

# **ANÁLISIS DE FRECUENCIA DE PALABRAS**

**Carlos Navarro Martínez**

**Fecha:** 29 de January de 2026

**Documento:** AdanYEva.txt

**Análisis:** Procesamiento de Lenguaje Natural

# 1. INTRODUCCIÓN

Este informe presenta un análisis de frecuencia de palabras realizado sobre el texto "*Adán y Eva en el Paraíso*" de Eça de Queiroz. El objetivo del análisis es identificar los términos más utilizados en el texto mediante técnicas de procesamiento de lenguaje natural (NLP) y representarlos visualmente a través de una nube de palabras.

## 2. METODOLOGÍA

El análisis se realizó siguiendo los siguientes pasos:

### 1. Descarga del libro:

The screenshot shows the Project Gutenberg website interface. At the top, there's a navigation bar with links for 'About', 'Frequently Downloaded', 'Main Categories', 'Reading Lists', and 'Search Options'. A 'Quick search' bar is also present. Below the navigation, the title 'Adán y Eva en el paraíso by Eça de Queirós' is displayed. To the left of the title is a small thumbnail image of the book cover. The main content area contains a summary of the book, followed by a 'Read or download for free' section. This section includes a 'How to read' link, a 'Read now!' button, and a list of download options. The list includes: EPUB3 (E-readers incl. Send-to-Kindle), EPUB (older E-readers), EPUB (no images, older E-readers), Kindle, older Kindles, Plain Text (UTF-8), and Download HTML (.zip). Each item has a file size listed next to it. There's also a note at the bottom stating 'There may be more files related to this item.'

- **2. Preprocesamiento del texto:** Se leyó el archivo de texto original y se aplicaron técnicas de limpieza.
- **3. Limpieza de texto:** Se eliminaron signos de puntuación, números y caracteres especiales. Todo el texto fue convertido a minúsculas.
- **4. Eliminación de stopwords:** Se removieron palabras vacías (stopwords) del español que no aportan significado semántico significativo (artículos, pronombres, conjunciones, etc.).
- **5. Análisis de frecuencias:** Se contabilizó la frecuencia de aparición de cada palabra única en el texto procesado utilizando la librería *collections.Counter*.
- **6. Generación de nube de palabras:** Se creó una representación visual de las frecuencias utilizando la librería *WordCloud*, donde el tamaño de cada palabra es proporcional a su frecuencia de aparición.

## 2.1. Librerías Utilizadas

Para el análisis se utilizaron las siguientes librerías de Python:

- **NLTK (Natural Language Toolkit):** Para tokenización y manejo de stopwords
- **Collections:** Módulo Counter para el conteo de frecuencias
- **WordCloud:** Para la generación de la nube de palabras
- **Matplotlib:** Para la visualización y guardado de la imagen

### 3. RESULTADOS

Después del procesamiento y limpieza del texto, se obtuvieron los siguientes resultados generales:

#### 3.1. Estadísticas Generales

Métrica	Valor
Total de palabras únicas	12,338
Total de palabras en el texto	36,969
Palabra más frecuente	tan (206 veces)
Promedio de apariciones por palabra	~3 veces

#### 3.2. Top 20 Palabras Más Frecuentes

Pos.	Palabra	Frecuencia
1	tan	206
2	hombre	133
3	don	120
4	dos	115
5	ojos	111
6	después	110
7	macario	109
8	ruy	107
9	toda	104
10	entonces	93
11	señor	92
12	noche	87
13	solo	81
14	vida	81
15	amigo	81
16	bajo	79
17	siempre	79
18	casa	79
19	matías	77
20	jacinto	76

### 3.3. Nube de Palabras

La siguiente imagen muestra la nube de palabras generada a partir del análisis de frecuencias. En esta visualización, el tamaño de cada palabra es proporcional a su frecuencia de aparición en el texto. Las palabras más grandes son las que aparecen con mayor frecuencia.

