

ODESIA: Space for Observing the Development of Spanish in Artificial Intelligence

Enrique Amigó¹, Jorge Carrillo-de-Albornoz^{1,2}, Andrés Fernández¹, Julio Gonzalo¹, Miguel Lucas³, Guillermo Marco¹, Roser Morante¹, Jacobo Pedrosa¹, Laura Plaza^{1,2}, Ibo Sanz³ and Beñat San-Sebastián³

¹Universidad Nacional de Educación a Distancia, Calle Juan del Rosal, 16, 28040 Madrid, Spain

²RMIT University, 124 La Trobe St, Melbourne VIC 3000, Australia

³Llorente y Cuenca Madrid, S.L., Calle Lagasca 88, 28001 Madrid, Spain

Abstract

In this paper we present the ODESIA project (Space for Observing the Development of Spanish in Artificial Intelligence). ODESIA is a research collaborative project between the UNED University and the public institution RED.es, financed by the UE through the NextGenerationEU funds. The main objective of the project is the development of an annual index that quantitatively and qualitatively measures the gap between the language technologies in Spanish and English in terms of the state of the art, market solutions, the level of adoption of technologies and user experience. Other goals of the project are the production of Natural Language Processing resources in Spanish, such as datasets, a benchmark for evaluating the performance of language models, a website that provides information related to the progress of NLP in Spanish, and a tool for facilitating the evaluation of NLP systems.

Keywords

ODESIA, NLP gap, market solutions, adoption level, benchmarks, evaluation benchmark, NLP portal

1. Introduction and Motivation

ODESIA (Espacio de Observación del Desarrollo del Español en la Inteligencia Artificial) is a collaborative research project between UNED University and the public institution RED.es. It is funded by the European Union through the NextGenerationEU funds. The project consortium includes several partners: the National Observatory of Technology and Society (ONTSI),¹ the NLP & IR UNED research group at UNED,² and the company Llorente & Cuenca.³

Currently, Artificial Intelligence (AI) plays a significant role in all economic sectors and social activities, particularly in English-speaking countries. The Spanish language is crucial in making Spain a significant player in AI. In turn, AI is essential for defending the competitive advantage that Spain has due to its language. The

National Artificial Intelligence Strategy (ENIA) has identified seven objectives, one of which is to position the Spanish language as a leader in the development of tools, technologies, and applications for the use of AI in various fields, both in Spain and globally.

In this context, it is essential to establish an observation space that can analyze and measure the distance between the level of development of the AI in Spanish and the level of development of the AI in English, which is the dominant digital language. Specifically, the project will focus on Natural Language Processing (NLP). ODESIA will provide the ENIA with the necessary infrastructure to measure the gap in NLP and monitor the evolution of the gap throughout the execution of the plan. This project can be a vital contribution to help Spanish become one of the leading languages in the deployment of AI and can contribute to develop a more diverse, plural, and ethical AI.

Additional information about the ODESIA project is available in the project website at: <http://odesia.uned.es>.

2. Goals

The primary goal of ODESIA is to create resources that allow to analyze and measure the gap and level of development between the AI in Spanish and the AI in English. The project aims to develop an index that can quantitatively and qualitatively establish the difference between the development of language technologies in Spanish

SEPLN-PD 2023: Annual Conference of the Spanish Association for Natural Language Processing 2023: Projects and System Demonstrations

✉ enrique@lsi.uned.es (E. Amigó); jcalbornoz@lsi.uned.es (J. Carrillo-de-Albornoz); afernandez@lsi.uned.es (A. Fernández); julio@lsi.uned.es (J. Gonzalo); mlucas@llorenteycuenca.com (M. Lucas); gmarco@lsi.uned.es (G. Marco); r.morant@lsi.uned.es (R. Morante); jacobopedrosa@lsi.uned.es (J. Pedrosa); lplaza@lsi.uned.es (L. Plaza); isanz@llorenteycuenca.com (I. Sanz); benat.sansebastian@llorenteycuenca.com (B. San-Sebastián)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

¹<https://www.ontsi.es>

²<https://sites.google.com/view/nlp-uned>

³<https://www.llorenteycuenca.com/>

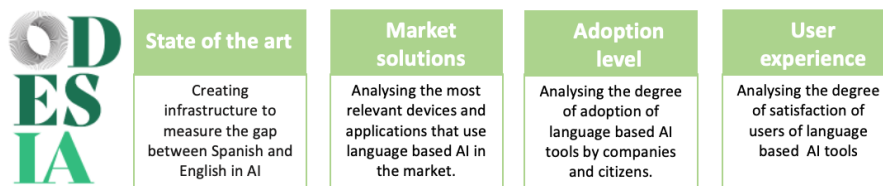


Figure 1: Project goals.

and English. This index will consider the state of the art, market solutions, adoption of technologies, and user experience as its key factors. The creation and periodic update of this index should contribute to raise awareness among major decision-makers and citizens about the relevance of promoting the development of language technologies in Spanish, which is crucial for Spain to become a prominent player in the AI landscape. Figure 1 summarizes the main goals of the ODESIA project.

ODESIA also pursues the following specific objectives:

1. Developing a leaderboard that allows for comparison between different pre-trained language models in Spanish and English for various NLP tasks. The information provided by this leaderboard can be utilized by policymakers, researchers, and the industry to monitor and comprehend the progress of language technologies in Spanish.
2. Developing bilingual datasets and evaluation methodologies to compare pre-trained language models in Spanish and English. This will also enable the comparison of NLP systems for tasks with the top practical applications.
3. Creating a website that contains information about and allows to track the progress of the state of the art in NLP for the Spanish language.
4. Developing a web application for evaluation that allows to perform a comprehensive and comparable evaluation of NLP systems for different tasks and evaluation contexts.

3. Results

3.1. Language Gap Index

The main output of ODESIA is an index that measures the gap in the development and adoption of language technologies between English and Spanish. This index is updated annually, providing a dynamic assessment of the progress made in language technologies in Spanish compared to English. The index is composed of four main areas: (1) state of the art, (2) availability and functionality of NLP market solutions, (3) adoption of NLP market

solutions, and (4) user experience and satisfaction. By computing the gap in these areas, it will be possible to quantify the differences between the two languages and monitor progress over time.

3.1.1. Gap in the State of the Art

The goal is to evaluate the state of language technologies in Spanish and English by examining the dissemination of research results in scientific media, the availability of resources for NLP research and their effectiveness.

To assess the gap in the state of the art, three indexes are used: The **dissemination index** measures the gap in the publication of scientific works in top NLP conferences as well as in funded NLP projects in both languages. The **resource index** measures the difference in the availability of pre-trained language models, annotated data, and language processing tools in English and Spanish. The **effectiveness index** quantifies the gap in performance between models and systems in both languages, compared to baseline systems that do not use any linguistic knowledge.

3.1.2. Gap in the Availability and Functionality of NLP market Solutions

The gap is calculated by identifying strategic families of NLP applications (e.g., chat bots, reputation/sentiment analysis) and selecting NLP products and services within such applications, both in English and Spanish. The map of functionalities associated with each family of applications and the coverage of existing products in both languages will be studied. The **functionality index** quantifies the language gap in terms of the coverage of the different functionalities offered by the products found in each application family.

3.1.3. Gap in the Adoption of NLP market Solutions

The objective is to quantify the level of adoption and value offered by language technologies in the market, providing added value to both companies and end users.

Two different indexes are calculated: The **adoption index** estimate the level of implementation/adoption of language technologies in Spanish/English by companies and by citizens, through the analysis of mentions in corporate reports and (social) media. The **impact index** estimates the reduction of costs or the increase in income for companies that are related to the implementation of NLP systems in any of their processes.

3.1.4. User experience and satisfaction

In this area, we aim to measure the user experience regarding NLP products and services in both English and Spanish, using semi-automatic analysis of the reputation of the products in social networks and user surveys. To estimate this gap, we compute two types of indexes: the reputational polarity index and the user satisfaction index.

The **reputational polarity index** estimates the overall polarity of mentions of different NLP services and products in both English and Spanish across various social networks, as well as the polarity of various attributes of the products/services. The **user satisfaction index** measures the degree of adoption of the products/services, both for personal (citizen adoption) and professional (business adoption) use, the level of user satisfaction, and the limitations encountered during use.

3.2. The ODESIA Leaderboard

The ODESIA leaderboard aims to provide the AI research community with a tool to evaluate and compare the performance of NLP models, in an easy and homogeneous manner.

The ODESIA leaderboard consists of two distinct benchmarks: one for Spanish and another for English, where identical tasks are proposed and models can be applied to compare their performance in both languages.

Currently, the ODESIA leaderboard is under development. It will include tasks such as named entity recognition, sexism detection and categorization, and propaganda detection and categorization. However, it will be frequently updated with new tasks, models, and results, making it a valuable resource for tracking progress in NLP research in both Spanish and English.

3.3. Datasets

The following datasets have been developed (completely or partially) during the first year of the ODESIA project, and are part of the ODESIA benchmark.

3.3.1. EXIST 2022

The EXIST (EXism Identification in Social Networks) 2022 dataset aims to facilitate research on automatic detection of sexism, containing texts from Twitter and Gab that are labeled based on whether they express or describe sexist attitudes or behaviors, and the type of sexism expressed or described. A detailed overview of the dataset is available at [1]. The development and training datasets were initially created for the EXIST 2021 evaluation campaign [2] held as part of IberLEF 2021. However, the test set has since been labeled as part of the ODESIA project and will be kept confidential as part of the ODESIA leaderboard.

3.3.2. DIPROMATS

The DIPROMATS dataset is designed to train and test models for identifying and characterizing propaganda techniques used by diplomats. It consists of labeled tweets, indicating the presence or absence of propaganda techniques, and if present, the type of propaganda used. The DIPROMAT dataset was used in the IberLef 2023 task on Automatic Detection and Characterization of Propaganda Techniques from Diplomats.⁴

3.3.3. DIANN

The DIANN (Disability Annotation) dataset was developed at UNED to be used in the DIANN task at IberLEF 2018.⁵ Its main objective is to annotate mentions of disabilities in scientific biomedical documents. A complete description of the dataset can be found in [3]. A new test set has been labeled as part of the ODESIA project and will be kept private as part of the ODESIA leaderboard.

3.4. EvALL 2.0

EvALL 2.0 (Evaluate ALL) is an evaluation tool designed for information systems, offering an extensive set of metrics that cover various evaluation contexts, including classification, ranking, and clustering, with or without disagreement. The tool is designed with three core concepts in mind: persistence, replicability, and effectiveness. Persistence enables users to store and retrieve past evaluations, while replicability ensures that all evaluations are carried out using the same methodology, making them strictly comparable. Effectiveness means that all metrics are based on measurement theory, and have been double-implemented and evaluated.

One of EvALL's key features is its universality. It offers a single format that allows for the evaluation of multiple metrics, even from different evaluation contexts.

⁴<https://sites.google.com/view/dipromats2023/home>

⁵<http://nlp.uned.es/diann/>

Additionally, EvALL 2.0 is easy to use, thanks to its user-friendly interface that enables users to generate personalized or guided evaluations with just a few clicks.

EvALL 2.0 provides a range of features to its users, allowing them to: (i) evaluate their information systems by providing the gold standard and their systems' predictions; (ii) access past evaluations to validate the effectiveness of new models; (iii) customize the selection of metrics and adjust their parameters; (iv) choose from a variety of input formats; (v) add new datasets to the EvALL 2.0 repository; and (vi) publish results related to datasets included in the repository, making them available to the wider scientific community.

3.5. ODESIA Portal

The objective of the ODESIA portal⁶ (see Figure 2) is to provide information on various NLP tasks in Spanish that have been tackled by the research community, including the results obtained for these tasks and the datasets available for training and evaluating NLP systems. Version 1.0 of the portal contains information on 128 tasks and 95 datasets from 125 competitions held between 2013 and 2022.

The information has been gathered from the main evaluation forums in the NLP area: SemEval⁷, CLEF⁸, CoNLL⁹, COLING¹⁰, IberLEF¹¹, IberEval¹² and PAN¹³.

The information in the ODESIA Portal is intended to satisfy the needs of different stakeholders:

- Researchers seeking information on the different NLP tasks that have been tackled in Spanish, their datasets and evaluation measures; the existing data sets, with information about the domain, linguistic variety of texts, type of texts, annotations, etc.; and the main evaluation forums or competitions in the Spanish NLP scene.
- Companies that want quick access to information about the variety of NLP tasks that may incorporate into their systems, the performance that they may expect to obtain for a task, the existing data sets and their characteristics.
- Public entities and decision-makers financing activities related to NLP and need to detect research gaps in NLP in Spanish, evaluate the novelty and relevance of project proposals and compare the results of financed projects with the state of the

art, as well as to understand the evolution of the state of the NLP in Spanish over time.

4. Future work

Our future work will concentrate on improving all the components that are necessary to carry out the project. We will compute various indexes in order to measure the gap between the development of natural language processing in Spanish and English. We plan to recalculate and release these indexes yearly in the next two years, at least.

In addition, we aim to expand the data and functionalities available in our ODESIA tools (Leaderboard, Portal, and EvALL), and create more datasets in Spanish.

Acknowledgments

ODESIA has been financed by the European Union (NextGenerationEU funds) through the “Plan de Recuperación, Transformación y Resiliencia”, by the Ministry of Economic Affairs and Digital Transformation and by the UNED University (C038/21-OT). However, the points of view and opinions expressed in this document are solely those of the author(s) and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the European Commission can be considered responsible for them. Laura Plaza and Jorge Carrillo-de-Albornoz are also financed by the Ministry of Universities and the European Union through the EuropeNextGenerationEU funds and the “Plan de Recuperación, Transformación y Resiliencia”.

References

- [1] F. Rodríguez-Sánchez, J. Carrillo-de Albornoz, L. Plaza, A. Mendieta-Aragón, G. Marco-Remón, M. Makeienko, M. Plaza, D. Spina, J. Gonzalo, P. Rosso, Overview of EXIST 2022: sexism identification in social networks, *Procesamiento del Lenguaje Natural* 69 (2022) 229–240.
- [2] F. Rodríguez-Sánchez, J. Carrillo-de Albornoz, L. Plaza, J. Gonzalo, P. Rosso, M. Comet, T. Donoso, Overview of EXIST 2021: sexism identification in social networks, *Procesamiento del Lenguaje Natural* 67 (2021) 195–207.
- [3] H. Fabregat, J. Martínez-Romo, L. Araujo, Overview of the DIANN Task: Disability annotation task, *Proceedings of the Third Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018) co-located with 34th Conference of the Spanish Society for Natural Language Processing (SEPLN 2018)* (2018).

⁶<http://portal.odesia.uned.es>

⁷<https://semeval.github.io>

⁸<http://www.clef-initiative.eu>

⁹<https://www.signll.org/conll>

¹⁰<https://coling2022.org>

¹¹<https://sites.google.com/view/iberlef2022>

¹²<https://sites.google.com/view/ibereval-2018>

¹³<https://pan.webis.de>

Portal del Estado de Arte del Procesamiento del Lenguaje Natural en Español

ESTADO DEL ARTE EN CIFRAS

<p>140</p> <p>Tareas</p>	<p>97</p> <p>Datasets</p>	<p>8</p> <p>Foros</p>	<p>125</p> <p>Competiciones</p>
---------------------------------	----------------------------------	------------------------------	--

TAREAS

<p>NON-CONTEXTUAL BINARY CLASSIFICATION OF OFFENSIVE COMMENTS MEX</p> <p>IBERLEF</p> <p>NLP topic: hate detection Dataset: MeOffendES Forum: IberLEF Competición: MeOffendES Detection of offensive language Idioma(s):</p> <p>Ver resultados >></p>	<p>L2 WRITING ASSISTANT</p> <p>SEMEVAL</p> <p>NLP topic: machine translation Dataset: L2 writing assistant-EN-ES Forum: SEMEVAL Competición: SemEval-2014 Task 5: L2 Writing Assistant Idioma(s): Español, English</p> <p>Ver resultados >></p>	<p>PROFILING FAKE NEWS SPREADERS</p> <p>CLEF</p> <p>NLP topic: fake news detection Dataset: PAN-AP-2020-ES Forum: CLEF Competición: 8th Author Profiling Task at PAN 2020: Profiling Fake News Spreaders on Twitter Idioma(s): Español, English</p> <p>Ver resultados >></p>	<p>EMOTION DETECTION</p> <p>IBERLEF</p> <p>NLP topic: sentiment analysis Dataset: EmoEVENT Forum: IberLEF Competición: Semantic Analysis at SEPLN (TASS) Idioma(s):</p> <p>Ver resultados >></p>
---	--	---	---

Figure 2: ODESIA Portal main page.