

# MID-TERM PROJECT

Presented by: Carmen, Kathleen, Selina, and Haripriya

# INTRODUCTION

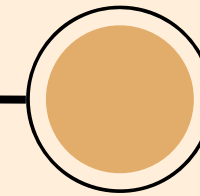
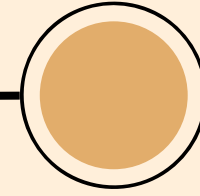
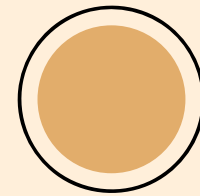
This study used AI and machine learning to identify rhetorical language in ancient Spanish and Catalan literature. This project consisted of three XML files containing ancient scriptures. Each of the files has been partially labeled and annotated with tags. Our group used machine learning to identify the following rhetorical devices for the rest of the text: metaphor, anaphora, pathos, and parallelism.

# MODELS

- **LLMs:** models trained using deep learning algorithms capable of a broad range of natural language processing (NLP) tasks. These models allow us to do few-shot learning with very little data.

- **NLP:** We used the BETO model (a BERT model) specifically trained on Spanish, for the rhetorical devices with more available data (ex. metaphor).

# PREPROCESSING



## **1. Parsing XML files for Rhetorical Devices**

We created a function that iterates through each `<seg>` element, checks for the type and subtype attributes, and extracts the text between them.

## **2. Tokenization**

We loaded a Spanish SpaCy model “`es_core_news_sm`”, which would break the text into smaller chunks(tokens).

## **3. Part of Speech (POS) feature extraction**

We used the SpaCy library to import a function that analyzes every word and finds the appropriate part of speech for that word.

The logo for Gemini 1.5 Flash is displayed on a dark blue gradient background. The word "Gemini" is in a light blue font with a small star above the 'i', and "1.5 Flash" is in a white font below it.

**Gemini**  
**1.5 Flash**

# PATHOS

- Pathos are sentences that purposefully evokes a strong emotion in the readers by using some emotional words or descriptions.
- There are only 13 prelabeled examples.
- Gemini 1.5 Flash
- Prompt engineering using LLMs allows users to do few-shot learning - perfect for very small datasets.
- We inputted 4 examples in the prompt and asked the model to identify the rest in the documents.
- Gemini has quota limit

# PATHOS

Here is the prompt I fed into the Gemini model

You are an expert in rhetorical analysis.

The following text is written in Spanish or Catalan and may contain multiple examples of emotional appeals (pathos).

Your task is to:

1. Read the full text.
2. Identify all specific segments that use emotional language to appeal to the reader's feelings.
3. Extract and list each of those emotional segments (in their original language).
4. For each, explain briefly in English why it is an emotional appeal.

Here are a few examples of emotional text:

1. "el qual conpuso Teresa de Cartajena seyendo apasyonada de graues dolenças"  
This an example of pathos.
2. "Salid, señores, yveréis la más desventurada, desamparada y más maldita mujer del mundo"  
This is an example of pathos.
3. "Y yo le dije, con muchas lágrimas"  
This is an example of pathos.
4. "Los que morauan en tinieblas y en sonbra de muerte, luz les es demostrada."  
This is an example of pathos



# CHECKING ACCURACY

```
if __name__ == "__main__":  
    document = """  
    Y yo estaba tan traspasada de  
pesar que no podía hablar de la humillación que me hacían aquellos señores. Y el  
triste de mi hijo decía: "Decidle a mi señora doña Teresa Fernández Carrillo que no  
me mande echar, que ahora saldrá mi alma hacia el cielo".  
Si tan ansios son de sentir la angustia mía. Versión a la lengua castellana actual: Mari  
señoras. Nobles y burguesas que escriben, en Anna Caballé, ed., La vida escrita por  
las mujeres, 1: "por mi alma os digo. De la Edad Media a la Ilustración", Barcelona,  
Círculo de Lectores, 2003, 33-41, y Barcelona, Lumen, 2004, t. 4, 28-40.  
E en aquesta caritatiua pregaria staua la senyora la major part de la nit ab continues lag  
Anna quel veu plorar no pogue tenir que lo cor seu no rompes en moltissimes lagrimes. La  
de 1363, en la casa de su primo el rey Pedro I de Castilla. Sancha Carrillo era  
sobrina de Alfonso XI, en cuya casa se había criado.  
E axi adaz e los fills seus continuant les sues clamors e pregaries ab molta feruor e deu  
    """
```

I input a text that includes 5 prelabeled other pathos from the XML files and hide them with other non-pathos text that are also from the documents. To test the accuracy, I check if the model correctly identifies all the pathos. Here, the model seems to be pretty accurate.

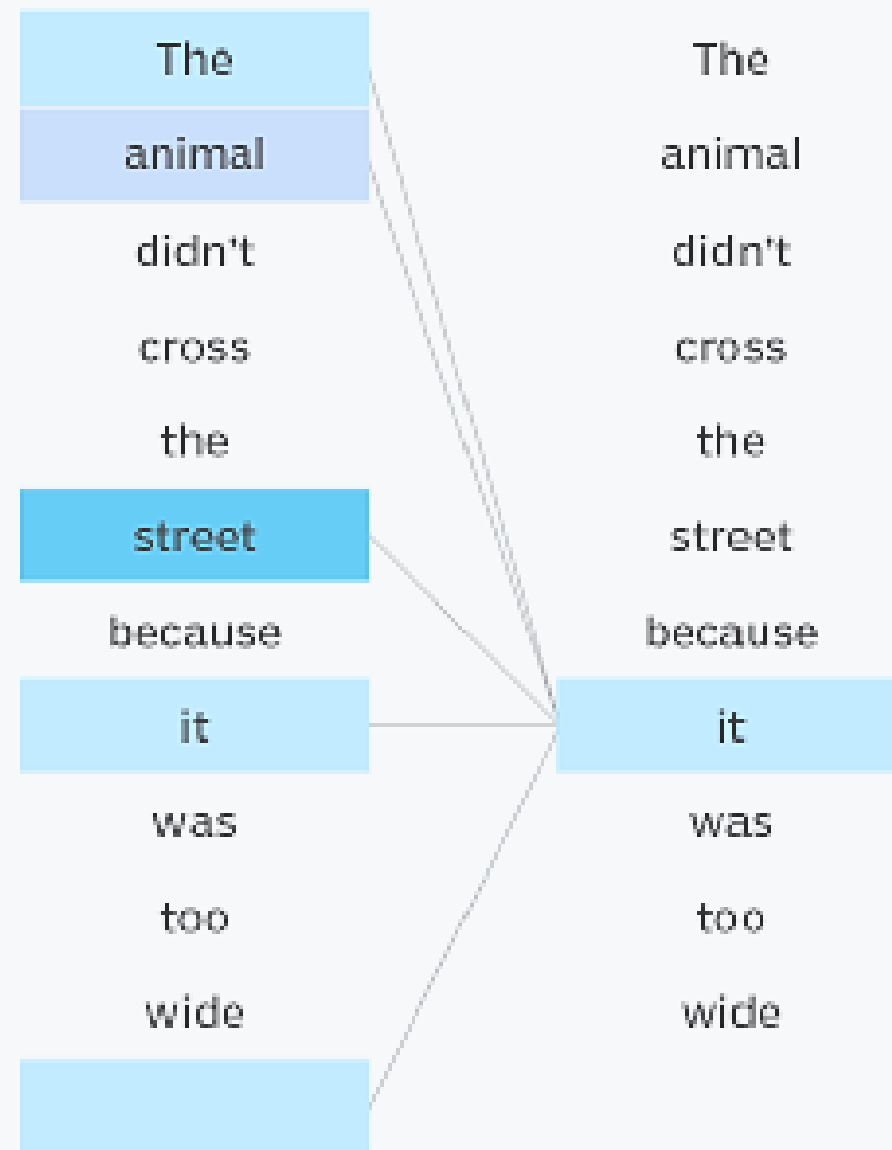
The provided text contains several examples of pathos, appeals to emotion. Let's analyze them:

1. **\*\*\*"Y yo estaba tan traspasada de pesar que no podía hablar de la humillación que me hacían aquellos señores."\*\*\*** This is pathos. The phrase "tan traspasada de"
2. **\*\*\*"Y el triste de mi hijo decía: "Decidle a mi señora doña Teresa Fernández Carrillo que no me mande echar, que ahora saldrá mi alma hacia el cielo".\*\*\*** This
3. **\*\*\*"Si tan ansios son de sentir la angustia mía."\*\*\*** This is pathos. The phrase "angustia mía" (my anguish) is a direct appeal to the reader's empathy. The q
4. **\*\*\*"E en aquesta caritatiua pregaria staua la senyora la major part de la nit ab continues lagrimes secretes e molt amagades: hauent ferma fe de obtenir lo que**
5. **\*\*\*"Anna quel veu plorar no pogue tenir que lo cor seu no rompes en moltissimes lagrimes."\*\*\*** This is pathos. The statement that Anna, upon hearing the speak
6. **\*\*\*"E axi adaz e los fills seus continuant les sues clamors e pregaries ab molta feruor e deuocio."\*\*\*** This is pathos. The phrase "ab molta feruor e deuocio" (w

In summary, all the listed segments utilize emotional language (despair, sorrow, suffering, tears, devotion) to directly connect with the reader's feelings and c

# METAPHOR

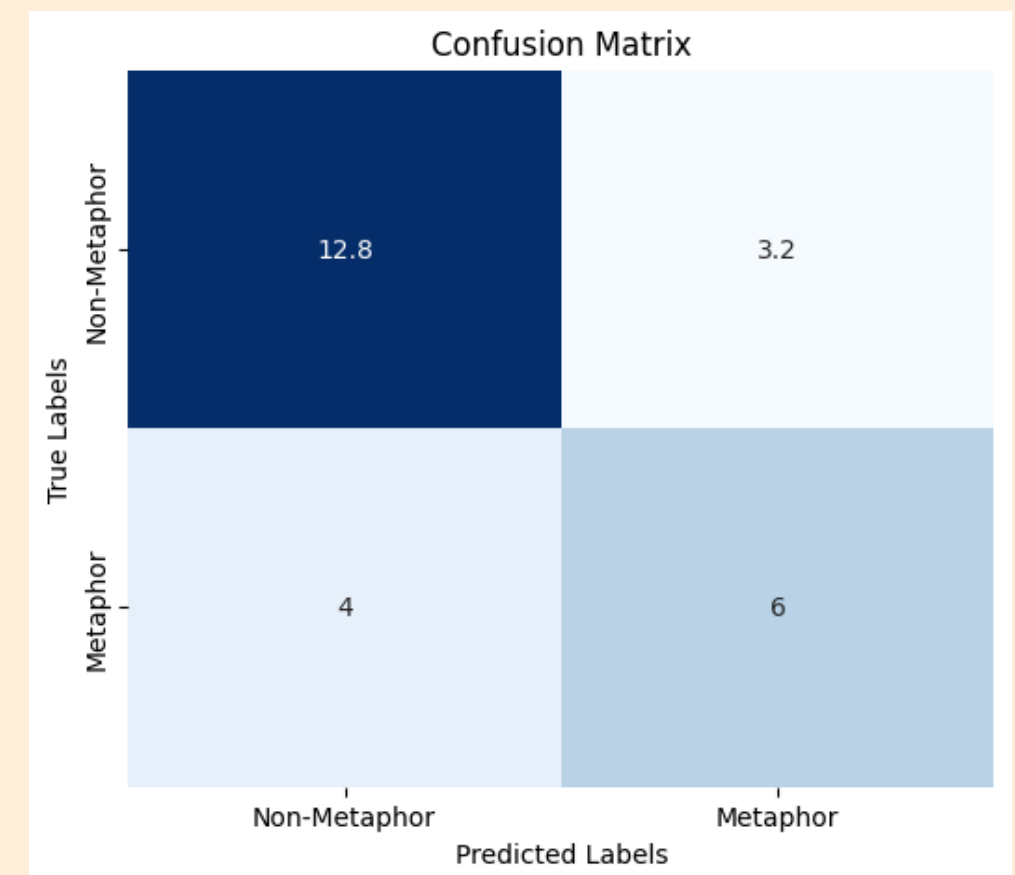
## BERT chart



©2020 TECHTARGET. ALL RIGHTS RESERVED

- Parse through the XML files for labeled metaphor examples (found 30) and combine with 100 additional AI-generated metaphor examples
- Preprocessed the text and input it into the Pre-trained BERT Model
- BERT has a better understanding of context, since it can take into account the words before and after a given word.
- Fine-tuned the model for this specific binary classification task
- The final model achieved 84.6% accuracy

Detailed Test Set Results:  
Accuracy: 0.8462  
Precision: 0.8769  
Recall: 0.8462  
F1-Score: 0.8355





# Metaphor Identified!

## Number of metaphors detected: 138

1. **La libertad de decir y la originalidad de esta autora son fruto de la escuela del dolor.**
  - a. "The freedom to speak and the originality of this author are the fruit of the school of pain."
2. **"La escritura femenina desplaza barreras de lo simbólico irrumpiendo en ellas con un disparate."**
  - a. "Feminine writing displaces barriers of the symbolic by bursting into them with a nonsense."
3. **"El espejo pasa entonces a reflejarla a ella borrosa o deformada."**
  - a. "The mirror then begins to reflect her, blurry or deformed."
4. **"Un ámbito que pretende sustituir la vida con la vida social, inaugurando la percepción de lo político que harán renacer siglos después las mentes más especulativas de las cortes europeas del Humanismo."**
  - a. "An area that aims to replace life with social life, inaugurating the perception of the political that will be reborn centuries later in the speculative minds of the European courts of Humanism."
5. **"La escritura le redimió y, de este modo, le permitió hacer en su vida un nuevo inicio."**
  - a. "Writing redeemed her and, in this way, allowed her to make a new beginning in her life."

metaphor_sentences_by_file	
Metaphor Sentence	Source File
Antes de , en esa etapa de su vida anterior a los recuerdos, Leonor López, su madre, sus hermanas y las hijas del monarca fueron llevadas a Bayona como rehenes por Martín López y por Pedro I, en el marco d	Memorias
La acogió entonces en Córdoba su tía abuela materna María García Carrillo, señora de Aguilar, en unas casas cercanas a las suyas junto a la colegiata de San Hipólito	Memorias
O tal vez Leonor solicitara el documento con el fin de evitar en el futuro conflictos patrimoniales con su marido, si la dote no había llegado a pagarse a consecuencia de las dificultades familiares, económicas	Memorias
Su marido, sin embargo, desaparece del relato de las Memorias	Memorias
Comenzó, en cambio, una estrecha relación con la Virgen, que le inspira y le ayuda en sus asuntos económicos y familiares	Memorias
Unos años antes en y en , el obispo de Córdoba, previa consulta real, había concedido permiso al prior y canónigos de San Hipólito para que vendieran unos corrales entre la iglesia y el muro de la ciudad	Memorias
por esos años, movida por la piedad y el amor de Dios, Leonor había adoptado y criado un niño judío al que bautizó como Alonso, que se había quedado huérfano durante el asalto a la judería de Córdoba a pri	Memorias
El jabón producía una renta interesante	Memorias
Cuando ya ha obtenido sus primeros deseos y estaba instalada en su casa junto a San Hipólito de Córdoba, precisamente donde ella quería y donde le indicó la Virgen, estalló la peste de , que afectó a la ciuda	Memorias
Entre cambios de humor y amarguras incontables, entre las que destaca la enemistad con sus primas las hijas de su protectora, Leonor relata los estragos de la peste, que entró en Aguilar y en su familia por m	Memorias
Leonor fue entonces expulsada de hecho de Aguilar por sus parientes y regresó a Córdoba, donde su tía y protectora rompió con ella	Memorias
Entre y mediados de , cuando tenía unos cuarenta años, Leonor dictó sus Memorias a un escribano público o notario de Córdoba	Memorias
Era hija de Juan de Gante, duque de Lancáster, y de Constanza de Castilla, hija de María de padilla y de Pedro I, madrina y compañera de infancia de Leonor López de Córdoba	Memorias
Tuvo dificultades para entenderse con la reina desde el inicio de la regencia	Memorias
En dio un golpe de estado contra la reina y su válida, sin conseguir destruir la relación entre ambas, pero erigiéndose de hecho en regente único	Memorias
El golpe incluyó, sin embargo, una grave campaña de difamación contra Leonor López de Córdoba, que él mismo explicó en una carta escrita en Guadalajara el de septiembre de y dirigida a la ciudad de Murc	Memorias
pero, en , en un contexto confuso en el que Leonor intentó volver al lado de Catalina, con permiso del omnipotente regente, y la reina fue inducida a temer que Leonor López se hubiera aliado con Fernando de	Memorias
La reina declaró que si se acercaba la mandaría quemar	Memorias
En estas casas, llamadas del rey Almanzor, se hizo en el siglo XVIII el Hospital Salazar, hoy Facultad de Filosofía y Letras de la Universidad de Córdoba	Memorias
A la capilla de la Trinidad, hoy se le llama de la Virgen del Rosario, un culto muy antiguo dentro de esa capilla	Memorias
Las Memorias Las Memorias son un relato a viva voz de las extraordinarias vivencias de la autora en sus primeros cuarenta años	Memorias
Terminan con la muerte de su hijo mayor y la ruptura con su tía y protectora, sin ninguna alusión a la etapa brillante de válida de la reina ni a su caída en desgracia	Memorias
Leonor dictó su texto con la intención explícita de que su versión de la verdadera historia de su vida perdurara en la memoria de la gente que oyera leer su relato	Memorias
Su estilo es extraordinariamente libre, escueto y certero, adecuado siempre a la expresión de los sentimientos más dispares y de más variada intensidad sentimientos que van de la dulzura a la ira, del amor a la	Memorias
Describe con precisión y belleza acontecimientos de su tiempo como las epidemias de peste de y , o algunas consecuencias del asalto a la judería de Córdoba en también, las relaciones que civilizan la com	Memorias
Su vida es testimonio de la época en que el reino de LeónCastilla cambió sus sentimientos de admiración hacia el Oriente andalusí por Europa	Memorias
En el centro de las Memorias está la figura de Leonor, rodeada de las mujeres de su familia y de la Virgen, en la que depositó su esperanza de recobrar su fortuna y de rehacer su vida	Memorias
La libertad de decir y la originalidad de esta autora son fruto de la escuela del dolor, de sus talentos y de la educación femenina que recibió en la infancia una infancia pasada entre su madre, sus tres hermanas	Memorias
Al estar presa entre los ocho y los dieciséis años, no tuvo preceptores que le inculcaran los estereotipos de género femenino propios de su clase y época, ni el que defendían los pedagogos misóginos que ente	Memorias
El espejo pasa entonces a reflejarla a ella borrosa o deformada	Memorias
El mito recoge una verdad conocida, verdad conocida que presenta y transmite en forma de fábula cuando otras formas de transmisión del conocimiento no son capaces de acoger esa verdad	Memorias
Un ámbito que pretende sustituir la vida con la vida social, inaugurando la percepción de lo político que harán renacer siglos después las mentes más especulativas de las cortes europeas del Humanismo	Memorias
La crónica de Alvar García de Santa María h	Memorias
El palacio o Casa del Bailío es del siglo XVI tenía contiguo a él el Hospital de incurables de San Jacinto, en el que está hoy la capilla de la Esperanza, en la plaza del Cristo de los Faroles	Memorias
Se celebraron un total de setenta y cuatro audiencias, unas por la mañana de seis a nueve, otras por las tardes, desde las cuatro hasta el toque de oraciones	Memorias
En el proceso de formación de su patrimonio destacan las operaciones que le llevaron a la posesión del cortijo de Teba y de las dos Huertas de la Reina, llamadas de Adentro y de Afuera	Memorias
Un marido que pertenecía, como ella, a la más alta nobleza de Castilla y que era, entre otras cosas, primo carnal de María de padilla, hija de María de Hinestrosa y abuela de la reina Catalina de Lancáster	Memorias
Había tenido, además, conflictos económicos con su madre	Memorias
A una criatura humana, la dificultad de poner libremente en palabras realidad vivida le genera sufrimiento	Memorias
Se muestra con los delicados síntomas de la histeria que es una manera sabia de convertir el cuerpo en texto a falta de palabras para decir algo vivido que las exige para dejarme en paz o, en casos extremos,	Memorias
Es así porque la criatura humana es un ser simbólico, un ser que habla, que necesita sentido y lengua materna para seguir viviendo	Memorias
En cambio, el desplazamiento de barreras de lo simbólico trae felicidad	Memorias
La trae a la gente en general porque ensancha su mundo, abriendo el respiro y renovando la esperanza	Memorias
La trae a quien desplaza la barrera porque le da ocasión de poner a disposición de quien lo quiera lo nuevo que ella o él trae al mundo, ya que cada criatura humana que nace trae algo nuevo al mundo común	Memorias
El algo nuevo es un legado de sentido, que puede estar hecho de palabras, palabras que, a veces, son relatos, son historia	Memorias
La escritura femenina desplaza barreras de lo simbólico irrumpiendo en ellas con un disparate	Memorias
Recientemente, Luisa Muraro ha añadido El Dios de las mujeres	Memorias
Podría decirse que la escritura es femenina cuando el Dios de las mujeres acaece en ella cuando acaece en las relaciones que se dan entre quien escribe, lo real y la capacidad de decir de quien escribe	Memorias
Y nace una invención simbólica	Memorias
La escritura femenina, al desplazar barreras de lo simbólico, trae al mundo sentido indisponible, sentido hasta entonces no disponible	Memorias
De la importancia del sentido para la existencia humana tomamos conciencia en el siglo XX, al darnos cuenta de que lo necesitábamos para vivir tanto como el bienestar económico	Memorias
En otras palabras, el ensayo histórico no expresa ya o no expresa apenas el acaecer del Dios de las mujeres en las relaciones humanas y sociales, relaciones profundamente modificadas por la revolución fente	Memorias
Las ideologías rellenan lo que Luisa Muraro ha llamado el centro vacío de cada palabra, es decir, el agujero que, en tiempos o en sitios sin Dios de las mujeres, es la huella que de él o ella queda, un agujero y u	Memorias
Son ejemplos épocas como el final de la Edad Media con su Guerra de los Cien Años, terminada por Juana de Arco o el siglo XX, terminado con el final del patriarcado	Memorias
En la distancia que la objetividad impone, se pierde el vínculo un vínculo que, en la vida, es indisoluble entre lo que acontece y las mujeres u hombres que viven y relatan los acontecimientos	Memorias
La escritura femenina de la historia explora la vivencia personal femenina o masculina de los acontecimientos junto con su acontecer, sin separaciones	Memorias
Son textos que nos tocan con su realidad puesta finalmente en palabras	Memorias
Para poner en palabras realidad vivida, las historiadoras han usado mucho a Dios y a lo sobrenatural	Memorias
Como desconciertan una desventura o una felicidad, sin dejar por ello de ser historia	Memorias
Un ejemplo son las leyendas fundacionales de bastantes monasterios de clarisas fundados en el siglo XIII, que ha estudiado, entre otras, Núria Jornet Benito	Memorias
Las Memorias de Leonor López de Córdoba Hay un texto fascinante de la historia de la Castilla de Pedro I, Enrique II, Juan I y Enrique III que expresa lo que es la escritura femenina de la historia	Memorias
Son las Memorias que dictó en Córdoba a un escribano público de esta ciudad, entre y mediados de , Leonor López de Córdoba Calatayud Córdoba	Memorias
Este texto es la primera autobiografía conocida en lengua castellana	Memorias
En cambio, la historia social propia del siglo XX perdió el interés por esta breve pero abigarrada obra, en parte por la pertenencia de clase de Leonor la alta nobleza pero sobre todo pienso precisamente por la	Memorias
La historia social tuvo el talento de desplazar la atención principal de la historiografía de lo individual los grandes hombres a lo colectivo, con el fin de servir a la concienciación y a la lucha para reducir las eno	Memorias
Esto le llevó a tener como principal referente hermenéutico el poder social y las estrategias para obtenerlo o abolirlo por unos, y a conservarlo o aumentarlo por otros	Memorias
Este es el motivo por el que la parte más interesante de la experiencia humana femenina excede y desborda la historia social, que carece de palabras para decirla	Memorias
Las Memorias de Leonor López de Córdoba testimonian estas tres prácticas	Memorias
Es esta teología la que es capaz de entender que Leonor López de Córdoba llevara al mercado su absoluta necesidad y su enorme deseo, necesidad y deseo que eran de realidad, de dejar de seguir viviendo co	Memorias
Por eso, su escritura está llena de momentos de temor y temblor, que son los signos que señalan el presentarse de la experiencia mística	Memorias
Basta leer su obra en clase para comprobarlo año tras año	Memorias
El temor y el temblor se hacen insoportables cuando a Leonor le llega el momento de llevar al mercado la vida de su hijo primogénito, Juan Fernández de Hinestrosa	Memorias
En el tiempo, es el momento en el que, estando en Santaella durante la epidemia de peste de , el niño que tenía entonces doce años y cuatro meses acude a su madre para decirle que no hay quien vele esa no	Memorias
Pero incluso con el recurso a Dios, el diálogo nos deja temblando	Memorias

stored in a CSV file, categorized by the three texts

# PARALLELISM

- A RHETORICAL DEVICE DEFINED BY HAVING A SIMILAR GRAMMATICAL STRUCTURE
- HAD NO TRAINING DATASET AVAILABLE
- TRANSLATED THE ANCIENT TEXT INTO MODERN SPANISH
- USED SPACY LIBRARY TO GET THE POS OF EACH WORD
- USED N-GRAMS TO FIND THE SAME POS STRUCTURES



# PARALLELISM CODE

(THIS IS NOT ALL OF THE CODE)

```
def process(file):
    #transforms the xml file into a list of lists
    vec= DictVectorizer()
    file_dict = parse_tei_xml(file)
    document = file_dict['text']
    count = 0
    for seg_text in document:
        count += 1
        words = seg_text.split(" ")
        if len(words) <= 4:
            continue
        pos_window = []
        doc = convert(seg_text)
        #adds the parts of speech for each word
        pos_window.append([{ w.text, w.pos_} for w in doc])
        eys_string = ''.join(str(w.pos_ + " ") for w in doc)
        #checks to see how many times the pos is repeating in a sentence
        vectorizer = CountVectorizer(stop_words=None,ngram_range=(2, 5), token_pattern=r'(?u)\b\w+\b')
        X = vectorizer.fit_transform([eys_string],)
        largest=X.toarray().max()
        #prints the line if the pos structure repeats multiple times
        if largest > 1 :
            print(repr(seg_text)+"- has parallelism.")
    print("Total number fo records in " + file + " : " + str(count) + "\n")
```

```
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction import DictVectorizer

Run cell (Ctrl-Enter)
cell executed since last change 11 mn
executed by Haripriya T 7:40 PM (2 hours ago)
re-executed in 3.556s
functional as F
as optim

from sklearn.feature_extraction.text import CountVectorizer
import en_core_web_sm
from lxml import etree

# Load the Spanish language model
nlp = spacy.load("es_core_news_sm")

archaic_to_modern = {
    "fizolo": "lo hizo", "llamadole": "le llamó", "dixo": "dijo", "qual": "cuál", "delante": "enfrente", "non": "no", "por que": "porque",
    "pluguiese": "gustase", "archaic": "modem", "vno": "una", "ynsula": "isla", "dixo": "dijo", "fablar": "hablar", "ovo": "huevo", "non": "no",
    "fizo": "hizo", "delante": "enfrente", "ca": "porque", "otroquo": "también", "falló": "encontró", "fallar": "encontrar", "pluguiese": "agradase",
    "sabidor": "sabio", "casacón": "matrimonio", "andava": "andaba", "fue": "fue", "tenió": "tenía", "desque": "después que", "al": "caballo",
    "porfia": "discusión", "gran": "grande", "demás": "además", "asaz": "mucho", "enojó": "se enojó", "pensó": "se imaginó",
    "asi mesmo": "asimismo", "dél": "de él", "aquella": "esa", "ciudad": "ciudad", "cierto": "cierto", "cerca": "cerca", "cielo": "cielo",
    "cima": "cima", "cena": "cena", "cien": "cien", "anciano": "anciano", "añedad": "ansiedad", "correr": "correr", "correría": "correría", "cunir": "sumir",
    "cedo": "pronto", "caña": "ira", "celebro": "cerebro", "coplar": "soplar", "combra": "sombra", "ciego": "ciego", "cuerpo": "cuerpo", "ciencia": "ciencia",
    "contava": "cuenta", "cifra": "cifra", "cierto": "cierto", "cerro": "cerro", "cerca": "cerca", "caña": "enojo", "cajado": "salfado",
    "cedo": "rápido", "ceñir": "ceñir", "cegar": "llegar", "cierto": "verdadero", "cima": "altura", "cunir": "hundir", "cerco": "cerca", "cima": "pico",
    "ceniza": "ceniza", "cebra": "cebra", "caña": "colera", "cántico": "cántico", "campo": "campo", "ceremonia": "ceremonia", "cielo": "firmamento", "cigarro": "cigarro",
    "cierto": "exacto", "cien": "cientos", "cerca": "al lado", "callo": "halló", "cantar": "cantar", "correr": "correr", "contar": "contar", "conciente": "consciente", "condado": "condado",
    "cincuenta": "cincuenta", "cambio": "cambio", "cuenda": "cuerda", "cupido": "cupido", "conforme": "conforme", "copa": "sopa", "curco": "surco",
    "cisión": "sesión", "cecilia": "cecilia", "como": "como", "caló": "saló", "cupieron": "supieron", "cambio": "sabio", "seguridad": "seguridad", "certamen": "certamen",
    "cillón": "sillon", "cilla": "silla", "cada": "cada", "joyzio": "juycio", "mugeril": "mujeril", "haze": "hace"
}

def parse_tei_xml(file_path):
    try:
        tree = etree.parse(file_path)
        root = tree.getroot() # the root element is referring to <TEI>
        # Initialize lists to store results
        plain_text = []
        rhetorical_devices = []
        lexical_segments = []
        # Iterate over all <seg> elements and returns rhetorical or lexis
        for seg in root.xpath("../seg"):
            seg_type = seg.get('type')
            seg_subtype = seg.get('subtype', 'No subtype') # Default to 'No subtype' if missing
            seg_text = seg.text
            # If text exists, process it
            if seg_text:
                plain_text.append(seg_text.strip())
                # Separate by type (rhetorical or lexical)
                if seg_type == 'rhetoric':
                    rhetorical_devices.append((seg_text.strip(), seg_subtype))
                elif seg_type == 'lexis':
                    lexical_segments.append((seg_text.strip(), seg_subtype))
            # Join and return the results
        return {
            #'text': " ".join(plain_text),
            'text': plain_text, # list of seg texts
            'rhetorical_devices': rhetorical_devices,
            'lexical_tags': lexical_segments
        }
    except Exception as e:
        print(f"Error processing file {file_path}: {e}")
        return None

#function that converts ancient Spanish to modern Spanish
def convert(text):
    normalized_text = text.lower()
    for key in archaic_to_modern:
        normalized_text = normalized_text.replace(key, archaic_to_modern[key])
    doc = nlp(normalized_text)
    return doc

def process(file):
    #transforms the xml file into a list of lists
    vec= DictVectorizer()
    file_dict = parse_tei_xml(file)
    document = file_dict['text']
    count = 0
    for seg_text in document:
        count += 1
        words = seg_text.split(" ")
        if len(words) <= 4:
            continue
        pos_window = []
        doc = convert(seg_text)
        #adds the parts of speech for each word
        pos_window.append([w.text, w.pos_ for w in doc])
        eys_string = ''.join(str(w.pos_ + " ") for w in doc)
        #checks to see how many times the pos is repeating in a sentence
        vectorizer = CountVectorizer(stop_words=None,ngram_range=(2, 5), token_pattern=r'(?u)\b\w+\b')
        X = vectorizer.fit_transform([eys_string],)
        largest=X.toarray().max()
        #prints the line if the pos structure repeats multiple times
        if largest > 1 :
            print(repr(seg_text)+"- has parallelism.")

def main():
    #inputs files
    files=[
        "/content/drive/My Drive/Classroom/AI_Mid_Term_XML_files/Arboleda119.xml",
        "/content/drive/My Drive/Classroom/AI_Mid_Term_XML_files/Memorias Final.xml",
        "/content/drive/My Drive/Classroom/AI_Mid_Term_XML_files/Vital119.xml"
    ]
    for file in files:
        #calls the function
        process(file)

if __name__ == "__main__":
    main()
```

# ANAPHORA

- Repetition of a word or phrase at the beginning of neighboring clauses.
- HuggingFace datasets and BETO (Spanish BERT) model
- Had to have `use_safetensors=True` for security reasons.

# ANAPHORA

```
Accuracy: 0.2308  
Precision: 0.8836  
Recall: 0.2308  
F1 Score: 0.3013
```

- My accuracy results.
- Room for improvement.

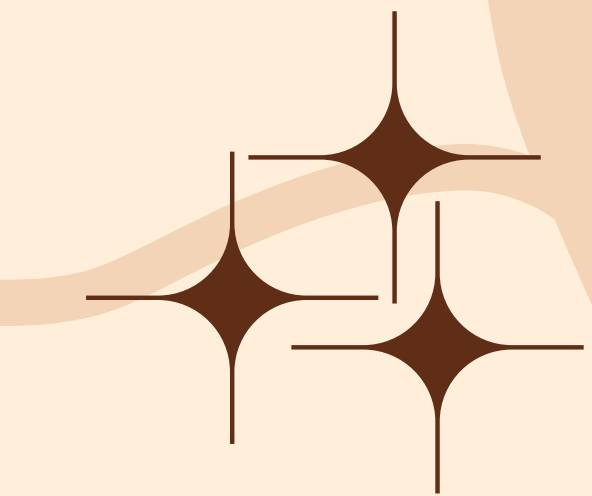
# **LIMITATIONS**

- **Very limited amount of data**
- **The annotated text may not contain certain rhetorical devices at all**
- **Checked for only a few rhetorical devices**
- **No mature models to process old Spanish**
- **The model assumes a connection with words**
- **Uncertain accuracy of LLMs**
- **There might be dialects that ML can not understand**
- **ML lacks the common sense humans have while looking for these devices**



# CONCLUSION

In conclusion, our team believes that machine learning could help improve the efficiency of this research to some extent, but the accuracy of the outputs should not be trusted completely. Some of the more complicated devices might need to be done manually. We look forward to seeing more mature models for processing ancient literature in different languages emerge in the future.





**THANK  
YOU!**

**ANY QUESTIONS?**