

Análisis Comparativo de Textos: Nuevo Testamento y el Libro de Mormón

Autor:

Christian Servando Garza González

Programa:

Procesamiento y Clasificación de Datos

Fecha:

19 de enero de 2025

Resumen

En el presente reporte se describe el proceso y los resultados de un análisis comparativo entre dos textos religiosos: el Nuevo Testamento y el Libro de Mormón. Dicho análisis se realizó mediante técnicas de Procesamiento de Lenguaje Natural (PLN) en *Python*, incorporando procesos de limpieza de texto, tokenización, eliminación de *stopwords*, *stemming* y *lemmatización*. Además, se incluyen nubes de palabras y análisis de n-gramas para observar las palabras y secuencias más frecuentes en ambos textos.

Índice

1. Introducción	2
2. Marco Teórico	2
2.1. Textos Religiosos Analizados	2
2.1.1. Nuevo Testamento	2
2.1.2. Libro de Mormón	3
2.2. Procesamiento de Lenguaje Natural (PLN)	3
3. Metodología	5
3.1. Descripción de los Datos	5
3.2. Herramientas	5
3.3. Flujo de Trabajo	6
4. Resultados	7
4.1. Estadísticas Generales	7
4.2. Frecuencia de Palabras	7
4.3. Nubes de Palabras	8
4.4. Análisis de Signos de Puntuación	8
5. Análisis e Interpretación	10
6. Conclusión	10
7. Referencias	11

1. Introducción

El estudio de textos religiosos mediante técnicas de Procesamiento de Lenguaje Natural (PLN) ofrece una perspectiva interesante sobre la frecuencia de palabras y conceptos fundamentales que contienen. En este reporte, se comparan dos importantes textos del cristianismo: el **Nuevo Testamento** y el **Libro de Mormón**.

El objetivo principal es explorar las similitudes y diferencias en el vocabulario empleado en ambos textos y, a partir de estadísticas descriptivas y visualizaciones, obtener información cuantitativa que ayude a entender mejor las particularidades de cada uno.

2. Marco Teórico

2.1. Textos Religiosos Analizados

2.1.1. Nuevo Testamento

El **Nuevo Testamento** es una de las secciones fundamentales de la Biblia cristiana. Abarca un conjunto de escritos que incluyen:

- *Los cuatro Evangelios* (Mateo, Marcos, Lucas y Juan), que narran la vida y enseñanzas de Jesucristo.
- *Hechos de los Apóstoles*, que describe la expansión temprana del cristianismo tras la resurrección de Jesús.
- *Las Epístolas o Cartas Apostólicas*, escritas por diversos autores (entre ellos el apóstol Pablo) dirigidas a distintas comunidades cristianas para instruir y orientar en la fe.
- *El Apocalipsis o Libro de la Revelación*, que contiene visiones proféticas atribuidas al apóstol Juan.

Estos documentos, escritos mayoritariamente en griego koiné durante el siglo I d.C., han sido traducidos a numerosos idiomas a lo largo de la historia. Se consideran testimonios clave de la doctrina cristiana, ya que relatan eventos centrales de la fe, tales como la vida, enseñanza, muerte y resurrección de Jesús. En conjunto, el Nuevo Testamento ha influido de manera decisiva en la cultura, la teología y la moral del mundo occidental.

2.1.2. Libro de Mormón

El **Libro de Mormón** es un texto que los miembros de La Iglesia de Jesucristo de los Santos de los Últimos Días consideran escritura sagrada, a la par de la Biblia. Según la tradición de esta denominación, fue traducido y publicado por Joseph Smith en 1830 a partir de planchas de oro escritas por profetas de la antigua América.

Este texto relata la historia espiritual de diversos pueblos que habitaron el continente americano, enlazando su narrativa con doctrinas y enseñanzas sobre Jesucristo. Consta de varios libros o secciones, entre los que destacan:

- *1 Nefi*, *2 Nefi*, entre otros, que describen el viaje de una familia desde Jerusalén hasta el continente americano.
- *Mosíah*, *Alma*, *Helamán*, libros que presentan acontecimientos y enseñanzas de profetas y gobernantes en diferentes épocas.
- *3 Nefi*, que relata la aparición de Jesucristo a los pobladores tras su resurrección y ascensión en Jerusalén.
- *El Libro de Mormón* propiamente dicho, que da nombre a toda la obra y donde se recogen enseñanzas y una crónica breve de su pueblo.

Al igual que el Nuevo Testamento, el Libro de Mormón está plagado de referencias divinas y doctrinales, pero con un enfoque y trasfondo distintos, reflejando creencias y revelaciones particulares de la tradición mormona. La existencia de múltiples traducciones de este libro ha facilitado su difusión global, llegando a desempeñar un papel fundamental en la teología y práctica de los Santos de los Últimos Días.

2.2. Procesamiento de Lenguaje Natural (PLN)

El Procesamiento de Lenguaje Natural (PLN) es un campo de la inteligencia artificial que se enfoca en la interacción entre computadoras y el lenguaje humano. Para lograr que las máquinas *lean* y *comprendan* textos de gran extensión, se emplean diversos métodos y técnicas que ayudan a limpiar, estructurar y analizar la información. A continuación, se describen algunas de las técnicas más relevantes utilizadas en este estudio:

- **Tokenización:** Consiste en dividir el texto en unidades mínimas de análisis denominadas *tokens*. Generalmente, un token corresponde a una palabra, pero también puede tratarse de otros elementos, como signos de puntuación o dígitos, dependiendo de la configuración de la herramienta de tokenización. Este proceso facilita la manipulación computacional del texto, al convertirlo en una lista o secuencia de tokens.
- **Stopwords:** Durante el análisis de texto, las *stopwords* son aquellas palabras de alta frecuencia, como artículos (“el”, “la”, “los”), conjunciones (“y”, “o”) y preposiciones (“en”, “de”), que por lo general no aportan un significado relevante en el contexto. Eliminarlas ayuda a reducir el ruido en los resultados y a centrar el análisis en términos más sustanciales.
- **Stemming:** El stemming se basa en recortar las palabras a su *raíz* o *stem*, sin tener en cuenta la forma léxica o la categoría gramatical. Por ejemplo, las palabras “cantar”, “cantando” y “cántico” podrían terminar reducidas al mismo *stem* “cant”. La ventaja principal del stemming es su rapidez y facilidad de implementación. Sin embargo, puede causar que diferentes palabras con sentidos distintos terminen teniendo la misma raíz, lo que en ocasiones reduce la precisión semántica.
- **Lematización:** La lematización, en cambio, transforma las palabras a su *forma base* o *lemma*, considerando la categoría gramatical y el contexto lingüístico. Por ejemplo, si el término es “hablando”, el lematizador podría determinar que la forma base es “hablar”. Este proceso suele requerir diccionarios o modelos del idioma para identificar correctamente la forma canónica de la palabra. A diferencia del stemming, la lematización tiende a ser más precisa, pero a la vez más costosa computacionalmente y dependiente del lenguaje.
- **Diferencia entre *Stemming* y *Lematización*:**
 - *Stemming*: Es una aproximación más simple, rápida y basada en reglas de recorte de sufijos. Puede producir raíces de palabras que no sean formas léxicas reales.
 - *Lematización*: Utiliza conocimiento del lenguaje (diccionarios, morfología, etc.) para obtener la forma base de cada término. Suele

arrojar resultados más cercanos a la forma real de las palabras, aunque requiere mayor capacidad de cómputo.

- **Análisis de n-gramas:** Un *n-grama* es una secuencia de n palabras consecutivas en un texto. Por ejemplo, los bigramas ($n = 2$) resultan de agrupar cada par de palabras contiguas. Este análisis permite descubrir patrones o expresiones frecuentes, así como la relación que existe entre palabras que suelen aparecer juntas.
- **Nubes de palabras:** Las nubes de palabras son representaciones visuales en las que las palabras más frecuentes aparecen con mayor tamaño y peso. Ofrecen una visión rápida e intuitiva de los términos principales de un texto, permitiendo identificar de un vistazo los conceptos más relevantes o repetidos.

3. Metodología

3.1. Descripción de los Datos

Para este análisis, se utilizaron los siguientes documentos en formato PDF:

- `LibrodelMormon.pdf`
- `Biblia_nuevotestamento.pdf`

El primer paso fue cargar estos documentos y extraer su contenido textual.

3.2. Herramientas

- **Python:** Lenguaje de programación principal.
- **Bibliotecas utilizadas:**
 - PyPDF2 (o similar) para la lectura de archivos PDF.
 - `re` para manejo de expresiones regulares.
 - `nltk` (Natural Language Toolkit) para tokenización, *stopwords*, *stemming* y *lemmatización*.
 - `wordcloud` para generar nubes de palabras.

- `matplotlib` para visualizaciones (si se requiere).
- `collections.Counter` para el conteo de palabras y n-gramas.

3.3. Flujo de Trabajo

1. **Carga de PDFs:** Se subieron los archivos `LibrodelMormon.pdf` y `Bilbia_nuevotestamento.pdf` a un entorno colaborativo (ej. Google Colab).
2. **Extracción de texto:** Con una función `extract_text_from_pdf` se recorrieron las páginas de cada PDF y se concatenó su contenido en una cadena de texto.
3. **Preprocesamiento del texto:** Incluyó:
 - Conversión a minúsculas.
 - Eliminación de caracteres especiales y números mediante expresiones regulares.
 - *Tokenización* con `nltk.word_tokenize`.
 - Eliminación de *stopwords* en español.
 - Filtrado adicional de tokens muy cortos (longitud menor a 4 caracteres).
4. ***Stemming* y *Lematización*:** Se crearon funciones separadas para:
 - Aplicar *Stemming* con `PorterStemmer`.
 - Aplicar *Lematización* con `WordNetLemmatizer`.
5. **Cálculo de estadísticas descriptivas:** Número total de palabras, número de palabras únicas y longitud promedio de palabra.
6. **Frecuencia de palabras:** Mediante la función `FreqDist` de `nltk` y objetos `Counter`.
7. **Análisis de n-gramas:** Extracción de **bigramas** ($n=2$) más frecuentes en ambos textos.
8. **Generación de nubes de palabras:** Utilizando la biblioteca `wordcloud` se visualizaron las palabras más frecuentes.

9. **Análisis de signos de puntuación:** Uso de expresiones regulares para contar la frecuencia de distintos signos.

4. Resultados

4.1. Estadísticas Generales

Con los textos tokenizados, se calcularon valores estadísticos básicos que proporcionan una primera aproximación cuantitativa:

- **Número total de palabras:** El *Libro de Mormón* resultó ser más extenso, con alrededor de 251,443 palabras, mientras que el *Nuevo Testamento* registró aproximadamente 86,504.
- **Número de palabras únicas:** El *Libro de Mormón* presentó cerca de 24,124 palabras únicas, por encima de las 13,472 encontradas en el *Nuevo Testamento*, reflejando una mayor diversidad léxica.
- **Longitud promedio de palabra:** Se observó un valor ligeramente mayor en el *Libro de Mormón* (6.62 caracteres) frente al *Nuevo Testamento* (6.47 caracteres), diferencia que, si bien no es muy amplia, indica una posible variación en la complejidad o riqueza morfológica del texto.

4.2. Frecuencia de Palabras

Tras el preprocesamiento, se llevó a cabo un conteo de las palabras más comunes para cada documento, junto con el análisis de *n-gramas* (en particular, bigramas). Entre los hallazgos principales, destacan:

- **Libro de Mormón:** Palabras como *dios*, *señor* y *tierra* lideran las frecuencias; asimismo, los bigramas relevantes (*josé smith*, *todas cosas*, *alma alma*) apuntan a la presencia de personajes y conceptos característicos de la tradición mormona.
- **Nuevo Testamento:** Además de *dios* y *señor*, aparecen con fuerza menciones a *jesús* y a figuras como *juan*. Entre los bigramas más mencionados se encuentran *jesús dijo*, *espíritu santo* y *todas cosas*, reflejando la narrativa evangélica y doctrinal del canon cristiano.

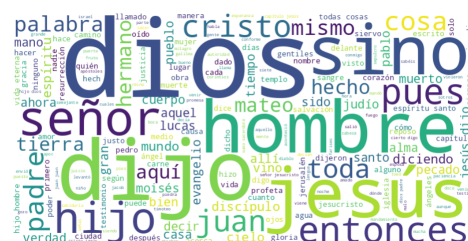
Estos resultados ofrecen una visión del énfasis temático en cada texto. Por un lado, el *Libro de Mormón* contiene referencias específicas a personajes de su propia tradición (p.ej., *josé smith* y *alma*), mientras que el *Nuevo Testamento* pone el foco en *jesús* y *sus discípulos*.

4.3. Nubes de Palabras

Para cada texto, se elaboró una nube de palabras que mostró de forma visual las palabras con mayor frecuencia, incrementando el tamaño de las más reiteradas. Dicha representación confirmó lo observado en la lista de frecuencias, poniendo en evidencia la centralidad de términos como *dios*, *señor* y *jesús* (en el caso del *Nuevo Testamento*) o *alma* y *tierra* (en el *Libro de Mormón*). Además, las nubes permiten apreciar la proporción en la que aparecen estas palabras en comparación con otras.



(a) Nube de Palabras del *Libro de Mormón*



(b) Nube de Palabras del *Nuevo Testamento*

Figura 1: Comparación de Nubes de Palabras entre el *Libro de Mormón* y el *Nuevo Testamento*

4.4. Análisis de Signos de Puntuación

Se realizó un análisis detallado de la frecuencia relativa de los signos de puntuación presentes en el *Libro de Mormón* y el *Nuevo Testamento*, calculando el porcentaje que representa cada signo respecto al total de signos de puntuación utilizados en cada texto. Esto permite identificar patrones en el estilo de redacción y en la presentación editorial de ambos textos.

En la Figura 2 se muestra una comparación de los signos de puntuación más frecuentes en ambos textos. Los resultados principales son los siguientes:

- Las **comas** (',') y los **puntos** ('.') son los signos más utilizados en ambos textos, representando aproximadamente el 28-30 % del total de los signos en cada texto. Sin embargo, el *Nuevo Testamento* presenta un uso ligeramente mayor de estos signos.
- Los **dos puntos** (':') también tienen una presencia significativa, especialmente en el *Libro de Mormón*, donde representan el 13.77 % frente al 12.02 % en el *Nuevo Testamento*.
- Los **puntos y coma** (';') y los **guiones** ('-', '—') muestran diferencias importantes en su frecuencia. El *Libro de Mormón* utiliza estos signos con mayor frecuencia relativa, lo que podría indicar un estilo de redacción más complejo o con mayor segmentación de ideas.
- Otros signos, como los paréntesis ('(', ')'), tienen una frecuencia menor pero relativamente similar en ambos textos, alrededor del 1.6 %.

Estas diferencias en el uso de signos de puntuación reflejan variaciones en el estilo de escritura y la estructura narrativa de cada texto. Mientras que el *Nuevo Testamento* parece tener un estilo más directo y lineal, el *Libro de Mormón* utiliza una mayor diversidad de signos, lo que podría estar relacionado con su mayor extensión y complejidad narrativa.

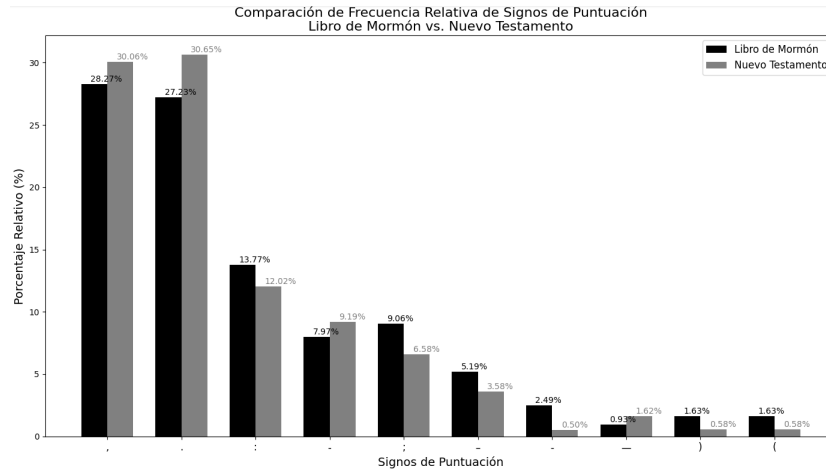


Figura 2: Comparación de Frecuencia Relativa de Signos de Puntuación entre el *Libro de Mormón* y el *Nuevo Testamento*

5. Análisis e Interpretación

El **Libro de Mormón** presenta un cuerpo textual más extenso, con más palabras totales y más palabras únicas, lo que puede indicar mayor diversidad léxica. Las constantes referencias a *josé smith* y *nefi* reflejan la temática particular de este texto en comparación con el Nuevo Testamento, que se centra más en *jesús*, *cristo* y el relato de sus discípulos.

La presencia de palabras como *espíritu santo* se mantiene en ambos textos, lo cual sugiere un nexo teológico compartido en la doctrina. No obstante, los nombres propios dominantes difieren: *nefi* y *josé smith* en el Libro de Mormón versus *juan*, *mateo* y *jesús* en el Nuevo Testamento.

6. Conclusión

El presente análisis comparativo entre el *Nuevo Testamento* y el *Libro de Mormón* ha permitido identificar similitudes y diferencias significativas tanto en el uso del lenguaje como en los patrones narrativos y editoriales de ambos textos.

Desde un punto de vista cuantitativo, se observó que el *Libro de Mormón* presenta una mayor extensión y diversidad léxica en comparación con el *Nuevo Testamento*. Esto podría atribuirse a su estructura más detallada, que abarca múltiples narrativas y eventos a lo largo de varios libros internos. Por otro lado, el *Nuevo Testamento* se caracteriza por una mayor simplicidad léxica, con un enfoque temático centrado en la figura de Jesucristo y sus discípulos.

El análisis de palabras más frecuentes y bigramas destaca la centralidad de términos como *dios* y *señor* en ambos textos, lo que refleja su carácter teológico compartido. Sin embargo, los nombres propios dominantes en cada texto —*nefi* y *josé smith* en el *Libro de Mormón* frente a *jesús*, *juan* y *mateo* en el *Nuevo Testamento*— subrayan las diferencias temáticas y doctrinales que los distinguen.

En cuanto a los signos de puntuación, se identificó un uso más diversificado en el *Libro de Mormón*, mientras que el *Nuevo Testamento* tiende a un estilo más directo y menos segmentado. Esto sugiere un contraste en la estructura narrativa: un texto más complejo y descriptivo en el caso del *Libro de Mormón*, frente a uno más sintético y lineal en el *Nuevo Testamento*.

Por último, el uso de técnicas de Procesamiento de Lenguaje Natural, co-

mo la tokenización, el análisis de n-gramas y las nubes de palabras, demostró ser una herramienta eficaz para obtener insights cuantitativos y cualitativos sobre textos religiosos de gran extensión. Estos métodos no solo ofrecen una perspectiva novedosa sobre los datos textuales, sino que también pueden aplicarse a otros estudios similares para explorar patrones lingüísticos, narrativos y temáticos en diversas fuentes literarias.

En resumen, este análisis ha demostrado que el lenguaje y la estructura editorial de los textos reflejan las diferencias culturales, doctrinales y narrativas de las tradiciones religiosas de las que emergen, proporcionando una base sólida para futuras investigaciones en este campo.

7. Referencias

Referencias

- [1] Santa Biblia. Nuevo Testamento. Reina-Valera 1960. Sociedad Bíblica, 1960.
- [2] Smith, Joseph. *El Libro de Mormón: Otro Testamento de Jesucristo*. La Iglesia de Jesucristo de los Santos de los Últimos Días, 1830.