

Traductor español-francés para jóvenes de 12 a 15 años.

Por Mayra Alejandra Sanchez Roblez y Víctor Gaytán Quevedo.



Introducción

*"Cuando hablas a una persona en un idioma que entiende, le hablas a su cabeza.
Cuando le hablas en su propio idioma, le hablas entonces a su corazón."*
– Nelson Mandela –.

Un idioma es una herramienta bidireccional que facilita la inmersión en una cultura y, a su vez, permite compartir la identidad y la herencia cultural propias. Esta es la motivación principal detrás del presente proyecto, que jóvenes estudiantes tengan la oportunidad de acercarse a dos culturas que, a pesar de su distancia geográfica, tienen una rica historia en común.

Este proyecto se centra en el desarrollo de un traductor francés-español dirigido a estudiantes de 12 a 15 años. Una motivación adicional fue la observación de que muchas de las herramientas de traducción disponibles están orientadas casi exclusivamente hacia el inglés y generalmente implicando un costo, limitando así su accesibilidad.

El impacto de la Inteligencia Artificial en la humanidad ha sido profundo en áreas como el análisis de datos, de imágenes, de audio y de video. Su potencial de aplicación es tan vasto, limitado únicamente por la imaginación humana. En particular, en el campo de la traducción, la IA ha tenido avances importantes como el desarrollo de arquitecturas de codificación y decodificación de frases, así como el desarrollo del algoritmo de **la Autoatención**, que es el que utiliza el modelo que elegimos.

Desarrollo

Inicialmente se había considerado la adaptación del traductor inglés-español disponible en Keras Hub. Desafortunadamente, surgieron dificultades especialmente al momento de entrenar el traductor con los datasets disponibles en <https://www.manythings.org/anki/>. Todos los datasets están disponibles en inglés y no se encontró ninguno que conectara directamente los dos idiomas de interés: el español y el francés. Se intentó incluso fusionar los datasets inglés-español e inglés-francés para construir uno español-francés, pero en el proceso se perdía información debido a la disparidad en el número de frases entre los datasets.

Tras la investigación de alternativas, se identificó el modelo **NLLB-200-distilled-600M** desarrollado por Meta IA como una opción prometedora. Cabe mencionar que este modelo ya ha sido entrenado con una gran cantidad de textos provenientes de fuentes como *Common Crawl* y *recopilación de información de distintas redes sociales de Meta*, abarcando 200 idiomas distintos.

En la literatura disponible en la red se encontró que es un modelo con un buen desempeño, ya que procesa secuencias de entrada de hasta 512 tokens de manera rápida y eficiente, aunque su desempeño disminuye para secuencias de traducción más largas arrojando traducciones de menor calidad.

El desarrollo de una interfaz de usuario intuitiva fue otro aspecto clave en el desarrollo del proyecto. Se exploraron diversas opciones hasta determinar que la biblioteca **Gradio** de Python ofrecía la mejor solución para diseñar una interfaz amigable y accesible. Desafortunadamente, a pesar de la facilidad de programación para crear la interfaz, al

querer mejorarla utilizando una plataforma de pago, Hugging Face, para tener una GPU más poderosa, se presentaron desafíos inesperados, ya que, la GPU de pago básica utiliza un método de funcionamiento donde enciende y apaga la GPU cada vez que se ejecutaba el traductor, por lo que, tomaba tiempos muy largos en arrojar el resultado.

Así que, la idea quedó desestimada, y seguimos usando Google Colab.

Resultados y conclusiones

Se invitó a un grupo de 30 estudiantes de 12 años con un nivel A1 de francés para que probaran el traductor. La retroalimentación fue mayormente positiva, resaltando la facilidad de uso, la paleta de colores y el diseño de la interfaz. No obstante, después de varias pruebas, todos expresaron que las traducciones parecían muy robotizadas, que carecían de naturalidad y fluidez.

En conclusión, el desarrollo de este traductor representó una valiosa experiencia en el campo de la traducción automática neuronal. El proceso inicial de adaptar un modelo preexistente, seguido de explorar y adoptar alternativas más apropiadas como el modelo **NLLB-200-distilled-600M** proporcionó importantes lecciones sobre los desafíos y las consideraciones prácticas involucradas en este tipo de proyectos. Si bien el traductor logró cumplir con su objetivo de proporcionar una herramienta accesible para la traducción de frases cortas entre español y francés, los resultados subrayan la complejidad de generar traducciones que capturen plenamente los matices del lenguaje natural.

Sitografía

- [Englis-to-Spanish translation with KerasHub.](#)
- [NLLB-200-distilled-600M - Multilingual translator.](#)
- [200 languages within a single AI model.](#)