

# Tarea 2. Análisis de sentimiento

Helena Patricia Carrillo Soto

Facultad de Ciencias Fisico-Matemáticas, UANL

Maestría Ciencia de Datos: Procesamiento y Clasificación de Datos

## Resumen

Se analizaron dos traducciones, una adaptación y una edición en el idioma original de la novela inglesa *Orgullo y Prejuicio*. De dicho análisis se obtienen y comparan estadísticas descriptivas de texto de la novela original como continuación al artículo previo. Además se realiza un análisis de sentimiento de los cuatro textos.

**Keywords:** Palabras; Frecuencias; Análisis de texto, Análisis de sentimiento

## 1. Introducción

**Análisis a realizar** En este artículo se realizará un análisis estadístico de la novela en el idioma original *Orgullo y Prejuicio* de la autora Jane Austen. Se analizaron las palabras más utilizadas en la novela. Posteriormente se elabora un análisis de sentimiento y emociones de las traducciones y la adaptación presentadas en el artículo pasado así como de la novela en el idioma original. Se realizaron comparativas de los resultados obtenidos.

**Obra elegida** *Orgullo y Prejuicio* publicada en 1813 se ha convertido en una de las mayores referentes de la literatura inglesa con el paso del tiempo. Se considera una de las primeras comedias románticas publicadas y ha tenido múltiples adaptaciones literarias y cinematográficas a lo largo de la historia .

## 2. Descripción de los datos

Los textos a analizar son los siguientes:

- Traducción realizada por **Ana María Rodríguez**. Editorial Penguin Clásicos.
- Traducción realizada por **Marta Salís**. Editorial Alba.
- Adaptación realizada por **Lourdes Íñiguez**. Editorial Clásicos a medida.
- Novela en el idioma original publicada por **George Allen**.

## 3. Metodología

Para realizar el análisis del texto en inglés se limpió dicho texto quitando su introducción y las notas de los editores que se encuentran al final. Se obtuvieron las palabras claves de cada párrafo al tiempo que se eliminan *stop-words*. Para esto se utiliza la librería *nlTK* y se agregaron algunas palabras a las *stop-words* de la librería después de un análisis inicial.

Posteriormente para realizar el análisis de sentimiento se dividieron todos los textos a analizar en sus respectivos capítulos. Inicialmente se planeaba hacer un análisis completo de los textos usando todos los capítulos, sin embargo por cuestiones de tiempo y eficiencia se decidió analizar solo algunos de ellos.

Los capítulos a analizar son los siguientes:

- **Capítulo 1**. Introducción de personajes y se establece el tono de la obra.
- **Capítulo 34**. Primera proposición de Mr. Darcy
- **Capítulo 36**. Realización de la verdad por parte de Elizabeth.
- **Capítulo 56**. Segunda proposición de Mr. Darcy

## 4. Resultados

**Análisis de texto en inglés** Los primeros estadísticos que se obtienen son el número de palabras y párrafos. Los resultados se pueden observar en el cuadro junto con los resultados de cada una de las traducciones [1](#).

De los resultados de dicho cuadro observamos que las traducciones de Ana María Rodríguez y Marta Salís tienen longitudes muy similares al texto en inglés.

Cuadro 1: Número de palabras y párrafos por texto analizado.		
Texto	Número de palabras	Número de párrafos
Trad. Ana María Rodríguez	640,758	2,224
Trad. Marta Salís	685,406	2,225
Adap. Lourdes Íñiguez	174,096	671
<b>Idioma original</b>	<b>686,466</b>	<b>2,244</b>

Después de este análisis se revisaron las palabras que más se repiten y se grafica el top quince. Podemos ver los resultados en la figura [1](#).

En esta gráfica podemos observar que los nombres de varios personajes forman parte del top. Observamos nuevamente que los protagonistas y los títulos que se usan para referirse a ellos tales como *mr*, *mrs* y *miss* forman gran parte del top. Esto refuerza la idea de que la novela esta basada fuertemente en sus personajes y las relaciones entre ellos.

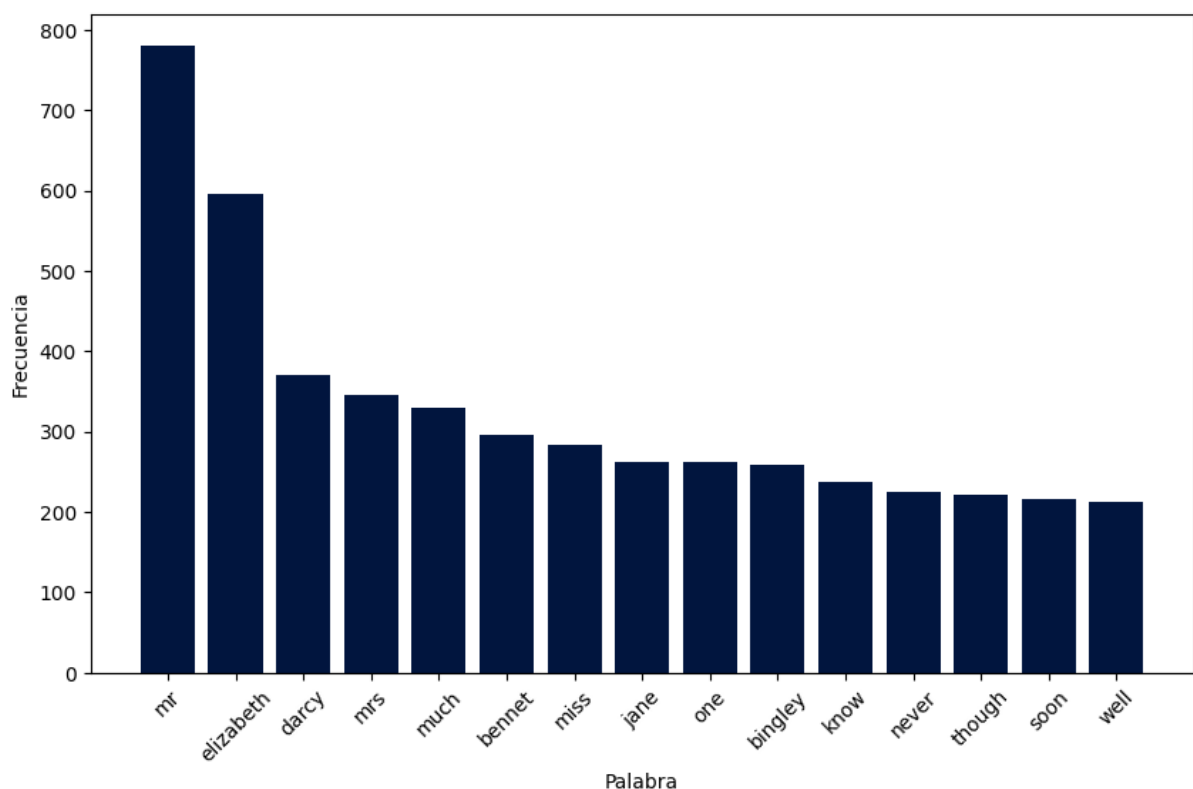


Figura 1: Palabras más repetidas en texto idioma original

Otra manera de visualizar las palabras más frecuentes de es a través de nubes de palabras. Esta se pueden encontrar en la figura [2](#).



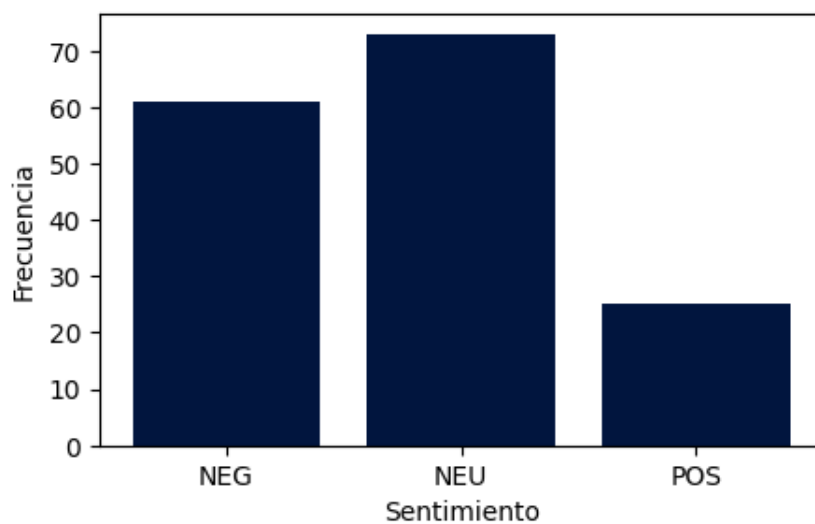


Figura 4: Frecuencias de sentimientos de la traducción de Marta Salís

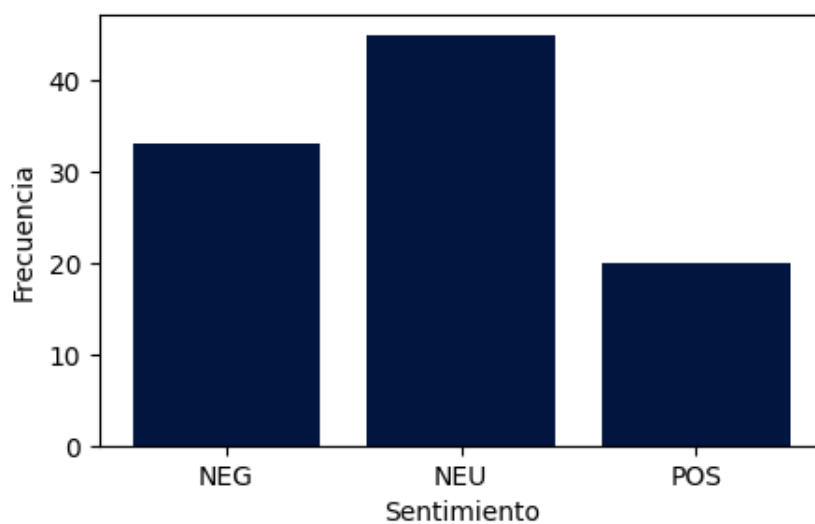


Figura 5: Frecuencias de sentimientos de la adaptación de Lourdes Íñiguez

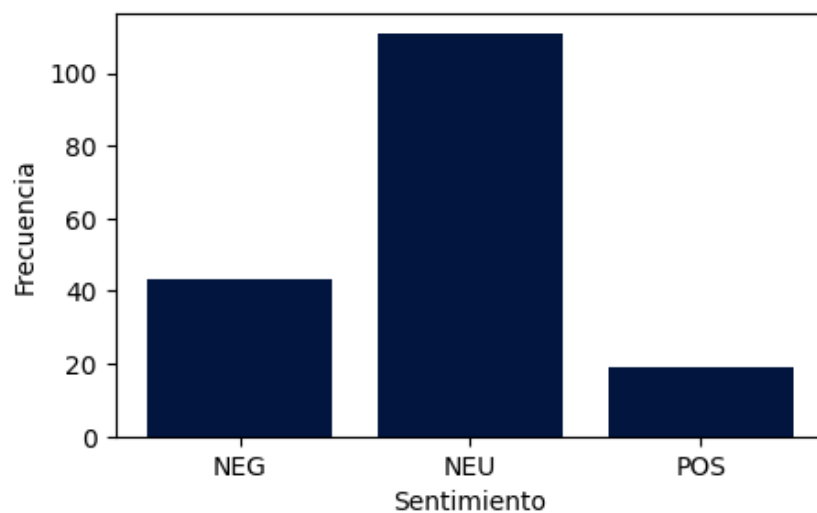


Figura 6: Frecuencias de sentimientos de la novela en el idioma original

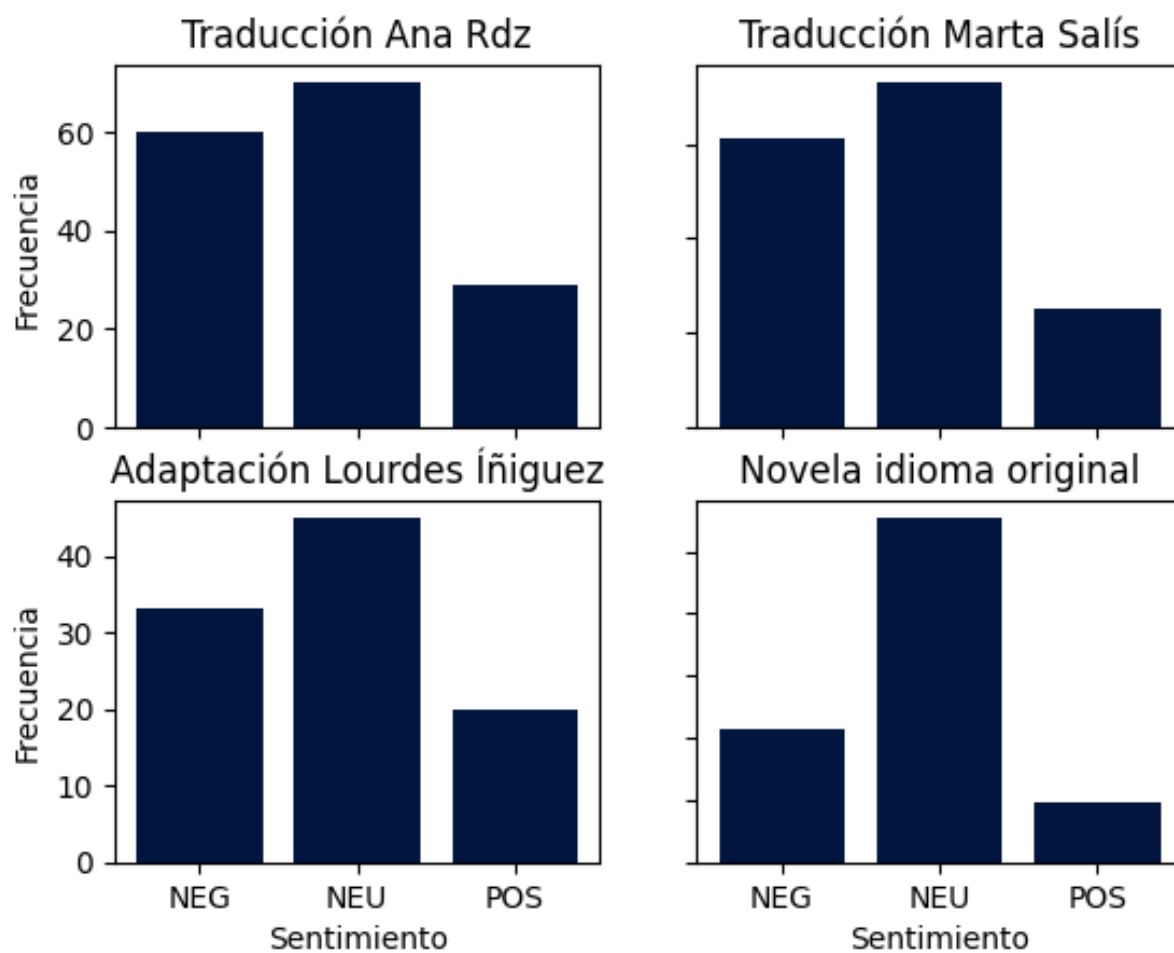


Figura 7: Frecuencias de sentimientos de los cuatro textos

También es posible graficar estas frecuencias en conjunto para facilitar su comparación. Esta gráfica se puede encontrar en la figura 8.

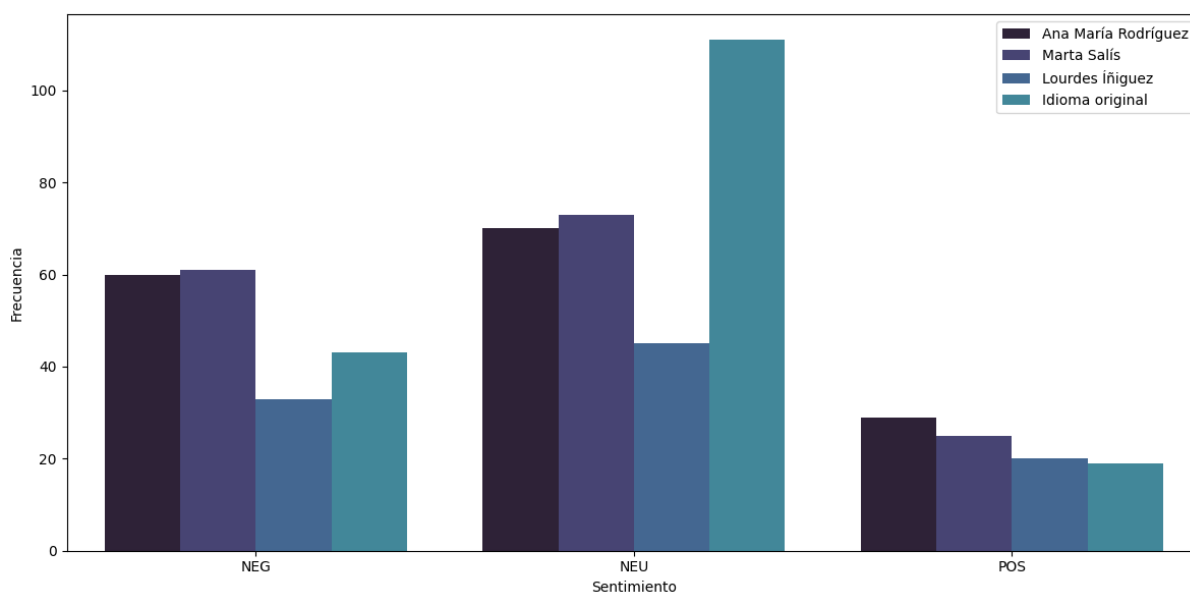


Figura 8: Frecuencias de sentimientos de los cuatro textos analizados

Comparando la gráfica podemos ver que en la novela original en inglés se percibe un lenguaje marcadamente más neutro. A pesar de que en todos los textos analizados se tiene que el resultado neutro es el más común la proporción entre los resultados muestra que esta es mayor en el idioma original. Podríamos teorizar que esto se debe a que el lenguaje español es más expresivo que el inglés.

**Análisis de emociones** Por otro lado se realizó también un análisis de las emociones expresadas en el texto con la ayuda de la misma librería. La librería cuenta con siete posibles resultados: **alegría, tristeza, ira, sorpresa, disgusto, miedo y otras.**

Una vez más se analizaron los párrafos de cada capítulo y se obtuvo la frecuencia con la que el resultado cae en cualquiera de las siete categorías. Se graficaron estas frecuencias y los resultados para cada texto se pueden encontrar en las figuras 9, 10, 11 y 12. De igual modo en la figura 13 también podemos observar las frecuencias en conjunto para facilitar su comparación.

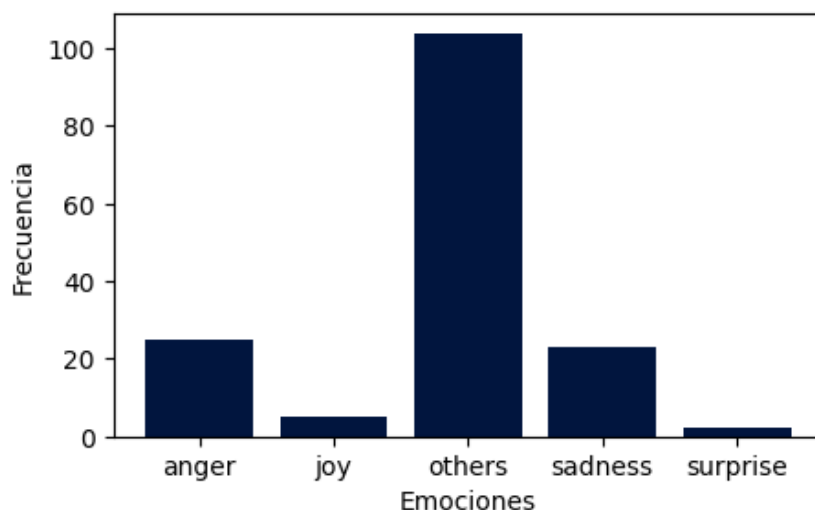


Figura 9: Frecuencias de emociones de la traducción de Ana Maria Rodriguez

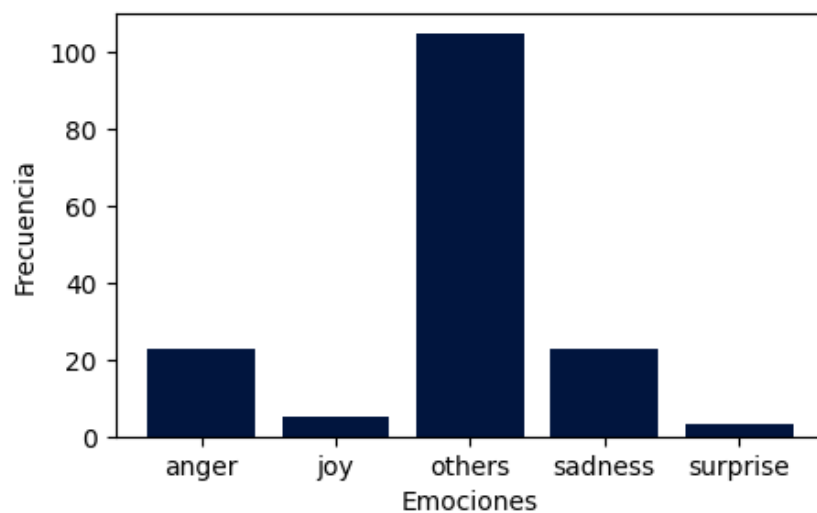


Figura 10: Frecuencias de emociones de la traducción de Marta Salís

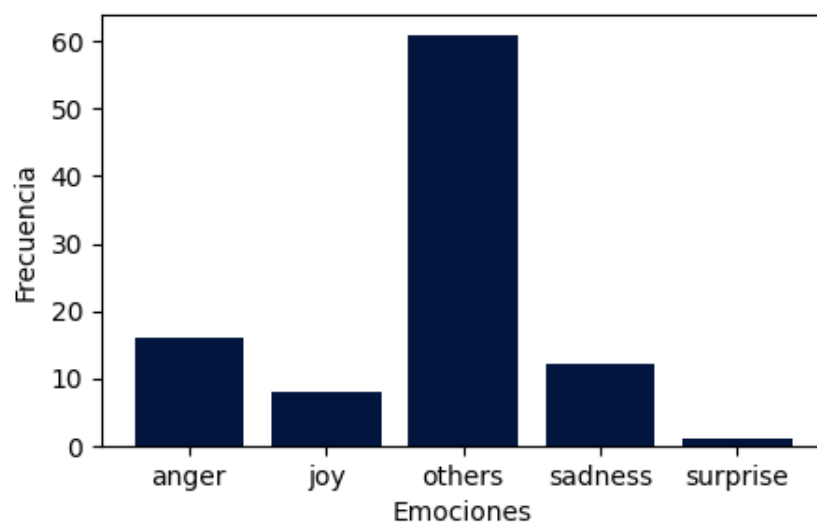


Figura 11: Frecuencias de emociones de la adaptación de Lourdes Íñiguez

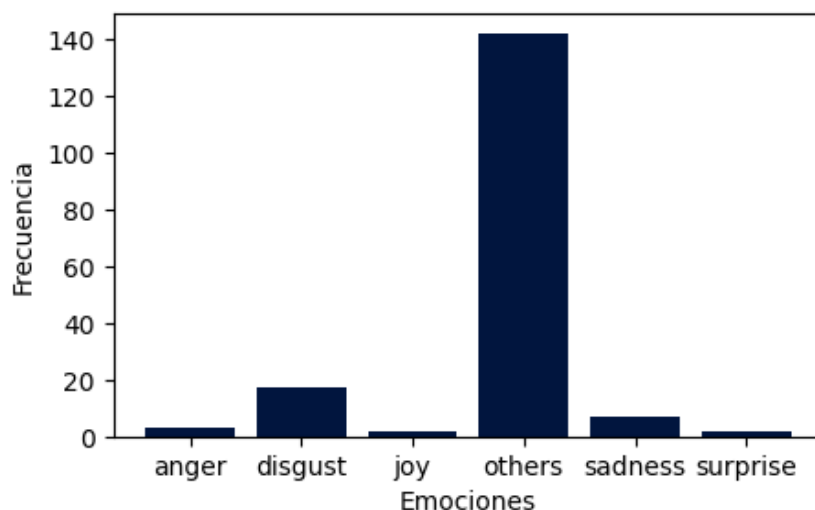


Figura 12: Frecuencias de emociones de la novela en el idioma original

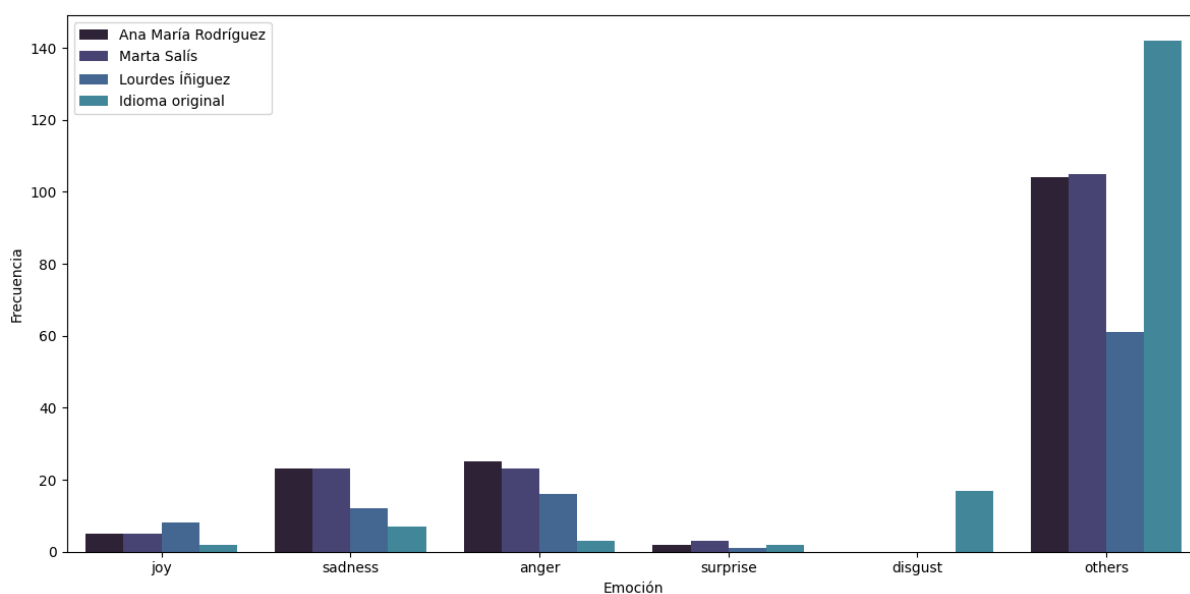


Figura 13: Frecuencias de emociones de los cuatro textos analizados

Llama la atención que el único texto en reflejar emociones de disgusto es la obra en el idioma original a la vez que esta muestra una falta considerable en las emociones de tristeza e ira comparado con los otros textos en español. Una vez más podemos teorizar que en el español las emociones de ira y tristeza son más fácilmente identificables y utilizan más lenguaje en común con el disgusto que en el idioma inglés.

## 5. Conclusiones

El análisis de sentimientos y emociones y su posterior comparación de obras en dos idiomas, en este caso inglés y español, inminentemente da paso a interesantes teorías y conclusiones respecto a la manera en que las personas se expresan en cada uno de ellos. Las diferencias culturales dentro de las cuales estos idiomas han evolucionado dan paso a interesantes comparativas entre la intensidad de las emociones percibidas en obras escritas dentro de cada uno de estos contextos.