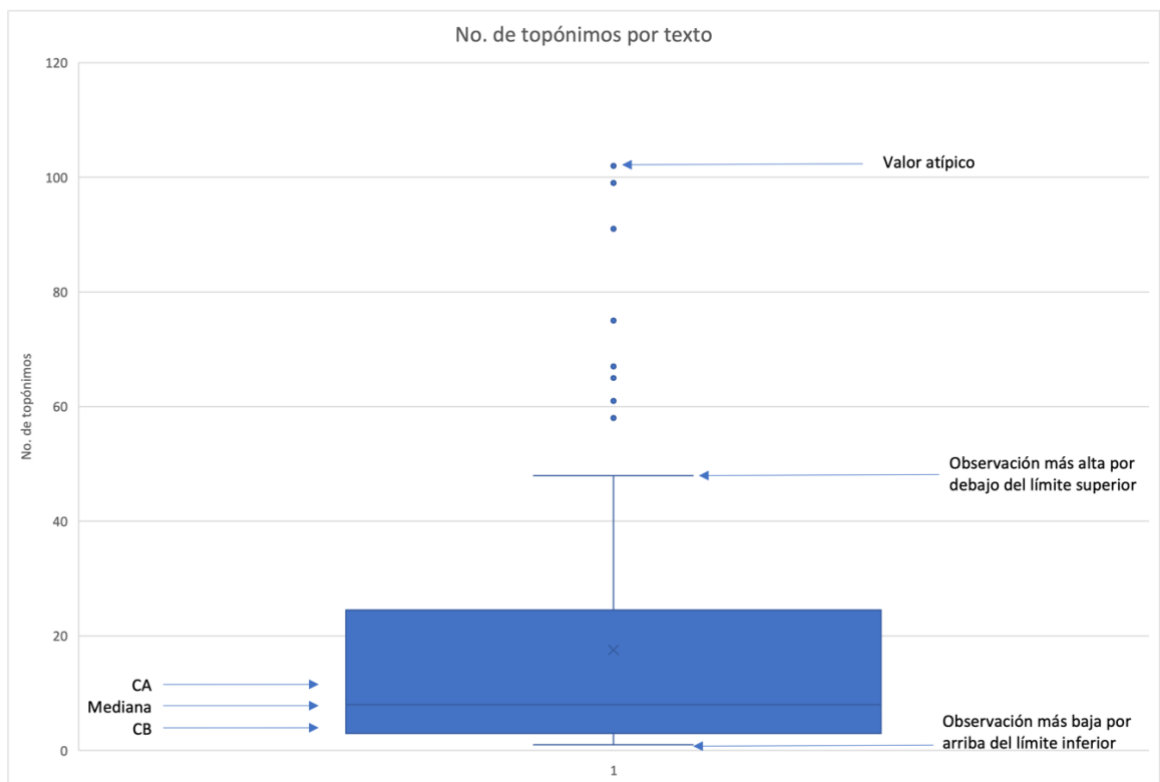


Figura 13. Diagrama de cajas y bigotes del núm. de topónimos por texto

xii. Interpretación del gráfico: como puede observarse, los datos están distribuidos de manera irregular. La mayoría de los datos se encuentran concentrados en la parte inferior del gráfico, lo que indica que la mayoría de los textos tienen entre uno y 23 topónimos. Sin embargo, hay numerosos valores atípicos con valores altos (superiores al límite superior), lo que explica la distribución irregular de los datos.

Después de calcular cómo se agrupan y se dispersan los datos del corpus y de identificar los textos que tienen valores atípicos para número de topónimos, se descartaron estos textos para suavizar los datos y evitar que unos pocos

textos con muchos topónimos afecten los resultados de los tests estadísticos por su gran número de observaciones.

Pese al suavizado de datos, el corpus todavía tiene valores extremos, aunque menos y con menor valor (particularmente en el número de topónimos, donde encontrábamos textos con 113 y hasta 6657 topónimos) (Fig. 14).

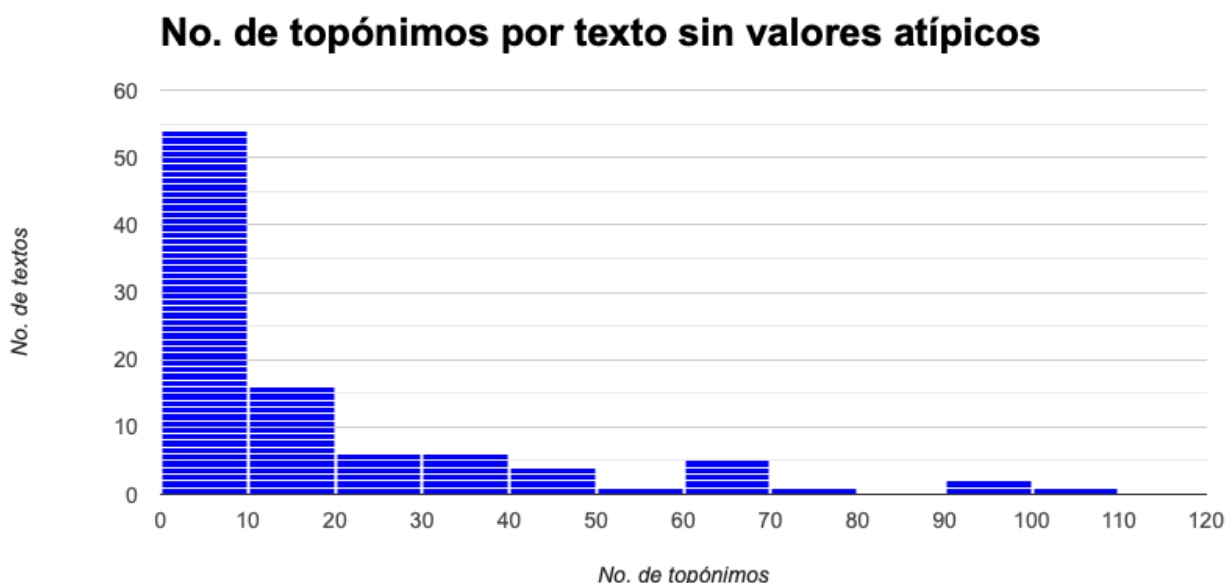


Figura 14. Núm.de topónimos por texto sin valores atípicos

### 2.2.2 CORPUS SIN VALORES ATÍPICOS.

Al descartar los textos con valores atípicos para topónimos quedan 96 textos, con un total de 862 páginas y 1685 topónimos (Anexo B), de los cuales 503 corresponden a *México* y sus variantes *Mexico*, *Mexiko*, *Mejico*, *Mejiko* y *Neuspanien*, y a formas derivadas como composiciones (*Neu-Mexico*) gentilicios (*Mexikaner*) y adjetivos (*mexikanischen*), mismas que se consideran para en análisis, a sabiendas de que estas formas derivadas coinciden con la del topónimo del que provienen.

Este es el topónimo más frecuente en el corpus y también el que más variedades nominales tiene. Por esto se harán las pruebas estadísticas para este

topónimo. Se espera que las pruebas confirmen o rechacen las hipótesis de que hay variables que determinan en qué forma se encontrará cierto topónimo en un texto.

La lista con los textos excluidos del corpus está en el Anexo C. La lista con los diferentes topónimos encontrados en este corpus, incluidas sus variantes nominales y adjetivos derivados, puede encontrarse en el Anexo D.

## 2.3 SELECCIÓN Y APLICACIÓN DE PRUEBAS ESTADÍSTICAS

Los textos que comprenden el corpus no cuentan con una distribución normal pese al suavizado de datos, así que se deben usar pruebas que no consideren la forma de la población, como las no paramétricas de Kolmogorov-Smirnov y la  $\chi^2$  de Bondad de Ajuste.

### 2.3.1 PRUEBA KOLMOGOROV-SMIRNOV

Esta prueba es útil para datos de intervalo, tales como el año y datos ordinales, como la clase de topónimo, aunque se realizará la prueba solo para el 1826. Esta prueba compara las frecuencias observadas contra las frecuencias esperadas (frecuencias teóricas); en el caso de la variable año, la frecuencia esperada sería que la variante *Mexiko*, por ejemplo, se encuentre con la misma frecuencia a lo largo de los años. Se agrupan en lustros los años para obtener un total de 8 grupos.

### 2.3.2 PRUEBA $\chi^2$ DE BONDAD DE AJUSTE

Esta prueba es útil para los datos nominales, como autor, revista y tipología, aunque en este trabajo solo se usará la prueba para revista y tipología. Esta

prueba, así como la de Kolmogorov-Smirnov, se basa en la comparación de las frecuencias observadas y las frecuencias esperadas (teóricas). Contrastaremos nuevamente las frecuencias de ocurrencia de cada variedad nominal con sus apariciones en cada revista y en los textos para cada tipología textual.

Se aplicará primero la prueba para la variable publicación. Esta variable tiene 9 categorías: Libro, *Taschenbuch für die gesammte Mineralogie* (Libro de bolsillo de toda la mineralogía, etiqueta: "T"), *Jahrbuch für Mineralogie, Geognosie, Geologie und Petrefaktenkunde* (Anuario de mineralogía, geognosia, geología y paleontología; en esta categoría se incluye la segunda serie de esta revista, que lleva el mismo nombre, pero con "Neues", es decir, *Nuevo anuario de...*, etiqueta: "J"), *Archiv für Bergbau und Hüttenwesen* (Archivo de minería y metalurgia; en esta categoría se incluye la segunda serie de esta revista, que lleva el nombre de *Archiv für Mineralogie, Geognosie, Bergbau und Hüttenkunde*, Archivo de mineralogía, geognosia, minería y metalurgia, etiqueta: "A"), *Hertha, Zeitschrift für Erd-, Völker und Staatenkunde* (Hertha, periódico de geografía física, etnología y geografía política; en esta categoría se incluye el suplemento anual del periódico llamado *Annalen der Erd-, Völker- und Staatenkunde*, Anales de la geografía física, etnología y geografía política; etiqueta: "H"), *Berg- und Hüttenmännischen Zeitung* (Periódico de minería y siderurgia; etiqueta: "B"), *Annalen der Physik und Chemie* (Anales de la física y química; etiqueta: "P"), *Journal für Chemie und Physik* (Revista de química y física; etiqueta: C), y *Fortschritte der Geographie und Naturgeschichte* (Avances de la geografía e historia natural; etiqueta: "F").

## 2.4 PROCESAMIENTO DE TEXTOS PARA LA EXTRACCIÓN DE DATOS DE ARCHIVOS CSV Y LOCALIZACIÓN DE TOPÓNIMOS DE ARCHIVOS PDF CON PYTHON

Python es un lenguaje de programación inventado en 1991 que ha ganado gran popularidad en los últimos años por su sintaxis sencilla y que ha probado ser bastante útil para las tareas relacionadas con el procesamiento de lenguaje natural (Bird, S., Klein, E., y Loper, E., 2019).

Con este lenguaje de programación es posible realizar tareas que normalmente forman parte de cualquier proceso de tratamiento de corpora, entre ellas la lectura de archivos. Es posible leer distintos formatos con Python y en esta investigación se usa para leer los archivos CSVs de los registros de topónimos en los textos y extraer informaciones específicas de ellos para analizarlos con las pruebas estadísticas.

### 2.4.1 AMBIENTE

Para garantizar que el código escrito para esta investigación sea de acceso libre se ha escrito en Google Colaboratory. Google Colaboratory es un producto de Google Research que permite a cualquier usuario escribir, ejecutar y compartir código en Python en el navegador. El proyecto se ha dividido en dos partes: el programa para la extracción de información de archivos en formato CVS y el programa para la localización de topónimos en archivos con formato PDF.

Debido a que el tamaño de los archivos en PDF es de casi un gigabyte, la carga y procesamiento de la totalidad de estos archivos puede ser muy tardada en Google Colaboratory, por lo que también se pone a disposición del usuario el código de Python para su uso en otros ambientes, como Jupyter Notebook,

aunque esto conlleve la instalación de Python y un mayor conocimiento de su uso. Los códigos y textos pueden encontrarse en el siguiente repositorio de Github: [https://github.com/Vosatovocko/Corpus\\_VillanuevaErick](https://github.com/Vosatovocko/Corpus_VillanuevaErick).

#### 2.4.2 IMPORTACIÓN DE MÓDULOS

Python permite a los usuarios importar módulos externos, es decir, conjuntos de funciones que pueden usarse sin necesidad de escribir nuevamente el código necesario para ejecutar estas funciones.

Para esta investigación se han utilizado los siguientes módulos para la lectura, extracción y análisis de datos de archivos csv:

- collections: módulo para servir como contenedor de colecciones de datos, como listas, diccionarios, sets y tuplas.
- csv: módulo para leer, escribir y modificar archivos en formato csv.
- matplotlib: módulo para visualización de datos y ploteo de gráficos.
- NumPy: módulo para la gestión de arreglos (estructuras de datos).
- Pandas: módulo para la gestionar y analizar estructuras de datos.
- google.colab: módulo para importar archivos del sistema local a Google Colaboratory.

Para la localización de topónimos en el corpus se han utilizado los siguientes módulos:

- pdfminer: módulo para la extracción de información de archivos en formato pdf.
- io: módulo para la transformación de datos para la comunicación entre la computadora y el usuario.

- Spacy: módulo para realizar diversas tareas de procesamiento de lenguaje natural, incluido el uso de Matchers y reconocimiento de entidades.
- re: módulo para crear expresiones regulares.
- google.colab: módulo para importar archivos del sistema local a Google Colaboratory.

### 2.3.3 PROCESAMIENTO DE TEXTOS PARA LA LOCALIZACIÓN DE TOPÓNIMOS EN PDFS CON PYTHON

La última parte de la investigación consiste en desarrollar un programa con el lenguaje de programación Python que nos permita aplicar los conocimientos recabados de los marcos teóricos anteriores a la localización automática de los topónimos y sus variedades nominales.

El investigador puede especificar una serie de características con Python que refinen su búsqueda. En el caso de esta investigación se consideran principalmente las capacidades de Python para usar expresiones regulares y matchers para poder hacer uso de caracteres no definidos que permitan la búsqueda de una cadena de caracteres con alguno de estos caracteres, o más, sin especificar. Un ejemplo sencillo de esto sería “Guana\*uato”, donde “\*” representa cualquier carácter; un ejemplo utilizado de matcher es declarar *Nueva España* como `([Nueva]+[-]+[España])+( [Neu]+[ ]+[Spanien])+([Neuspanien].)`

### 2.3.3.1 EVALUACIÓN DE RESULTADOS DE LA LOCALIZACIÓN CON PYTHON

Para medir el desempeño del buscador de topónimos se calcula la precisión (*precision*), exhaustividad (*recall*) y exactitud (*accuracy*) para calcular el F1-Score. El F1-Score es una forma de medir el rendimiento de sistemas en procesamiento de lenguaje natural usando una escala del 0 al 1, siendo 1 un resultado perfecto (Derczynsk, 2016, pp. 261-264).

Estos valores se calculan a partir del número de verdaderos positivos y negativos, y los falsos positivos y negativos. Los verdaderos positivos son los topónimos que están presentes en el texto y que el programa también ha encontrado y los verdaderos negativos son los topónimos que no están en el texto y que el programa tampoco ha reconocido; los falsos positivos son los topónimos que no están presentes en el texto y que el programa sí ha reconocido y los falsos negativos son los topónimos que sí están en el texto pero que el programa no ha reconocido.

Los valores que hay que calcular primeramente son:

- La precisión: indica qué tan preciso es el sistema al encontrar los positivos.

$$\text{Precisión} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos positivos}} = \frac{\text{Verdaderos positivos}}{\text{Total de verdaderos predichos}}$$

- La exhaustividad: indica cuántos positivos reales ha capturado el sistema.

$$\text{Exhaustividad} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos negativos}} = \frac{\text{Verdaderos positivos}}{\text{Total de verdaderos (P o N)}}$$

- La exactitud indica cuántos casos ha acertado el sistema.

$$\text{Exactitud} = \frac{\text{Verdaderos positivos} + \text{Verdaderos negativos}}{\text{Verdaderos positivos} + \text{Verdaderos negativos} + \text{Falsos positivos} + \text{Falsos negativos}}$$



El cálculo de la precisión y la exhaustividad nos permite calcular el F1-Score, la función para expresar el equilibrio entre estas, con la siguiente fórmula:

$$F1 = 2 * \frac{Precisión * Exhaustividad}{Precisión + Exhaustividad}$$

Para evaluar el rendimiento del programa se han elegido los textos de 1826 porque es el año con más textos registrados en el corpus y año en el que se encuentran todas las variantes nominales de *México*.

# RESULTADOS

## 3.1 DESCRIPCIÓN DEL CORPUS

La compilación inicial, edición y selección de textos descritas en el capítulo de metodología ha resultado en el corpus de esta investigación. Este comprende los 96 textos (Apéndice A), 862 páginas y 1685 topónimos.

La distribución de páginas y topónimos entre los textos del corpus es irregular, pese al suavizado de datos. La extensión mínima de los textos del corpus es de una página, mientras que el texto más largo es de 205 páginas; por otra parte, el texto que menos topónimos tiene cuenta con un topónimo, mientras que el que más topónimos tiene cuenta con 102.

La diferencia más notable entre los textos, con relación a su número de páginas y topónimos, puede caracterizarse en términos de su formato: el corpus cuenta, por una parte, con textos con formato de artículo publicado en una revista, anuario o boletín, y, por otra parte, con textos con formato de libro independiente. Es en este último formato en el que encontramos mayor concentración de topónimos por página, además de mayor número de páginas en general.

## 3.2 EVALUACIÓN DE RESULTADOS DE LA LOCALIZACIÓN DE TOPÓNIMOS

Los resultados de la localización de topónimos por texto se describen en la tabla a continuación:

<b>Texto: 22</b>	
<b>Topónimos</b>	<b>4</b>
<b>Encontrados</b>	<b>4</b>
Total palabras	2779
TP	4
TN	2772
FP	0
FN	0
<b>Total</b>	<b>2776</b>
Exactitud	1
Exhaustividad	1
Precisión	1
<b>F1-Score</b>	<b>1</b>

<b>Texto: 23</b>	
<b>Topónimos:</b>	<b>3</b>
<b>Encontrados:</b>	<b>3</b>
Total palabras	736
TP:	3
TN:	733
FP:	0
FN:	0
<b>Total</b>	<b>736</b>
Exactitud	1
Exhaustividad	1
Precisión	1
<b>F1-Score</b>	<b>1</b>

<b>Texto: 24</b>	
<b>Topónimos</b>	<b>8</b>
<b>Encontrados</b>	<b>8</b>
Total palabras	1854
TP	8
TN	1846
FP	0
FN	0
<b>Total</b>	<b>1854</b>
Exactitud	1
Exhaustividad	1
Precisión	1
<b>F1-Score</b>	<b>1</b>

<b>Texto: 26</b>	
<b>Topónimos</b>	<b>10</b>
<b>Encontrados</b>	<b>10</b>
Total palabras	2350
TP	10
TN	2340
FP	0
FN	0
<b>Total</b>	<b>2350</b>
Exactitud	1
Exhaustividad	1
Precisión	1
<b>F1-Score</b>	<b>1</b>

<b>Texto: 27</b>	
<b>Topónimos</b>	<b>3</b>
<b>Encontrados</b>	<b>3</b>
Total palabras	1027
TP	3
TN	1024
FP	0
FN	0
<b>Total</b>	<b>1027</b>
Exactitud	1
Exhaustividad	1
Precisión	1
<b>F1-Score</b>	<b>1</b>

<b>Texto: 28</b>	
<b>Topónimos</b>	<b>3</b>
<b>Encontrados</b>	<b>3</b>
Total palabras	1725
TP	3
TN	1722
FP	0
FN	0
<b>Total</b>	<b>1725</b>
Exactitud	1
Exhaustividad	1
Precisión	1
<b>F1-Score</b>	<b>1</b>

<b>Texto: 29</b>	
<b>Topónimos</b>	<b>1</b>
<b>Encontrados</b>	<b>1</b>
Total palabras	659
TP	1
TN	658
FP	0
FN	0
<b>Total</b>	<b>659</b>
Exactitud	1
Exhaustividad	1
Precisión	1
<b>F1-Score</b>	<b>1</b>

<b>Texto: 30</b>	
<b>Topónimos</b>	<b>1</b>
<b>Encontrados</b>	<b>1</b>
Total palabras	263
TP	1
TN	262
FP	0
FN	0
<b>Total</b>	<b>263</b>
Exactitud	1
Exhaustividad	1
Precisión	1
<b>F1-Score</b>	<b>1</b>

<b>Texto: 31</b>	
<b>Topónimos</b>	<b>2</b>
<b>Encontrados</b>	<b>2</b>
Total palabras	245
TP	2
TN	243
FP	0
FN	0
<b>Total</b>	<b>245</b>
Exactitud	1
Exhaustividad	1
Precisión	1
<b>F1-Score</b>	<b>1</b>

Texto: 32	
<b>Topónimos</b>	<b>2</b>
<b>Encontrados</b>	<b>2</b>
Total palabras	720
TP	2
TN	718
FP	0
FN	0
<b>Total</b>	<b>720</b>
Exactitud	1
Exhaustividad	1
Precisión	1
<b>F1-Score</b>	<b>1</b>

Texto: 33	
<b>Topónimos</b>	<b>1</b>
<b>Encontrados</b>	<b>1</b>
Total palabras	459
TP	1
TN	458
FP	0
FN	0
<b>Total</b>	<b>459</b>
Exactitud	1
Exhaustividad	1
Precisión	1
<b>F1-Score</b>	<b>1</b>

Texto: 34	
<b>Topónimos</b>	<b>1</b>
<b>Encontrados</b>	<b>1</b>
Total palabras	190
TP	1
TN	189
FP	0
FN	0
<b>Total</b>	<b>190</b>
Exactitud	1
Exhaustividad	1
Precisión	1
<b>F1-Score</b>	<b>1</b>

Tabla 3. Resultados de la localización de topónimos

### 3.2.1 CONCLUSIONES DE LA EVALUACIÓN DE RESULTADOS DE LA LOCALIZACIÓN DE TOPÓNIMOS

En todos los textos se alcanzó el mejor resultado de F1-Score (1). Esto indica que el programa ha reconocido todos los topónimos del texto. Este es un muy buen resultado, aunque debemos considerar también que las condiciones de los textos son muy particulares: nos hemos asegurado manualmente de que los topónimos estén correctamente escritos en los textos y esto pudo haber sesgado también los resultados. El uso de este programa con otros textos que no hayan sido editados no garantiza su localización y todavía es dependiente de la legibilidad de los textos. Por otra parte, una vez garantizada la legibilidad de los textos, el programa es capaz de localizar los topónimos de *México* y sus variantes nominales y formas derivadas.

### 3.3 RESULTADOS DE LAS PRUEBAS ESTADÍSTICAS

En este apartado se muestran los procedimientos y resultados de la aplicación de las pruebas estadísticas a los datos brutos obtenidos de Python. Las pruebas

se basan en la comparación de frecuencias teóricas y frecuencias observadas y es necesario aplicar las pruebas para cada variante que se quiera analizar.

### 3.3.1 Kolmogorov Smirnov - Año

Para tener una primera aproximación a la relación entre la alonimia de *México* y el año de aparición de sus variantes se grafica su serie de tiempo (Fig. 15). Las hipótesis, el procedimiento y las conclusiones se plantean para cada variante nominal de la siguiente manera:

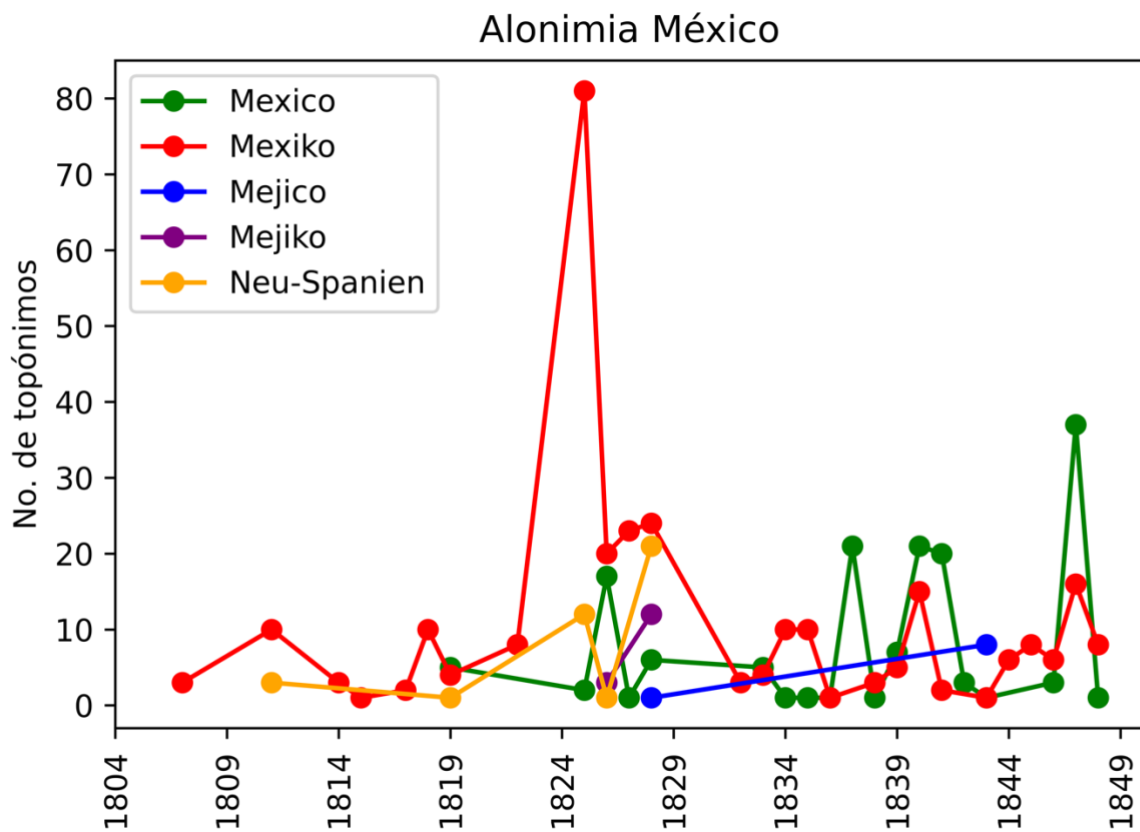


Figura 15. Serie de tiempo de la alonimia de México

VARIABLE DEPENDIENTE: *Mexico*

1. Hipótesis:

- i.  $H_0 = F_o(x) = F_e(x)$  o Hipótesis nula: la frecuencia observada del topónimo *Mexico* a lo largo de los años es la misma que la frecuencia esperada que asume una distribución normal.
- ii.  $H_1 = F_o(x) \neq F_e(x)$  o Hipótesis alterna: la frecuencia observada del topónimo *Mexico* a lo largo de los años es diferente de la frecuencia esperada.

2. Cálculo de frecuencias acumulativas para las distribuciones observadas y teóricas y del valor crítico: la variante *Mexico* y sus derivados ocurre 154 veces a lo largo del periodo del corpus, al dividirlo entre los 10 periodos del corpus se obtiene que *Mexico* ocurre 15.4 veces por grupo, es decir, por cada intervalo de 5 años, esta es la frecuencia esperada ( $F_e$ ). El procedimiento del cálculo se presenta en el Tabla 4. Para un nivel de significación  $\alpha=0.01$ , unidireccional, y  $n=154$ , el valor crítico (VC, calculado mediante la fórmula  $VC=1.63/\sqrt{n}$ ) =0.131.

<b><i>Mexico</i> por lustro</b>					
GRUPOS	NÚM. DE APARICIONES		Fo(x)	Fe(x)	Fo(x)-Fe(x)
	Fo	Fe			
1800-1804	0	15.4	0	0.1	-0.1
1805-1809	0	15.4	0	0.2	-0.2
1810-1814	0	15.4	0	0.3	-0.3
1815-1819	5	15.4	0.03246753	0.4	-0.3675325
1820-1824	0	15.4	0.03246753	0.5	-0.4675325
1825-1829	26	15.4	0.2012987	0.6	-0.3987013
1830-1834	6	15.4	0.24025974	0.7	-0.4597403
1835-1839	31	15.4	0.44155844	0.8	-0.3584416
1840-1844	45	15.4	0.73376623	0.9	-0.1662338
1845-1849	41	15.4	1	1	0
<b>n=</b>	<b>154</b>			<b> D =</b>	<b>0.4675</b>
				<b><math>\alpha</math>=</b>	<b>0.1313</b>

Tabla 4. Cálculo de frecuencias acumulativas para las distribuciones observadas y teóricas de *Mexico* por año

3. Conclusiones: ya que  $|D| > VC$  podemos rechazar la hipótesis nula y concluir que existe una diferencia en la frecuencia de aparición del topónimo “Mexico” a lo largo de los años.

VARIABLE DEPENDIENTE: *Mexiko*

1. Hipótesis:

- i.  $H_0 = F_o(x) = F_e(x)$  o Hipótesis nula: la frecuencia observada del topónimo “Mexiko” a lo largo de los años es la misma que la frecuencia esperada.
- ii.  $H_1 = F_o(x) \neq F_e(x)$  o Hipótesis alterna: la frecuencia observada del topónimo “Mexiko” a lo largo de los años es diferente de la frecuencia esperada.

2. Cálculo de frecuencias acumulativas para las distribuciones observadas y teóricas y del valor crítico: la variante *Mexiko* y sus derivados ocurre 287 veces a lo largo del periodo del corpus, al dividirlo entre los 10 periodos del corpus se obtiene que *Mexiko* ocurre 28.7 veces por grupo, esta es la  $F_e$ . El procedimiento del cálculo se presenta en el Tabla 5. Para un nivel de significación  $\alpha=0.01$ , unidireccional, y  $n=287$ , el valor crítico=0.0962.

<b>Mexiko</b> por lustro					
GRUPOS	NÚM. DE APARICIONES		Fo(x)	Fe(x)	Fo(x)-Fe(x)
	Fo	Fe			
1800-1804	0	28.7	0	0.1	-0.1
1805-1809	3	28.7	0.01045296	0.2	-0.189547
1810-1814	13	28.7	0.05574913	0.3	-0.2442509
1815-1819	17	28.7	0.11498258	0.4	-0.2850174
1820-1824	8	28.7	0.14285714	0.5	-0.3571429
1825-1829	148	28.7	0.65853659	0.6	0.05853659
1830-1834	17	28.7	0.71777003	0.7	0.01777003
1835-1839	19	28.7	0.78397213	0.8	-0.0160279
1840-1844	24	28.7	0.86759582	0.9	-0.0324042
1845-1849	38	28.7	1	1	0
<b>n=</b>	<b>287</b>			<b> D =</b>	<b>0.3571</b>
				<b>α=</b>	<b>0.096</b>

Tabla 5. Cálculo de frecuencias acumulativas para las distribuciones observadas y teóricas de *Mexiko* por año

3. Conclusiones: ya que  $|D| > VC$  podemos rechazar la hipótesis nula y concluir que existe una diferencia en la frecuencia de aparición del topónimo *Mexiko* a lo largo de los años.

VARIABLE DEPENDIENTE: *Mejico*

1. Hipótesis:

- i.  $H_0 = F_o(x) = F_e(x)$  o Hipótesis nula: la frecuencia observada del topónimo "Mejico" a lo largo de los años es la misma que la frecuencia esperada.
- ii.  $H_1 = F_o(x) \neq F_e(x)$  o Hipótesis alterna: la frecuencia observada del topónimo "Mejico" a lo largo de los años es diferente de la frecuencia esperada.



2. Cálculo de frecuencias acumulativas para las distribuciones observadas y teóricas y del valor crítico: la variante *Mejico* y sus derivados ocurre 9 veces a lo largo del periodo del corpus, al dividirlo entre los 10 periodos del corpus se obtiene que *Mejico* ocurre 0.9 veces por grupo, esta es la  $F_e$ . El procedimiento del cálculo se presenta en el Tabla 6. Para un nivel de significación  $\alpha=0.01$ , unidireccional, y  $n=9$ , el valor crítico (consultando la tabla de valores críticos de  $D$  porque  $n<35$ )  $=0.514$ .

<b><i>Mejico</i> por lustro</b>					
GRUPOS	NÚM. DE APARICIONES		Fo(x)	Fe(x)	Fo(x)-Fe(x)
	Fo	Fe			
1800-1804	0	0.9	0	0.1	-0.1
1805-1809	0	0.9	0	0.2	-0.2
1810-1814	0	0.9	0	0.3	-0.3
1815-1819	0	0.9	0	0.4	-0.4
1820-1824	0	0.9	0	0.5	-0.5
1825-1829	1	0.9	0.11111111	0.6	-0.4888889
1830-1834	0	0.9	0.11111111	0.7	-0.5888889
1835-1839	0	0.9	0.11111111	0.8	-0.6888889
1840-1844	8	0.9	1	0.9	0.1
1845-1849	0	0.9	1	1	0
<b>n=</b>	<b>9</b>			<b> D =</b>	<b>0.6888</b>
				<b><math>\alpha=</math></b>	<b>0.514</b>

Tabla 6. Cálculo de frecuencias acumulativas para las distribuciones observadas y teóricas de *Mejico* por año

3. Conclusiones: ya que  $|D|>VC$  podemos rechazar la hipótesis nula y concluir que existe una diferencia en la frecuencia de aparición del topónimo *Mejico* a lo largo de los años.

VARIABLE DEPENDIENTE: *Mejiko*

1. Hipótesis:

- i.  $H_0 = F_o(x) = F_e(x)$  o Hipótesis nula: la frecuencia observada del topónimo *Mejiko* a lo largo de los años es la misma que la frecuencia esperada.
- ii.  $H_1 = F_o(x) \neq F_e(x)$  o Hipótesis alterna: la frecuencia observada del topónimo *Mejiko* a lo largo de los años es diferente de la frecuencia esperada.

2. Cálculo de frecuencias acumulativas para las distribuciones observadas y teóricas y del valor crítico: la variante *Mejiko* y sus derivados ocurre 15 veces a lo largo del periodo del corpus, al dividirlo entre los 9 periodos del corpus se obtiene que *Mejiko* ocurre 1.5 veces por grupo, esta es la  $F_e$ . El procedimiento del cálculo se presenta en el Tabla 7. Para un nivel de significación  $\alpha=0.01$ , unidireccional, y  $n=14$ , el valor crítico (consultando la tabla de valores críticos de  $D$  porque  $n < 35$ )  $= 0.404$ .

<b>Mejiko por lustro</b>					
GRUPOS	NÚM. DE APARICIONES		Fo(x)	Fe(x)	Fo(x)-Fe(x)
	Fo	Fe			
1800-1804	0	1.5	0	0.1	-0.1
1805-1809	0	1.5	0	0.2	-0.2
1810-1814	0	1.5	0	0.3	-0.3
1815-1819	0	1.5	0	0.4	-0.4
1820-1824	0	1.5	0	0.5	-0.5
1825-1829	15	1.5	1	0.6	0.4
1830-1834	0	1.5	1	0.7	0.3
1835-1839	0	1.5	1	0.8	0.2
1840-1844	0	1.5	1	0.9	0.1
1845-1849	0	1.5	1	1	0
<b>n=</b>	<b>15</b>			<b> D =</b>	<b>0.5</b>
				<b><math>\alpha</math>=</b>	<b>0.404</b>

Tabla 7. Cálculo de frecuencias acumulativas para las distribuciones observadas y teóricas de *Mejiko* por año

3. Conclusiones: ya que  $|D| > VC$  podemos rechazar la hipótesis nula y concluir que existe una diferencia en la frecuencia de aparición del topónimo *Mejiko* a lo largo de los años.

VARIABLE DEPENDIENTE: *Neuspanien*

1. Hipótesis:

- i.  $H_0 = F_o(x) = F_e(x)$  o Hipótesis nula: la frecuencia observada del topónimo *Neuspanien* a lo largo de los años es la misma que la frecuencia esperada que asume una distribución normal.
- ii.  $H_1 = F_o(x) \neq F_e(x)$  o Hipótesis alterna: la frecuencia observada del topónimo *Neuspanien* a lo largo de los años es diferente de la frecuencia esperada.

2. Cálculo de frecuencias acumulativas para las distribuciones observadas y teóricas y del valor crítico: la variante *Neuspanien* y sus derivados ocurre 38 veces a lo largo del periodo del corpus, al dividirlo entre los 10 periodos del corpus se obtiene que *Neuspanien* ocurre 3.8 veces por grupo; esta es la  $F_e$ . El procedimiento del cálculo se presenta en el Tabla 8. Para un nivel de significación  $\alpha=0.01$ , unidireccional, y  $n=38$ , el valor crítico=0.618.

<b>Neuspanien</b> y variantes por lustro					
GRUPOS	NÚM. DE APARICIONES		Fo(x)	Fe(x)	Fo(x)-Fe(x)
	Fo	Fe			
1800-1804	0	3.8	0	0.1	-0.1
1805-1809	0	3.8	0	0.2	-0.2
1810-1814	3	3.8	0.07894737	0.3	-0.2210526
1815-1819	1	3.8	0.10526316	0.4	-0.2947368
1820-1824	0	3.8	0.10526316	0.5	-0.3947368
1825-1829	34	3.8	1	0.6	0.4
1830-1834	0	3.8	1	0.7	0.3
1835-1839	0	3.8	1	0.8	0.2
1840-1844	0	3.8	1	0.9	0.1
1845-1849	0	3.8	1	1	0
<b>n=</b>	<b>38</b>			<b> D =</b>	<b>0.3947</b>
				<b>α=</b>	<b>0.264</b>

Tabla 8. Cálculo de frecuencias acumulativas para las distribuciones observadas y teóricas de *Neuspanien* por año.

3. Conclusiones: ya que  $|D| > VC$  podemos rechazar la hipótesis nula y concluir que existe una diferencia en la frecuencia de aparición del topónimo *Neuspanien* a lo largo de los años.

### 3.3.1.1 CONCLUSIONES GENERALES DE LA APLICACIÓN DE LA PRUEBA KOLMOGOROV-SMIRNOV PARA LA VARIABLE INDEPENDIENTE AÑO

Se rechazaron todas las hipótesis nulas, lo que indica que todas las variedades nominales tienen una distribución particular a lo largo del periodo del corpus. Se ha podido comprobar que el año es una variable que influye en la variedad nominal encontrada en el corpus. La variante nominal más utilizada es la de *Mexiko*, misma que ocurre en casi todos los grupos.

### 3.3.2 $\chi^2$ DE BONDAD DE AJUSTE - FUENTE DE PUBLICACIÓN

Esta prueba es útil para los datos nominales, como autor, fuente de publicación y tipología, aunque en este trabajo solo se usará la prueba para fuente de publicación y tipología. Esta prueba, así como la de Kolmogorov-Smirnov, se basa en la comparación de las frecuencias observadas y las frecuencias esperadas (teóricas). Contrastaremos nuevamente las frecuencias de ocurrencia de cada variedad nominal con sus apariciones en cada revista y en los textos para cada tipología textual.

Se aplicará primero la prueba para la variable revista. Esta variable tiene 9 categorías: Libro (etiqueta: "L"), *Taschenbuch für die gesammte Mineralogie* (Libro de bolsillo de toda la mineralogía, etiqueta: "T"), *Jahrbuch für Mineralogie, Geognosie, Geologie und Petrefaktenkunde* (Anuario de mineralogía, geognosia, geología y paleontología; en esta categoría se incluye la segunda serie de esta revista, que lleva el mismo nombre, pero con "Neues", es decir, *Nuevo anuario de...*, etiqueta: "J"), *Archiv für Bergbau und Hüttenwesen* (Archivo de minería y metalurgia; en esta categoría se incluye la segunda serie de esta revista, que lleva el nombre de *Archiv für Mineralogie, Geognosie, Bergbau und Hüttenkunde*, Archivo de mineralogía, geognosia, minería y metalurgia, etiqueta: "A"), *Hertha, Zeitschrift für Erd-, Völker und Staatenkunde* (Hertha, periódico de geografía física, etnología y geografía política; en esta categoría se incluye el suplemento anual del periódico llamado *Annalen der Erd-, Völker- und Staatenkunde*, Anales de la geografía física, etnología y geografía política; etiqueta: "H"), *Berg- und Hüttenmännischen Zeitung* (Periódico de minería y siderurgia; etiqueta: "B"), *Annalen der Physik und Chemie* (Anales de la física y química; etiqueta: "P"), *Journal für Chemie und Physik* (Revista de química y física; etiqueta: C), y

*Fortschritte der Geographie und Naturgeschichte* (Avances de la geografía e historia natural; etiqueta: “F”).

Para tener una primera aproximación a la relación entre las variedades nominales de *México* y las revistas en las que aparecen se presentan estos datos en una gráfica de barras apilada (Fig. 16). Las hipótesis, el procedimiento y las conclusiones se plantean para cada variedad nominal de la siguiente manera:

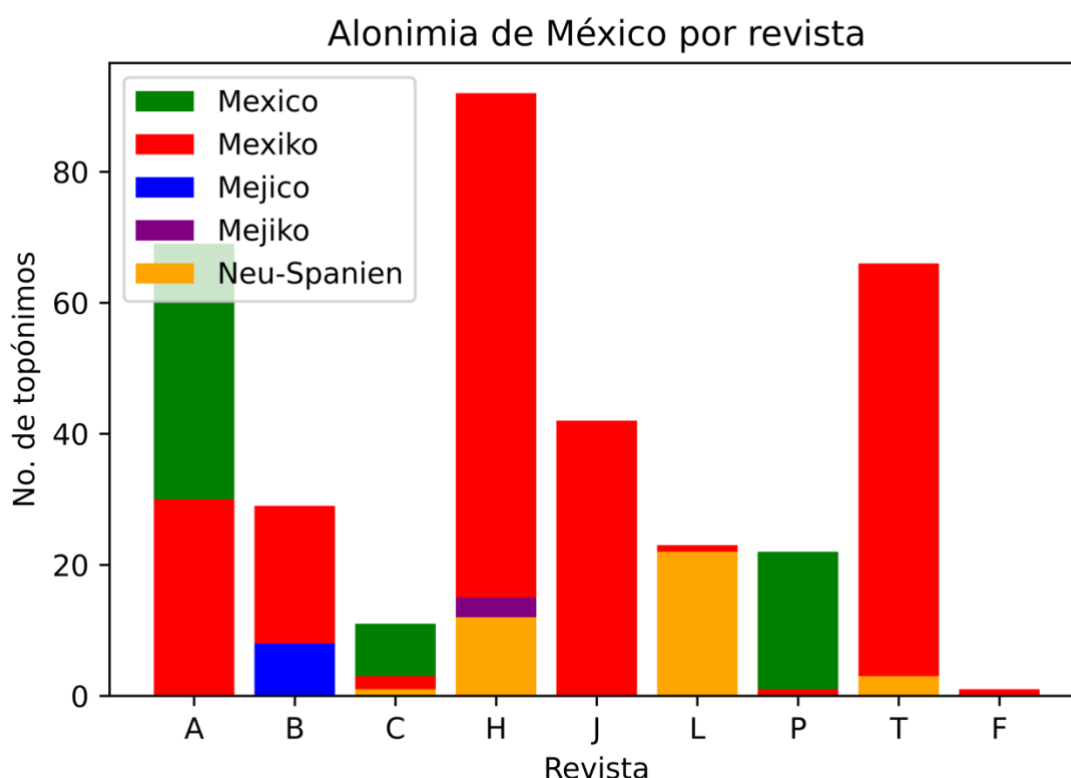


Figura 16. Variedades nominales de *México* por fuente

VARIABLE DEPENDIENTE: *Mexico*

1. Hipótesis:

- i. H0= la frecuencia observada del topónimo *Mexico* es la misma entre las revistas del corpus.
- ii. H1= la frecuencia observada del topónimo *Mexico* es diferente entre las revistas del corpus.

2. Cálculo de frecuencias esperadas para cada frecuencia observada y del valor crítico: la variante *Mexico* y sus derivados ocurre 154 veces en los textos del corpus, al dividirlo entre las 9 fuentes del corpus se obtiene que *Mexico* ocurre 17.11 veces por grupo; esta es la  $F_e$ . El procedimiento del cálculo se presenta en el Tabla 9. Para un nivel de significación  $\alpha=0.01$ , unidireccional, y  $gl=8$ , el valor crítico (consultando la tabla de  $\chi^2$ ) =20.09.

Prueba $\chi^2$ de bondad de ajuste										
<i>Mexico</i> por fuente de publicación										
	T	L	J	A	H	B	P	C	F	TOTALES
Fo	2	11	4	11	22	69	30	5	0	154
Fe	17.11	17.11	17.11	17.11	17.11	17.11	17.11	17.11	17.11	153.99

$\chi^2_{obt} =$										
$\sum((f_o - f_e)^2 / f_e)$	13.34	2.18	10.04	2.18	1.39	157.36	9.71	8.57	17.11	<b>221.88</b>
										$\alpha =$ <b>20.09</b>

Tabla 9. Cálculo de frecuencias esperadas y frecuencias observadas por revista para *Mexico*

3. Conclusiones: ya que  $\chi^2_{obt} > \chi^2_{crit}$  podemos rechazar la hipótesis nula y concluir que existe una diferencia en la frecuencia de aparición del topónimo *Mexico* entre las fuentes del corpus.

VARIABLE DEPENDIENTE: *Mexiko*

1. Hipótesis:

- i.  $H_0$ = la frecuencia observada del topónimo *Mexiko* es la misma entre las revistas del corpus.
- ii.  $H_1$ = la frecuencia observada del topónimo *Mexiko* es diferente entre las revistas del corpus.

2. Cálculo de frecuencias esperadas para cada frecuencia observada y del valor crítico: la variante *Mexiko* y sus derivados ocurre 287 veces en los textos del corpus, al dividirlo entre las 9 fuentes del corpus se obtiene que *Mexiko* ocurre 31.88 veces por grupo; esta es la  $F_e$ . El procedimiento del cálculo se presenta en el Tabla 10. Para un nivel de significación  $\alpha=0.01$ , unidireccional, y  $gl=8$ , el valor crítico (consultando la tabla de  $\chi^2$ ) =20.09.

Prueba $\chi^2$ de bondad de ajuste										
<i>Mexiko</i> por fuente de publicación										
	T	L	J	A	H	B	P	C	F	TOTALES
Fo	66	23	92	3	1	30	42	29	1	287
Fe	31.88	31.88	31.88	31.88	31.88	31.88	31.88	31.88	31.88	286.92

$\chi^2_{obt} =$											
$\sum((f_o - f_e)^2)/f_e$	36.51	2.47	113.37	26.16	29.91	0.11	3.21	0.26	29.91	<b>241.93</b>	
										$\alpha =$	<b>20.09</b>

Tabla 10. Cálculo de frecuencias esperadas y frecuencias observadas por revista para *Mexiko*

3. Conclusiones: ya que  $\chi^2_{obt} > \chi^2_{crit}$  podemos rechazar la hipótesis nula y concluir que existe una diferencia en la frecuencia de aparición del topónimo *Mexiko* entre las fuentes del corpus.

VARIABLE DEPENDIENTE: “Mejico”

1. Hipótesis:

- i.  $H_0$ = la frecuencia observada del topónimo *Mejico* es la misma entre las revistas del corpus.
- ii.  $H_1$ = la frecuencia observada del topónimo *Mejico* es diferente entre las revistas del corpus.



2. Cálculo de frecuencias esperadas para cada frecuencia observada y del valor crítico: la variante *Mejico* y sus derivados ocurre 9 veces en los textos del corpus, al dividirlo entre las 9 fuentes del corpus se obtiene que *Mejico* ocurre 1 vez por grupo; esta es la  $F_e$ . El procedimiento del cálculo se presenta en la Tabla 11. Para un nivel de significación  $\alpha=0.01$ , unidireccional, y  $gl=8$ , el valor crítico (consultando la tabla de  $\chi^2$ ) =20.09.

		Prueba $\chi^2$ de bondad de ajuste									
		<i>Mejico</i> por fuente de publicación									
		T	L	J	A	H	B	P	C	F	TOTALES
Fo		0	0	1	0	0	0	0	8	0	9
Fe		1	1	1	1	1	1	1	1	1	9

$\chi^2_{obt} =$										
$\sum((f_o - f_e)^2)/f_e$	1	1	0	1	1	1	1	49	1	<b>56</b>
										$\alpha =$ <b>20.09</b>

Tabla 11. Cálculo de frecuencias esperadas y frecuencias observadas por revista para *Mejico*

3. Conclusiones: ya que  $\chi^2_{obt} > \chi^2_{crit}$  podemos rechazar la hipótesis nula y concluir que existe una diferencia en la frecuencia de aparición del topónimo *Mejico* entre las fuentes del corpus.

VARIABLE DEPENDIENTE: *Mejiko*

1. Hipótesis:

- i.  $H_0$ = la frecuencia observada del topónimo *Mejiko* es la misma entre las revistas del corpus.
- ii.  $H_1$ = la frecuencia observada del topónimo *Mejiko* es diferente entre las revistas del corpus.

2. Cálculo de frecuencias esperadas para cada frecuencia observada y del valor crítico: la variante *Mejiko* y sus derivados ocurre 15 veces en los textos del corpus, al dividirlo entre las 9 fuentes del corpus se obtiene que *Mejiko* ocurre 1.66 veces por grupo; esta es la  $F_e$ . El procedimiento del cálculo se presenta en el Tabla 12. Para un nivel de significación  $\alpha=0.01$ , unidireccional, y  $gl=8$ , el valor crítico (consultando la tabla de  $\chi^2$ ) =20.09.

		Prueba $\chi^2$ de bondad de ajuste								
		<i>Mejiko</i> por fuente de publicación								
	T	L	J	A	H	B	P	C	F	TOTALES
Fo	0	0	15	0	0	0	0	0	0	15
Fe	1.66	1.66	1.66	1.66	1.66	1.66	1.66	1.66	1.66	14.94

$\chi^2_{obt} =$											
$\sum((f_o - f_e)^2)/f_e$	1.66	1.66	107.2	1.66	1.66	1.66	1.66	1.66	1.66	1.66	<b>120.48</b>
										$\alpha =$	<b>20.09</b>

Tabla 12. Cálculo de frecuencias esperadas y frecuencias observadas por revista para *Mejiko*

3. Conclusiones: ya que  $\chi^2_{obt} > \chi^2_{crit}$  podemos rechazar la hipótesis nula y concluir que existe una diferencia en la frecuencia de aparición del topónimo “Mejiko” entre las fuentes del corpus.

VARIABLE DEPENDIENTE: *Neuspanien*

1. Hipótesis:

- i.  $H_0$ = la frecuencia observada del topónimo *Neuspanien* es la misma entre las revistas del corpus.
- ii.  $H_1$ = la frecuencia observada del topónimo *Neuspanien* es diferente entre las revistas del corpus.

2. Cálculo de frecuencias esperadas para cada frecuencia observada y del valor crítico: la variante *Neuspanien* y sus derivados ocurre 38 veces en los textos del corpus, al dividirlo entre las 9 revistas del corpus se obtiene que *Neuspanien* ocurre 4.22 veces por grupo; esta es la  $F_e$ . El procedimiento del cálculo se presenta en el Tabla 13. Para un nivel de significación  $\alpha=0.01$ , unidireccional, y  $gl=8$ , el valor crítico (consultando la tabla de  $\chi^2$ ) =20.09.

		Prueba $\chi^2$ de bondad de ajuste								
		<i>Neuspanien</i> por fuente de publicación								
	T	L	J	A	H	B	P	C	F	TOTALES
Fo	3	22	12	1	0	0	0	0	0	38
Fe	4.22	4.22	4.22	4.2 2	4.2 2	4.2 2	4.2 2	4.2 2	4.2 2	37.98

$\chi^2_{obt} =$										
$\sum((f_o - f_e)^2)/f_e$	0.35	74.91	14.34	2.4 5	4.2 2	4.2 2	4.2 2	4.2 2	4.2 2	<b>113.15</b>
									$\alpha =$	<b>20.09</b>

Tabla 13. Cálculo de frecuencias esperadas y frecuencias observadas por revista para *Neuspanien*

3. Conclusiones: ya que  $\chi^2_{obt} > \chi^2_{crit}$  podemos rechazar la hipótesis nula y concluir que existe una diferencia en la frecuencia de aparición del topónimo *Neuspanien* y sus variantes entre las fuentes del corpus.

### 3.3.2.1 CONCLUSIONES GENERALES DE LA APLICACIÓN DE LA PRUEBA $\chi^2$ DE BONDAD DE AJUSTE PARA LA VARIABLE FUENTE

Todas las hipótesis nulas fueron rechazadas en favor de la existencia de una diferente frecuencia de aparición de los topónimos según la revista. Todos los valores están considerablemente por encima del valor crítico, por lo que

podemos concluir que sí hay una relación entre la revista de publicación y la forma del topónimo.

### 3.3.3 $\chi^2$ DE BONDAD DE AJUSTE - TIPOLOGÍA TEXTUAL

Se aplicará nuevamente esta prueba pero para el dato nominal de tipología textual. Esta variable tiene solo 3 categorías (tanto en el corpus final como en el corpus general): textos con una secuencia predominantemente narrativa (N), descriptiva (D) o explicativa (E).

Para tener una primera aproximación a la relación entre las variedades nominales de *México* y las tipologías textuales de los textos en que aparecen se presentan estos datos en una gráfica de barras apilada (Fig. 17). Las hipótesis, el procedimiento y las conclusiones se plantean para cada variedad nominal de la siguiente manera:

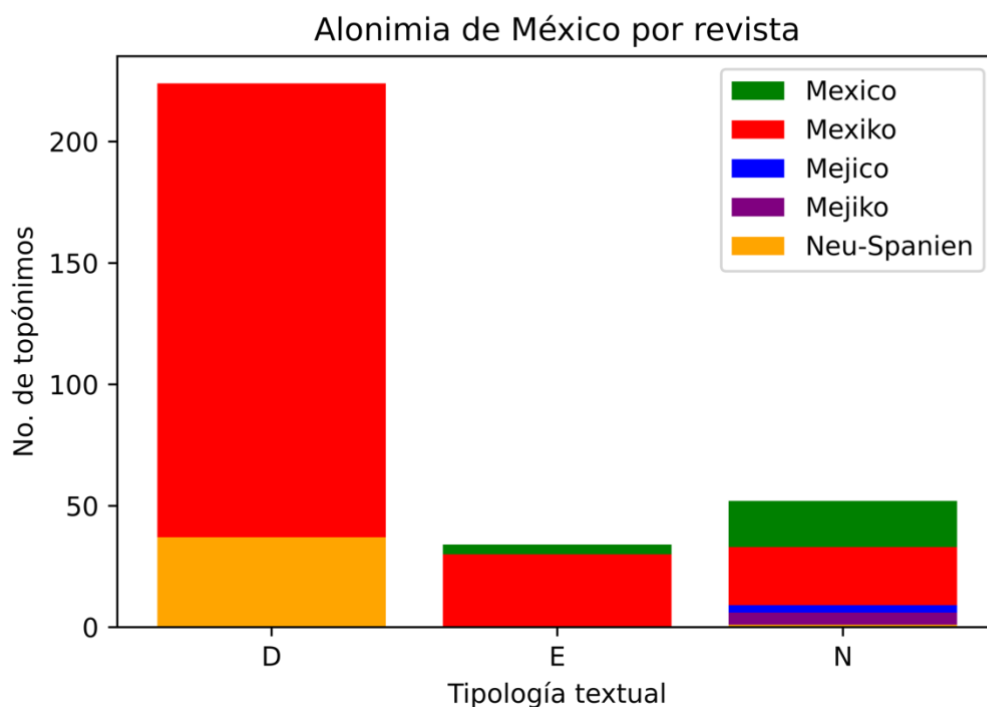


Figura 17. Variedades nominales de *México* por tipología textual

VARIABLE DEPENDIENTE: *Mexico*

1. Hipótesis:

- i. H0= la frecuencia observada del topónimo *Mexico* es la misma entre las tipologías textuales del corpus.
- ii. H1= la frecuencia observada del topónimo *Mexico* es diferente entre las tipologías textuales del corpus.

2. Cálculo de frecuencias esperadas para cada frecuencia observada y del valor crítico: la variante *Mexico* y sus derivados ocurre 154 veces en los textos del corpus, al dividirlo entre las 3 tipologías del corpus se obtiene que *Mexico* ocurre 51.33 veces por grupo; esta es la Fe. El procedimiento del cálculo se presenta en el Tabla 14. Para un nivel de significación  $\alpha=0.01$ , unidireccional, y  $gl=2$ , el valor crítico (consultando la tabla de  $\chi^2$ ) =9.21.

Prueba $\chi^2$ de bondad de ajuste				
<i>Mexico</i> por fuente de publicación				
	N	D	E	TOTALES
Fo	52	68	34	154
Fe	51.33	51.33	51.33	153.99

$\chi^2_{obt} =$				
$\sum ((f_o - f_e)^2) / f_e$	0.00874537	5.41377167	5.85094292	<b>11.27346</b>
			$\alpha =$	<b>9.21</b>

Tabla 14. Cálculo de frecuencias esperadas y frecuencias observadas por tipología textual para *Mexico*

3. Conclusiones: ya que  $\chi^2_{obt} > \chi^2_{crit}$  podemos rechazar la hipótesis nula y concluir que existe una diferencia en la frecuencia de aparición del topónimo *Mexico* entre las tipologías textuales del corpus.

VARIABLE DEPENDIENTE: *Mexiko*

1. Hipótesis:

- i. H0= la frecuencia observada del topónimo *Mexiko* es la misma entre las tipologías textuales del corpus.
- ii. H1= la frecuencia observada del topónimo *Mexiko* es diferente entre las tipologías textuales del corpus.

2. Cálculo de frecuencias esperadas para cada frecuencia observada y del valor crítico: la variante *Mexiko* y sus derivados ocurre 287 veces en los textos del corpus, al dividirlo entre las 3 tipologías del corpus se obtiene que *Mexiko* ocurre 95.66 veces por grupo; esta es la Fe. El procedimiento del cálculo se presenta en el Tabla 15. Para un nivel de significación  $\alpha=0.01$ , unidireccional, y  $gl=2$ , el valor crítico (consultando la tabla de  $\chi^2$ ) =9.21.

Prueba $\chi^2$ de bondad de ajuste				
<i>Mexiko</i> por fuente de publicación				
	N	D	E	TOTALES
Fo	33	224	30	287
Fe	95.66	95.66	95.66	286.98

$\chi^2_{obt} =$				
$\sum((f_o - f_e)^2)/f_e$	41.0440686	172.184357	45.0683211	<b>258.296747</b>
	$\alpha =$			<b>9.21</b>

Tabla 15. Cálculo de frecuencias esperadas y frecuencias observadas por tipología textual para *Mexiko*

3. Conclusiones: ya que  $\chi^2_{obt} > \chi^2_{crit}$  podemos rechazar la hipótesis nula y concluir que existe una diferencia en la frecuencia de aparición del topónimo *Mexiko* entre las tipologías textuales del corpus.

VARIABLE DEPENDIENTE: *Mejico*

1. Hipótesis:

- i. H0= la frecuencia observada del topónimo *Mejico* es la misma entre las tipologías textuales del corpus.
- ii. H1= la frecuencia observada del topónimo *Mejico* es diferente entre las tipologías textuales del corpus.

2. Cálculo de frecuencias esperadas para cada frecuencia observada y del valor crítico: la variante *Mejico* y sus derivados ocurre 9 veces en los textos del corpus, al dividirlo entre las 3 revistas del corpus se obtiene que *Mejico* ocurre 3 veces por grupo; esta es la  $F_e$ . El procedimiento del cálculo se presenta en el Tabla 16. Para un nivel de significación  $\alpha=0.01$ , unidireccional, y  $gl=2$ , el valor crítico (consultando la tabla de  $\chi^2$ ) =9.21.

Prueba $\chi^2$ de bondad de ajuste				
<i>Mejico</i> por fuente de publicación				
	N	D	E	TOTALES
Fo	9	0	0	9
Fe	3	3	3	9

$\chi^2_{obt} =$				
$\sum ((f_o - f_e)^2 / f_e)$	12	3	3	<b>18</b>
			$\alpha =$	<b>9.21</b>

Tabla 16. Cálculo de frecuencias esperadas y frecuencias observadas por tipología textual para *Mejico*

3. Conclusiones: ya que  $\chi^2_{obt} > \chi^2_{crit}$  podemos rechazar la hipótesis nula y concluir que existe una diferencia en la frecuencia de aparición del topónimo *Mejico* entre las tipologías textuales del corpus.

VARIABLE DEPENDIENTE: *Mejiko*

1. Hipótesis:

- i. H0= la frecuencia observada del topónimo *Mejiko* es la misma entre las tipologías textuales del corpus.
- ii. H1= la frecuencia observada del topónimo *Mejiko* es diferente entre las tipologías textuales del corpus.

2. Cálculo de frecuencias esperadas para cada frecuencia observada y del valor crítico: la variante *Mejiko* y sus derivados ocurre 15 veces en los textos del corpus, al dividirlo entre las 3 tipologías del corpus se obtiene que *Mejiko* ocurre 5 veces por grupo; esta es la Fe. El procedimiento del cálculo se presenta en el Tabla 17. Para un nivel de significación  $\alpha=0.01$ , unidireccional, y  $gl=2$ , el valor crítico (consultando la tabla de  $\chi^2$ ) =9.21.

Prueba $\chi^2$ de bondad de ajuste				
<i>Mejiko</i> por fuente de publicación				
	N	D	E	TOTALES
Fo	6	9	0	15
Fe	5	5	5	15

$\chi^2_{obt} =$				
$\sum ((f_o - f_e)^2 / f_e)$	0.2	3.2	5	<b>8.4</b>
			$\alpha =$	<b>9.21</b>

Tabla 17. Cálculo de frecuencias esperadas y frecuencias observadas por tipología textual para *Mejiko*

3. Conclusiones: ya que  $\chi^2_{obt} < \chi^2_{crit}$  no podemos rechazar la hipótesis nula. No hay evidencia de una diferencia en la frecuencia de aparición del topónimo *Mejiko* entre las tipologías textuales del corpus.



VARIABLE DEPENDIENTE: *Neuspanien*

1. Hipótesis:

- i. H0= la frecuencia observada del topónimo *Neuspanien* es la misma entre las tipologías textuales del corpus.
- ii. H1= la frecuencia observada del topónimo *Neuspanien* es diferente entre las tipologías textuales del corpus.

2. Cálculo de frecuencias esperadas para cada frecuencia observada y del valor crítico: la variante *Neuspanien* y sus derivados ocurre 38 veces en los textos del corpus, al dividirlo entre las 3 tipologías del corpus se obtiene que *Neuspanien* ocurre 12.66 veces por grupo; esta es la Fe. El procedimiento del cálculo se presenta en el Tabla 18. Para un nivel de significación  $\alpha=0.01$ , unidireccional, y  $gl=2$ , el valor crítico (consultando la tabla de  $\chi^2$ ) =9.21.

Prueba $\chi^2$ de bondad de ajuste				
<i>Neuspanien</i> por fuente de publicación				
	N	D	E	TOTALES
Fo	1	37	0	38
Fe	12.66	12.66	12.66	37.98

$\chi^2_{obt} =$				
$\sum ((f_o - f_e)^2 / f_e)$	12	3	3	<b>18</b>
	$\alpha =$			<b>9.21</b>

Tabla 18. Cálculo de frecuencias esperadas y frecuencias observadas por tipología textual para *Neuspanien*

3. Conclusiones: ya que  $\chi^2_{obt} > \chi^2_{crit}$  podemos rechazar la hipótesis nula y concluir que existe una diferencia en la frecuencia de aparición del topónimo *Neuspanien* entre las tipologías textuales del corpus.

### 3.3.3.1 CONCLUSIONES GENERALES DE LA APLICACIÓN DE LA PRUEBA $\chi^2$ DE BONDAD DE AJUSTE PARA LA VARIABLE TIPOLOGÍA TEXTUAL

Todas menos una hipótesis nula fueron rechazadas. Esto indica que hay una relación entre el topónimo y la tipología textual del texto en el que se encuentra. El hecho que esta variable tenga pocas categorías nos permite observar más claramente la manera en la que se distribuyen los datos.

Una hipótesis inicial al compilar el corpus fue que los textos con una secuencia textual predominantemente narrativa se caracterizarían por usar técnicas de traducción extranjerizantes, propio de la literatura de viajes científica popularizada por Humboldt, uno de los autores más citados del corpus. Otra hipótesis fue que los textos narrativos tendrían una mayor extensión. Esta suposición se basa en que los textos narrativos de la literatura de viajes buscan acercar al lector a la otra cultura, empleando entre otros recursos el uso de los endónimos.

Por otra parte, se esperaba que en los textos descriptivos y explicativos, que son en su mayoría tratados científicos concentrados en la minería, procesos de extracción de metales y ciencias de la tierra, y reportes de estados financieros de minas y de actividades económicas relacionadas, ocurriera más *Mexiko* por ser la variante que representa el proceso traductor más sencillo (usando la técnica de adaptación fonético-gráfica), lo que sí se ha observado.

Esta suposición se basaba en que los textos descriptivos y explicativos científicos buscan describir o explicar un fenómeno, mineral, material, estado financiero, entre otros, prestando poca atención al contexto donde ocurre o se encuentra su objeto de estudio, es decir, sin preocuparse en dar más información acerca de México y recurriendo al exónimo más difundido y aceptado.

Después de observar los resultados de la graficación de los datos y de las pruebas para comprobar su significación estadística considero importante corroborar que estas hipótesis parecen comprobarse con las observaciones del análisis. Puede hacerse una aproximación a esto al elaborar un listado de los textos excluidos del corpus de este trabajo que dé cuenta de la tipología textual del texto y de la variante nominal de México que aparece. Se presenta el cuadro en el Anexo C.

Después de la primera aproximación a los datos a través del listado podemos suponer que las intuiciones que se han expresado arriba podrían ser correctas. En los textos descriptivos y explicativos ocurre con más frecuencia *Mexiko*, el exónimo más “simple” en términos de las modificaciones que sufre el endónimo para llegar a la lengua meta; en los narrativos encontramos dos endónimos de México: *Mexico* y *Mejico*, ambos que pueden leerse en documentos oficiales y cuya alternancia podría explicarse por las fuentes en las que se basaron los investigadores.

Otra predicción inicial relacionada con los textos narrativos fue que estos presentarían más información contextual y de los topónimos mismos que mencionan. Al revisar los textos podemos comprobar esto: “Ersteigung des Popocatepetl (Volcan grande de Mejico ó de Puebla) im Jahre 1827” de Berghaus, Heinrich, describe la etimología de Popocatépetl; *Versuch einer getreuen Schilderung der Republik Mejico* de Mühlepfordt, Eduard, describe la etimología de Mejico, Citlatépetl, Nauhcampatepetl, Popocatépetl e Iztaccihuatl.

Otros factores importantes que pueden apreciarse al observar el listado es que la variable que hemos excluido del análisis de este trabajo de autor puede jugar un papel importante. Podemos ver que tanto en los textos excluidos del

corpus de este trabajo como en los textos del corpus general hay autores, como Gerolt, que emplean una misma variante del topónimo constantemente; otros autores, como Burkart, emplean numerosas variedades, que pueden variar según la revista y el año, por ejemplo, Burkart, 1826 (*Mejiko*), 1827 (*Mexico*) y 1832 (*Mexiko*). Esto puede deberse a que este autor estuvo numerosos años presente en México y que adoptó los nombres de México conforme se oficializaban o socializaban *in situ* con los agentes locales con los que colaboró, como ya se ha mostrado en otra investigación (Azuela y Villanueva, 2022).

## DISCUSIÓN Y CONCLUSIONES FINALES

Al concluir esta investigación podemos corroborar que el fenómeno de la alonimia es un fenómeno complejo debido a los diversos factores que intervienen en él. En el caso del topónimo *México* estos factores podrían determinar la variante del topónimo que aparece en ciertos textos.

La variable de año nos ha permitido comprobar que la variante *Neuspanien* ocurre en un periodo muy específico de nuestro corpus, al comienzo de él, lo que puede explicarse por razones históricas, particularmente por los periodos de gobierno, primero por las comunidades indígenas, segundo por el Imperio Español y tercero el de México independiente. Por otra parte, las demás variables parecen traslaparse de 1824 a 1844, lo que lleva a buscar una explicación en las otras variables del estudio.

La tipología textual parece mostrar tendencias más concretas: en textos narrativos se encuentran los endónimos y en textos descriptivos y explicativos los exónimos. Es en estos dos géneros en los que encontramos mayor estabilidad en el uso de una variante, *Mexiko*, mientras que en el género narrativo se encuentra una proliferación de variantes nominales a partir de la independencia de México, donde encontramos *Mexico*, *Mexiko* y *Mejico*.

Lo anterior puede explicarse en el hecho de que el proceso de consumación de independencia implicó la difusión de numerosos documentos de tipo promulgación de independencia, constitución, llamado popular, entre otros, tanto en América como en España, donde se encuentran diversas variantes nominales para México.

El uso de una u otra variante nominal en el caso de los textos narrativos parece mostrar una intención de comunicar el nombre del país desde cierta

perspectiva o tomando como base cierta fuente, lo que podría apuntar incluso a un posicionamiento político por parte de estos autores: el reconocimiento de México independiente.

La revista también es uno de los factores que influye en la ocurrencia de cierta forma del topónimo: *Taschenbuch...* y *Jahrbuch...* muestran consistencia en el uso del exónimo, aunque esto también puede atribuirse al tipo textual predominante en ambas revistas, que es el descriptivo.

El uso de pruebas estadísticas ha permitido comprobar que hay una distribución heterogénea de las variantes nominales de *México* con respecto a las variables de año, tipología textual y fuente. Al usar medidas de agrupación y dispersión se pudo caracterizar de mejor manera la distribución de los datos del corpus, lo que permitió reconocer el efecto de los valores atípicos y elegir las pruebas adecuadas para analizar los datos.

Pese a que hemos descartado estos textos del análisis estadístico, la extrapolación de las conclusiones de las pruebas aplicadas al corpus reducido ha permitido comprobar ciertas tendencias relacionadas al año de publicación de los textos y la tipología textual. En este último caso vale mencionar que la tipología textual que más se redujo fue la narrativa: se redujo en un 30% mientras que la descriptiva 15% y la explicativa no sufrió cambios.

Esto nos indica que los textos narrativos tienen, en lo general, una mayor concentración de topónimos, esto en concordancia con lo mencionado anteriormente de que estos textos se caracterizan por ser más profusos en los detalles del contexto, en oposición a los descriptivos y explicativos.

La reducción de los textos del corpus también nos ha llevado a reconocer la importancia de otras variables, como la de autor. Ya que el 40% de los textos

descartados son de Joseph Burkart, podemos confirmar que este autor ha cumplido un papel importante en la investigación científica en geología y mineralogía mexicana de este periodo.

El análisis de estas variables ha probado ser de utilidad para el reconocimiento de los factores que pueden intervenir en la ocurrencia de variables nominales de la toponimia mexicana. Esto ha sido una buena aportación, por una parte, a la toponomástica mexicana y, por otra parte, la metodología empleada ha sido también un buen acercamiento al uso de herramientas computacionales para el estudio de fenómenos lingüísticos basados en corpora.

El análisis de datos, su representación gráfica y su divulgación a través de Python ha probado ser de utilidad para la investigación lingüística y considero que la metodología presentada en este trabajo puede ser de ayuda para futuros estudiantes e investigadores que trabajen con grandes cantidades de texto.

Estas herramientas computacionales están siempre en constante desarrollo y los procedimientos y módulos empleados para esta investigación pueden no ser de utilidad para una investigación lingüística con otros objetivos o en otro momento pueden no estar disponibles. Considero que Python es un lenguaje de programación muy flexible y con numerosas aplicaciones y que la metodología, los módulos y las interfaces descritas en esta investigación sirven de punto de partida para el desarrollo de programas para la investigación lingüística con otros objetivos.

Para cerrar con esta investigación se recuperan las preguntas de investigación planteadas al inicio del trabajo y retomadas en la metodología:

a) ¿Cuáles son los topónimos y las variantes nominales de ellos a los que hacen referencia los autores alemanes en los textos coleccionados del corpus constituido?

Los topónimos encontrados en el corpus han sido reportados en el Anexo D junto con sus respectivas variantes nominales. El análisis del topónimo *México* indica que las variaciones se caracterizan, principalmente, por la alternancia entre grafías:

- Alternancia entre <x> y <j>: en esta alternancia la <x> es la grafía más frecuente (aparece en 441 de los 503 casos correspondientes a *México*), la que es usada casi exclusivamente en textos explicativos y descriptivos (de los 402 topónimos correspondientes a estas dos tipologías, solo hay 9 ocurrencias del topónimo con la grafía <j>), y la que puede encontrarse a lo largo de todo el periodo; la grafía <j> ocurre solo en 24 topónimos de los 503, la mayoría de ellas en la tipología narrativa (15 veces), y ocurre solo en los años de 1825-1829 y 1840-1844.
- Alternancia entre <c> y <k>: en esta alternancia la <k> es la grafía más frecuente (aparece en 302 de los 503 topónimos correspondientes a “México”), la que es usada en las tres tipologías textuales, aunque con un mayor número de ocurrencias en la tipología descriptiva, y la que puede encontrarse a lo largo de todo el periodo; la <c> ocurre en 163 de los 503 topónimos, casi la mitad de ocurrencias que la <k>, ocurre proporcionalmente mayor veces en la tipología narrativa, y ocurre constantemente a partir del año de 1815.



Al contrastar estas alternancias con la lista de topónimos encontrados del Anexo D puede observarse que estas alternancias son constantes a lo largo de otros topónimos. Podemos encontrar *Guanajuato* y *Guanaxuato*, *Zacatecas* y *Zakatecas*, *Atotonilco* y *Atotonilko*, *Biscaina* y *Biskaina*, *Guadalajara* y *Guadalaxara*, *Jalapa* y *Xalapa*, *Campeche* y *Kampeche*, *Californien* y *Kalifornien*, *Oaxaca* y *Oajaca*, *Tlalpujahuá* y *Tlalpuxahua*, entre otros.

Además de estas alternancias a nivel gráfico, se encuentra también la variación de diversidad de palabras de *Neu-Spanien* o *Neuspanien*.

- b) ¿Cuáles son las motivaciones de las variaciones internas y externas de los topónimos?

En el análisis del topónimo *México* se ha reconocido como motivación interna de variación el uso de un grafema *c*: este grafema del topónimo original es adaptado a la lengua alemana con el grafema <k>, presente con mucha más frecuencia en el léxico del alemán, particularmente cuando se encuentra en textos descriptivos y explicativos. Esto significa que, estamos frente a una variación que responde a una coincidencia fonética, el mismo fonema con grafía diferente en cada lengua. La coincidencia del fonema /Ko/ en español, correspondiente a la grafía <c> *México*, con el fonema /Ko/ en alemán, correspondiente a la grafía 'Ko' para *México* en alemán.

Por otra parte, como motivaciones de cambio externas se han podido comprobar las variables de tipología textual, año y, aunque con

menos contundencia, la fuente de publicación, como se puede observar en los datos del Anexo C.

c) ¿Cómo pueden localizarse todas las entradas de un topónimo y sus respectivas variantes?

El desarrollo de un programa con Python ha permitido reconocer la utilidad de los módulos Pdfminer y io para extraer el texto de los documentos en PDF, CSV para la extracción de documentos en formato CSV, Spacy y Re para la creación de patrones de búsqueda, Collections, NumPy y Pandas para el análisis de datos cuantitativos y Matplotlib para la representación gráfica de estos datos. De esta manera, el uso de estas diferentes librerías ha permitido la localización de variantes nominales del topónimo “México” con resultados satisfactorios.

Al empezar esta investigación se planteó como objetivo identificar los topónimos de un corpus delimitado sobre minería y ciencias de la tierra en alemán para describir las consecuencias de su alonimia y crear una propuesta para su localización automática asistida por computadora.

El objetivo específico fue analizar lingüísticamente los topónimos y reconocer algún tipo de sistematicidad en sus procesos de traducción y/o adaptación dentro de los parámetros del corpus para postular reglas que permitan la localización asistida por computadora de los topónimos y sus respectivas variedades nominales.

Para lograr este objetivo considero que el marco teórico fue suficiente en tanto que se lograron reconocer procesos sistemáticos en la

adaptación/traducción de topónimos que pudieron ser agrupados en cuatro categorías: técnica de adaptación superficial o fonético-gráfica, técnica de variación de raíz, técnica de concordancia semántica y técnica de diversidad de palabras; el reconocimiento de estas técnicas de adaptación permitió identificar las herramientas computacionales específicas necesarias para la localización de estos topónimos en el corpus.

La metodología empleada en esta investigación puede ser evaluada, consecuentemente, de manera positiva y considero que esta es la mayor aportación de la investigación: se describieron los procesos de compilación y preparación del corpus, incluido su edición para la mejor legibilidad por computadora, se demostró el uso de pruebas estadísticas para comprobar la significación de los datos del corpus y se presentaron las herramientas computacionales útiles para el procesamiento de datos y localización de los topónimos del corpus que consistieron en el uso de módulos especializados para la extracción de texto de documentos en formato PDF, la identificación basada en cadenas de caracteres no completamente definidas para poder recuperar las variantes nominales y la presentación de los datos en forma de gráficas..

Los resultados cogidos de esta investigación pueden dividirse en dos: resultados del análisis manual y resultados del análisis automático. Los resultados del análisis manual son, por una parte, la lista de topónimos presentes en el corpus y, por otra, la clasificación de técnicas de traducción/adaptación; los resultados del análisis automático fueron el conteo de frecuencias que nos permitió contar con datos para el análisis estadístico y la localización exacta de los topónimos en el corpus.

El análisis con Python nos permitió la manipulación de los datos generados manualmente y también la visualización gráfica de estos datos, además de comprobar que esta herramienta es útil en las tareas de procesamiento y búsqueda de palabras en grandes volúmenes de texto, lo que representa un ahorro de tiempo para el investigador.

Como en toda investigación, hay limitaciones debido al tiempo o a los recursos que nos han impedido explotar el máximo potencial de esta herramienta computacional y hacer un análisis manual más profuso.

En este trabajo no fue posible incorporar más textos al corpus que habrían sido de gran interés para conocer en qué medida los topónimos llegan a una forma más estable a lo largo del tiempo al establecerse o normalizarse una variante por encima de otras. Haber incorporado textos del periodo del Segundo Imperio Mexicano, en el que también hubo una gran producción científica por parte de investigadores alemanes y austriacos, hubiera permitido contar con datos importantes para esto. Tampoco fue posible realizar un análisis más extenso que permitiera incorporar los textos de mayor extensión que fueron descartados de la versión final del corpus, ni realizar un análisis que contrastara los procesos de traducción y adaptación entre diferentes tipos de topónimos (micro- y macrotopónimos, especialmente).

En un futuro pueden explorarse otras posibles líneas de investigación y pueden aprovecharse áreas de oportunidad para refinar la metodología y el programa con Python a través de la incorporación de módulos para la traducción de palabras con Python que permitirían generar posibles variantes nominales que, si bien no fueron el objeto del estudio principal, estaban presentes en el

corpus, como Engelberg-Cerro del Ángel, Neu-León-Nuevo León y San Eugenio-Sankt Eugen.

También sería posible identificación general de todos los topónimos del corpus y su agrupación automática basada en la similitud de texto. Esto es posible usando el reconocimiento de entidades nombradas de Spacy y la Distancia de Levinshtein.

Uno de los principales problemas relacionados con el rendimiento del programa tampoco ha podido ser resuelto: el problema de la legibilidad de los textos. Todos los textos del corpus han sido editados para garantizar el correcto reconocimiento de los topónimos en ellos, pero en textos no editados no es posible garantizar esto. Es posible incorporar al programa un sistema de reconocimiento óptico de caracteres usando el módulo OpenCV, por ejemplo, que permita entrenar el programa para el correcto reconocimiento de caracteres con tipografías raras, como la Fraktur, o para reconocer texto en páginas oscurecidas o con tinta difuminada, como fue el caso de muchos textos de esta investigación. Estas son áreas de oportunidad para futuras investigaciones.

Considero que el fenómeno de la alonimia es uno sumamente relevante y uno que está presente cada vez más en nuestro mundo globalizado. El desarrollo de metodologías para el reconocimiento de variedades nominales para un mismo objeto geográfico translingüísticamente permitiría acceder a fuentes de información en distintas lenguas; si este proceso, además, se conjunta con la extracción automática de información de internet a través de técnicas de Web scraping, se contaría con una herramienta muy poderosa para el acceso rápido y actualizado a la información en distintas lenguas.

## Bibliografía

- Adam, J. (1991). *Les textes: types et prototypes*. París, Francia: Nathan
- Allen, R. (1993). *Linguistic Wars*. Oxford, R.U.: Oxford University Press.
- Azuela, L. y Villanueva, E. Joseph Burkart's geological research in Mexico (1825-1834): a paradigmatic case of knowledge co-production. *Physis, LVI*, 135-148.
- Back, O. (2002). *Übersetzbare Eigennamen: Eine synchronische Untersuchung von interlingualer Allonymie und Exonymie*. Viena, Austria: Edition Praesens.
- Back, O. (2006). *Buchstäblich geschrieben: Aufsätze über Schrifttheorie, Orthographie und Transkription*. Viena, Austria: Edition Praesens.
- Bahr, C. (2009). *Zur Übersetzung von Eigennamen am Beispiel der Entwicklung und Übersetzung der Ortsnamen Amerikas* [Diploma, Universität Leipzig].  
[https://ul.qucosa.de/landing-page/?tx\\_dlfid=https%3A%2F%2Ful.qucosa.de%2Fapi%2Fqucosa%253A12156%2Fmets](https://ul.qucosa.de/landing-page/?tx_dlfid=https%3A%2F%2Ful.qucosa.de%2Fapi%2Fqucosa%253A12156%2Fmets)
- Bird, S., Klein, E. y Loper, E. (2009). *Natural Language Processing with Python*. California, EUA: O' Reilly Media, Inc.
- Derczynski, L. (2016). Complementarity, F-score, and NLP Evaluation. *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, 261-266.
- García, J. (2011). Los aspectos semánticos de la toponimia. *Actes de la IV Jornada d'Onomàstica*, 177-188.
- Halliday, M. (2005). *Computational and Quantitative Studies*. Cornwall, Reino Unido: Continuum.
- Heath, S. (1972). *Telling Tongues: Language Policy in Mexico, Colony to Nation*. Columbia, EUA: Columbia University.
- Hough, C. (2016). *The Oxford Handbook of Names and Naming*. Oxford, Reino Unido: Oxford University Press.
- Hurtado, A. (2001). *Traducción y Traductología: introducción a la Traductología*. Madrid, España: Cátedra.

- Iglesias, Á. (1991). Nombres Propios: para una tentativa de clasificación. *Dictionnaire historique des noms de famille romans. Actes du III Colloqui PATROM*, 227-241.
- Lara, L. (2006). *Curso de Lexicología*. Ciudad de México, Mexico: El Colegio de México
- Lefebvre, K. y Paredes, C. (2017). *La memoria de los nombres: la toponimia en la conformación histórica del territorio. De Mesoamérica a México*. Morelia, México: CIGA: UNAM
- Lemnitzer, L. y Zinsmeister H. (2010). *Korpuslinguistik. Eine Einführung*. Alemania: Narr Verlag.
- Leon D. (2016). Complementarity, F-score, and NLP Evaluation. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 261-266.
- Mentz, B. von. (2017). Topónimos y cronología: notas sobre una puerta distinta al estudio del pasado. *Historia Mexicana*, 67(1), 7-59.
- Nübling, D. (2015). *Namen: Eine Einführung in die Onomastik*. Tübinga, Alemania: Narr Francke Attempto Verlag.
- Pocklington, R. (2022). Introducción a la toponomástica - 1. Aspectos universales de la toponimia [Archivo PDF]. [https://www.academia.edu/21713377/INTRODUCCIÓN\\_A\\_LA\\_TOPONOMÁSTICA\\_1\\_ASPECTOS\\_UNIVERSALES\\_DE\\_LA\\_TOPONIMIA](https://www.academia.edu/21713377/INTRODUCCIÓN_A_LA_TOPONOMÁSTICA_1_ASPECTOS_UNIVERSALES_DE_LA_TOPONIMIA)
- Pratt, M. (1992). *Imperial Eyes: Travel Writing and Transculturation*. Londres, Reino Unido: Routledge.
- Simón, J. R. (2006). Propuesta para la determinación de los tipos de textos. *Sapiens. Revista Universitaria de Investigación*, 7(1), 163-179.
- Wermke, M. et al. (2009). *Duden: Die Grammatik*. Mannheim, Alemania: Bibliographisches Institut AG.
- Wright, D. (2007). La Política Lingüística en la Nueva España. *Acta Universitaria*, 17(3), 5-19.