

## Introducción al Corpus Global Interlingüístico Chino

Lili Wang<sup>1</sup> y Like Fan<sup>2</sup>

<sup>1</sup>Universidad de Las Palmas de Gran Canaria, España.

<sup>2</sup>Shanghai International Studies University, China.

### Resumen

El Corpus Global Interlingüístico Chino fue creado por el Ministerio de Educación de China en 2019. Tiene una gran importancia para la investigación sobre la enseñanza del chino como lengua extranjera. Por ello, los objetivos establecidos para este trabajo son los siguientes:

1. Presentar al Corpus interlingüístico global del chino;
2. Analizar el contenido y las características del Corpus interlingüístico global del chino;

Para lograr los objetivos de la investigación, los métodos que utilizamos son principalmente la búsqueda de plataforma y el análisis cuantitativo y cualitativo.

A través de nuestra investigación, encontramos que la escala de diseño total de este corpus es de 50 millones de palabras. A noviembre de 2021, el corpus incluye más de 23,67 millones de palabras de datos originales de 139 países, y datos anotados de varios niveles que suman aproximadamente 126 millones de palabras. El contenido de la anotación del corpus incluye 10 niveles: caracteres chinos, palabras chinas, frases, oraciones y componentes de oraciones, discurso, estilo, retórica, puntuación, anotación fonética de corpus hablado y de vídeo y anotación del lenguaje corporal de corpus de vídeo.

Tomando España como ejemplo, el corpus contiene un total de 139 composiciones chinas escritas por estudiantes españoles. El número total de palabras chinas ocupa el puesto 31 entre los 139 países.

En resumen, el Corpus Interlingüístico Global Chino es un corpus interlingüístico con un contenido muy rico. Pero, al mismo tiempo, también hemos encontrado algunos problemas con este corpus. Por ejemplo, la recopilación de este corpus no está equilibrada: la cantidad total de corpus de los países asiáticos es significativamente mayor que la de otras regiones, y la velocidad y la cantidad de recopilación de corpus han disminuido significativamente en los últimos años. Finalmente, esperamos que nuestro trabajo pueda aportar algunas sugerencias útiles para la construcción de este corpus.