

# IGARRITZ: el predictor de palabras para el euskera basado en la inteligencia artificial y su evaluación en el entorno escolar

**Unai Atutxa**  
**Iker de la Iglesia**  
**Mikel Iruskieta**



# 1. INTRODUCCIÓN



Alumna de 4º de ESO con parálisis cerebral; comunicación escrita mediante la mirada; el euskera lengua materna y la principal de su sistema educativo

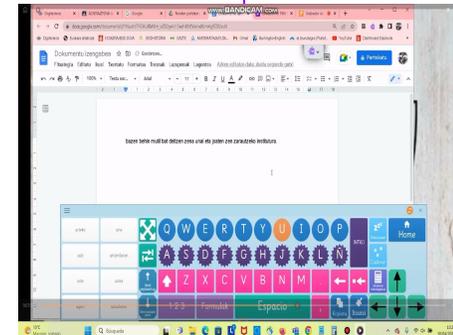


Eliminar o minimizar las barreras que limitan el aprendizaje y la participación de todo el alumnado. Luna (2013)

**Sistemas Aumentativos y Alternativos de Comunicación (SAAC)**  
aumentar la autonomía del alumnado y mejorar su calidad de vida (Gonzalez, 2003)

**Sistema de seguimiento ocular (eye-tracking)**  
Escribir con el iris es más lento y laborioso, dificulta la comunicación y el realizar trabajos

**Sistema actual: GRID3**  
Aumenta las posibilidades de interacción y comunicación de aquel que lo usa



**¿Qué podemos aportar?**  
**Predicción de palabras en euskera**  
Lista de 10000 palabras

## 2. OBJETIVOS

- ▶ **Evaluar IGARRITZ: un entorno web adaptado para predecir palabras en euskera:**
  - ▶ Empleando inteligencia artificial, utilizando modelos de lenguaje
  - ▶ Diseñado para su uso con programas de control mediante el seguimiento ocular
- ▶ **Evaluar su capacidad de ahorrar pulsaciones de teclado**
  - ▶ Comparándolo con el sistema empleado hasta el momento (GRID3)
  - ▶ Modelos de lenguaje VS Diccionarios
- ▶ **Evaluación en un contexto real: la primera toma de contacto de la alumna con IGARRITZ**

# 3. METODOLOGIA

- ▶ **Entrenamiento del modelo de lenguaje:**
  - ▶ modelo de lenguaje basado en HiTZ/roberta-euscrawl-base-cased (Artetxe et al. 2022), con la arquitectura transformer RoBERTa (Vaswani et al., 2017, Liu et al. 2019).
  - ▶ reentrenado con textos sobre educación, para ajustar las predicciones al lenguaje de una alumna de la ESO

Rol	Corpus	Documentos	Token	URL
Modelo de lenguaje	Euscrawl	12528k	423M	<a href="https://www.ix.eus/euscrawl/">https://www.ix.eus/euscrawl/</a>
Reentrenamiento para el ajuste	Gizapedia	5688	714k	<a href="https://gizapedia.org/">https://gizapedia.org/</a>
	Wikipedia	1962	3,3M	<a href="https://eu.wikipedia.org/wiki/Kategoria:Hezkuntza_Programa_ko_artikuluak">https://eu.wikipedia.org/wiki/Kategoria:Hezkuntza_Programa_ko_artikuluak</a>
	Berria	354	161k	<a href="https://www.berria.eus/ikasgela/gaiak/hezkuntza">https://www.berria.eus/ikasgela/gaiak/hezkuntza</a>
	Elkar	43	1,1M	-
	Redacciones ESO	1.469	194k	-

## ► Servicio web IGARRITZ:

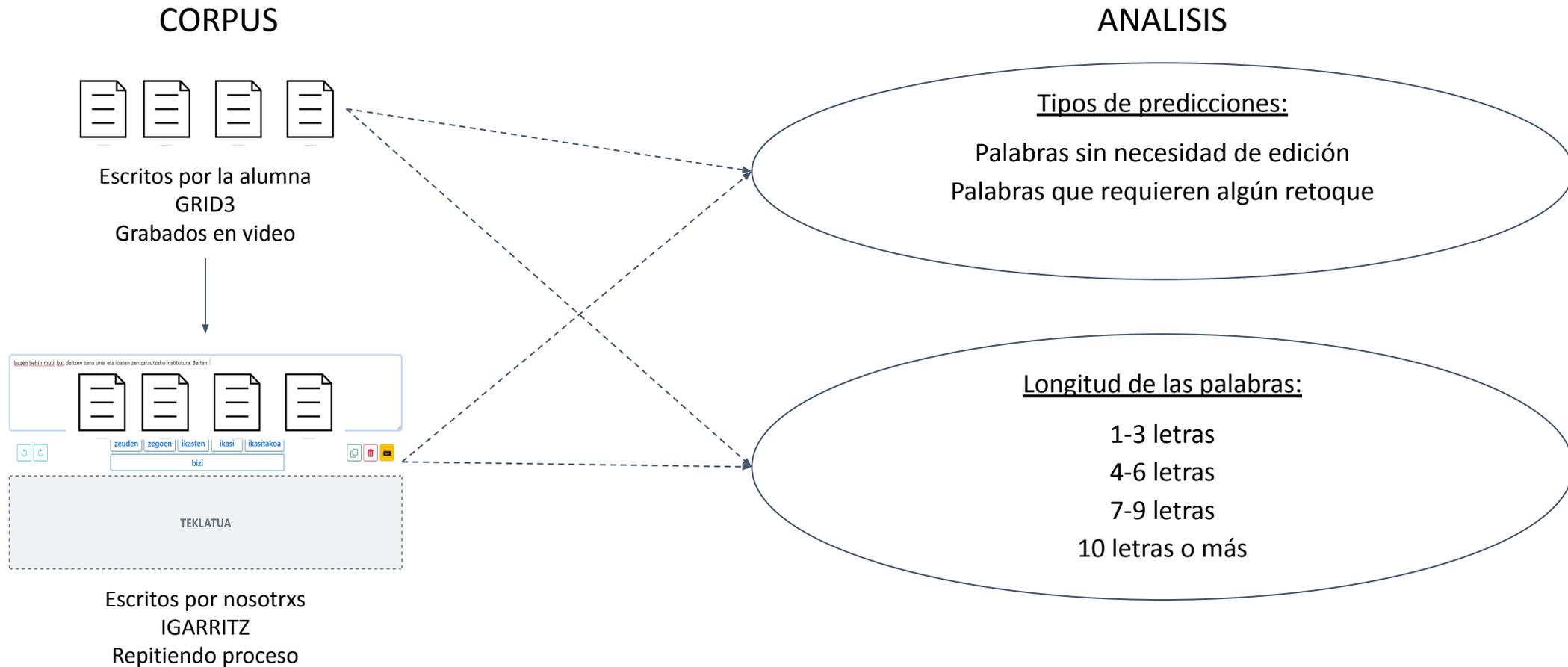
- disponible en la dirección <http://igarritz.clariah.eu>
- utilizando el explorador Firefox



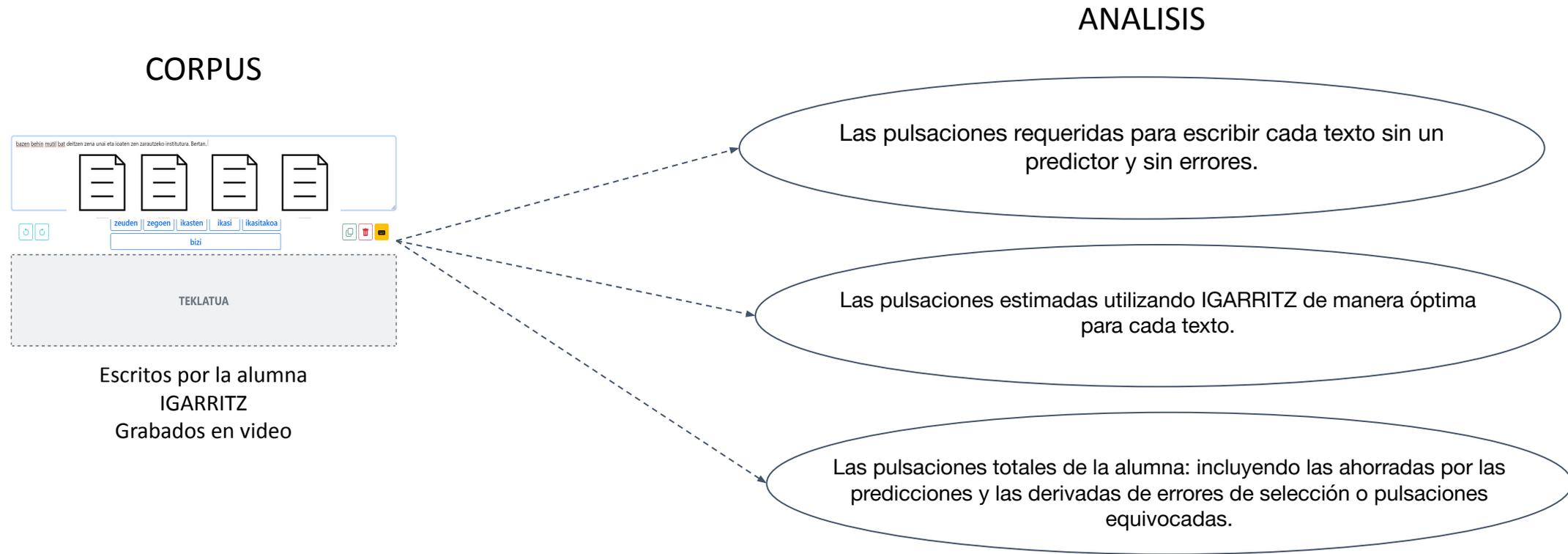
The image shows a screenshot of the IGARRITZ web service interface. The browser address bar displays [igarritz.clariah.eu](http://igarritz.clariah.eu). The main text area contains the sentence: "Ongi etorri geologi klasera gaur azalduko dizuet zelan sortzen ari den flyschak, zer da arroka, arroken motak eta zer den ikusten dugun laino beltz hori. Arroak min". A blue box highlights the text "Arroak min". Below the text area, there are several buttons: "Deshacer operaciones" (Undo operations), "Predicciones" (Predictions), and "Esconder teclado Copiar contenido Borrar contenido" (Hide keyboard Copy content Delete content). The prediction buttons are "mineral", "min", "mintzagai", "mintzo", and "minduta". A hand cursor is pointing at the "mineral" button. Below the prediction buttons, the word "mintza" is displayed. The interface is annotated with several callouts: "Texto escrito hasta el momento" (Text written up to now) pointing to the main text area; "Palabra que se está escribiendo" (Word being written) pointing to the highlighted text; "Deshacer operaciones" (Undo operations) pointing to the undo buttons; "Predicciones" (Predictions) pointing to the prediction buttons; "Esconder teclado Copiar contenido Borrar contenido" (Hide keyboard Copy content Delete content) pointing to the keyboard control buttons. A dashed box labeled "TEKLATUA" (KEYBOARD) is shown below the main interface, and a smaller inset window shows the same interface with the keyboard area highlighted.

## ► Evaluación de IGARRITZ:

- la capacidad de ahorrar letras de IGARRITZ VS GRID3



► Evaluación del uso de la alumna:

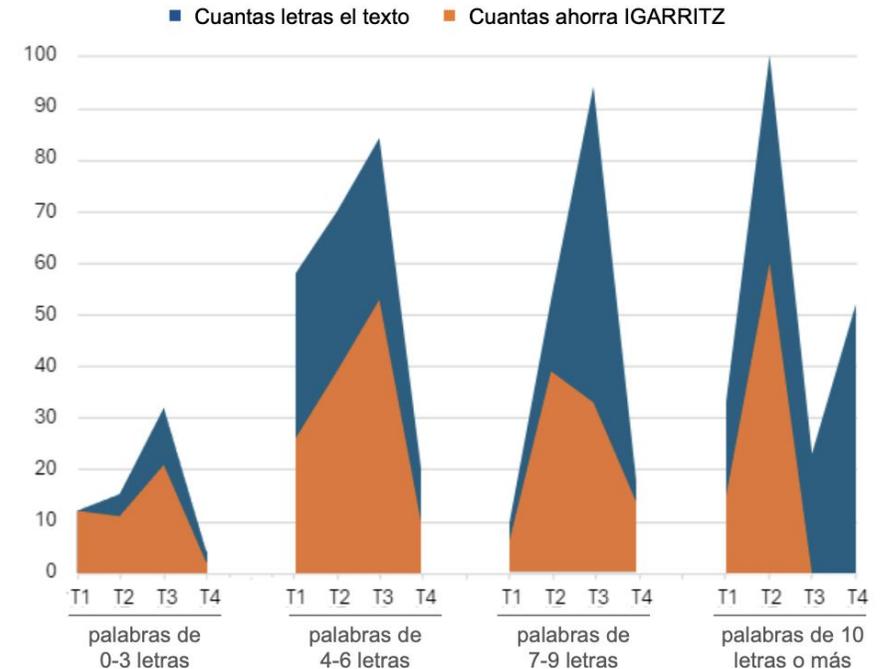


# 4. RESULTADOS

- ▶ IGARRITZ supera significativamente en todos los textos las predicciones realizadas por GRID3, aun adecuandolo con diccionarios
  - ▶ Diferencia sensible entre los textos: contexto previo y la influencia del tema y contenido del texto
  - ▶ Diferencia entre palabras de diferente longitud. Margen de mejora en palabras largas.

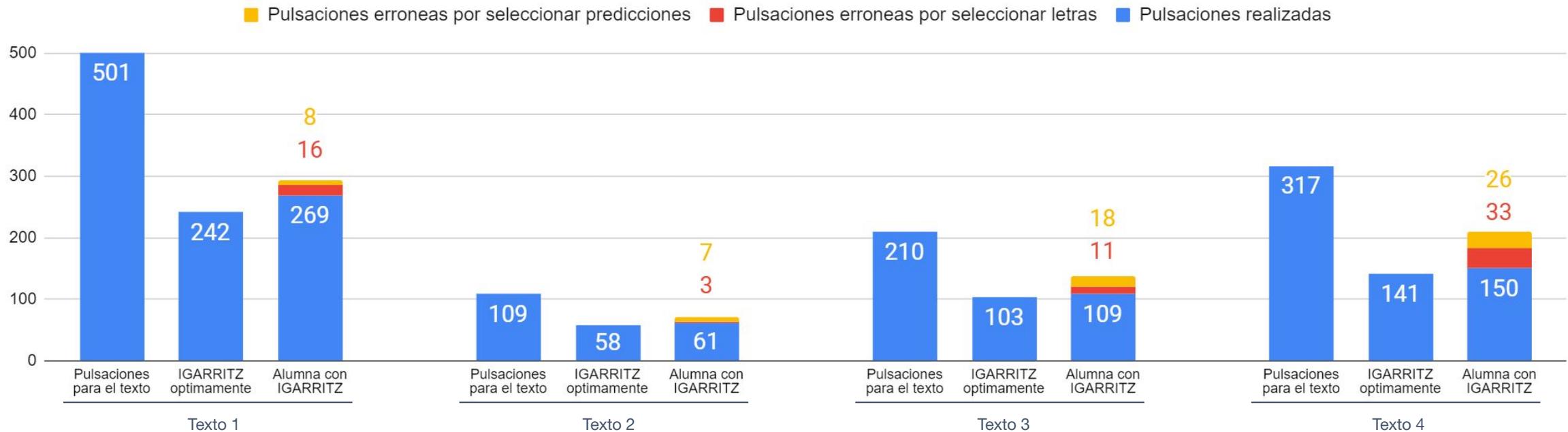
	Letras del texto	Letras ahorradas		Ahorradas con retoques	
		Original	Igarritz	Original	Igarritz
Texto 1	108	20	55	16	16
Texto 2	238	45	149	54	9
Texto 3	233	56	107	22	20
Texto 4	90	0	24	4	7
<b>Total</b>	<b>669</b>	<b>121</b>	<b>335</b>	<b>93</b>	<b>52</b>
<b>Ahorrado</b>		<b>18,09%</b>	<b>50,07%</b>	<b>13,90%</b>	<b>7,77%</b>

Evaluación de IGARRITZ por tipología de palabras (sin retoques)						
	Texto 1	Texto 2	Texto 3	Texto 4	Total	Ahorro
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 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 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 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%	-

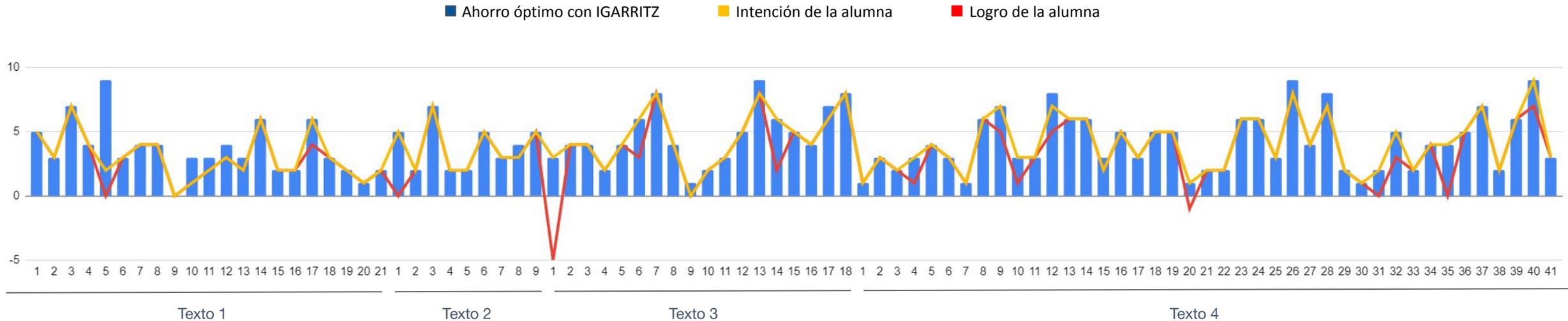


► La alumna ha utilizado IGARRITZ para escribir casi todas las palabras

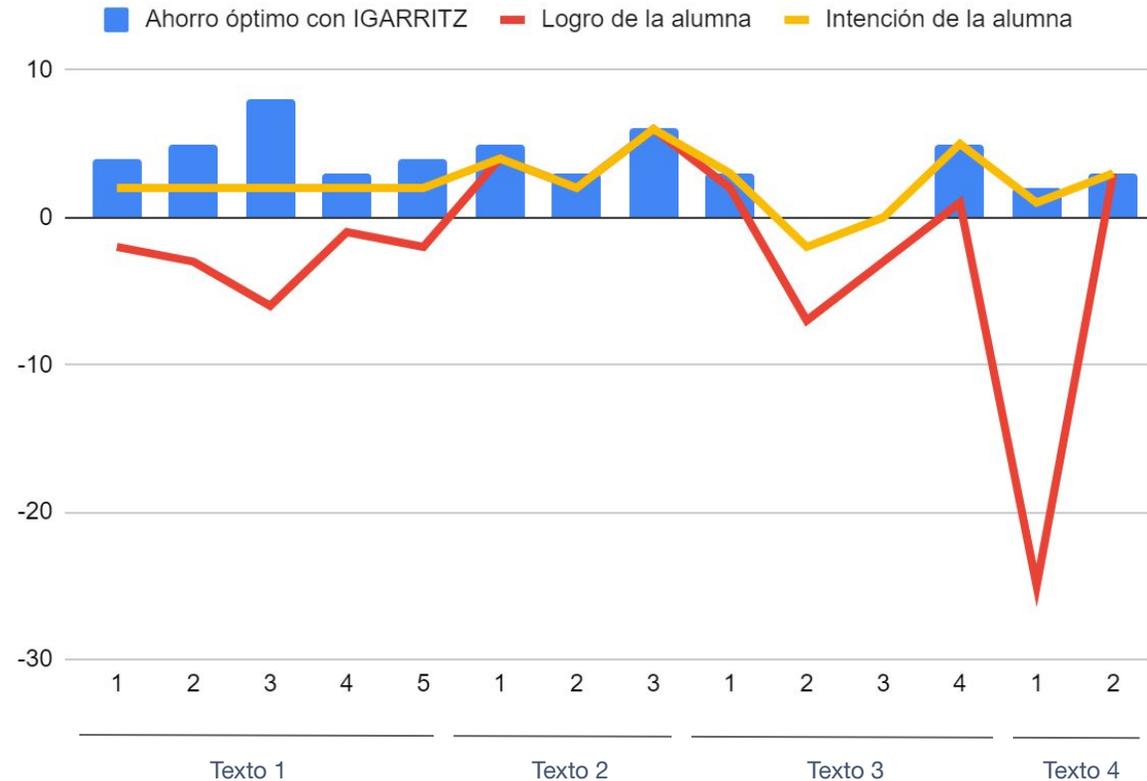
- En 202 de 216 (93,5%)
- De manera muy eficaz: ahorrando en todos los textos más del 90% de las pulsaciones que permite ahorrar IGARRITZ



- ▶ Emplea de manera muy eficaz cuando IGARRITZ propone la predicción que necesita



- ▶ Al intentar retocar las predicciones, comete errores frecuentemente



## 5. CONCLUSIONES

- ▶ IGARRITZ: primer entorno web adaptado para predecir palabras en euskera y diseñado para programas de seguimiento ocular, fácilmente accesible utilizando el explorador Firefox en <http://igarritz.clariah.eus>.
- ▶ IGARRITZ ofrece predicciones más relevantes de palabras en euskera que las herramientas empleadas hasta la fecha.
- ▶ La eficacia de la alumna al utilizar IGARRITZ demuestra su potencial en accesibilidad.
- ▶ Futuras investigaciones y mejoras:
  - ▶ dificultades observadas al corregir predicciones
  - ▶ mejorar las predicciones en las palabras de mayor longitud
  - ▶ controlar el movimiento ocular en la interfaz

# Bibliografía

Artetxe, M., Aldabe, I., Agerri, R., Perez-de-Viñaspre, O., & Soroa, A. (2022). Does Corpus Quality Really Matter for Low-Resource Languages?. *arXiv preprint arXiv:2203.08111*.

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.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Luna, M.R. (2013): Tecnología y discapacidad: Una mirada pedagógica. *RDU, Revista Digital Universitaria*, 14(12), 1-19, <http://www.ru.tic.unam.mx/handle/123456789/2177> [26-04-2024].

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

ESKERRIK ANITZ

**MILA ESKER**

*Esker anitz*

**ESKERRIK ASKO**

*Mil Esker*

*Esker mila*

Con la ayuda del



EUSKO JAURLARITZA  
GOBIERNO VASCO

HEZKUNTZA SAILA

DEPARTAMENTO DE EDUCACIÓN

CinteCongress.eus\_2024