

Indigenous Languages Spoken in Argentina: A Survey of NLP and Speech Resources

Belu Ticona^{1,2}, Fernando Carranza^{3,4}, and Viviana Cotik^{2,5}

¹ George Mason University, United States

mticona@dc.uba.ar,

² Departamento de Computación, Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires (UBA), Argentina

³ Departamento de Letras, FFyL, UBA, Argentina

⁴ Instituto de Filología y Literaturas Hispánicas “Dr. Amado Alonso”,
UBA, Argentina

⁵ Instituto de Investigación en Ciencias de la Computación (ICC),
CONICET-UBA, Argentina

Publicado en 31st International Conference on Computational Linguistics (COLING). Proceedings pp: 8449–8461.
<https://aclanthology.org/2025.coling-main.565/>

Resumen. Argentina tiene una diversa, aunque aún poco conocida, herencia de lenguas indígenas. La mayoría de estas están en riesgo de desaparecer, lo que resultaría en una pérdida significativa para el patrimonio mundial y el conocimiento cultural. Actualmente, no hay información unificada sobre la cantidad de hablantes ni de las herramientas computacionales para estas lenguas. En este trabajo, presentamos una sistematización de las lenguas indígenas habladas en Argentina, junto con los datos demográficos nacionales de la población indígena del país. Clasificamos las lenguas en siete familias: Mapuche, Tupí-Guaraní, Guaycurú, Quechua, Mataco-Mataguaya, Aymara y Chon. También proveemos un panorama introductorio de los recursos computacionales disponibles para estas lenguas, ya sea que estén especialmente desarrolladas para las variedades de Argentina o no.

Palabras clave: Argentina, lenguas indígenas, PLN, habla

Abstract. Argentina has a diverse, yet little-known, Indigenous language heritage. Most of these languages are at risk of disappearing, resulting in a significant loss of world heritage and cultural knowledge. Currently, no unified information on speakers and computational tools is available for these languages. In this work, we present a systematization of the Indigenous languages spoken in Argentina, along with national demographic data on the country’s Indigenous population. The languages are classified into seven families: Mapuche, Tupí-Guaraní, Guaycurú, Quechua, Mataco-Mataguaya, Aymara, and Chon. We also provide an introductory survey of the computational resources available for these languages, whether or not they are specifically developed for Argentine varieties.

Keywords: Argentina, Indigenous languages, NLP, speech