ESTUDIO FILOLÓGICO-COMPUTACIONAL DEL *MEMORIAL* DE FRANCISCO NÚÑEZ MULEY: COMPARACIÓN SEMÁNTICA ENTRE LA EDICIÓN CRÍTICA HUMANA Y LA INTERPRETACIÓN AUTOMÁTICA MEDIANTE GPT-4

د. علياء عبدالعزيز عبدالسلام الشريف (*)

Resumen

Este artículo realiza un análisis semántico comparativo entre la edición crítica humana del *Memorial* de Francisco Núñez Muley (1566) y su interpretación automática mediante el modelo GPT-4. A través del estudio de dos fragmentos del manuscrito, se examinan las capacidades y limitaciones del modelo para interpretar textos históricos sin mediación contextual. El análisis revela desviaciones temáticas y discursivas significativas, producto de mecanismos estadísticos de predicción lingüística y no de comprensión semántica real. Se evidencian fenómenos como la interpolación simbólica y el *overfitting* contextual. El estudio concluye que, aunque GPT-4 puede generar textos plausibles, su uso en filología debe ser estrictamente supervisado. Se propone su empleo como herramienta complementaria, no sustitutiva, en tareas exploratorias o didácticas.

Palabras clave

Filología computacional, comparación semántica, interpretación automática, edición crítica de manuscritos.

PHILOLOGICAL-COMPUTATIONAL STUDY OF THE *MEMORIAL* BY FRANCISCO NÚÑEZ MULEY: SEMANTIC COMPARISON BETWEEN THE HUMAN CRITICAL EDITION AND THE AUTOMATIC INTERPRETATION BY GPT-4

(*) أستاذ مساعد بقسم اللغة الإسبانية وآدابها، كلية الآداب – جامعة القاهرة.

Abstract

This article presents a semantic comparative analysis between the human critical edition of *Francisco Núñez Muley's Memorial* (1566) and its automatic interpretation generated by the GPT-4 language model. Through the study of two manuscript fragments, the paper evaluates the model's ability to interpret historical texts without external contextual guidance. The analysis reveals significant thematic and discursive deviations, caused by statistical language prediction mechanisms rather than real semantic understanding. Phenomena such as symbolic interpolation and contextual overfitting are observed. The study concludes that although GPT-4 can produce linguistically plausible texts, its use in philological research requires strict supervision. It is proposed as a complementary, not substitutive, tool for exploratory or educational purposes.

Keywords

Computational philology, semantic comparison, automatic interpretation, critical edition of manuscripts.

دراسة فيلولوجية-حاسوبية لمخطوط "الذكرى" لفرانسيسكو نونييث مولاي: مقارنة دلالية بين التحقيق النقدي البشري والتأويل الآلي باستخدام GPT-4

المستخلص

يقدّم هذا المقال تحليلًا دلاليًا مقارنًا بين النسخة النقدية البشرية من مخطوط *"الذكرى"* لفرانسيسكو نونييث مولاي (١٥٦٦) والتأويل الآلي له باستخدام نموذج اللغة GPT-4. ومن خلال دراسة مقطعين منه، يقيّم البحث قدرة النموذج على تفسير النصوص التاريخية دون سياق خارجي داعم. ويُظهر التحليل وجود انحرافات دلالية وخطابية ملحوظة، ناتجة عن آليات تنبؤ لغوي إحصائية لا عن فهم دلالي حقيقي. وقد تم رصد ظواهر مثل الإدخال الرمزي المبالغ فيه والتكيّف المفرط مع الأنماط السياقية. ويخلص البحث إلى أنGPT–4، رغم قدرته على توليد نصوص محتملة لغويًا، يتطلب إشرافًا نقديًا صارمًا عند استخدامه في الدراسات الفيلولوجية. ويُقترح استخدامه كأداة مكمّلة، لا بديلة، في السياقات الاستكشافية أو التعليمية.

الكلمات المفتاحية:

الفيلولوجيا الحاسوبية، مقارنة دلالية، تأويل آلي، النسخة النقدية للمخطوطات.

Introducción

La edición crítica de manuscritos constituye una de las tareas centrales de la filología, orientada a recuperar el texto original con la mayor fidelidad posible mediante la comparación de variantes, la interpretación contextual y el juicio experto (Fantini, 1950). Con el avance de la inteligencia artificial (IA), han surgido modelos de lenguaje capaces de generar textos complejos a partir de entradas lingüísticas mínimas. Esto ha abierto nuevas líneas de investigación en torno a su posible aplicación en la crítica textual, especialmente en el tratamiento de documentos históricos.

Este estudio se sitúa en el cruce entre la filología tradicional y la IA generativa, con el objetivo de evaluar las capacidades interpretativas del modelo GPT-4 (versión de pago, actualizada en 2024) en el análisis de textos patrimoniales. Para ello, se ha escogido el *Memorial* de Francisco Núñez Muley, documento morisco del siglo XVI redactado como respuesta al Edicto de 1567 que prohibía el uso de la lengua árabe, el traje tradicional y las festividades islámicas. Esta obra representa una defensa cultural estratégica dentro del contexto de asimilación forzosa en la España cristiana (Martín Ruiz, 1995; Real Academia de la Historia, 2022).

El corpus se ha limitado a dos fragmentos representativos del manuscrito conservado en la Biblioteca Nacional de España (BNE, ms. 6176, fº 325v), cuya imagen paleográfica ha sido reproducida en Zalalo (2022), y cuya transcripción crítica ha sido tomada de la edición de Juan Martín Ruiz (1995, en *Política y moral en el Siglo de Oro*, pp. 391–402). Esta selección responde a un enfoque exploratorio: se priorizó el análisis profundo de casos específicos que permitieran observar de forma cualitativa los mecanismos interpretativos del modelo, en lugar de una cobertura exhaustiva del texto. Asimismo, se consideraron las dificultades de legibilidad paleográfica, el alcance deliberadamente acotado del estudio y el propósito metodológico de seleccionar fragmentos temáticamente afines que permitieran una observación comparativa del desempeño interpretativo del modelo.

La elección de GPT-4 se justifica por su rendimiento superior en tareas de comprensión semántica, coherencia discursiva e inferencia contextual, tal como evidencian evaluaciones como MMLU (Massive Multitask Language Understanding) y HELM (Holistic Evaluation of Language Models) (OpenAI, 2023; Liang et al., 2022). Se opta deliberadamente por trabajar únicamente con la versión de pago de GPT-4 (actualizada en 2024), dado que en el momento de la investigación constituye el modelo de referencia más avanzado y estable. La comparación con otros modelos se reserva para futuros trabajos, pues el objetivo aquí es observar en profundidad los

mecanismos interpretativos de un solo sistema bajo condiciones controladas. Su uso permite explorar tanto su potencial como sus límites en contextos donde el conocimiento histórico, la sensibilidad lingüística y la interpretación cultural son fundamentales.

El análisis se articula a través de una comparación entre la interpretación automática generada por el modelo (sin glosa ni contextualización externa) y la versión crítica humana. El procedimiento se documenta a partir de un prompt único y explícito, diseñado para garantizar consistencia en todas las ejecuciones: "*Ofrece una reformulación comprensible del texto en español contemporáneo, sin alterar su estructura argumentativa ni su extensión aproximada*". Con ello se busca mantener la replicabilidad metodológica, evitando sesgos derivados de variaciones en la instrucción. Las divergencias se clasifican según criterios léxicos, pragmáticos y retóricos, siguiendo una metodología filológico-computacional. Este enfoque no busca validar técnicamente el modelo, sino reflexionar sobre su aplicabilidad real en contextos de complejidad temporal y cultural.

Aunque la edición crítica asistida por IA es un campo emergente, existen estudios preliminares que exploran la interacción entre algoritmos generativos y textos históricos (Jockers & Underwood, 2016; Pierazzo, 2022). Este trabajo se inserta en ese marco, proponiendo una evaluación crítica del modelo como herramienta complementaria en la edición de textos patrimoniales. En consecuencia, se propone una estructura en cuatro partes: marco teórico, procedimiento metodológico de comparación, análisis de resultados y reflexión final sobre los desafíos y posibilidades que abre la IA en la filología del siglo XXI. Si bien el presente estudio adopta un enfoque cualitativo centrado en el contraste detallado de casos, se reconoce que en investigaciones posteriores puede complementarse con métricas cuantitativas (por ejemplo, índices de exactitud semántica o análisis estadísticos de desviaciones). En esta fase exploratoria se prioriza la observación de casos representativos como vía para identificar patrones interpretativos.

2. Crítica Textual y Edición Filológica de Manuscritos

La crítica textual es una disciplina central en los estudios filológicos, cuyo objetivo es reconstruir, a partir de los testimonios disponibles, la forma más fiel posible de un texto original. Esta tarea exige tanto un análisis riguroso de las variantes textuales como una comprensión profunda de los factores lingüísticos, históricos y culturales que condicionaron su transmisión (Fantini, 1950). En el caso de los manuscritos de los siglos XV y XVI, cuya conservación y circulación dependían aún de la copia manual, la edición

crítica requiere una atención doble: por un lado, al análisis interno del texto —rasgos lingüísticos, estilísticos y retóricos—, y por otro, a una lectura contextual que considere la posible intervención del copista, así como los efectos de la censura ideológica o institucional sobre la forma final del documento.

La edición filológica, por tanto, no solo restituye palabras, sino que rescata significados. De ahí viene la relevancia del editor crítico, cuya labor va más allá de la transcripción para adentrarse en una interpretación históricamente informada. En este estudio se ha trabajado sobre la edición crítica de José María Martín Ruiz (1995), quien reproduce y analiza el *Memorial* de Francisco Núñez Muley situándolo en el marco de las políticas represivas hacia los moriscos y destacando su dimensión discursiva como alegato en defensa de prácticas culturales amenazadas por el Edicto de 1566.

3. Modelos de Lenguaje e Inteligencia Artificial en el Análisis Textual

Con el desarrollo de modelos de lenguaje como GPT-4, se han abierto nuevas posibilidades para la interpretación textual en campos como la literatura, la lingüística y la filología. Estos modelos, entrenados sobre grandes volúmenes de datos lingüísticos, generan respuestas coherentes mediante la predicción probabilística de secuencias de palabras, basándose en patrones frecuentes del lenguaje natural. Sin embargo, esta capacidad no equivale a una comprensión semántica en sentido estricto, sino a una simulación estadística que reproduce estructuras verosímiles sin captar plenamente el significado subyacente (Xataka, 2023; Cloudflare, 2024).

Aunque su rendimiento ha sido notable en tareas contemporáneas de procesamiento del lenguaje, su aplicación a textos históricos plantea desafíos específicos: ¿pueden estos modelos reconocer arcaísmos, identificar estructuras retóricas del pasado o interpretar referencias culturales propias del siglo XVI? Tales interrogantes han comenzado a abordarse en el ámbito de la filología digital, donde las herramientas de IA se emplean para explorar textos patrimoniales desde nuevas perspectivas (Centro de Estudios Mudéjares & Universidad de Alicante, 2021).

En este estudio se ha experimentado con GPT-4 en condiciones no supervisadas, proporcionándole fragmentos históricos sin intervención ni glosa humana. Esta estrategia permite observar el funcionamiento autónomo del modelo ante documentos de alta complejidad contextual, con el objetivo de identificar tanto sus capacidades como sus limitaciones en la interpretación semántica de textos transmitidos por tradición manuscrita.

4. El *Memorial* de Francisco Núñez Muley como Texto Patrimonial y Argumentativo

El *Memorial* redactado por Francisco Núñez Muley en 1566 es uno de los testimonios más valiosos del pensamiento morisco antes de la rebelión de las Alpujarras. El texto responde al Edicto real que prohibía expresiones culturales islámicas entre los neoconversos, incluyendo el uso de la lengua árabe, las vestimentas tradicionales y las celebraciones festivas. Su autor, descendiente de una familia noble de origen marroquí, no solo denuncia el carácter opresivo de estas medidas, sino que construye un alegato racional, respetuoso y profundamente estratégico en defensa de la diversidad cultural (Martín Ruiz, 1995; Real Academia de la Historia, 2022).

Este texto, breve en extensión pero intenso en significado, combina recursos retóricos, apelaciones a la legalidad, alusiones religiosas y argumentos de utilidad social. Su lectura exige una sensibilidad filológica que comprenda tanto el contenido como la forma del discurso. Precisamente por ello, resulta un material idóneo para evaluar las posibilidades —y limitaciones— de una interpretación automática mediante IA, que debe enfrentarse no solo a un léxico arcaico, sino a una arquitectura discursiva compleja y cargada de implicaciones históricas.

5. Metodología de Contraste entre Edición Crítica Humana e Interpretación Automática

Este estudio adopta un enfoque cualitativo y comparativo, centrado en el análisis semántico de dos fragmentos seleccionados del *Memorial* de Francisco Núñez Muley. Su propósito es evaluar la capacidad interpretativa del modelo de lenguaje GPT-4 ante un texto histórico del siglo XVI, contrastando su salida automática con la interpretación ofrecida por una edición crítica especializada. La metodología combina herramientas de la filología tradicional —como la segmentación temática y el análisis léxico-contextual— con criterios adaptados al estudio computacional del discurso.

5.1. Selección y preparación del corpus

El texto base utilizado proviene de la edición crítica de José María Martín Ruiz (1995), publicada en *Baetica: Estudios de historia moderna y contemporánea* (núm. 17, pp. 391–402), que ofrece una transcripción cuidada del *Memorial* a partir del manuscrito conservado en la Biblioteca Nacional de España (BNE, ms. 6176, fº 325v). Esta versión fue seleccionada por su reconocida fidelidad filológica y por ser una de las más citadas en estudios recientes sobre el pensamiento morisco.

A partir de esta edición, se eligieron dos fragmentos representativos siguiendo criterios de densidad argumentativa, riqueza léxica y estructura retórica. Se priorizaron pasajes con carga ideológica relevante que permitieran evaluar la capacidad interpretativa del modelo, descartando aquellos de carácter puramente narrativo o anecdótico.

5.2. Generación del texto automático con GPT-4

Los fragmentos fueron introducidos en GPT-4 (versión actualizada en 2024) sin ningún tipo de glosa, nota editorial ni contextualización adicional. La instrucción proporcionada fue interpretar el contenido del texto y reformularlo de manera comprensible en español contemporáneo, sin alterar su estructura argumentativa ni su extensión aproximada. Esta estrategia metodológica permite observar el comportamiento espontáneo del modelo frente a un texto con alta carga histórica y cultural, sin intervención humana ni enriquecimiento previo.

5.3. Criterios de comparación

Se llevó a cabo una lectura comparada entre la versión crítica y la respuesta automática, tomando como unidad de análisis el enunciado o segmento argumentativo. Sobre esta base, se construyó una tabla con tres columnas: (1) transcripción según edición crítica, (2) paráfrasis generada por GPT-4 y (3) observaciones filológicas.

Las divergencias se clasificaron en tres categorías:

- Léxicas: sustituciones, simplificaciones o reinterpretaciones de vocabulario arcaico o culturalmente específico.
- Pragmáticas: alteraciones de la intención comunicativa, omisión de matices o pérdida del tono retórico original.
- Retóricas: modificaciones en la organización del argumento, pérdida de conectores lógicos o supresión de progresiones persuasivas.

Cada fragmento fue evaluado en función de su fidelidad semántica, la preservación de la intención del autor y la coherencia discursiva. Asimismo, se identificaron los recursos automáticos empleados por el modelo para suplir vacíos interpretativos, tales como generalizaciones, anacronismos o reformulaciones neutras.

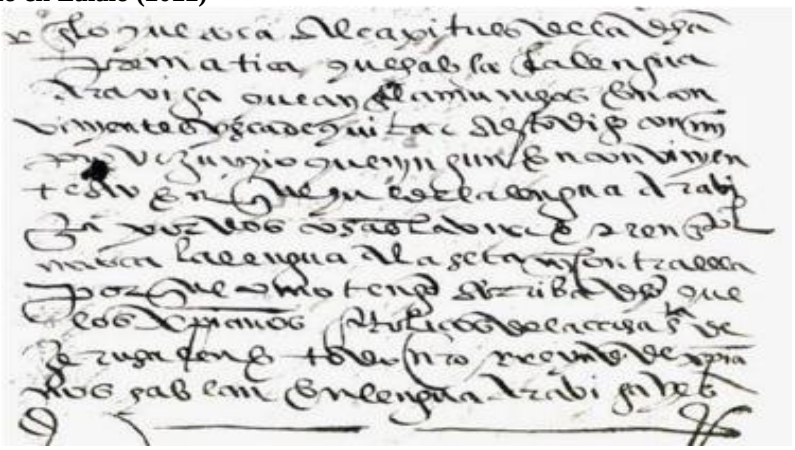
6. Análisis Semántico Comparativo entre la Edición Crítica Humana y la Interpretación Automática del *Memorial* de Francisco Núñez Muley (siglo XVI)

6.1. Fragmento Seleccionado del Manuscrito

La imagen paleográfica del fragmento seleccionado, objeto de análisis en este estudio, ha sido reproducida en el artículo de Zalalo (2022), sin que se haya recurrido a la transcripción ni al comentario editorial que allí se

ofrecen. Tal como se indicó previamente, dicho fragmento corresponde al *Memorial* de Francisco Núñez Muley, conservado en la Biblioteca Nacional de España bajo la signatura BNE, ms. 6176, folio 325v:

Imagen1. **Fragmento 1 del *Memorial* de Fco. Núñez Muley (BNE, ms. 6176, fº 325v), reproducido en Zalalo (2022)**



6.2. Transcripción Crítica según Martín Ruiz (1995)

El fragmento correspondiente, citado en la edición crítica de Martín Ruiz (1995, p. 399), alude a un intento frustrado de imponer la prohibición del traje morisco durante el mandato del arzobispo Gaspar de Ávalos. La medida, iniciada en zonas rurales (las alcarías), fue detenida por la oposición de diversas autoridades locales, y el autor la presenta como un precedente de resistencia institucional frente a las políticas represivas recientes:

Dende a ciertos años, don Gaspar de Ávalos, siendo arzobispo de Granada, de hecho quiso quitarnos el hábito, comenzando por los de las alcarías, y trayendo aquí algunos de Güéjar sobre ello. El presidente que estaba en el lugar que está agora vuestra señoría, y los oidores desta audiencia, y el marqués de Mondéjar y el Corregidor se lo contradijeron, y paró por las mesmas razones.

6.3. Interpretación Automática del Fragmento Manuscrito mediante GPT-4

Se solicitó al modelo una interpretación del contenido del fragmento seleccionado, mediante un prompt directo que indicaba: *"Ofrece una reformulación comprensible del texto en español contemporáneo, sin alterar su estructura argumentativa ni su extensión aproximada"*.

La respuesta generada por el modelo fue la siguiente:

Tabla 1. **Interpretación Automática – Fragmento 1**

A continuación, realizo una transcripción tentativa del texto original (en castellano del siglo XVI) seguida de una reformulación en español contemporáneo. El manuscrito es difícil de leer en algunas partes, pero he reconstruido lo que se puede distinguir con mayor claridad: **Transcripción tentativa (castellano antiguo):** Lo que toca al capítulo desta suya forma... ... que habla de la lengua arábiga, es cosa que de antiguo estaba muy en uso y se usaba comúnmente entre los moriscos... ... y que no viene a este tiempo en que se habla lengua de Castilla. Y en esto cesa... (*Varias partes están deterioradas o poco legibles, por lo que la transcripción es incompleta y tentativa.*) **Reformulación en español contemporáneo:** En lo que respecta al capítulo de esta forma suya, que trata sobre la lengua árabe, se dice que antiguamente era de uso común y frecuente entre los moriscos, pero que ya no corresponde a esta época, en la que se habla castellano. Por lo tanto, dicho uso ha cesado.

6.4. Análisis Semántico Comparativo entre la Edición Crítica Humana y la Interpretación Automática

Tabla 2. **Comparación Semántica – Fragmento 1**

Edición crítica humana (Martín Ruiz, 1995, p. 399)	**Interpretación automática generada por GPT-4**	**Observaciones filológicas según criterios de comparación**
"Dende a ciertos años, don Gaspar de Ávalos, siendo arzobispo de Granada, de hecho quiso quitarnos el hábito, comenzando por los de las alcarías, y trayendo aquí algunos de Güéjar sobre ello. El presidente que	"En lo que respecta al capítulo de esta forma suya, que trata sobre la lengua árabe, se dice que antiguamente era de uso común y frecuente entre los moriscos, pero que ya no	**Léxicas:** Se observa una sustitución completa de los términos específicos del fragmento original —como "hábito", "Güéjar", "oidores" o "marqués de Mondéjar"— por expresiones genéricas de naturaleza temática ("lengua árabe", "esta época", "uso común") que, si bien son coherentes con otros pasajes del *Memorial*, no tienen

estaba en el lugar que está agora vuestra señoría, y los oidores desta audiencia, y el marqués de Mondéjar y el Corregidor se lo contradijeron, y paró por las mesmas razones.”	corresponde a esta época, en la que se habla castellano. Por lo tanto, dicho uso ha cesado.”	correlato semántico directo con el contenido particular de este enunciado. **Pragmáticas:** La intención del texto original era denunciar un intento concreto de imponer un cambio cultural, protagonizado por una figura eclesiástica específica (Gaspar de Ávalos). Esta dimensión crítica desaparece en la versión generada por el modelo, que presenta el contenido de forma neutral, sin reflejar el conflicto. Se pierde el tono de resistencia que caracteriza al original, y se omiten tanto los personajes históricos como las instituciones involucradas. El resultado es un discurso plano, que ya no transmite la tensión política y social del pasaje original. **Retóricas:** En la versión generada por el modelo, desaparece la estructura argumentativa del texto original, que presentaba claramente una acción (el intento de prohibir el hábito morisco) y una respuesta institucional (la oposición de varias autoridades). Esta relación causa-efecto se pierde, ya que la interpretación automática ofrece una formulación general sin conectores lógicos ni contraste

		argumentativo. Además, se modifica el orden del discurso y se elimina la progresión que en el original guiaba al lector a través de un razonamiento persuasivo. **Recursos automáticos observados:** El modelo genera una paráfrasis basándose en temas comunes del *Memorial*, como la lengua árabe, aunque este tema no aparece en el fragmento analizado. Esta interpretación errónea se produce por un proceso de inferencia automática que no está respaldado por el contenido real del texto. En lugar de interpretar lo que efectivamente dice el fragmento, el modelo utiliza mecanismos de completamiento semántico basados en asociaciones estadísticas. Como resultado, reemplaza el conflicto histórico concreto por una formulación genérica, suaviza el tono argumentativo y omite detalles clave como nombres, cargos o acciones específicas. Esto demuestra que la salida generada no se apoya en una lectura filológica, sino en una reconstrucción probabilística del sentido.

El caso recogido en la tabla anterior constituye un ejemplo paradigmático de cómo un modelo de lenguaje automático como GPT-4 puede desviarse del sentido original al interpretar un texto histórico del siglo XVI. A partir de

esta discrepancia, se propone un análisis lingüístico articulado en tres niveles complementarios: (1) cognitivo-computacional, que examina los mecanismos estadísticos que guían la generación de texto; (2) discursivo, enfocado en la organización argumentativa del contenido producido; y (3) pragmático-semántico, centrado en la fidelidad interpretativa con respecto al sentido y a la intención original del fragmento.

6.4.1. Dependencia de patrones estadísticos, no de comprensión real

GPT-4 genera secuencias lingüísticas en función de la probabilidad de aparición de ciertos términos y construcciones aprendidas durante su entrenamiento, sin operar con una comprensión semántica en sentido humano. En este caso, una posible ambigüedad en la lectura paleográfica del inicio —“Lo que toca al capítulo desta suya forma…”— activa en el modelo asociaciones frecuentes relacionadas con la lengua árabe, tema presente en otros pasajes del *Memorial*. El resultado es una paráfrasis coherente discursivamente, pero desacertada en relación con el contenido específico del fragmento.

6.4.2. Reconstrucción verosímil a partir de un estímulo incompleto

La interpretación automática —"En lo que respecta al capítulo de esta forma suya, que trata sobre la lengua árabe..."— no corresponde al contenido del manuscrito. Se trata de una inferencia generada a partir de patrones discursivos aprendidos, no de una lectura real del fragmento. En contextos con input parcial, ilegible o ambiguo, el modelo tiende a rellenar vacíos con formulaciones plausibles, pero históricamente inexactas. Esta "ilusión de coherencia" puede ser convincente si no se contrasta con una edición crítica fiable.

6.4.3. Activación de tópicos frecuentes y confusión temática

El modelo ha sido probablemente entrenado con textos alusivos al mundo morisco, lo que explica su tendencia a reconstruir contenidos esperables al detectar términos como “forma”, “lengua” o “moriscos”. Esta activación de esquemas dominantes, propia de la semántica cognitiva, genera una interpretación basada en regularidades contextuales —no en el fragmento puntual—, lo que puede derivar en *overfitting* semántico-contextual, es decir, en un sobreajuste interpretativo que impone significados no contenidos en el texto base[1].

6.4.4. Implicaciones metodológicas

Este caso pone de manifiesto tres fenómenos fundamentales: (1) desplazamiento temático, al interpretar el contenido sobre la vestimenta

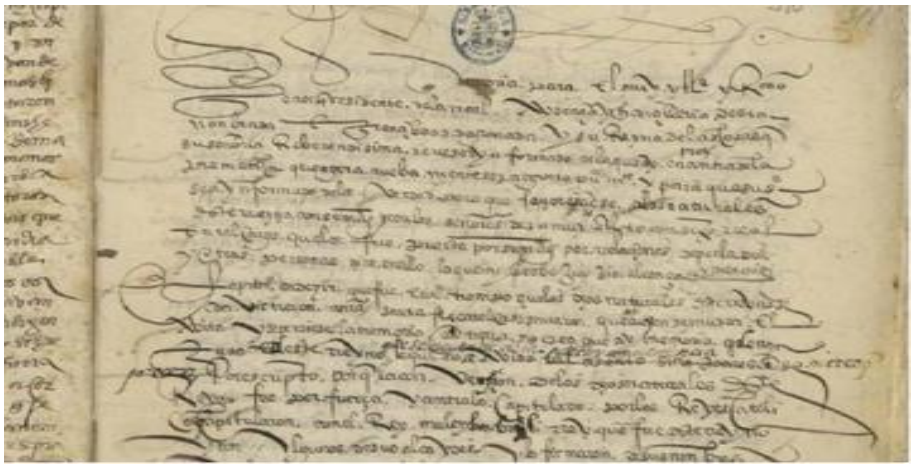
morisca como si tratara de la lengua árabe; (2) generalización improcedente, al generar un texto coherente con la obra global, pero desvinculado del pasaje analizado; y (3) reconstrucción heurística, basada en simulación estadística más que en comprensión textual. Todo ello subraya la necesidad de un acompañamiento experto cuando se emplean modelos generativos en estudios filológicos. Aunque GPT-4 puede ofrecer resultados lingüísticamente elaborados, su fiabilidad interpretativa depende en gran medida del control humano, especialmente en el tratamiento de textos con alta densidad cultural o retórica.

7. Análisis de un Segundo Caso: Interpretación Automática de un Nuevo Fragmento

Con el fin de ampliar la evaluación del desempeño interpretativo del modelo GPT-4 frente a textos patrimoniales, se analizó un segundo fragmento del *Memorial* de Francisco Núñez Muley. Este pasaje proviene del mismo folio del manuscrito conservado en la Biblioteca Nacional de España (BNE, ms. 6176, fº 325v), y ha sido igualmente reproducido en el artículo de Zalalo (2022).

A diferencia del caso anterior, el fragmento seleccionado se sitúa inmediatamente después del alegato inicial en defensa de las costumbres moriscas. Esta cercanía estructural y temática permite examinar si el modelo mantiene coherencia en la continuidad discursiva o si reproduce los mismos errores interpretativos ya detectados, lo que contribuirá a delimitar mejor los alcances y límites de su rendimiento semántico.

Imagen2. **Fragmento 2 del *Memorial* de Fco. Núñez Muley (BNE, ms. 6176, fº 325v), reproducido en Zalalo (2022)**



Aquí, la edición crítica de Martín Ruiz (1995, p. 399) recoge con claridad un fragmento en el que Núñez Muley rememora los intentos, durante los reinados de la reina Juana y del emperador Carlos V, de prohibir el uso del traje morisco. No obstante, dichas medidas fueron finalmente suspendidas ante los inconvenientes que generaron. El pasaje mantiene el tono característico del *Memorial*: una defensa firme pero mesurada de las costumbres moriscas, entendidas como expresiones culturales legítimas, ajenas a lo religioso:

Después, reinando la reina doña Juana, su hija, pareciendo convenir (no sé por cierto á quién), se mandó que dejásemos el traje morisco; y por algunos inconvinientes que se representaron, se suspendió, y lo mesmo viniendo á reinar el cristianísimo emperador don Carlos.

Al introducir este mismo fragmento manuscrito en GPT-4, utilizando la misma instrucción que en el primer caso —"*ofrece una reformulación comprensible del texto en español contemporáneo, sin alterar su estructura argumentativa ni su extensión aproximada*"—, el modelo generó una interpretación que, si bien resulta lingüísticamente coherente, vuelve a apartarse del contenido original, tal como ocurrió con el primer fragmento:

Tabla 3. **Interpretación Automática – Fragmento 2**

Lectura aproximada del fragmento (parcial, debido a la caligrafía complicada y las abreviaturas): A partir de las primeras líneas y del contexto general del *Memorial*, se puede identificar que este fragmento menciona: La defensa de ciertas prácticas culturales (como la lengua, los vestidos, la música, etc.) que los moriscos consideraban parte de su identidad, no como elementos religiosos sino como costumbres heredadas. El autor insiste en que estas prácticas no son contrarias a la religión cristiana ni deben interpretarse como un acto de rebeldía. También señala que los viejos cristianos tienen prácticas similares en ciertas regiones, con lo cual intenta desmontar la idea de que lo morisco es necesariamente enemigo de lo cristiano. **Reformulación comprensible en español contemporáneo:** …a esta ciudad, para que el marqués no conociera el valor ni la excelencia de nuestra lengua árabe, ni la riqueza de los libros que están escritos en ella, ni el saber de los sabios del pasado. Por esta razón se vieron forzados a abandonarla, y muchos jóvenes —tanto hombres como mujeres— que la dominaban y la escribían bien, dejaron de usarla debido a la prohibición. Así, como si fuera una lengua muerta, se ha ido olvidando. Ahora, después de casi cincuenta años, son muy pocos los que aún la conocen, y ni siquiera saben escribirla. De este modo, ha desaparecido porque no se permitió que se siguiera aprendiendo.

La paráfrasis automática se centra en la lengua árabe y en su supuesta desaparición a raíz de una prohibición oficial. Aunque dichos temas están presentes en otras secciones del *Memorial*, no constituyen el contenido de este fragmento concreto, lo que confirma un nuevo caso de desplazamiento temático por parte del modelo.

Tabla 4. **Comparación Semántica – Fragmento 2**

Edición crítica humana (Martín Ruiz, 1995, p. 399)	**Interpretación automática generada por GPT-4**	**Observaciones filológicas según criterios de comparación**
“Después, reinando la reina doña Juana, su hija, pareciendo convenir (no sé por cierto á	“…a esta ciudad, para que el marqués no conociera el valor ni la excelencia de nuestra lengua árabe, ni la riqueza de los libros	**Léxicas:** El modelo sustituye léxico relacionado con decretos sobre vestimenta (“traje morisco”) por un campo semántico ajeno al fragmento original

quién), se mandó que dejásemos el traje morisco; y por algunos inconvinientes que se representaron, se suspendió, y lo mesmo viniendo á reinar el cristianísimo emperador don Carlos.”	que están escritos en ella, ni el saber de los sabios del pasado. Por esta razón se vieron forzados a abandonarla, y muchos jóvenes —tanto hombres como mujeres— que la dominaban y la escribían bien, dejaron de usarla debido a la prohibición. Así, como si fuera una lengua muerta, se ha ido olvidando. Ahora, después de casi cincuenta años, son muy pocos los que aún la conocen, y ni siquiera saben escribirla. De este modo, ha desaparecido porque no se permitió que se siguiera aprendiendo.”	(lengua árabe, libros, olvido generacional). **Pragmáticas:** La intención crítica del autor sobre medidas suspendidas es reemplazada por un relato nostálgico sin anclaje histórico real en el fragmento. Se introduce un tono emocional no presente en el original. **Retóricas:** El modelo omite la referencia explícita a las autoridades reales (Juana y Carlos V) y la lógica causal del texto (mandato → suspensión), construyendo en su lugar una narrativa abstracta sobre pérdida cultural. **Recursos automáticos observados:** El modelo recurre a temas recurrentes del *Memorial* —como la pérdida de la lengua árabe o la decadencia cultural— que no están presentes en el fragmento concreto. En lugar de interpretar el contenido real, extrae ideas de otros pasajes similares del texto y las traslada de forma generalizada, utilizando imágenes simbólicas conocidas para compensar la falta de comprensión del contexto específico.

El desajuste reflejado en la tabla anterior confirma nuevamente la tendencia del modelo a proyectar contenidos frecuentes del texto global sobre estímulos parciales o ambiguos. En este caso, la interpretación automática combina elementos de distintas secciones del *Memorial* y produce una versión homogénea pero descontextualizada del discurso morisco. Además, introduce recursos retóricos y afirmaciones —como la revalorización de la lengua árabe, su olvido progresivo o la pérdida intergeneracional del conocimiento— que no están presentes en este fragmento específico.

Desde una perspectiva metodológica, este segundo caso refuerza las observaciones previas: GPT-4 genera textos con coherencia superficial, pero su interpretación se basa en inferencias estadísticas, no en una lectura filológica del pasaje concreto. Esto puede derivar en reconstrucciones formalmente plausibles, pero inexactas desde el punto de vista histórico y textual. Por tanto, el uso de modelos generativos en el análisis de documentos patrimoniales requiere un acompañamiento experto que garantice la fidelidad semántica, la contextualización crítica y el respeto a la lógica discursiva original.

8. Discusión de los Resultados

El análisis de los dos fragmentos revela un patrón convergente de desviación interpretativa por parte del modelo GPT-4, sustentado en la activación de temas recurrentes del *Memorial* de Francisco Núñez Muley. En ambos casos, el modelo no responde al contenido específico del fragmento paleográfico, sino que proyecta sobre él sentidos generales del texto, activados por términos ambiguos o por fragmentos aislados.

No obstante, la naturaleza del error varía en cada caso. En el primer fragmento, se observa un desplazamiento temático explícito: el modelo transforma un pasaje sobre la prohibición del hábito morisco en un discurso genérico sobre la lengua árabe, sin justificación contextual ni apoyo léxico. Este giro temático elimina la dimensión institucional y política documentada en el original, y reconstruye un tópico ausente en el texto base.

En el segundo caso, la desviación resulta más sutil, aunque no menos significativa: el modelo identifica superficialmente el tema central del fragmento —una orden real sobre el traje morisco—, pero en lugar de desarrollarlo, proyecta contenidos simbólicos y emocionales tomados de otros pasajes del texto. Así, construye una narrativa nostálgica en torno a la pérdida de la lengua árabe. Se trata, en definitiva, de una interpolación discursiva que amalgama referencias dispersas y añade elementos ajenos al fragmento, desdibujando su coherencia argumentativa.

Ambas desviaciones responden al mismo mecanismo de fondo: la tendencia del modelo a completar contextos mediante estrategias de verosimilitud lingüística, priorizando la coherencia superficial sobre la fidelidad semántica. Sin embargo, difieren en su manifestación: el primer caso implica una sustitución completa del tema, mientras que el segundo conserva parcialmente la referencia temática, pero la diluye con adiciones simbólicas ajenas.

La siguiente tabla sintetiza estas diferencias:

Tabla 5. **Tipología de Desviaciones Interpretativas Cometidas por GPT-4 en Dos Fragmentos del *Memorial* de Fco. Núñez Muley**

Aspecto Analizado	**Fragmento 1: Desplazamiento Temático**	**Fragmento 2: Interpolación Simbólica**
Tema original	Prohibición del hábito morisco por parte del arzobispo Gaspar de Ávalos (iniciativa local en el contexto rural de Granada).	Suspensión de decretos reales (de la reina Juana y el emperador Carlos V) contra el uso del traje morisco.
Interpretación de GPT-4	Discurso general sobre la lengua árabe como tradición morisca.	Reflexión nostálgica sobre la pérdida generacional de la lengua árabe.
Tipo de desviación	Sustitución completa del tema: se omite la referencia al hábito y se introduce otro tópico no presente en el fragmento.	Mezcla de referencias dispersas: se conserva el tema del decreto, pero se contamina con elementos simbólicos ajenos al fragmento.
Mecanismo del error	Activación de un tópico frecuente en el *Memorial* (la lengua árabe), no presente en este pasaje concreto.	Inserción de contenidos simbólicos derivados del tono general del texto (decadencia, olvido cultural).
Error léxico	Sustitución de nombres propios e instituciones por términos genéricos.	Reemplazo del campo semántico de la vestimenta por el de la lengua y el conocimiento cultural.
Efecto	Borrado del conflicto	Dilución de la intención

interpretativo	político-institucional entre autoridades locales y el arzobispo.	argumentativa original del fragmento, centrada en el precedente de resistencia institucional.
Recursos automáticos usados	Generalización temática basada en patrones frecuentes sobre la arabidad morisca en el *Memorial*.	Inclusión de imágenes simbólicas recurrentes (pérdida cultural, juventud que no aprende, olvido de raíces lingüísticas).
Tono discursivo	El original presenta un tono sobrio y argumentativo, con apelaciones a figuras institucionales; la interpretación automática adopta un tono expositivo y generalista, sin anclaje histórico claro.	El original mantiene un tono mesurado y persuasivo, basado en hechos; el modelo introduce un tono simbólico y nostálgico, con carga emocional no presente en el fragmento.
Fidelidad semántica	Baja: el contenido del fragmento se sustituye por completo con otro tópico.	Moderadamente baja (media tendiente a baja): se conserva una alusión temática (decreto real), pero fuertemente distorsionada por elementos simbólicos ajenos (referencias a la lengua y el conocimiento cultura).

De esta comparación se desprende una desviación temática sistemática en las interpretaciones automáticas generadas por GPT-4. En ambos fragmentos originales, centrados en la prohibición del traje morisco, el modelo sustituye el contenido por discursos sobre la lengua árabe. Esta recurrencia no parece accidental, sino que responde a una combinación de factores lingüísticos y computacionales.

Por un lado, la defensa de la lengua árabe constituye uno de los ejes discursivos más reiterados en el *Memorial* de Francisco Núñez Muley. Al estar entrenado sobre grandes corpus textuales en los que este tópico tiene una fuerte presencia, el modelo tiende a activar dicho marco conceptual cuando encuentra términos genéricos como *forma*, *mandar*, *moriscos* o *capítulo*. Esta tendencia se ve acentuada si el estímulo de entrada es

ambiguo, difícil de leer paleográficamente o proviene de una transcripción tentativa, favoreciendo así la proyección de significados estadísticamente verosímiles, aunque inapropiados para el caso concreto.

Por otro lado, la lengua árabe —en tanto símbolo de identidad cultural y herencia histórica— ofrece un campo semántico más rico y emocionalmente resonante que el del traje, lo que facilita su activación durante el proceso de completamiento predictivo. En lugar de interpretar el fragmento a partir de su contenido literal, el modelo genera una paráfrasis "coherente" con el tono general del *Memorial*, pero desvinculada de la situación discursiva específica. Este fenómeno, ya identificado en secciones anteriores, responde a lo que se denomina *overfitting* semántico-contextual: una sobreadaptación del modelo a patrones recurrentes del corpus de entrenamiento, que tiende a imponer interpretaciones esperadas incluso cuando el estímulo concreto requiere un análisis diferenciado[2].

En síntesis, el modelo no cambia el tema al azar, sino que reconstruye el sentido del texto a partir de regularidades aprendidas. Por ello, aunque ambos fragmentos giran en torno a la vestimenta morisca, la interpretación automática deriva hacia un discurso sobre la lengua, debido a su alta saliencia temática, su frecuencia en el corpus de entrenamiento y su fuerte potencial simbólico. Esta desviación convierte a la lengua árabe en una especie de "respuesta por defecto" del modelo ante entradas textuales poco definidas.

9. Conclusión

El presente análisis ha puesto de relieve que, si bien el modelo GPT-4 demuestra una notable capacidad para generar paráfrasis fluidas y estilísticamente plausibles, no siempre asegura una interpretación rigurosa de textos históricos complejos como el *Memorial* de Francisco Núñez Muley. Los casos estudiados evidencian una tendencia del modelo a desviarse del contenido original mediante extrapolaciones temáticas, interpolaciones simbólicas o reconstrucciones semánticas que responden más a patrones frecuentes de entrenamiento que a una lectura filológica del fragmento específico.

En este contexto, GPT-4 puede desempeñar un papel complementario en investigaciones filológicas, siempre que se emplee con criterio y bajo una supervisión crítica. Su valor reside, sobre todo, en la exploración preliminar de sentidos, la formulación de hipótesis interpretativas o su aplicación en entornos didácticos. Sin embargo, no sustituye el análisis humano,

especialmente en tareas que exigen fidelidad documental, comprensión contextual y conocimiento de referentes históricos.

Entre los principales riesgos detectados se encuentran: la descontextualización del contenido, la simplificación argumentativa, la omisión de actores históricos y la proyección de discursos típicos no pertinentes, todos ellos susceptibles de generar sesgos o falsos positivos interpretativos.

Dado que las investigaciones sobre la aplicación de modelos generativos en el tratamiento de textos patrimoniales aún son incipientes, este estudio ofrece una contribución preliminar al debate metodológico sobre su uso. Los resultados subrayan la necesidad de mantener una vigilancia crítica constante al integrar estas herramientas en entornos filológicos, especialmente cuando se trabaja con fuentes históricas propensas a ser malinterpretadas por sistemas entrenados con datos generalistas o contemporáneos.

De cara al futuro, se abren líneas de aplicación prometedoras: el análisis de otros textos del Siglo de Oro o de la literatura morisca, el entrenamiento de modelos con corpus históricos específicos que respeten variantes gráficas y sintácticas del español antiguo, así como usos didácticos en el aula de filología histórica, donde la interacción con la IA puede facilitar la comprensión de estructuras y significados arcaicos, siempre bajo la guía de docentes especializados.

[1] Uno de los riesgos más comunes en los modelos generativos actuales es el *overfitting semántico-contextual*, una tendencia a proyectar interpretaciones basadas en patrones frecuentes del corpus de entrenamiento, incluso cuando estos no se corresponden con el estímulo textual recibido. En consecuencia, el modelo puede generar paráfrasis aparentemente coherentes pero alejadas del contenido real del texto (Na8, 2017).

[2] Consúltese la nota anterior para más detalles.

10. Bibliografía

Algee-Hewitt, M., Allison, S., Gemma, M., Heuser, R., Moretti, F., & Walser, H. (2016). *Canon/Archive: Large-scale dynamics in the literary field* (Pamphlet 11). Stanford Literary Lab. https://litlab.stanford.edu/LiteraryLabPamphlet11.pdf

Centro de Estudios Mudéjares & Universidad de Alicante. (2021). *Sharq al-Andalus. Estudios Mudéjares y Moriscos*, (22), Teruel–Alicante (2017–2018). Instituto de Estudios Turolenses / Área de Estudios Árabes e Islámicos. https://rua.ua.es/server/api/core/bitstreams/c789a774-c435-41a5-abbf-b9c3f2d1e294/content

Cloudflare. (2024, 11 de agosto). *¿Qué es el procesamiento del lenguaje natural (PLN)?* https://www.cloudflare.com/es-la/learning/ai/natural-language-processing-nlp/?utm_source=chatgpt.com

Fantini, J. (1950). *Iniciación teórico-práctica a la crítica textual*. *Helmántica*, 1(1-4), 169–185. https://doi.org/10.36576/summa.2306

Garrido Palma, F. J. (2019). *Génesis del capitalismo andaluz en la primera modernidad europea*. Universidad de Coimbra, Centro de Estudos Interdisciplinares do Século XX (CEIS20). https://estudogeral.uc.pt/bitstream/10316/91051/2/G%C3%A9nesis%20del%20capitalismo%20andaluz%20en%20la%20primera%20modernidad%20europea.pdf

Hendrycks, D., Burns, C., Kadavath, S., et al. (2021). *Measuring Massive Multitask Language Understanding*. arXiv. https://arxiv.org/abs/2009.03300

Infobae. (2025, 28 de junio). *Glosario de tecnología: qué significa reconocimiento óptico de caracteres (OCR)*. https://www.infobae.com/tecno/2025/06/28/glosario-de-tecnologia-que-significa-reconocimiento-optico-de-caracteres-ocr/

Liang, P., Reynolds, L., & Ganguli, D. et al. (2022). *Holistic Evaluation of Language Models (HELM)*. Stanford Center for Research on Foundation Models. https://crfm.stanford.edu/helm/latest/

Martín Ruiz, J. M. (1995). *El Memorial del morisco Francisco Núñez Muley*. Baetica: Estudios de Historia Moderna y Contemporánea, *17*, 391–402. https://dialnet.unirioja.es/servlet/articulo?codigo=95331

Na8. (2017, 12 de diciembre). *Qué es overfitting y underfitting y cómo solucionarlo*. Aprende Machine Learning. https://www.aprendemachinelearning.com/que-es-overfitting-y-underfitting-y-como-solucionarlo/

Núñez Muley, F. (2024). *Memorial en defensa de las costumbres moriscas* (Serie Historia, nº 285). Linkgua Ediciones. 26 páginas. ISBN 978-84-900769-8-9. https://www.amazon.com/Memorial-defensa-costumbres-moriscas-Historia/dp/8490076987

_____ (2023). *Memorial (Granada, 1566)* (M. López López, Ed. y trad.). Open Iberia/América. https://openiberiaamerica.hcommons.org/memorial-de-francisco-nunez-muley-granada-1566/

~~(2016)~~. *Memorial redactado con motivo de la puesta en práctica del Edicto de 1566*. En R. Amrán & A. Cortijo Ocaña (Eds.), *Minorías en la España medieval y moderna (ss. XV–XVII)* (pp. [especificar páginas]). Santa Barbara: Publications of eHumanista, University of California. https://www.ehumanista.ucsb.edu/

OpenAI. (2023). *GPT-4 Technical Report*. https://doi.org/10.48550/arXiv.2303.08774

Pierazzo, E. (2016). *Digital Documentary Editions and the Others*. En M. J. Driscoll & E. Pierazzo (Eds.), *Digital Scholarly Editing: Theories and Practices* (pp. 61–77). Open Book Publishers. https://doi.org/10.11647/OBP.0095.03

Real Academia de la Historia (2022). *Francisco Núñez Muley*. En *Historia Hispánica: biografías*. https://historia-hispanica.rah.es/biografias/32657-francisco-nunez-muley

Stryker, C., & Holdsworth, J. (2024). *¿Qué es el procesamiento del lenguaje natural (NLP)?* IBM. https://www.ibm.com/mx-es/think/topics/natural-language-processing

Xataka. (2023, 7 de junio). *GPT-4: qué es, cómo funciona, cómo usarlo*. https://www.xataka.com/basics/gpt-4-que-como-funciona-como-usarlo-que-puedes-hacer-este-modelo-lenguaje-inteligencia-artificial?utm_source=chatgpt.com

Zalalo, M. A. A. R. (2022). *Una carta de un morisco granadino al rey de España: Defensa de la lengua árabe y de las tradiciones islámicas*. Dād. Revista de Lingüística y Literatura Árabes, 3(5), 205–231. https://www.daadjournal.com/daad/article/download/43/81/146