

16. IGARRITZ: EL PREDICTOR DE PALABRAS PARA EL EUSKERA BASADO EN LA INTELIGENCIA ARTIFICIAL Y SU EVALUACIÓN EN EL ENTORNO ESCOLAR

Unai Atutxa Barrenetxea

Iker de la iglesia Martínez

Mikel Irukieta Quintian

Universidad del País Vasco (UPV/EHU)

1. INTRODUCCIÓN

Cuando un o una alumna no puede comunicarse oralmente, los Sistemas Aumentativos y Alternativos de Comunicación (SAAC) son indispensables para aumentar la autonomía del alumnado y mejorar su calidad de vida (González, 2003). La Comunicación Aumentativa y Alternativa tiene como objetivo complementar o reemplazar el habla o la escritura en personas que sufren graves problemas de habla y movimiento. Los SAAC pueden ser de dos tipos (Garay-Vitoria y Abascal, 2006; Lloyd y Karlan, 1984;): i) aquellos sistemas que no necesitan de soporte físico o material que no sea el cuerpo, sistema sin ayuda (entre los gestualistas: el alfabeto dactilológico, la lengua de signos; entre los oralistas: lectura labiofacial...) y ii) sistemas con ayuda (comunicadores, tablet con pictogramas, programas de ordenador...). Una gran diversidad de discapacidades físicas o cognitivas entre las que se encuentran la parálisis cerebral, la esclerosis lateral amiotrófica y otras enfermedades neurodegenerativas, suelen ir acompañadas de trastornos del lenguaje asociados. Cualquiera que sea la razón, la comunicación oral es imposible para estas personas que también tienen serias dificultades para controlar físicamente su entorno. En particular, la mayoría de ellos y ellas no pueden utilizar los dispositivos de entrada estándar de una computadora o tienen grandes dificultades. En el caso de los y las alumnas con parálisis cerebral, cada vez es más habitual el uso de un hardware de seguimiento de la mirada con el que poder interactuar o teclear en el ordenador. Este tipo de herramientas adaptadas pueden llegar a ser la única vía de comunicación para ellos y ellas. Sin embargo, como bien indican

Khelil et al. (2023), aunque los SAAC facilitan las capacidades de comunicación de sus usuarios, la entrada de texto sigue siendo lenta y agotadora. Estos autores consideran dos enfoques complementarios para acelerar la escritura: i) selección rápida de teclas mediante un diseño de teclado optimizado y ii) ahorro de pulsaciones mediante predicción de palabras.

En este trabajo, una alumna (4º de ESO) con parálisis cerebral emplea el ordenador mediante un sistema de seguimiento ocular (eye-tracking), y para ello los programas que tiene a su disposición son: i) Grid3, Easyclick (Irisbond, 2024) o ii) AsTeRICS-Grid (Klaus et al., 2019). Gracias a la ayuda de dichos softwares, puede manejar el ordenador mediante el seguimiento de la mirada: abrir o cerrar archivos y escribir textos a través del teclado virtual de la pantalla. Sin embargo, escribir con el iris es más lento y laborioso, por consiguiente, es importante que estos sistemas sugieran palabras, permitiendo redactar textos más extensos de manera más rápida. Aunque estos sistemas predictivos a menudo permiten la escritura en euskera (un recurso habitual suele ser poder introducir un diccionario o listas de palabras) funcionan peor que en otras lenguas y difícilmente ayudan a aumentar la producción. En la actualidad, la alumna carece de un sistema eficaz que le facilite una escritura ágil y cómoda en euskera, su lengua materna y lengua vehicular en el sistema educativo vasco. En consecuencia, sus dificultades para escribir y la velocidad con la que escribe afectan directamente a la calidad de los textos; ya que, al tener que hacer un gran esfuerzo, tiende a utilizar frases más cortas o poco elaboradas. A nuestro entender, la incorporación de un predictor de palabras en euskera representaría un avance cualitativo. Un sistema de predicción adecuado ayudaría a minimizar las pulsaciones de teclado gracias a las predicciones que ofrece, y asimismo, evitaría errores de tecleo u ortografía. Con todo ello, pensamos que se lograría promover la educación inclusiva, puesto que tal y como indica Luna (2013) la inclusión se debe fundamentar en eliminar o minimizar las barreras que limitan el aprendizaje y la participación de todo el alumnado.

Por ello, en vías de facilitar la inclusión del alumnado, no es de extrañar que desde hace tiempo se haya visto la predicción de las palabras como una gran oportunidad. Palazuelos-Caigas (2001) aborda el estudio de la inclusión de información lingüística en la predicción de palabras en castellano, con el objetivo de mejorar los sistemas de ayuda a la escritura de personas que pueden tener distintos tipos de discapacidad. En dicho trabajo, podemos apreciar que para comienzos de siglo ya existían diversos sistemas que utilizan predicción de palabras: Aurora 3.0, Co-Writer, EZ Keys, PAL, PredictAbility, Prejudice, Prophet y Telepatic, entre otros. La mayoría de ellos son programas que muestran ventanas emergentes sobre el procesador de texto estándar. Estas ventanas contienen listas de palabras predichas. Otras herramientas son más completas, como editores o procesadores de texto, que incluyen emuladores de teclado para ser utilizados con el ratón (acceso directo) o implementan un barrido automático para ser utilizadas

con pulsadores. El avance ha sido muy significativo estas últimas décadas, y son varios los sistemas de predicción de palabras que se han desarrollado y se están desarrollando utilizando distintos métodos para diferentes idiomas (Ghayoomi y Momtazi, 2009). Tradicionalmente, estos sistemas se han basado en el modelado estadístico de lenguajes de n-gramas. Recientemente, se han desarrollado modelos de lenguaje más sofisticados para mejorar el rendimiento de estos modelos de lenguaje tradicionales (Cavaliere et al., 2016). Uno de los modelos de lenguaje más destacados es BERT (Bidirectional Encoder Representations from Transformers), creado por Google (2018). BERT se basa en la innovadora arquitectura Transformer. Lo que lo distingue de los modelos tradicionales es su capacidad bidireccional para procesar texto. Mientras que los modelos convencionales siguen un flujo unidireccional, ya sea de izquierda a derecha o de derecha a izquierda, BERT utiliza una atención bidireccional que le permite capturar el contexto de las palabras tanto en su contexto previo como posterior. Estos modelos han demostrado un rendimiento impresionante en una variedad de tareas de procesamiento de lenguaje natural, incluida la predicción de palabras.

No obstante, en el caso del euskera no ha habido tanto avance, y prueba de ello es la situación de nuestra alumna, y es que la única manera para hacer previsiones con el software utilizado hasta la fecha (GRID3) es introduciendo un diccionario de 10000 palabras. Sin embargo, desde hace poco la alumna tiene disponible la nueva herramienta IGARRITZ: el primer entorno web adaptado para predecir palabras en euskera basándose en la Inteligencia Artificial que, además, es compatible con programas de seguimiento ocular, fácilmente accesible utilizando el explorador Firefox en <http://igarritz.clariah.eus>.

En cuanto a la evaluación de este tipo de sistemas de predicción de textos, la mayoría de las evaluaciones suelen ser automáticas. Este tipo de evaluación puede proporcionarnos información muy importante, pero no podemos dejar de lado la aplicación de dichas herramientas en casos reales. Son muy pocos los trabajos que llevan a cabo la evaluación en un entorno real; a destacar el realizado por Palazuelos-Caigas (2001), donde se evalúa el uso del sistema Predice por parte de personas con discapacidad, con el fin de obtener datos sobre la influencia de la predicción en la cantidad y calidad de los textos que generan, así como su valoración subjetiva. Resulta sumamente relevante el estudio realizado por Raynal y Badr (2022), en el cual exploran el seguimiento ocular de los usuarios mientras llevan a cabo una tarea de introducción de texto utilizando un teclado virtual con una lista de palabras predictivas. El objetivo principal de esta investigación es analizar las estrategias empleadas por los usuarios al interactuar con dicha lista de predicción.

2. OBJETIVOS

En este estudio, nuestro objetivo es evaluar el sistema predictivo IGARRITZ en un entorno real. Para lograrlo, emplearemos textos redactados por la estudiante durante el horario escolar y realizaremos dos evaluaciones:

Evaluación de la capacidad de IGARRITZ para reducir la cantidad de letras necesarias, comparándola con el sistema anteriormente utilizado, GRID3.

Evaluación de la eficacia con la que la estudiante ha utilizado IGARRITZ, mediante el análisis de las pulsaciones de teclado requeridas.

3. METODOLOGÍA

La metodología se divide en cuatro secciones principales. En primer lugar, se detalla la herramienta IGARRITZ, la cual es empleada por la alumna para la redacción de textos. Se presentarán sus características principales y se explicará su funcionamiento. En la segunda sección, se describe el corpus creado y el proceso utilizado para su recopilación. En los dos últimos apartados, se expondrá el proceso seguido para llevar a cabo las dos evaluaciones planteadas en este trabajo.

3.1. Servicio web IGARRITZ:

Nuestra primera tarea ha sido desarrollar un sistema de predicción de palabras en euskera. Nuestro sistema se fundamenta en la Inteligencia Artificial, utilizando modelos de lenguaje. Por consiguiente, el primer paso es el entrenamiento del modelo. Este se logra mediante el análisis de grandes conjuntos de textos para establecer relaciones entre palabras y desarrollar modelos estadísticos. Dado que estos modelos pueden originarse de diversas formas, el proceso de creación de un modelo específico se conoce como entrenamiento. Cuantos más datos se utilicen para entrenar el modelo, más precisa será la representación del idioma. Nuestro sistema emplea el modelo de lenguaje HiTZ/roberta-euscrawl-base-cased (Artetxe et al., 2022), que se basa en la arquitectura transformer RoBERTa (Liu et al., 2019; Vaswani et al., 2017). Este modelo ha sido entrenado con una variedad de textos, lo que le permite a menudo predecir palabras y frases que no necesariamente coinciden con el lenguaje utilizado por una estudiante de la ESO. Por ende, para adaptar las predicciones al lenguaje de una estudiante de la ESO, hemos recopilado textos relacionados con la educación, como Gizapedia (Sarasola y Sarasola, 2023), redacciones de estudiantes y textos sobre educación de fuentes como Wikipedia, Elkar y Berria.

Con la metodología mencionada, hemos desarrollado el servicio web IGARRITZ, accesible en <http://igarritz.clariah.eus> mediante el navegador Firefox. En la parte superior de la interfaz, el usuario puede introducir texto, mientras que en el centro, el sistema ofrece seis predicciones en todo momento. Antes de comenzar a escribir una palabra, IGARRITZ mostrará las seis opciones más probables. Si ninguna de las sugerencias es la deseada, el usuario puede ingresar una letra utilizando el teclado físico o virtual, e IGARRITZ ajustará sus predicciones en consecuencia. Para seleccionar una predicción, el usuario puede hacer clic en ella con el ratón o mantener la mirada sobre la palabra durante 0,6 segundos. A medida que transcurre el tiempo, la predicción seleccionada se coloreará, facilitando al usuario la confirmación de su elección. Además, IGARRITZ incluye un teclado virtual que el usuario puede mostrar u ocultar según su preferencia. Para evitar la selección accidental de una palabra no deseada mientras se leen las predicciones, IGARRITZ cuenta con botones de deshacer operaciones. También ofrece la opción de copiar el texto escrito para pegarlo en otro lugar, así como la opción de eliminarlo. En caso de borrar el texto accidentalmente, se puede rectificar deshaciendo la acción.

3.2. Corpus

El corpus se compone de dos subcorpus distintos. El primero ha sido reunido con el propósito de evaluar la herramienta IGARRITZ, mientras que el segundo se ha recopilado específicamente para evaluar el desempeño de la estudiante.

3.2.1. Corpus para la evaluación de IGARRITZ

La estudiante ha redactado cuatro textos utilizando la herramienta Grid3, y hemos registrado todo el proceso en vídeo. Luego, con la asistencia de estos vídeos, hemos compilado las palabras empleadas en dichos textos y las hemos trasladado fielmente a IGARRITZ, respetando plenamente el contexto en el que fueron utilizadas originalmente por la estudiante. Estos textos fueron elaborados durante el horario de apoyo, y a continuación detallaremos sus características principales:

Texto 1: Consiste en 19 palabras y 108 letras.

Texto 2: Contiene 34 palabras y 238 letras.

Texto 3: Consta de 43 palabras y 233 letras.

Texto 4: Posee 12 palabras y 90 letras.

Debido a la complejidad de reunir este tipo de corpus, cada texto presenta características particulares. El segundo texto es la continuación de una carta, por lo que únicamente incluye la segunda parte de la misma. Sin embargo, esta segunda parte se redactó teniendo en cuenta el contexto proporcionado por la primera parte, que se integró en ambos sistemas como contexto previo.

Por otro lado, los textos 3 y 4 forman parte de un mismo cuerpo de texto, aunque fueron escritos en días diferentes. Por lo tanto, al redactar el cuarto texto, se utilizó el tercero como contexto previo.

3.2.2. Corpus para la evaluación del desempeño de la estudiante en cuanto al uso IGARRITZ

La estudiante ha redactado otros cuatro textos utilizando IGARRITZ, esta vez durante el horario escolar, y nuevamente se han grabado en vídeo para analizar todo el proceso de escritura. Las características principales de los textos son las siguientes:

Texto 1: Consta de 74 palabras y 410 letras.

Texto 2: Posee 15 palabras y 90 letras.

Texto 3: Contiene 27 palabras y 180 letras.

Texto 4: Tiene 47 palabras y 263 letras.

En esta ocasión, en el texto 3, la estudiante inicia la redacción en IGARRITZ con un contexto previo, ya que había escrito previamente tres frases más. Por otro lado, en el cuarto texto, la estudiante cuenta con un enunciado que utiliza como contexto en IGARRITZ, lo cual influye en la calidad de las predicciones. Además, es importante mencionar que la alumna comienza a escribir con el enunciado en la parte superior, pero en un momento decide moverlo hacia abajo, lo que probablemente haya afectado las predicciones, aunque no de manera significativa.

3.3. Evaluación de Igarritz

En esta evaluación, analizamos el ahorro de letras proporcionado por ambos predictores, sin considerar las pulsaciones de teclado. Esto se debe a que nuestro objetivo principal en esta primera evaluación es evaluar la calidad de las predicciones. En la siguiente evaluación, que evaluará el desempeño de la alumna, nos centraremos en las pulsaciones de teclado.

Para evaluar IGARRITZ, hemos empleado el subcorpus descrito en el subapartado 3.2.1. Los cuatro textos que conforman este subcorpus han sido escritos por la alumna utilizando las predicciones de GRID3. En primer lugar, analizamos las predicciones realizadas por esta herramienta. Luego, con la asistencia de los vídeos, hemos recopilado las palabras utilizadas en estos textos y las hemos trasladado fielmente a IGARRITZ, manteniendo completamente su contexto. En algunos casos, la alumna ha escrito una palabra completa y luego la ha eliminado; en estos casos, la palabra se ha contabilizado. Sin embargo, no hemos incluido letras individuales o cadenas de letras que no forman una palabra, ya que no es posible predecirlas. Con base en estas dos normas, la evaluación se ha realizado mediante el análisis de dos tipos de palabras:

Palabras que el predictor prevé en su totalidad: cuando la palabra escrita por la estudiante ha sido exactamente predicha por el predictor, hemos recogido en qué letra lo propone y cuántas letras permite ahorrar. Por ejemplo, cuando hemos querido escribir la palabra "irakasleak", IGARRITZ nos ha dado el resultado deseado después de haber escrito "ir", por lo que en este caso hemos etiquetado que hemos tenido que escribir 2 letras y que se pueden ahorrar 8.

Las palabras que necesitan ser editadas para formar la palabra deseada: Aunque el predictor no sugiere la palabra deseada, en ocasiones las predicciones se aproximan mucho a la palabra que se intenta escribir, y mediante una edición mínima en la palabra sugerida, podemos llegar a obtener la palabra deseada. En estos casos, registramos el número de letras que se pueden ahorrar gracias a las ediciones realizadas. Es posible que más de una predicción realizada por el predictor nos permita hacer ediciones, y en esos casos, seleccionamos la opción que permite ahorrar la mayor cantidad de letras. Por ejemplo, IGARRITZ no predice la palabra "oinarritzeko", pero al escribir "o", entre las seis predicciones aparecen "oinarrituta", "oinarriturik" y "oinarrituz". En este caso, seleccionamos "oinarrituz" para eliminar "uz" y luego añadir "zeko". De esta manera, de una palabra de 12 letras, se escribió "o" y "zeko" (5 letras). Como para borrar "uz" se requieren dos clicks, el ahorro total es de 5 letras. Las otras dos predicciones no se seleccionaron porque requerirían eliminar una letra adicional en ambos casos.

La evaluación se ha llevado a cabo de dos maneras: i) a nivel general y ii) a nivel de texto. Además, hemos explorado la variable de la cantidad de letras en las palabras y hemos analizado las predicciones de IGARRITZ en los cuatro textos. Para esto, hemos clasificado las palabras en cuatro grupos: i) 0-3 letras, ii) 4-6 letras, iii) 7-9 letras y iv) 10 o más letras.

Con todos estos datos, podemos analizar: i) la eficacia en el ahorro de letras de los sistemas IGARRITZ y GRID3. ii) El rendimiento del ahorro de letras de IGARRITZ en diferentes tipos de palabras según su longitud. iii) La cantidad de letras que podríamos ahorrar en cada tipo de palabras mediante las predicciones disponibles sin necesidad de escribir ninguna letra. iv) El porcentaje de la palabra que debe estar escrita en cada tipo para que IGARRITZ realice una predicción precisa.

3.4. Evaluación del uso realizado por la alumna

La estudiante ha redactado otros 4 textos utilizando IGARRITZ, los cuales componen el subcorpus 3.2.2. En primer lugar, calculamos el número de pulsaciones necesarias para escribir cada texto sin predictor alguno, evitando cometer errores, teniendo en cuenta las letras, los espacios, la puntuación y las pulsaciones especiales (como la activación de mayúsculas). Luego, calculamos el número de pulsaciones de teclado que se requerirían utilizando óptimamente IGARRITZ, considerando también las ediciones realizadas a las predicciones para escribir el texto deseado. Finalmente,

registramos el número total de pulsaciones realizadas por la alumna. Para ello, recopilamos las pulsaciones de teclado (letras y espacios) que se ahorraron gracias a las predicciones, así como las pulsaciones derivadas de los errores cometidos al seleccionar una predicción no deseada o al pulsar una letra equivocada.

4. RESULTADOS

En esta sección de resultados, examinaremos los resultados obtenidos en las dos evaluaciones llevadas a cabo.

4.1. Resultados de la evaluación de IGARRITZ

En relación a la evaluación manual de IGARRITZ, los resultados presentados en la Tabla 1 son bastante alentadores. En los cuatro textos analizados, IGARRITZ ha demostrado una calidad de predicciones significativamente superior a la proporcionada por GRID3. Al observar los datos generales de los textos, se destaca que con IGARRITZ se logra una reducción del 50,07% en las letras que podríamos omitir gracias a las predicciones precisas de palabras, en comparación con el modesto 18,09% de la herramienta GRID3. Aunque IGARRITZ ofrece menos oportunidades de ahorrar letras mediante la edición de las predicciones (7,7%), esta tendencia es normal. Al acertar más predicciones tal como las necesitábamos, es lógico que disminuya la necesidad de realizar ediciones.

Es relevante destacar que IGARRITZ ha mejorado consistentemente los resultados de GRID3 en todos los textos evaluados. No obstante, es importante tener en cuenta las diferencias entre los textos. Por ejemplo, en el segundo texto se observa un ahorro significativo del 62,61% (149 de 238 letras) gracias a las predicciones correctas, mientras que en el cuarto texto esta cifra disminuye considerablemente al 26,67% (24 de 90 letras). Es crucial considerar que el segundo texto es una continuación directa de una carta previamente escrita, la cual fue incorporada inicialmente en IGARRITZ. Esto sugiere que el contexto anterior puede influir positivamente en las predicciones. Sin embargo, el cuarto texto, aunque también es una continuación de un fragmento previo, arroja resultados menos satisfactorios. Por lo tanto, aunque el contexto parece ser un factor relevante, el contenido y el tema del texto también parecen incidir directamente en la precisión de las predicciones.

Tras examinar el potencial de ahorro de letras proporcionado por IGARRITZ, hemos investigado qué tipo de palabras reciben las mejores predicciones, como se muestra en la Tabla 1. Al analizar la cantidad de letras, se observa que las palabras de 1 a 3 letras son las que presentan menor número de caracteres (63 letras; 9,42% del total). Sin embargo, es en este grupo donde IGARRITZ demuestra su mayor eficacia predictiva, permitiendo un ahorro del 85,71% de las letras. En los tres grupos restantes, la distribución de letras es más equilibrada: un 34,38% (230) corresponden a palabras de 4 a 6 letras, un 25,26% (169) a palabras de 6 a 9

letras y un 30,94% (207) a palabras de más de 10 letras. En términos de potencial de ahorro, las palabras de 4 a 6 letras y de 7 a 9 letras muestran datos muy similares, con un 61,30% y un 63,31% respectivamente. Sin embargo, este porcentaje disminuye notablemente en el caso de las palabras de más de 10 letras, alcanzando un 36,71%.

Al analizar los resultados texto por texto, se evidencia una diferencia significativa. En el primer texto, las palabras de más de 10 letras representan 32 caracteres, de los cuales IGARRITZ permite ahorrar 13 (40,62%). En el segundo texto, los resultados son aún más prometedores, con un ahorro del 63% al permitir evitar 63 de las 100 letras correspondientes. Sin embargo, en los otros dos textos, los resultados son distintos; en ambos casos, aunque las palabras recopilan 23 y 52 letras respectivamente, no se logra ahorrar ninguna a menos que se edite la palabra. Por otra parte, se observa que a medida que aumenta el número de letras en el texto, IGARRITZ contribuye a disminuir las pulsaciones necesarias, manteniendo una alta tasa de predicciones correctas. No obstante, esta tendencia no es uniforme, ya que se encuentran textos con un número similar de letras que muestran variaciones en la reducción de las pulsaciones requeridas. Esta discrepancia en los resultados sugiere que no siempre se cumple una proporcionalidad directa. Además, de un texto a otro, encontramos una variación significativa entre palabras del mismo tipo, lo cual reafirma la influencia del contexto previo y del contenido o tema del texto en la efectividad de las predicciones de IGARRITZ.

Por otra parte, dado que nuestro principal objetivo es el de minimizar la cantidad de letras que se debe seleccionar, hemos evaluado el desempeño de IGARRITZ en predecir palabras sin necesidad de escribir ninguna letra. Con palabras de 1 a 3 letras, hemos logrado los mejores resultados, ahorrando un 85,71% de las letras sin tener que ingresar ninguna. Sin embargo, los resultados varían en los otros tipos de palabras. En palabras de 4 a 6 letras, el ahorro se reduce significativamente al 21,73%, y en palabras de 7 a 9 letras, al 29,58%. Para palabras con más de 10 letras, el ahorro es aún menor, del 11,11%. Aunque los resultados son excelentes para palabras cortas, es esencial destacar que este grupo representa sólo el 9,42% del total de letras. Por tanto, el trabajo futuro debería ser el de mejorar los porcentajes en los grupos restantes.

En la misma línea, hemos examinado qué porcentaje de cada palabra debemos escribir en IGARRITZ para lograr una predicción precisa. Como antes, observamos que a medida que las palabras se alargan, el rendimiento disminuye. Para palabras de 1 a 3 letras, se ha tenido que escribir en promedio el 18,18% de la palabra para que IGARRITZ acierte. Sin embargo, este porcentaje aumenta considerablemente en palabras de 4 a 6 letras, alcanzando el 38,51%, y en palabras de 7 a 9 letras, el 37,42%; se duplica la cifra del conjunto de las palabras más cortas. Finalmente, en palabras de más de 10 letras, se ha tenido que escribir más de la mitad de la palabra en promedio, el 61,30%. Estos resultados sugieren que las mejoras futuras en IGARRITZ deberían concentrarse especialmente en palabras largas, ya que aquí reside el mayor margen de optimización.

TABLA 1. Evaluación de IGARRITZ: Comparación con GRID3 y evaluación basada en la tipología de las letras.

Evaluación de IGARRITZ y GRID3						
	Texto 1	Texto 2	Texto 3	Texto 4	Total	Ahorro
Letras del texto	108	238	233	90	669	-
Letras ahorradas GRID3: predicción correcta	20	45	56	0	121	18,09%
Letras ahorradas IGARRITZ: predicción correcta	55	149	107	24	335	50,07%
Letras ahorradas GRID3: con ediciones	16	54	22	4	93	13,9%
Letras ahorradas IGARRITZ: con ediciones	16	9	20	7	52	7,7%
Evaluación de IGARRITZ por tipología de palabras (sin edición de texto)						
Letras en palabras de 1-3 letras	12	15	32	4	63	-
Letras ahorradas en palabras de 1-3 letras	11	12	27	4	54	85,71%
Letras ahorradas sin tener que escribir letras (1-3)	11	9	24	3	47	74,6%
Porcentaje escrito para predicción correcta (1-3)	%8,33	%20	%15,15	%50	%18,18	-
Letras en palabras de 4-6 letras	57	70	84	19	230	-
Letras ahorradas en palabras de 4-6 letras	24	45	59	13	141	61,30%
Letras ahorradas sin tener que escribir letras (4-6)	6	11	24	9	50	21,73%
Porcentaje escrito para predicción correcta (4-6)	%58,48	35,76%	29,7%	30%	38,51%	-
Letras en palabras de 7-9 letras	7	53	94	15	169	-
Letras ahorradas en palabras de 7-9 letras	4	48	41	14	107	63,31%
Letras ahorradas sin tener que escribir letras (7-9)	0	31	7	8	50	29,58%
Porcentaje escrito para predicción correcta (7-9)	42,85%	9,94%	56,45%	7,14%	%37,42	-
Letras en palabras de 10 letras o más	32	100	23	52	207	-
Letras ahorradas en palabras de 10 letras o más	13	63	0	0	76	36,71%
Letras ahorradas sin tener que escribir letras (10≤)	0	23	0	0	23	11,11%
Porcentaje escrito para predicción correcta (10≤)	59,09%	37,12%	100%	100%	61,74%	-

Fuente: elaboración propia

4.2. Resultados de la evaluación del uso realizado por la alumna

La estudiante ha redactado cuatro textos adicionales utilizando IGARRITZ, escritos durante el horario escolar y sin asistencia de ningún docente. Primero, calculamos la cantidad de pulsaciones de teclado necesarias para escribir cada texto sin cometer errores. Luego, calculamos las pulsaciones necesarias en caso de utilizar IGARRITZ de manera óptima. Finalmente, con la ayuda de los videos que registran todo el proceso llevado a cabo por la alumna, contabilizamos todas las pulsaciones realizadas, incluyendo los errores al seleccionar predicciones incorrectas o letras no deseadas. Estos fueron los resultados para cada texto:

Texto 1: Escribiendo el texto se necesitaron 501 pulsaciones. Utilizando IGARRITZ de manera óptima, se podría reducir a 242 pulsaciones (48,5%). La alumna realizó 293 pulsaciones, de las cuales 8 fueron selecciones erróneas de predicciones y 16 fueron errores al pulsar letras.

Texto 2: Escribiendo el texto se necesitaron 109 pulsaciones. Utilizando IGARRITZ de manera óptima, se podría reducir a 58 pulsaciones (53,21%). La alumna realizó 71 pulsaciones, de las cuales 7 fueron selecciones erróneas de predicciones y 3 fueron errores al pulsar letras.

Texto 3: Escribiendo el texto se necesitaron 210 pulsaciones. Utilizando IGARRITZ de manera óptima, se podría reducir a 103 pulsaciones (49,04%). La alumna realizó 138 pulsaciones, de las cuales 18 fueron selecciones erróneas de predicciones y 11 fueron errores al pulsar letras.

Texto 4: Escribiendo el texto se necesitaron 317 pulsaciones. Utilizando IGARRITZ de manera óptima, se podría reducir a 141 pulsaciones (44,47%). La alumna realizó 209 pulsaciones, de las cuales 26 fueron selecciones erróneas de predicciones y 33 fueron errores al pulsar letras.

Teniendo en cuenta que la estudiante ha utilizado IGARRITZ para escribir 202 de las 216 palabras (93,5%), podemos concluir que lo ha empleado de manera constante. Además, lo ha hecho con gran eficacia, ya que los resultados demuestran que ha intentado ahorrar en todos los textos más del 90% de las pulsaciones posibles con IGARRITZ. Sin embargo, no ha logrado ahorrar tantas pulsaciones como pretendía, ya que ha cometido algunos errores: 112 pulsaciones en total, lo que representa el 8,96% de sus pulsaciones. De estas, 59 pulsaciones (4,72% del total) fueron debidas a la selección de predicciones no deseadas. Aunque los resultados son muy satisfactorios, indican que mejorando la colocación de las predicciones en IGARRITZ o ajustando el tiempo de fijación en las predicciones, se podría mejorar aún más el desempeño de la alumna.

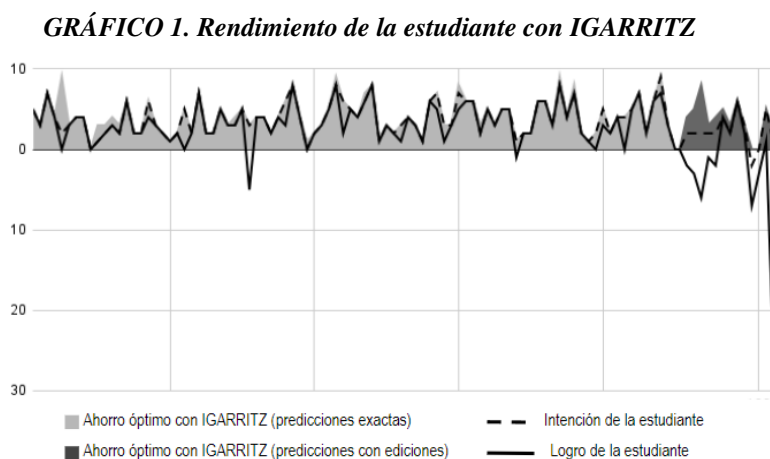
Dado que estos errores afectan significativamente la producción y el cansancio de la estudiante, hemos decidido analizarlos más a fondo. Para ello, hemos examinado cómo ha interactuado con cada tipo de predicción. Primero, hemos analizado detalladamente

cómo se ha desarrollado con las predicciones que coincidían exactamente con lo que necesitaba. Posteriormente, hemos evaluado su desempeño con las predicciones que requerían alguna edición.

Como se puede ver en el Gráfico 1, el uso de IGARRITZ minimiza la necesidad de realizar muchas ediciones en las predicciones, ya que el sistema anticipa correctamente la palabra requerida. Esto ha beneficiado a la estudiante, quien en la mayoría de los casos recibe la palabra adecuada de IGARRITZ. El gráfico muestra que cuando esto sucede, la alumna aprovecha la herramienta para ahorrar pulsaciones. Sin embargo, durante este proceso, comete algunos errores al presionar letras o predicciones innecesarias, lo que ralentiza su ritmo y puede aumentar la fatiga. A pesar de estos errores, en casi todas las operaciones ahorra pulsaciones.

Hay dos excepciones: en una, realizó una pulsación adicional y en otra, cinco pulsaciones más de las necesarias. No obstante, estos parecen ser casos aislados y no necesariamente causados por IGARRITZ, ya que los errores son comunes al escribir, especialmente con la mirada.

El gráfico revela que los desafíos más significativos para la alumna surgen al intentar editar una predicción. En ocasiones, ella opta por seleccionar una palabra similar a la que necesita con la intención de ahorrar pulsaciones. No obstante, este proceso se convierte en un obstáculo, especialmente porque requiere reubicar el cursor en una posición específica de la palabra. Es importante destacar que en uno de los casos, la alumna llega a realizar 26 pulsaciones adicionales para completar una palabra. Por ende, para facilitar este proceso, resulta crucial adaptar IGARRITZ en el futuro a esta necesidad y proporcionar a la estudiante estrategias adecuadas para editar las predicciones.



5. APORTE CIENTÍFICO Y CONCLUSIONES

En este estudio hemos evaluado IGARRITZ en un contexto real. IGARRITZ tiene dos características principales: i) hacer predicciones de palabras en euskera mediante inteligencia artificial y ii) permitir el uso del entorno web a través de la mirada. Siendo la primera herramienta de estas características, la evaluación que hemos realizado es crucial, ya que nos ha permitido observar su uso diario por parte de una estudiante que necesita esta herramienta. Los resultados han sido muy positivos. Hemos comprobado que las predicciones de IGARRITZ son significativamente mejores que las de GRID3 en euskera. Además, la estudiante ha utilizado las predicciones de IGARRITZ de manera constante y eficaz en la redacción de textos, lo cual es notable dado que acaba de empezar a utilizar esta herramienta.

De cara al futuro, la evaluación realizada nos ha permitido establecer una serie de objetivos. Algunos de estos objetivos deberán abordarse con vistas a mejorar las predicciones; por ejemplo, hemos notado que la estudiante ha tenido dificultades cuando ha intentado editar las predicciones, por lo que será importante adaptar el entorno web a sus necesidades o presentarle estrategias adecuadas. Por su parte, IGARRITZ ha mostrado muy buenos resultados al predecir palabras de 1 a 3 caracteres, pero tiene un margen de mejora a medida que las palabras se alargan. También prevemos realizar mejoras en la interfaz. Por ejemplo, hemos observado que a veces ha tenido dificultades para mover el cursor, por lo que habrá que hacerlo más accesible. Además podría ser muy beneficioso integrar TTS (Text-to-Speech) en IGARRITZ para que pueda utilizar como medio de comunicación oral en clase.

6. AGRADECIMIENTOS

Este trabajo se ha realizado gracias a una ayuda del Departamento de Educación del Gobierno Vasco/Eusko Jaurlaritzza y la colaboración tanto personal/familiar como escolar de la estudiante implicada.

IGARRITZ está disponible en la infraestructura de investigación CLARIAH-EUS financiada por la Diputación de Gipuzkoa y el Gobierno Vasco: <https://igarritz.clariah.eus/>.

REFERENCIAS

Artetxe, M., Aldabe, I., Agerri, R., Perez-de-Viñaspre, O., y Soroa, A. (2022). Does Corpus Quality Really Matter for Low-Resource Languages?. *arXiv preprint arXiv:2203.08111*. <http://dx.doi.org/10.18653/v1/2022.emnlp-main.499>

- Ben Khelil, C., Rayar, F., Antoine, J. Y., Hoiry, L., Raynal, M., y Halftermeyer, A. (2023, July). What you need is what you get: Adapting word prediction of Augmentative and Alternative Communication aids to youth language. In *International Conference on Human-Computer Interaction* (pp. 240-247). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35674-6_30
- Cavalieri, D. C., Palazuelos-Cagigas, S. E., Bastos-Filho, T. F., y Sarcinelli-Filho, M. (2016). Combination of language models for word prediction: An exponential approach. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(9), 1481-1494. <https://doi.org/10.1109/TASLP.2016.2547743>
- Garay-Vitoria, N., y Abascal, J. (2006). Text prediction systems: a survey. *Universal Access in the Information Society*, 4, 188-203.
- Ghayoomi, M., y Momtazi, S. (2009, octubre). An overview on the existing language models for prediction systems as writing assistant tools. En *2009 IEEE International Conference on Systems, Man and Cybernetics* (pp. 5083-5087). IEEE. <https://doi.org/10.1109/ICSMC.2009.534602>
- González, P. M. (2003). Sistemas alternativos y aumentativos de comunicación (SAAC) y accesibilidad: Bases teóricas de los SAAC. *Puertas a la lectura*, (4), 129-136.
- Google. (2018). BERT (Bidirectional Encoder Representations from Transformers) [Modelo de lenguaje]. Recuperado el 20 de junio de 2024, de <https://github.com/google-research/bert>
- Irisbond. (2024). *EasyClick* [Software]. Recuperado el 20 de junio de 2024, de <https://www.irisbond.com/en/aac-products/easyclick/>
- Klaus, B., Aigner, B., y Veigl, C. (2019). AsTeRICS Grid—a flexible web-based application for Alternative Communication (AAC), environmental and computer control. *Pervasive and Mobile Computing*, 55, 174-188.
- Lloyd, L. y Karlan, G. (1984). Non-speech communication symbols and systems: where have we been and where are we going? *Journal of mental deficiency research*, 28 (1), 3–20.
- Luna, M.R. (2013): Tecnología y discapacidad: Una mirada pedagógica. *RDU, Revista Digital Universitaria*, 14(12), 1–19.
- Palazuelos-Cagigas, S. E. (2001). Aportación a la predicción de palabras en castellano y su integración en sistemas de ayuda a personas con discapacidad física (Tesis doctoral, Universidad Politécnica de Madrid).
- Raynal, M., y Badr, G. (2022). Study of user behavior when using a list of predicted words. En *International Conference on Computers Helping People with Special Needs* (pp. 331–337). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-08648-9_38
- Sarasola, J., y Sarasola, E. (2023). Gizapedia, giza eta gizarte zientzien entziklopedia. *CLARIAH-EUS 2. workshop: azpiegitura eraikitzen*. Recuperado el 27 de junio de 2024, de <https://gizapedia.org/>
- Smartbox Assistive Technology. (n.d.). *Grid 3* [Software]. <https://thinksmartbox.com/product/grid-3/>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., y Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.