Digilec 13 (2026), pp. 82-99

Fecha de recepción: 17/04/2026
Fecha de aceptación: 29/06/2026
Fecha de publicación: 31/07/2026
DOI: 10.17979/digilec.2026.13.13509

e-ISSN: 2386-6691

HTR y evaluación de un corpus de educación primaria de textos manuscritos en Euskera

HTR and assessment of a handwritten Basque text corpus from primary education

Leire FERNÁNDEZ-ATORRASAGASTI *
Institución: Universidad del País Vasco / Euskal Herriko Unibertsitatea
Orcid: https://orcid.org/0000-0002-3838-7795

Irune IBARRA LIZUNDIA **
Institución: Universidad del País Vasco / Euskal Herriko Unibertsitatea
Orcid: https://orcid.org/0000-0002-6655-4327

Mikel IRUSKIETA ***
Institución: Universidad del País Vasco / Euskal Herriko Unibertsitatea
Orcid: https://orcid.org/0000-0002-6121-3902

Resumen
Las pruebas en Educación Primaria (EP) para la realización de la Evaluación Diagnóstica (ED) del Sistema Educativo Vasco en materia de lengua se recopilan en textos manuscritos. Si dichos textos pudieran digitalizarse, sería posible analizarlos tanto manualmente como con técnicas de Procesamiento del Lenguaje Natural (PLN) para su análisis. El objetivo de este estudio es crear un modelo de Reconocimiento de Texto Manuscrito (HTR) para la digitalización de textos manuscritos en euskera producidos por alumnado de 4.º de Educación Primaria, así como crear un corpus digitalizado de dichos textos. Aunque existen modelos de HTR en euskera de estudiantes con mayor edad, se desarrolló uno adaptado a las particularidades escriturales del alumnado. Se compilaron 870 textos manuscritos facilitados por el Instituto Vasco de Evaluación e Investigación Educativa (ISEI-IVEI) y se utilizó la plataforma Transkribus usando técnicas avanzadas de Inteligencia Artificial (IA) para la creación de modelos. La evaluación se realizó de forma manual y se llevó a cabo un análisis cualitativo de los errores del modelo mediante

* Email institucional: leire.fernandeza@ehu.eus
** Email institucional: irune.ibarra@ehu.eus
*** Email institucional: mikel.iruskieta@ehu.eus

un sistema de codificación. Los resultados muestran una tasa de error entre el 5% y el 7%, lo que representa un avance en la digitalización de texto de escritura infantil. Finalmente, se prevé poner a disposición de la comunidad científica y educativa tanto el modelo de HTR como el corpus, facilitando futuras investigaciones en este ámbito.

Palabras clave: reconocimiento de texto manuscrito; evaluación diagnóstica; escritura manuscrita; educación primaria

Abstract

The Primary Education (EP) test for the Diagnostic Evaluation (ED) of the Basque Educational System within the specific domain and field of language are collected as handwritten textual productions. If these texts were digitized, they could be analyzed both manually by experts and automatically with Natural Language Processing (NLP) techniques. The principal objective of this study is to develop a Handwritten Text Recognition (HTR) model for the digitization of handwritten texts in Basque produced by 4th grade Primary Education students, as well as to create a digitized corpus of these texts. Although Handwritten Text Recognition (HTR) models in Basque exist for older students, a model was developed specifically adapted to the young students' writing characteristics. A total of 870 handwritten texts provided by the Basque Institute of Educational Evaluation and Research (ISEI-IVEI) were collected, and the Transkribus platform was used, applying advanced Artificial Intelligence (AI) techniques to train the models. The evaluation process was carried out manually, and a qualitative analysis of the model's errors was conducted using a coding system. The results show an error rate between 5 % and 7 %, representing an advance in the digitization of children's handwritten texts. Finally, both the Handwritten Text Recognition (HTR) model and the corpus are expected to be made available to the scientific and educational community, facilitating future research in this area.

Keywords: handwritten text recognition; diagnostic assessment; handwriting; primary education

1. INTRODUCCIÓN

La escritura es una de las formas de expresión más utilizadas en la escuela y constituye un código de comunicación fundamental en los procesos educativos. Según González (2003), escribir es un proceso complejo que demanda un esfuerzo considerable, ya que requiere coordinar aspectos gramaticales, léxicos, ortográficos y estructurales, entre otros. En Educación Primaria (EP), la escritura manual representa un desafío particular, puesto que implica formar letras, números y otros caracteres mientras se coordinan simultáneamente procesos cognitivos y habilidades motoras, tanto finas como gruesas (Dinehart, 2015). Los autores Adak et al. (2017), en ese mismo año, destacaron que se había escrito mucho sobre las propiedades que afectan a la legibilidad en la escritura a mano, pero no tanto sobre la legibilidad en el reconocimiento de escritura a mano. Fogel et al. (2022) mencionaban que la formación de letras contribuye de manera significativa a la legibilidad de las muestras escritas, más que cualquier otra propiedad. Butler (2003) afirmaba que históricamente la legibilidad era el término que describía el efecto de los aspectos espaciales de un texto en su comprensibilidad y se veía afectada por diversas propiedades gráficas, como la forma o formación de las letras y de las palabras; el espaciado entre letras, palabras y líneas; la longitud de línea y tamaño de letra; y el contraste de las palabras respecto a la página. Siendo la forma una de las propiedades de la legibilidad, es una buena práctica la enseñanza de las letras agrupadas en familias (Ibarra et al., 2020; Government of South Australia, 2007).

Las pruebas en EP para la realización de la Evaluación Diagnóstica (ED) del Sistema Educativo Vasco en materia de lengua se recopilan en textos manuscritos. En este contexto, la competencia lingüística se concibe como la capacidad para utilizar textos escritos de manera eficaz y adecuada en contextos diversos, respetando la diversidad lingüística. En el ámbito educativo, esta competencia se evalúa mediante la ED, establecida por la Ley Orgánica 2/2006, de 3 de mayo, de Educación, que se aplica en 4º de Primaria y 2º de ESO con el objetivo de diagnosticar el desarrollo de las competencias básicas del alumnado. En la Comunidad Autónoma Vasca, la ED es gestionada por el Instituto Vasco de Evaluación e Investigación Educativa (ISEI-IVEI), que evalúa la competencia lingüística a través de la comprensión y la producción escrita, esta última recogida en EP mediante textos manuscritos (ISEI-IVEI, 2025). En relación con la producción escrita en euskera, Ibarra (2016) analizó textos de 174 estudiantes de 4º de Primaria escritos en 15 minutos y observó una gran variabilidad en la cantidad de palabras producidas, que oscilaba entre 32 y 136.

El carácter manuscrito de estos textos impide tanto el etiquetado lingüístico manual como su análisis automático mediante herramientas de Procesamiento del Lenguaje Natural (PLN), ya que ambos procesos requieren la digitalización previa del texto. Dicha digitalización puede realizarse mediante mecanografiado o mediante el uso de sistemas de reconocimiento de texto manuscrito, conocidos como *Handwritten Text Recognition* (HTR, por sus siglas en inglés). El HTR es el proceso automatizado de convertir texto escrito a mano en texto digital. Mediante una combinación de procesamiento de imágenes y sistemas de reconocimiento, los modelos HTR permiten decodificar caracteres y

palabras escritas a mano. Actualmente, estos sistemas se basan en técnicas de aprendizaje automático y los modelos eficaces requieren conjuntos de datos de entrenamiento diversos (Rakesh et al., 2024). El HTR constituye en la actualidad una de las principales aplicaciones de la IA en el ámbito de las Humanidades Digitales (Stokes y Kiessling, 2024). Estos sistemas son capaces de tender puentes entre el mundo analógico y el digital y se han utilizado principalmente en bibliotecas y archivos, facilitando tanto la búsqueda de texto completo como el análisis de textos históricos a gran escala. De este modo, los contenidos manuscritos se vuelven más accesibles, explorables y reutilizables en una amplia gama de investigaciones y aplicaciones. Por ello, la investigación y el desarrollo en este campo resultan necesarios y ofrecen grandes oportunidades para mejorar el rendimiento del reconocimiento de texto manuscrito (Mahadevkar et al., 2024).

El reconocimiento es más preciso cuando los textos manuscritos han sido creados por adultos, mediante tareas de copia y cuando los textos son organizados y limpios (Nisa et al., 2023). Según los mismos autores, los textos escritos a mano en una situación de examen contienen muchos errores (tachones, texto añadido después, 'ruido de fondo', etc.) que obstaculizan un reconocimiento eficaz. Así, tras recoger muestras de más de 500 estudiantes universitarios en un simulacro de examen, presentan una forma de anotar los errores de forma sistemática en su *Students Messy Handwritten Dataset* (SMHD). Con respecto al reconocimiento de escrituras manuscritas de estudiantes noveles o en edad escolar, la literatura científica es todavía muy limitada y no se han identificado estudios que analicen de manera específica este tipo de escritura en contextos educativos obligatorios, ya que la investigación existente se concentra mayoritariamente en manuscritos históricos y contextos patrimoniales (Nockels et al., 2022).

Para crear modelos de HTR, la plataforma Transkribus (https://www.transkribus.org/) se ha consolidado como una de las herramientas fundamentales y de fácil manejo. Aunque su origen se encuentra en el reconocimiento y la transcripción automatizada de textos históricos (Muehlberger et al., 2019), actualmente se ha convertido en una herramienta clave para investigadores en Humanidades Digitales. Son varios los retos que hay que solventar en HTR, entre ellas: la precisión obtenida, los diseños de *layout*, la variedad de estilos de escritura y la limitación de datos (dataset). Así, en cuanto a la precisión (accuracy), es el total de palabras correctamente reconocidas dividida entre el total de palabras del corpus *100 (Mahadevkar et al, 2024). Según la plataforma Transkribus, un sistema que alcanza un CER (tasa de error de caracteres) igual o menor al 10% se considera altamente eficaz para la transcripción automática y un CER adecuado para el texto manuscrito (Transkribus, 2025). Según Jaipukar et al (2025) hay diferencias en los estilos de escritura, las formas y el espaciado individuales. En esa plataforma se han publicado dos modelos de HTR para el euskera: i) HABE_All GT 24 con un CER de 2,9% basado en un corpus de 323.776 palabras con textos de persona de todas las edades a partir de secundaria y entrenado en 2,737 páginas con el identificador Model ID59603. ii) El Contemporary Basque Student Handwritten con un CER de 6.07% basado en un corpus de 51,195 palabras con textos de alumnado de secundaria y entrenado en 525 páginas con el identificador Model ID185185. En el presente estudio los textos analizados corresponden a alumnado de EP.

Una vez digitalizados, los textos pueden organizarse en corpus lingüísticos, que reflejan el uso auténtico del lenguaje y permiten estudiar la evolución del lenguaje desde múltiples factores, tales como el contexto de comunicación y la tipología textual. Hasta la fecha, no se ha desarrollado un corpus específico que analice la escritura manuscrita de estudiantes en la educación obligatoria, lo que representa una oportunidad para futuras investigaciones en este campo (Durrant, 2022).

El objetivo de este estudio es crear un modelo de HTR para digitalizar textos manuscritos de estudiantes de 4º de EP en euskera. Esta labor puede responder a las necesidades del Sistema Educativo Vasco. Para conseguir dicho objetivo, se plantean varios objetivos secundarios: 1.1. digitalizar textos manuscritos en euskera, combinando corrección manual con modelos automatizados de *layout* y transcripción automática (HTR); 1.2. crear un corpus digitalizado (texto legible por las máquinas e imágenes de texto) en el mismo idioma que sirva para investigar y evaluar la escritura de los estudiantes dentro del Sistema Educativo Vasco en la etapa de EP; 1.3. evaluar la eficacia y precisión de los modelos HTR de euskera creados en la plataforma Transkribus y analizar los resultados obtenidos en la digitalización de textos en euskera.

Para conseguir el objetivo principal, el procedimiento ha sido el siguiente: la obtención de los datos mediante ISEI-IVEI, garantizando el anonimato de los estudiantes. A continuación, se utilizó la plataforma Transkribus para el reconocimiento automático de plantillas y texto, entrenando modelos de IA adaptados a las características específicas de los documentos. Después, se corrigieron manualmente todos los textos para garantizar la fiabilidad de los textos digitales. Luego, se realizó un análisis comparativo entre las transcripciones automáticas y manuales para identificar y clasificar los errores del sistema de reconocimiento de texto. Acto seguido, se obtuvieron datos cuantitativos y cualitativos sobre la evaluación del modelo, así como información sobre los factores que influyeron en su rendimiento y optimizar el proceso de digitalización de textos en euskera. Finalmente, los resultados indican que es posible el reconocimiento de los textos de EP con una tasa de error entre 5%-7% y se ha procedido a publicar los modelos de HTR en la plataforma Transkribus (https://www.transkribus.org/public-models). En un futuro se depositará el corpus en infraestructuras europeas de Humanidades Digitales, como CLARIN-ERIC.

En cuanto al resto del artículo, este está compuesto por las siguientes secciones: la Sección 2, donde se describe en detalle el método empleado para la recogida de datos, el proceso de digitalización y la creación del modelo de HTR; la Sección 3, que presenta los resultados obtenidos en términos de precisión, errores y validación manual de las transcripciones; y, por último, la Sección 4, dedicada a la discusión de los resultados y a las conclusiones, en la que se reflexiona sobre las aportaciones del modelo, las limitaciones encontradas y las futuras líneas de investigación orientadas a la mejora de la evaluación de la competencia lingüística y a la creación de recursos digitales en euskera.

2. MÉTODO

Este trabajo se enmarca en un estudio metodológico aplicado, de diseño descriptivo-analítico, orientado a la evaluación del rendimiento de un sistema de

reconocimiento de texto manuscrito (HTR) en un contexto educativo real. En este apartado se detalla la metodología empleada para el análisis de textos manuscritos en euskera producidos por estudiantes de EP, incluyendo la recogida indirecta de datos, la caracterización de los participantes, la selección y uso de herramientas para digitalizar y transcribir los textos, el entrenamiento y evaluación de los modelos HTR, y la creación de un sistema de codificación desarrollado para categorizar los errores del modelo HTR.

2.1. Recogida indirecta de datos

En este estudio, se dispone de una muestra significativa de 870 textos escritos por estudiantes de 4ª de EP del Sistema Educativo Vasco. Estos textos han sido facilitados por el ISEI-IVEI, organismo encargado de realizar las ED en la Comunidad Autónoma Vasca.

2.2. Participantes

En el marco de la ED organizada por el ISEI-IVEI en Euskadi, los alumnos de 4.º de EP participan en una evaluación a mitad de etapa. Como se ha descrito en la introducción, este proceso tiene como objetivo valorar el nivel de desarrollo de las competencias clave y específicas adquiridas hasta ese momento, en este caso, la competencia en comunicación lingüística. Los participantes, con edades comprendidas entre los 9 y 10 años, representan una muestra diversa del alumnado vasco, incluyendo estudiantes de centros públicos y concertados, así como de los diferentes modelos lingüísticos (A, B y D). Esta evaluación considera la diversidad del alumnado, incluyendo aquellos con necesidades educativas especiales, para quienes se realizan las adaptaciones necesarias en las pruebas. La ED de este grupo de edad permite una intervención temprana y efectiva en el proceso educativo.

2.3. Herramientas

Para llevar a cabo la metodología se emplearon tres herramientas principales: Transkribus, una hoja de cálculo y sistema de codificación de errores. Mediante la plataforma Transkribus, una plataforma de reconocimiento automático de plantillas y hojas en imágenes, que utiliza técnicas de IA especializadas, se ha realizado la digitalización y la creación de modelos. Además, el proceso ha permitido entrenar modelos de IA adaptados a las características específicas de los documentos analizados, optimizando así la precisión del reconocimiento textual. Con la hoja de cálculo, se ha llevado a cabo una evaluación exhaustiva de los resultados generados por los modelos de HTR. La evaluación se realizó de forma manual, comparando el texto transcrito automáticamente por nuestros modelos creados en Transkribus con el texto transcrito manualmente (*ground truth*). Esta comparación permitió un análisis cualitativo de las diferencias y similitudes entre ambas versiones y de esa forma se pudo comprobar la eficacia del modelo de HTR entrenado. Mediante el sistema de codificación de errores, se desarrolló una categorización de los diferentes tipos de errores para optimizar este

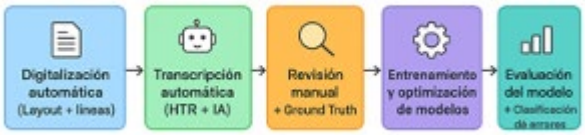
proceso. Esta categorización se diseñó a partir de la comparación sistemática entre las transcripciones automáticas y las transcripciones manuales, agrupando los errores según los patrones observados en estas discrepancias. De este modo, se identificaron las inexactitudes presentes en las transcripciones automáticas y, al mismo tiempo, se extrajeron datos estadísticos generales, proporcionando una visión global de la precisión y eficacia del modelo HTR empleado.

2.4. Proceso

El proceso seguido se ha compuesto de cuatro pasos: primero, ha habido un proceso de digitalización automática. Después, se han revisado las transcripciones y se ha creado *ground truth*. A continuación, se han creado y optimizado los modelos *layout* y HTR y finalmente, se ha llevado a cabo la evaluación manual del modelo HTR y sistema de codificación (ver Figura 1):

Figura 1

Proceso de creación y evaluación del modelo HTR para textos en euskera de estudiantes de EP



- Identificar el diseño de la hoja (*layout*): mediante el uso de técnicas avanzadas de IA, se ha desarrollado un modelo de diseño que permite extraer solo el texto manuscrito de la plantilla original y detectar automáticamente las líneas que contienen el texto manuscrito por los estudiantes, además de numerar las líneas correspondientes.
- Transcripción de textos: se han empleado modelos de reconocimiento de texto manuscrito (HTR) previamente desarrollados en proyectos anteriores, combinados con técnicas de IA.
- Revisión de transcripciones y creación de *ground truth*: las transcripciones generadas automáticamente han sido revisadas y corregidas manualmente, asegurando que reflejen fielmente el texto escrito por los estudiantes, para reducir al máximo la tasa de error que pueda presentar el modelo automático.
- Creación y optimización de modelos *layout* y HTR: se han entrenado nuevos modelos de IA para *layout* y HTR basados en los textos corregidos manualmente (*ground truth*). Este proceso iterativo ha optimizado los modelos existentes, reduciendo significativamente las tasas de error en la plataforma Transkribus.
- Evaluación manual del modelo HTR y sistema de codificación: se ha realizado un análisis comparativo del corpus de evaluación (10% del corpus total: 43 transcripciones automáticas de manuscritos) con las versiones corregidas manualmente, con el fin de identificar y clasificar los errores del sistema de reconocimiento de texto.

De esta manera, se organizaron y analizaron los datos manualmente, mediante la comparación de las transcripciones automáticas con sus respectivas versiones corregidas manualmente. Este procedimiento permitió analizar los errores que había en las transcripciones automáticas. Los errores se clasificaron según un sistema de codificación creado ad hoc, con el objetivo de obtener indicadores cuantitativos y cualitativos sobre el rendimiento del sistema. A continuación se puede apreciar cuál fue la codificación de los diversos tipos de errores (ver Tabla 1):

Tabla 1

Codificación de la evaluación según tipos de errores

Concepto	Código	Descripción
Confusión de letras	O	Errores en la detección de letras que no pertenecen a las familias de letras establecidas
Confusión entre las familias de letras	LF	Errores clasificados en seis grupos según la forma de las letras minúsculas o interrupciones en el trazo: LF1 (o, a, d, g, c, q); LF2 (e, l, b, f, h); LF3 (m, n, ñ, p); LF4 (r, s, x, z); LF5 (u, y, v, w); LF6 (i, j, k, t). Errores en las letras según pertenezcan a una de las seis familias de letras
Omisión de letra	ELI	Errores donde la transcripción automática no identifica una letra presente en el texto original, ya sea al inicio, en medio o al final de una palabra
Comisión de letra	LID	Errores en los que la transcripción automática añade una letra inexistente en el texto original, ya sea al inicio, en medio o al final de una palabra
Mayúscula/Minúscula	A	Errores de confusión de mayúsculas y minúsculas, es decir, confusión de alógrafos, ya sea de la misma letra o entre familias de letras
Puntuación	P	Incluye confusiones entre diferentes signos de puntuación, como detectar

		una coma en lugar de un punto o viceversa. También abarca la Puntuación No Identificada (PEI), donde la transcripción automática no detecta un signo de puntuación existente, y la Puntuación Identificada (PID) erróneamente, donde se detecta un signo de puntuación inexistente
Números	ZO	Errores en la identificación de números, incluyendo confusiones entre diferentes números, o entre números y letras en ambas direcciones
Tachadura	Z	Errores relacionados con la detección o no detección de tachaduras, como “xxx”. Puede ocurrir que la transcripción automática detecte una tachadura inexistente o no detecte una existente

- Publicación de los modelos de HTR: Transkribus permite la publicación de modelos de HTR que alcanzan un rendimiento excepcional. Estos modelos se consideran óptimos cuando su tasa de error se mantiene por debajo del umbral crítico, que oscila entre el 5% y el 7%. Los modelos que cumplen con este estricto criterio de precisión son elegibles para ser compartidos públicamente en la plataforma.

2.5. Declaración ética

Para llevar a cabo esta investigación, el acceso a los textos se ha realizado bajo estrictas medidas de confidencialidad y protección de datos, cumpliendo con las normativas vigentes de protección de datos personales. El ISEI-IVEI garantizó el anonimato completo del alumnado, utilizando los datos únicamente con fines de investigación.

3. Resultados

3.1. Análisis del corpus

El resultado principal ha sido obtener un corpus digitalizado de 4º de Primaria que permite su análisis automático y el desarrollo de modelos HTR. Así, el conjunto de datos consta de 792 textos, que en total suman 68.652 palabras, de las cuales 11.132 son únicas. La extensión de los documentos varía considerablemente: los más extensos contienen

entre 160 y 175 palabras, mientras que los más breves pueden contener de 5 a 10 palabras. Se puede observar que los textos de los alumnos de 4º de primaria, en general, han sido bastante variables en cuanto a la extensión del texto. En definitiva, aunque algunos textos son extensos, muchos otros son bastante reducidos, lo que refleja la diversidad en el proceso de aprendizaje. El análisis preliminar de los textos se desarrolla en las siguientes subsecciones.

3.1.1. Categorías gramaticales y errores ortográficos

Al examinar las palabras que aparecen con mayor frecuencia en el corpus, especialmente los sustantivos y verbos más comunes, se puede inferir que el tema principal de estos textos gira en torno a las vacaciones familiares. Esto se evidencia no solo en los ejemplos mencionados anteriormente, sino también en el análisis de frecuencia de palabras realizado. Cabe destacar que la elección de los temas para la evaluación es responsabilidad del ISEI-IVEI, que selecciona cuidadosamente textos y propuestas relevantes y cercanas al contexto de los estudiantes. Este proceso busca garantizar que los materiales sean representativos de situaciones comunicativas reales, contribuyendo así a una evaluación más auténtica y significativa de la comprensión escrita.

Por otro lado, el análisis del corpus ha revelado una alta incidencia de errores ortográficos (ortografía funcional) (véase Gráfico 1 y Gráfico 2). Un caso notable es el de las palabras ‘ondartza’ (playa) y ‘artu’ (agarrar), que en euskera se escriben correctamente con "h" inicial: ‘hondartza’ y ‘hartu’. Sin embargo, se ha observado que, en aproximadamente la mitad de las ocasiones, los alumnos han omitido esta letra. Este tipo de equivocaciones, junto con otros errores similares, pueden tener un impacto en los resultados del análisis de frecuencia de palabras, alterando ligeramente las estadísticas obtenidas.

Gráfico 1

Lemas de sustantivos y sus frecuencias en el corpus

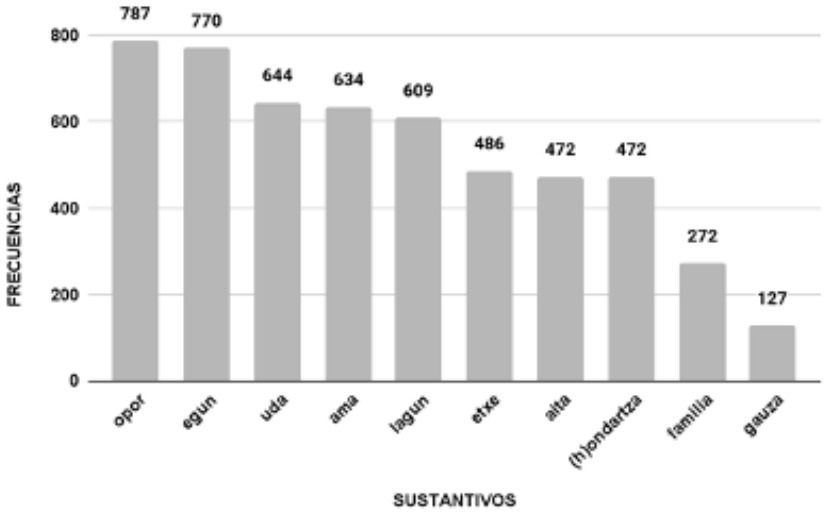
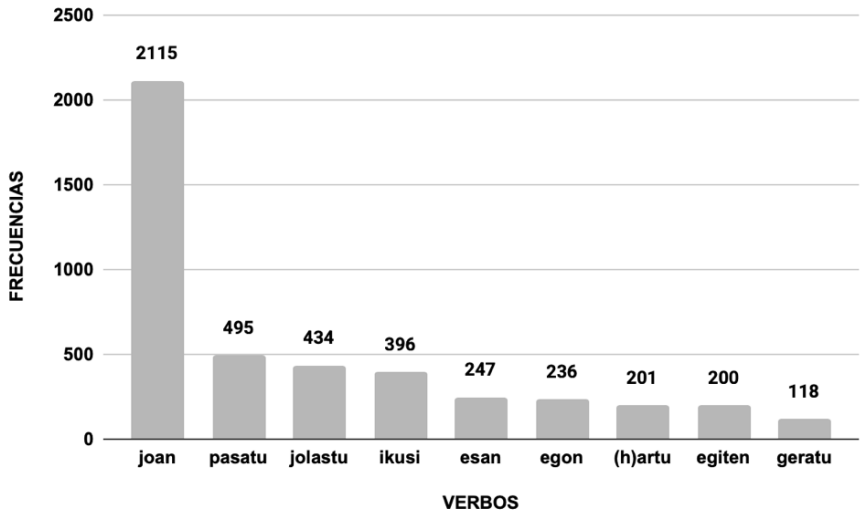


Gráfico 2

Lemas de verbos y sus frecuencias en el corpus



3.2. Análisis de los modelos de *layout* y HTR

Con el fin de interpretar los resultados obtenidos del modelo de *layout* y HTR (véase Tabla 2), se realizó el entrenamiento de modelos de IA orientados tanto al *layout* como al HTR. Este proceso comenzó con la preparación y segmentación cuidadosa del corpus, permitiendo que los modelos pudieran identificar correctamente las distintas regiones de los documentos, como encabezados, párrafos, notas al margen y otros elementos estructurales relevantes. Una vez generadas las primeras transcripciones automáticas mediante estos modelos, se procedió a una fase crucial de corrección manual. En esta etapa, se revisaron minuciosamente los textos producidos por el sistema, identificando y corrigiendo errores tanto en la interpretación de la estructura del documento como en la transcripción del contenido manuscrito. Gracias a la combinación de la corrección manual y las técnicas avanzadas de aprendizaje automático, se logró una mejora continua en el rendimiento de los modelos. Cada ciclo de corrección y reentrenamiento permitió reducir progresivamente la tasa de error, optimizando tanto el reconocimiento del *layout* como la precisión en la transcripción de textos manuscritos.

Analizando los resultados de este estudio, se observa que el modelo de identificación de *layout* con ID 77229, entrenado en un conjunto de 714 páginas, demostró una alta precisión, alcanzando una tasa de error en letras de tan solo el 2,28% en el corpus de evaluación. En cuanto al modelo de Reconocimiento de Texto Manuscrito (HTR) con ID 77829, entrenado en un conjunto ligeramente mayor de 749 páginas, se obtuvo una tasa de error en letras del 6,04%.

Tabla 2

Resultados de los modelos de layout y HTR

	Páginas entrenadas	Tasa de error en letras
Layout	714	2,28%
HTR	749	6,04%

El corpus de evaluación de HTR, con el 10% de los estudiantes contiene 43 documentos con 4.005 total de palabras y 1.638 palabras únicas. La extensión de los documentos varía considerablemente: los más extensos contienen entre 117 y 133 palabras, mientras que los más breves contienen de 72 a 75 palabras.

3.3. Análisis de los errores del sistema

El presente análisis compara los errores de escritura del 10%, según Transkribus la muestra más grande, categorizando y cuantificando los tipos de errores más frecuentes. Los resultados del análisis de errores de escritura entre las transcripciones automáticas y las transcripciones corregidas manualmente, en total, se han identificado 1202 errores, distribuidos en las categorías mencionadas que ofrecen una visión detallada de los errores más frecuentes.

El error más frecuente, con una marcada diferencia, es la confusión de letras, categorizado como "O" en el estudio (véase Tabla 3). Este tipo de error representa casi la mitad de todos los casos observados, con 587 erratas, lo que equivale al 48,84% del total. Con respecto al segundo tipo de error más común se relaciona con las familias de letras, agrupadas en las categorías LF1 a LF6. En conjunto, estos errores suman 219 casos, representando el 18,22% del total. Es interesante notar la distribución desigual dentro de esta categoría: LF1 muestra la mayor frecuencia con 87 casos, seguida por LF4 con 44 casos, mientras que LF3 y LF6 presentan 32 casos cada una. LF2 tiene 24 casos, y notablemente, LF5 no registra ningún error. Los errores de puntuación, categorizados como "P", ocupan el tercer lugar en frecuencia con 172 casos, constituyendo el 14,31% del total. Un análisis más detallado de esta categoría revela tres subcategorías específicas: Puntuación No Identificada (PEI) con 73 casos, Puntuación Identificada (PID) con 27 casos, y 72 casos adicionales que implican confusión entre diferentes tipos de signos de puntuación. Estos últimos casos se refieren a situaciones donde, por ejemplo, la transcripción automática ha detectado una coma cuando en realidad era un punto, o viceversa. En cuanto a la omisión de letras (ELI) y la comisión de letras (LID) representan el 7,24% y 5,74% de los errores respectivamente, con 87 y 69 casos. Sobre los errores en el uso de mayúsculas y minúsculas (A) son menos frecuentes, con 43 casos que representan el 3,58% del total. Finalmente, en relación con las categorías con menor incidencia son las tachaduras (Z) y los errores relacionados con números (ZO), que representan el 1,25% y 0,83% respectivamente, con 15 y 10 casos.

Tabla 3

Resultados de la evaluación del sistema de codificación

Código	Concepto	K	Porcentaje	Total
O	Confusión de letras	587	48,84%	587
LF	Familia de letras	219	18,22%	219
LF1		87	7,24%	
LF4		44	3,66%	
LF3		32	2,66%	
LF6		32	2,66%	
LF2		24	2,00%	
LF5		0	0,00%	
P	Puntuación	172	14,31%	172
PEI	Puntuación No Identificada	73	42,44%	
PID	Puntuación Identificada	27	15,70%	
ELI	Omisión de letra	87	7,24%	87
LID	Comisión de letra	69	5,74%	69
A	Mayúscula/Minúscula	43	3,58%	43
Z	Tachadura	15	1,25%	15
ZO	Números	10	0,83%	10
Errores en total			100%	1202

Teniendo en cuenta el análisis realizado, y sin extraer conclusiones definitivas, se ha evaluado el rendimiento del modelo automático HTR en la transcripción de textos manuscritos en euskera. Los datos de valoración, obtenidos a partir de un 10% del corpus total, muestran una tasa de error del 6,04% en letras para el modelo HTR. Dada la relevancia de este dato (el 48% de los errores del 6,04% corresponden a confusión de letras), se continúa investigando para optimizar el modelo, respetando en todo momento la naturaleza exploratoria de este análisis.

4. DISCUSIÓN

Los resultados de este estudio ponen de manifiesto las dificultades inherentes al reconocimiento automático de escritura manuscrita infantil, estrechamente vinculadas al propio proceso de adquisición de la escritura. En las primeras etapas educativas, la producción escrita se caracteriza por una notable variabilidad en la forma de las letras, en la regularidad del trazo y en la organización del texto, aspectos que han sido descritos en investigaciones previas desde una perspectiva cognitiva y neurolingüística (Berninger y Richards, 2010). Esta inestabilidad gráfica constituye un reto añadido para los sistemas de Reconocimiento de Texto Manuscrito (HTR), que generalmente alcanzan mejores resultados cuando se enfrentan a escrituras más homogéneas y normalizadas.

En este contexto, la tasa de error en letras del 6,04% obtenida por el modelo HTR puede considerarse un resultado satisfactorio. Estudios previos han demostrado que, incluso con arquitecturas avanzadas basadas en redes neuronales profundas, el

reconocimiento de escritura manuscrita no restringida continúa presentando dificultades cuando se trabaja con corpus heterogéneos (Graves et al., 2009; Ingle et al., 2019). Por tanto, el rendimiento alcanzado en este trabajo se sitúa en una línea coherente con los resultados descritos en la literatura y refuerza la validez del enfoque metodológico adoptado.

El análisis detallado de los errores muestra que la confusión de letras constituye la principal fuente de error del sistema. Este patrón coincide con lo descrito en estudios anteriores, que señalan la similitud visual entre grafías y la ambigüedad de determinados trazos manuscritos como factores clave en los errores de reconocimiento automático (Fischer et al., 2012). En el caso de la escritura infantil, estas dificultades se ven acentuadas por la falta de consistencia en la representación de las letras, característica propia de las etapas iniciales del aprendizaje de la escritura (Berninger y Richards, 2010).

De manera similar, los errores agrupados en familias de letras reflejan limitaciones del sistema a la hora de discriminar grafemas con rasgos gráficos próximos. Este tipo de errores ha sido documentado tanto en enfoques clásicos como en modelos más recientes de HTR, lo que sugiere que la variabilidad gráfica sigue siendo un factor determinante en el rendimiento de estos sistemas, independientemente de la técnica empleada (Fischer et al., 2012; Ingle et al., 2019).

En relación con los errores de puntuación, su presencia resulta coherente con investigaciones que indican que los signos de puntuación suelen presentar mayores dificultades para los sistemas automáticos, debido a su reducido tamaño y a la diversidad de formas manuscritas que pueden adoptar (Graves et al., 2009). Por el contrario, los errores relacionados con mayúsculas, números y tachaduras presentan una incidencia menor en el conjunto de errores analizados, lo que concuerda con trabajos que sitúan los principales desafíos del HTR en la identificación de grafías y en la variabilidad del trazo, más que en elementos gráficos aislados (Fischer et al., 2012).

Finalmente, el uso de la plataforma Transkribus para el entrenamiento y la evaluación de los modelos ha resultado adecuado para los objetivos de este estudio. Investigaciones previas han destacado su utilidad para la creación de modelos de HTR adaptados a corpus específicos y a documentos con características particulares (Kahle et al., 2017). En línea con lo señalado por Ingle et al. (2019), los resultados obtenidos indican que la combinación de corrección manual, ampliación progresiva del corpus y reentrenamiento iterativo constituye una estrategia eficaz para mejorar el rendimiento de los sistemas de reconocimiento, especialmente en contextos complejos como el de la escritura manuscrita infantil.

5. CONCLUSIONES, LÍMITES Y FUTURO TRABAJO

En el presente estudio se ha digitalizado una colección de 870 textos manuscritos de estudiantes de EP del Sistema Educativo Vasco. Para ello, se han desarrollado modelos de detección automática del layout y del texto manuscrito o HTR en la plataforma Transkribus, utilizando textos manuscritos de alumnos de EP. Dado que actualmente no existen modelos específicos para este nivel educativo, se han desarrollado modelos adaptados a las características de la escritura de los estudiantes de EP. Así, el layout ha

alcanzado una precisión del 97.72% en la identificación de líneas en el conjunto de evaluación, mientras que la precisión del modelo HTR ha sido de 93.96% en CER. Por lo tanto, los resultados obtenidos en este proyecto demuestran un nivel de precisión notable, ya que su tasa de error se sitúa entre el 5% y el 7% CER, indicando un avance significativo en la adaptación y eficacia de estos modelos para el contexto educativo específico. Es un logro importante para la comunidad educativa en general y los órganos de evaluación como el ISEI-IVEI en particular. Además, haberlo conseguido en un idioma en el que escasean los datos también es destacable.

El desarrollo de modelos para la digitalización de textos manuscritos de EP es complejo. Por un lado, los estilos caligráficos son muy heterogéneos; por otro lado, los textos tienden a ser breves y, por último, la presencia de errores ortográficos es una dificultad añadida. De todas maneras, nuestra tarea no ha sido la de corregir los errores ortográficos, sino más bien la de preservarlos. Que haya errores ortográficos en las muestras del alumnado provoca que la capacidad del modelo sea limitada, tanto para que el sistema aprenda el léxico como para que el sistema aplique las correcciones automáticas. La alta frecuencia de errores encontrados podría atribuirse a varios factores, como la variabilidad en la caligrafía de los estudiantes, la similitud visual entre ciertas letras, o las limitaciones en los algoritmos de reconocimiento. Este resultado destaca la necesidad de incrementar los datos de entrenamiento para distinguir sutilezas en la formación de letras individuales, especialmente en el contexto de la escritura en desarrollo de los estudiantes. La aplicación de la plataforma Transkribus, basada en IA, para el reconocimiento automatizado de plantillas y texto en imágenes, ha sido adaptada y optimizada para el euskera. Este proceso implicó el desarrollo de modelos de *layout* y de reconocimiento de texto manuscrito (HTR) específicos, que han demostrado mejorar la precisión y la eficiencia de la transcripción automática.

Adicionalmente, se diseñó un sistema de codificación personalizado para clasificar los errores en las transcripciones, lo que facilita la identificación de áreas de mejora y la optimización de los algoritmos de reconocimiento de texto. El error más común, la confusión de letras, de este modelo representa casi la mitad de todos los casos (48,84%). Este hallazgo sugiere que el modelo es altamente efectivo para su propósito, pero la reducción de la confusión entre letras similares constituye una oportunidad clara de mejora. Abordar este aspecto podría elevar aún más la precisión. Sin embargo, los errores relacionados con las familias de letras, que representan el 18,22% del total, revelan patrones específicos de confusión entre grupos de letras similares. Este hallazgo sugiere que ciertas familias de letras son más propensas a la confusión como, por ejemplo, lo que podría informar el desarrollo de estrategias de reconocimiento más específicas y eficaces. Además, la ausencia de errores en la familia LF5 (u, v, w, y) merece una investigación más profunda para entender qué características de estas letras las hacen más fácilmente reconocibles por el sistema HTR, teniendo en cuenta que, en euskera, en este caso, no se utilizan las letras v, w e y. Centrando en los errores de puntuación, que constituyen el 14,31% del total, revelan un área de mejora significativa para los sistemas HTR. La distribución de estos errores en subcategorías (puntuación no identificada, puntuación identificada incorrectamente, y confusión entre diferentes tipos de signos de puntuación) proporciona una visión detallada en este ámbito. Estos resultados sugieren que los

sistemas HTR actuales tienen dificultades no solo para detectar la presencia de signos de puntuación, sino también para distinguir entre diferentes tipos de signos. Este hallazgo es importante, ya que la puntuación juega un papel crucial en las herramientas de procesamiento del lenguaje natural que analizan la complejidad del lenguaje. En muchas métricas, se tiene en cuenta la estructura de la oración y otros elementos relacionados con la puntuación, lo que subraya su importancia en la comprensión y análisis del texto. Esta información es fundamental no solo para mejorar la precisión de los sistemas HTR en la captura de la estructura y el significado completo de los textos manuscritos, sino también para el desarrollo de herramientas de análisis lingüístico más sofisticadas.

No obstante, la mayor frecuencia de errores de omisión de letras (7,24%) en comparación con la comisión de letras (5,74%) sugiere que el sistema tiende más a no reconocer letras existentes que a añadir letras inexistentes. Este hallazgo podría estar relacionado con la sensibilidad del sistema a la variabilidad en la escritura manuscrita, donde las letras pueden ser difíciles de distinguir o estar mal formadas, como se ha podido observar en nuestros textos, las letras como a u o mal formadas, o la s y r.

El relativamente bajo porcentaje de errores en el uso de mayúsculas y minúsculas (3,58%) indica un buen rendimiento del sistema en este aspecto. Sin embargo, este resultado también sugiere que, aunque los sistemas HTR son capaces de distinguir entre mayúsculas y minúsculas con cierta precisión, aún existen áreas de mejora, especialmente en contextos donde la caligrafía es menos clara o está en desarrollo.

Los resultados detallados de este análisis de errores proporcionan una base sólida para mejorar los algoritmos de HTR, especialmente en el contexto de la escritura manuscrita de estudiantes. La identificación precisa de los tipos de errores más comunes puede informar el desarrollo de estrategias de reconocimiento más eficaces y personalizadas para este grupo específico de usuarios.

Para finalizar, centrando en los límites y posibles futuros trabajos, hay que enfocar en las categorías con menor incidencia (tachaduras y errores relacionados con números) y las categorías adicionales (PEI y PID) que señalan áreas específicas para futuras investigaciones y mejoras en los sistemas HTR. Estos errores, aunque menos frecuentes, indican que posiblemente se necesiten más datos para poder concluir mejor sobre el tema.

6. AGRADECIMIENTOS

Agradecemos a ISEI-IVEI la cesión de los datos anónimos utilizados en esta investigación.

7. FINANCIACIÓN

Este trabajo ha sido posible gracias a la beca predoctoral del Gobierno Vasco/Eusko Jaurlaritza (Ref. PRE-2024_1_0273), a la financiación de la Diputación de Gipuzkoa a través del PROYECTO IDATZIZ: análisis de muestras de escritura en euskera mediante herramientas tecnológicas en 5 centros (P17) y CLARIAH-EUS-gArA (2024-CIE2-000003-01).

REFERENCIAS

Adak, C., Chaudhuri, B. B. y Blumenstein, M. (2017). Legibility and aesthetic analysis of handwriting. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* (Vol. 1, pp. 175-182). IEEE. doi: 10.1109/ICDAR.2017.37.

Berninger, V. W. y Richards, T. L. (2010). Inter-relationships among behavioral markers, genes, brain, and treatment in dyslexia and dysgraphia. *Future Neurology*, *5*(4), 597–617. https://doi.org/10.2217/fnl.10.22

Butler, T. S. (2003). *Human interaction with digital ink: legibility measurement and structural analysis* [Tesis de Doctorado]. University of Hertfordshire. http://hdl.handle.net/2299/14158

Dinehart, L. H. (2015). Handwriting in early childhood education: Current research and future implications. *Journal of Early Childhood Literacy*, 15, 97–118. DOI: https://doi.org/10.1177/1468798414522825

Durrant, P. (2022). Studying children's writing development with a corpus. *Applied Corpus Linguistics*, *2*(3), 100026. https://doi.org/10.1016/j.acorp.2022.100026

Fischer, A., Frinken, V., Bunke, H. y Fornés, A. (2012). Lexicon-free handwritten word spotting using character HMMs. *Pattern Recognition Letters*, *33*(7), 934–942. https://doi.org/10.1016/j.patrec.2011.09.009

Fogel, Y., Rosenblum, S. y Barnett, A. L. (2022). Handwriting legibility across different writing tasks in school-aged children. *Hong Kong journal of occupational therapy: HKJOT, 35*(1), 44–51. https://doi.org/10.1177/15691861221075709

González, R. M. (2003). Propuesta de intervención en los procesos cognitivos y estructuras textuales en niños con DAE. *Psicothema*, *15*(3), 458-463.

Government of South Australia, Department of Education and Children's Services (2007). *Handwriting in the South Australian Curriculum* (2nd ed.; G. Groves, Ed.). Hyde Park Press. https://www.australianschoolfonts.com.au/assets/resources/handwriting-2nd-ed-2007-sa.pdf

Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H. y Schmidhuber, J. (2009). A novel connectionist system for unconstrained handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(5), 855–868. https://doi.org/10.1109/TPAMI.2008.137

Ibarra, I. (2016). *Esku-idazketa eta testu eleanitzen arteko loturak*. [Tesis de Doctorado]. Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU). https://addi.ehu.eus/bitstream/handle/10810/26312/TESIS_IBARRA_LIZUNDIA_IRUNE.pdf?sequence=1&isAllowed=y

Ibarra, I., Ortube, M., eta Muxika, I. (2020). *Letra xeheak, bai!* Booktegi. Recuperado de https://www.booktegi.eus/liburuak/letra-xeheakbai/

Ingle, R., Fujii, Y., Deselaers, T., Baccash, J. y Popat, A. (2019). A scalable handwritten text recognition system. Proceedings of ICDAR 2019, 17–24. https://doi.org/10.48550/arXiv.1904.09150

Instituto Vasco de Evaluación e Investigación Educativa (ISEI-IVEI). (2025). *Evaluación diagnóstica de mitad de etapa. Competencia en comunicación lingüística*. Gobierno Vasco. https://isei-ivei.euskadi.eus/es/

Jaipukar, P.B., Pal V.S., Varade, S.M. y Mendhule, S.P. (2025). Interpreting Handwriting Text Recognition. *International Journal of Progressive Research in Engineering Management and Science (IJPREMS) (Int Peer Reviewed Journal)*, *5*(8), 1745-1748. https://www.ijprems.com/uploadedfiles/paper//issue_8_august_2025/43592/final/fin_ijprems1756717555.pdf

Kahle, P., Colutto, S., Hackl, G. y Mühlberger, G. (2017). Transkribus—A service platform for transcription, recognition and retrieval of historical documents. Proceedings of ICDAR 2017, 19–24. https://doi.org/10.1109/ICDAR.2017.307

Ley Orgánica 2/2006, de 3 de mayo, de Educación. (BOE núm. 106, de 4 de mayo de 2006, pp. 1-112).

Mahadevkar, S., Patil, S. y Kotecha, K. (2024). Enhancement of handwritten text recognition using AI-based hybrid approach. *MethodsX*, 12, 1-11. DOI: https://doi.org/10.1016/j.mex.2024.102654

Muehlberger, G., Seaward, L., Terras, M., Ares Oliveira, S., Bosch, V., Bryan, M., Colutto, S., Déjean, H., Diem, M., Fiel, S., Gatos, B., Greinoecker, A., Grüning, T., Hackl, G., Haukkovaara, V., Heyer, G., Hirvonen, L., Hodel, T., Jokinen, M., ... Zagoris, K. (2019). Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study. *Journal of Documentation*, *75*(5), 954-976. https://www.emerald.com/insight/content/doi/10.1108/jd-07-2018-0114/full/html

Nisa, H., Thom, J., Ciesielski, V. y Tennakoon, R. (2023). *Student Messy Handwritten Dataset (SMHD)*. RMIT University. Dataset. https://doi.org/10.25439/rmt.24312715.v1

Nockels, J., Gooding, P., Ames, S., & Terras, M. (2022). Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of Transkribus in published research. *Archival science*, *22*(3), 367–392. https://doi.org/10.1007/s10502-022-09397-0

Rakesh, S., Kushal Reddy, P., V. Prashanth, V. y Srinath Reddy, K. (2024). Handwritten text recognition using deep learning techniques: A survey. *MATEC Web of Conferences*, *392*(01126), 1-8. DOI: https://doi.org/10.1051/matecconf/202439201126

Stokes, P., y Kiessling, B. (2024). Sharing Data for Handwritten Text Recognition (HTR). *Digital Humanities in Practice*. ffhal-04444641ff

Transkribus. (2025). *Character Error Rate and Learning Curve*. Transkribus Help Center. https://help.transkribus.org/character-error-rate-and-learning-curve,