# Proyecto de grado

 $\bullet \bullet \bullet$ 

2023

### Guaraní



Aunque se habla muchísimo no se usa tanto **formalmente\*** en las redes y la web (difícil recolectar datos)

Guaraní :: Escasos recursos

**Datos** 

"Crawlers" web

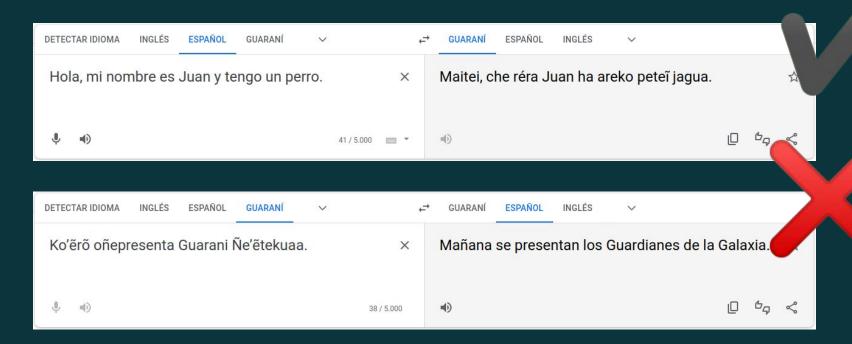


Identificadores de lengua

Extracción de información

### **Guaraní** :: Escasos recursos

# Sin embargo:



## Guaraní :: En qué hemos trabajado

### Development of a Guarani - Spanish Parallel Corpus

### Luis Chiruzzo<sup>1</sup>, Pedro Amarilla<sup>2</sup>, Adolfo Ríos<sup>2</sup>, Gustavo Giménez Lugo<sup>3</sup>

<sup>1</sup>Universidad de la República, Montevideo, Uruguay
<sup>2</sup>Universidad Nacional de Asunción, San Lorenzo, Paraguay
<sup>3</sup>Universidade Tecnológica Federal do Paraná, Curitiba, PR - Brasil
luischir@fing.edu.uy, pj.amarilla@gmail.com, adolforinu@gmail.com, gustavo@dainf.ct.utfpr.edu.br

### Jojajovai: A Parallel Guarani-Spanish Corpus for MT Benchmarking

Luis Chiruzzo<sup>1</sup>, Santiago Góngora<sup>1</sup>, Aldo Alvarez<sup>2</sup>, Gustavo Giménez-Lugo<sup>3</sup>, Marvin Agüero-Torales<sup>4,5</sup>, Yliana Rodríguez<sup>1</sup>

<sup>1</sup>Universidad de la República, Montevideo, Uruguay
 <sup>2</sup>Universidad Nacional de Itapúa, Encarnación, Paraguay
 <sup>3</sup>Universidade Tecnológica Federal do Paraná, Curitiba, PR - Brasil
 <sup>4</sup>Universidad de Granada, Granada, Spain

<sup>5</sup>Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya, Barcelona, Spain luischir@fing.edu.uy, sgongora@fing.edu.uy, aldo.alvarez@fiuni.edu.py, qustavoql@utfpr.edu.br, maquero@correo.ugr.es, yrodriquez@fhuce.edu.uy

### Using Guarani Verbal Morphology on Guarani-Spanish Machine Translation Experiments

Uso de la Morfología Verbal del Guaraní en Experimentos de Traducción Automática Guaraní-Español

> Yanina Borges, Florencia Mercant, Luis Chiruzzo Universidad de la República Montevideo, Uruguay {yanina.borges.mijailovich, florencia.mercant, luischir}@fing.edu.uy

### Experiments on a Guarani Corpus of News and Social Media

#### Santiago Góngora, Nicolás Giossa, Luis Chiruzzo

Facultad de Ingeniería Universidad de la República Montevideo, Uruguay

{sgongora, nicolas.giossa, luischir}@fing.edu.uy

#### Can We Use Word Embeddings for Enhancing Guarani-Spanish Machine Translation?

#### Santiago Góngora, Nicolás Giossa, Luis Chiruzzo

Facultad de Ingeniería Universidad de la República Montevideo, Uruguay

{sgongora, nicolas.giossa, luischir}@fing.edu.uy

### Initial Experiments for Building a Guarani WordNet

### Luis Chiruzzo<sup>1</sup> Marvin M. Agüero-Torales<sup>2,3</sup> Aldo Alvarez<sup>4</sup> Yliana Rodríguez<sup>1</sup>

<sup>1</sup>Universidad de la República, Montevideo, Uruguay
 <sup>2</sup>University of Granada, Granada, Spain
 <sup>3</sup>Global CoE of Data Intelligence, Fujitsu, Madrid, Spain
 <sup>4</sup>Universidad Nacional de Itapúa, Encarnación, Paraguay

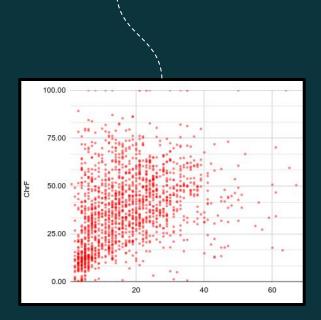
luischir@fing.edu.uy,maguero@correo.ugr.es,aldo.alvarez@fiuni.edu.py,yrodriguez@fhuce.edu.uy

## Guaraní :: Posibles líneas de trabajo

Trabajar a fondo en un sistema de recolección de datos de la web

Diseñar una gramática que genere ejemplos sencillos de oraciones (y usarlas para entrenar) Evaluar word embeddings nuevamente (más datos y otros algoritmos)





Otras ideas (más lingüísticas)
POS tagging
NER
Parsing