Procesamiento del Lenguaje Natural, Revista nº 75, septiembre de 2025, pp. 513-522 recibido 03-07-2025 revisado 18-07-2025 aceptado 01-08-2025

# Overview of SatiSPeech at IberLEF 2025: Multimodal Audio-Text Satire Classification in Spanish

## *Resumen de SatiSPeech en IberLEF 2025: Reconocimiento de Sátira Multimodal Basado en Texto y Audio en Español*

**Ronghao Pan, José Antonio García-Díaz, Tomás Bernal-Beltrán, Francisco García-Sánchez, Rafael Valencia-García**

Facultad de Informática, Universidad de Murcia
{ronghao.pan, joseantonio.garcia8, tomas.bernalb, frgarcia, valencia}@um.es

**Abstract:** This article provides an overview of the SatiSPeech 2025 shared task, organized as part of the IberLEF 2025 workshop, held in conjunction with the XLI International Congress of the Spanish Society for Natural Language Processing (SEPLN 2025). The main goal of this task is to advance research on the automatic recognition of satire—a complex form of communication that presents unique challenges for natural language processing, particularly in areas such as subjectivity analysis, emotion recognition, and deep language understanding. The task is divided into two independent subtasks. The first subtask focuses on satire detection using transcriptions from YouTube videos, distinguishing between satirical and non-satirical content through a text classification approach. The second subtask introduces a multimodal perspective, combining textual and acoustic information, which requires addressing challenges in data representation and the design of models capable of modality fusion. Eleven teams participated in SatiSPeech 2025, each proposing and evaluating different strategies to tackle these problems. This overview analyzes the proposed approaches, the techniques employed, the results obtained, and the key insights gained from this edition.
**Keywords:** Satire Identification, Multimodality, Audio Speech Recognition.

**Resumen:** Este artículo presenta una visión general de la tarea SatiSPeech 2025, organizada en el marco del taller IberLEF 2025, celebrado conjuntamente con la XLI Congreso Internacional de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN 2025). El objetivo principal de esta iniciativa es avanzar en el estudio y reconocimiento automático de la sátira, una forma compleja de comunicación que plantea retos específicos para el procesamiento del lenguaje natural, especialmente en tareas relacionadas con la subjetividad, el análisis emocional y la comprensión profunda del discurso. La tarea se ha dividido en dos subtareas independientes. La primera subtarea se centra en la detección de contenido satírico a partir de transcripciones de vídeos procedentes de canales satíricos y no satíricos de YouTube, planteando el problema como una clasificación textual. La segunda subtarea introduce un enfoque multimodal, combinando información textual y acústica, lo que implica desafíos adicionales tanto en la representación de los datos como en el diseño de modelos adecuados para la fusión de modalidades. SatiSPeech 2025 ha contado con la participación activa de 11 equipos, que han desarrollado y evaluado diferentes estrategias para abordar estas tareas. A lo largo del artículo se analizan los enfoques propuestos, las técnicas empleadas y los resultados obtenidos, así como las principales conclusiones derivadas de esta edición.
**Palabras clave:** Identificación de Sátira, Multimodalidad, Reconocimiento de Audio.

### *1 Introduction*

Satire is a rich and nuanced form of communication blending humor and irony to highlight or challenge political, cultural, or societal issues. Unlike more direct forms of humor, satire often relies on subtle cues, such

ISSN 1135-5948 DOI 10.26342/2025-75-38

as tone, exaggeration, and contextual incongruity. These cues can be difficult for human audiences to detect. These complexities are further amplified in multimodal scenarios, where meaning arises from the interaction between language, speech, and prosodic features (Jiang, Li, and Hou, 2019). Misinterpreting satire can lead to confusion or incorrect interpretations, especially when addressing controversial or sensitive topics. Despite these difficulties, the ability to automatically detect and interpret satire is becoming increasingly relevant in areas such as media monitoring, misinformation detection, and sentiment analysis. Multimodal approaches that integrate textual, acoustic, and contextual information offer a promising path toward systems capable of understanding the layered, context-sensitive nature of satirical content.

The SatiSPeech 2025 shared task, which is part of the IberLEF 2025 workshop (González-Barba, Chiruzzo, and Jiménez-Zafra, 2025), is designed to investigate the challenges of satire detection from a multimodal perspective by combining textual and acoustic information. Satire, which is rich in irony, ambiguity, and cultural references, poses particular difficulties for automatic systems, especially when multiple communication modalities are involved (del Pilar Salas-Zárate et al., 2020). The task is framed as a binary classification problem of distinguishing satirical content from non-satirical content by leveraging linguistic and vocal features that convey the subtleties of satirical discourse. Key challenges include identifying salient satire indicators, such as lexical patterns, prosodic cues, and conversational dynamics. Another major hurdle is the limited availability of multimodal datasets that reflect the complexity and diversity of real-world satire; existing resources often lack authentic, context-rich examples. Confronting these limitations advances multimodal classification methods and fosters novel strategies for integrating text and speech in satire detection.

Satire detection, particularly in multimodal settings, is a relatively understudied research area. Most existing studies have focused on unimodal approaches, primarily text-based or, to a lesser extent, visual content-based. For example, the authors in (Li et al., 2020) proposed a system that combines text and images to detect satire in news headlines. Their study showed that multimodal models outperform unimodal ones in capturing the subtle cues that are characteristic of satirical content. Similarly, (Wick-Pedro et al., 2024) investigated the capabilities of large language models (LLMs) for identifying satirical news in Brazilian Portuguese, achieving competitive F1 scores and shedding light on how these models handle satirical language. These studies highlight the potential of using multiple modalities and advanced models to address the complexity of satire.

Compared to other languages, research on satire detection in Spanish has received comparatively less attention, although recent efforts have begun to fill this gap. One notable contribution is SatiCorpus 2021 (García-Díaz and Valencia-García, 2022), a dataset of satirical and non-satirical texts evaluated using linguistic features and deep learning–based embeddings. This approach yielded strong accuracy results. The authors of (Ortega-Bueno, Rosso, and Pagola, 2022) introduced MvAttLSTM, a model integrating handcrafted linguistic features, universal sentence embeddings, and contextualized representations from pre-trained models. This model achieved state-of-the-art performance on irony and satire detection tasks in Spanish tweets. These studies underscore the necessity of creating language-specific resources and customized methodologies to accurately capture the cultural and linguistic nuances inherent in satirical discourse.

The remainder of this paper is structured as follows. Section 2 describes the two subtasks proposed in SatiSPeech 2025. Section 3 introduces the dataset created for the task, including its sources and characteristics. Section 4 outlines the methodologies adopted by participating teams to address each subtask. The official results and a discussion of the main findings are presented in Section 5. Finally, Section 6 summarizes the key outcomes and outlines directions for future work.

## 2 Task description

Detecting satire in a multimodal setting presents many challenges. Satirical content often relies on a combination of subtle linguistic cues, cultural references, and subtle vocal nuances, which makes it difficult to capture through a single modality. Com-

bining textual and acoustic information adds another layer of complexity because each modality conveys meaning in distinct and sometimes asynchronous ways. Additionally, designing effective fusion strategies and addressing the computational demands of multimodal processing are significant obstacles. These challenges underscore the need for robust approaches that can model the complex, cross-modal nature of satirical communication. To achieve this goal, the following subtasks have been defined.

1. Task 1: Text Satire Detection. The goal of this subtask is to determine whether a text is satirical or non-satirical. Participants must rely exclusively on textual information to identify linguistic cues and discourse features that signal satirical intent.
2. Task 2: Multimodal Satire Detection. This subtask introduces a multimodal approach to satire detection by combining textual and auditory information. The objective is to determine whether a given text-audio pair represents satirical or non-satirical content. Participants are required to develop systems capable of effectively integrating both modalities, addressing the inherent challenges of cross-modal alignment, representation, and fusion in order to accurately capture the subtleties of satirical expression.

The competition was hosted on the CodaLab platform and is available at the following link: `https://codalab.lisn.upsaclay.fr/competitions/21501`. It was structured into three phases: Practice, Evaluation, and Post-Evaluation, allowing participants to iteratively develop, test, and refine their systems throughout the task timeline.

During the practice phase, participants were given access to a subset of the training data in order to become familiar with the data format and task requirements. A baseline notebook featuring example systems for both subtasks based on support vector machines and acoustic features was also provided. Participants used this baseline as a starting point to develop and refine their own approaches. Later, the full training set was released to support system development. During this phase, participants were allowed up to 100 submissions on the CodaLab platform. During the evaluation phase, the test set was released, and participants submitted their final systems for official evaluation. Each team was permitted a maximum of ten submissions, and they had to select one for the final ranking. Rankings were based on the macro F1 score for both tasks, which allowed participants to compete in either one or both subtasks independently.

## *3 Dataset*

This task's dataset was specifically compiled to address the challenges of satire detection in a multimodal context by combining text and audio. Data were collected from a variety of Spanish-language YouTube channels, including popular satirical programs such as El Intermedio, Zapeando, Homo-Zapping, and El Mundo Today, as well as non-satirical news sources like Antena 3 Noticias, El Mundo, and BBC News. This selection ensures broad coverage of regional language varieties and communication styles across the Spanish-speaking world.

| label | train | test | total |
|---|---|---|---|
| no-satire | 3168 | 1404 | 4572 |
| satire | 2832 | 596 | 3428 |
| total | 6000 | 2000 | 8000 |

Table 1: Corpus statistics per label.

The compilation process involved extracting videos from these channels and using a speaker diarization tool to segment them into manageable audio units (Bredin and Laurent, 2021). To maintain consistency, segments longer than 25 seconds were excluded. We generated transcriptions of the audio segments using Whisper (Radford et al., 2023), providing high-quality textual counterparts for the audio data.

Annotation was carried out using a semi-supervised approach that combined manual labeling by three domain experts with automatic classification techniques, enhancing both efficiency and reliability. The final labels were validated and refined through a manual review process conducted by the organizers to ensure high annotation quality. The dataset captures a wide range of Spanish dialects and cultural references, which re-

duces regional bias and enhances generalizability.

As can be seen in Table 1, a total of 8,000 audio segments were selected for this contest, with 80% used for training and 20% for testing.

## 4 Participant approaches

A total of 30 users registered for the SatiSPeech 2025 shared task. Eleven of these teams have submitted their results and working notes describing their systems, and have participated in the unimodal and multimodal tasks.

Next, we will present a brief summary of the participants' systems.

- **Ferrara** (Bortolotti, 2025). For the textual modality, the system utilizes BETO (Cañete et al., 2023), a Spanish pre-trained BERT-based model, while HuBERT (Hsu et al., 2021), a self-supervised speech representation model, is employed for the audio modality. These representations were initially integrated through a late fusion architecture that combined class probability distributions. After the competition, enhancements included exploring weighted averaging and implementing a Multimodal Attention Network for fusion. Ferrara's system achieved an F1-Score of 83.704 for the Multimodal Task (Task 2) and 83.210 for the Textual Task (Task 1), ranking 3rd in Task 2 and 5th in Task 1.
- **FlowIn** (Bao et al., 2025). For text, it utilized BETO embeddings and TF-IDF features with various classifiers such as Logistic Regression, Support Vector Machine (SVM), and XGBoost, employing a soft voting ensemble for Task 1. Audio features were extracted using Mel Frequency Cepstral Coefficients (MFCCs) (Zheng, Zhang, and Song, 2001) processed by a Convolutional Neural Network (CNN), which were then concatenated with text embeddings and fed to a Multi-Layer Perceptron (MLP), Logistic Regression, or SVM for multimodal classification. A key aspect was the independent fine-tuning of the CNN for 100 epochs to produce more stable and refined audio features, significantly enhancing its contribution to the final hard voting ensemble. FlowIn ranked 4th in the official evaluation phase with an F1-score of 83.274 in Task 1 and 83.274 in Task 2.
- **ITST** (Paredes-Valverde and del Pilar Salas-Zárate, 2025). For text-only satire detection (Task 1), RoBERTa-BNE (Gutiérrez-Fandiño et al., 2022) was used to extract contextual sentence embeddings, which were then classified by a SVM. For multimodal satire detection (Task 2), acoustic features extracted using Wav2Vec 2.0 (Conneau et al., 2021) were concatenated with the RoBERTa text embeddings, representing an early fusion approach. This combined 1792-dimensional vector was then fed into a second SVM classifier, with hyperparameters optimized via grid search for both tasks. ITST achieved 2nd place in Task 1 (Textual) with an F1-Score of 84.546 and 5th place in Task 2 (Multimodal) with 83.271.
- **LACELL** (Almela et al., 2025). For textual analysis, the system employed linguistic features extracted using UMUTextStats (García-Díaz et al., 2022) and sentence embeddings derived from various fine-tuned Spanish LLMs like MarIA (Gutiérrez-Fandiño et al., 2022) and BETO. Acoustic features were captured using MFCCs. These diverse features were integrated through an ensemble learning approach, specifically a Knowledge Integration (KI) strategy utilizing a shallow neural network on concatenated feature vectors. The system achieved a macro F1-score of 81.470 in both tasks, 9th place in Task 1 and 8th in Task 2, outperforming the baseline.
- **ngocan0987**. This team placed the 7th position in Task 1 and the 6th position in Task 2. This team did not submit the working notes describing their proposed approach.
- **SINAI** (Espin-Riofrio, Ortiz-Zambrano, and Montejo-Ráez, 2025). The SINAI team employed supervised classifiers, primarily a Voting Classifier, trained on stylometric and linguistic features like n-grams, polarity, irony, syntactic structures, and textual complexity. Text data underwent preprocessing including cleaning, stopword

| id | label | transcription |
|---|---|---|
| `ff6093f7-2afab470` | satire | "Y vámonos con el señor que nunca en su vida ha usado un traje a la medida o sin manchas de mole, Ricardo Monreal, también conocido como el pendejo que podía ser presidente. Monri no solo tiene el último puesto en las encuestas, también tiene la peor estrategia, ya que intentó hacer un rap. Se editó a sí mismo en un vídeo de Star Wars y contrató al sobrino del gab, Gibran Ramírez, como asesor." |
| `c4619d02-a108f443` | no-satire | "En Francia, se juega una nueva carta en favor de su política de austeridad. A pedido de Manuel Valls, que forme otro gobierno. El actual ha dimitido en pleno, en medio de agudas críticas internas contra los recortes que defienden el presidente y su primer ministro." |
| `d545938e-d4107ff` | satire | "Gracias! Ya caíste. Hoy, debido a una plaga en Nueva York, implementarán anticonceptivos para ratas. Con razón el maestro Splinter adoptó cuatro tortugas." |
| `43d7d437-0fd21a4` | no-satire | "En Japón al menos 13 personas murieron a causa de las inundaciones y deslaves causados por el tifón Guifa. La tormenta, una de las más fuertes en años en el país asiático, azotó la costa al sur de la capital, Tokio." |

Table 2: Examples of the dataset.

removal, and lemmatization. While the Voting Classifier achieved perfect scores (1.0 F1-score) on an internal test set, its performance on the official challenge test set resulted in an F1-score of 79.479, securing 10th place in both tasks. This drop highlighted generalization issues when facing more diverse data, but affirmed the value of combining stylometric and linguistic features for capturing satirical nuances.

- **UAE** (Lagos-Ortiz, Medina-Moreira, and Apolinario-Arzube, 2025). For textual analysis (Task 1), BETO was utilized to extract contextual sentence embeddings, which were then classified by a SVM. In the multimodal task (Task 2), Wav2Vec 2.0 extracted acoustic features, which were then concatenated with the BETO text embeddings in an early fusion approach before being fed into a second SVM classifier. UAE achieved macro F1-scores of 81.631 for the text-based task (Task 1), 8th place, and 81.496 for the multimodal task (Task 2), 7th place, both surpassing the respective baselines. Despite high performance on internal validation sets, the slight decrease in F1-score when adding audio suggested that simple feature concatenation might not fully leverage prosodic cues or that the models overfit the training data.

- **UKR** (Gladun, Rogushina, and Martínez-Béjar, 2025). For text-based satire detection (Task 1), the UKR team employed a supervised fine-tuning approach using the Spanish monolingual BERT model (BETO), extracting its contextual `[CLS]` token embeddings. For multimodal detection (Task 2), they extended this by extracting MFCCs from audio samples, which were then concatenated with the BETO text embeddings in an early fusion strategy. A custom classification head was used for binary satire prediction in both tasks. On their internal validation split, the system achieved strong macro F1 scores of 96.480 for Task 1 and 96.990 for Task 2, demonstrating that audio features offered complementary cues. Officially, the system ranked 6th in Task 1 with an F1-score of 83.202 and 9th in Task 2 with an F1-score of 80.131, successfully outperforming the challenge baseline.

- **UMU-Ev** (Valero-Vilella, 2025). For textual analysis, RoBERTa-BNE was primarily used for contextual embeddings, while HuBERT-based representations combined with MFCCs and

prosodic features were employed for audio. These modalities were integrated through a late fusion architecture, with the final classification handled by a SVM. The approach achieved top-tier performance, securing 3rd place in the monomodal text task (Task 1) with a Macro F1-Score of 84.455. Notably, it attained 1st place in the multimodal task (Task 2) with a Macro F1-Score of 88.340, proving strong performance with computationally lightweight architectures.

- **UPV-ELiRF** (Barceló-Milkova et al., 2025). For text-only satire detection (Task 1), the UPV-ELiRF team employed fine-tuned transformer-based language models (like BETO, BERTIN (de la Rosa et al., 2022), and RoBERTa-BNE) using the LoRA technique (Hu et al., 2022), and also explored few-shot prompting with large language models such as Qwen2.5, Llama-3.1, and Mistral. For multimodal satire detection (Task 2), they extracted audio embeddings using Whisper, Wav2Vec2, or HuBERT, and text embeddings from XLM-RoBERTa (Liu et al., 2019), which were then concatenated and fed into various classifiers like Support Vector Classifiers (SVC) and MLP. Their comprehensive approach achieved remarkable results, securing 1st place in the text-only task (Task 1) with a Macro F1-Score of 85.638 and 2nd place in the multimodal task (Task 2) on the official leaderboard with a Macro F1-Score of 86.444.

- **UTP** (Cedeño-Moreno, Vargas-Lombardo, and Delgado-Herrera, 2025). For text-only satire detection (Task 1), they employed FastText (Joulin et al., 2016) embeddings with a Random Forest classifier. For multimodal satire detection (Task 2), this architecture was extended by concatenating Wav2Vec2 acoustic embeddings with the FastText textual embeddings, which were then classified by a second Random Forest model. On the official test set, the system achieved a macro F1-score of 77.936 for Task 1 and 76.476 for Task 2. The UTP team ranked 11th out of 11 participants in both tasks, not outperforming the baseline, which suggested that the simple integration of modalities introduced inconsistencies in classification.

## *5 Results and discussion*

The official leaderboards for the SatiSPeech 2025 shared task are shown in Table 3 for Task 1 and in Table 4 for Task 2.

| Ranking | Team | Macro F1-Score |
|---|---|---|
| 01 | UPV-ELiRF | 85.638 |
| 02 | ITST | 84.546 |
| 03 | UMU-Ev | 84.455 |
| 04 | FlowIn | 83.274 |
| 05 | Ferrara | 83.210 |
| 06 | UKR | 83.202 |
| 07 | ngocan0987 | 82.041 |
| 08 | UAE | 81.631 |
| 09 | LACELL | 81.470 |
| 10 | SINAI | 79.479 |
| – | BASELINE | 79.373 |
| 11 | UTP | 77.936 |
| | *Mean* | *82.822* |

Table 3: SatiSPeech 2025 official leaderboard for Task 1, including the official ranking and result.

| Ranking | Team | Macro F1-Score |
|---|---|---|
| 01 | UMU-Ev | 88.340 |
| 02 | UPV-ELiRF | 86.444 |
| 03 | Ferrara | 83.704 |
| 04 | FlowIn | 83.274 |
| 05 | ITST | 83.271 |
| 06 | ngocan0987 | 82.776 |
| 07 | UAE | 81.496 |
| 08 | LACELL | 81.470 |
| 09 | UKR | 80.131 |
| – | BASELINE | 79.924 |
| 10 | SINAI | 79.479 |
| 11 | UTP | 76.476 |
| | *Mean* | *82.188* |

Table 4: SatiSPeech 2025 official leaderboard for Task 2, including the official ranking and result.

The textual component was consistently the most critical for satire detection. Many teams, including Ferrara, FlowIn, UAE, and UKR, heavily relied on BETO, which pre-trained on large Spanish corpora like Wikipedia and news articles, captures

language-specific nuances crucial for understanding idiomatic expressions, tone, and cultural references in Spanish satire. The ITST team used the Spanish RoBERTa-BNE model for contextual sentence embeddings, while UMU-Ev also evaluated it. UPV-ELiRF also explored BERTIN (a RoBERTA model pre-trained on Spanish texts) and XLM-RoBERTa.

The primary method for text representation involved extracting fixed-size sentence vectors, often from the `[CLS` token of the final hidden layer of transformer models. FlowIn utilized TF-IDF features as a *strong and interpretable baseline* for initial satire classification, demonstrating its effectiveness for capturing surface-level lexical cues. The LACELL team employed the UMUTextStats tool, specifically designed for Spanish, to extract morphosyntactic features. Their analysis showed that *satirical texts tend to contain a higher frequency of affix-related features (e.g., suffixes in nominative nouns), longer words, and a higher syllable-per-word ratio, possibly due to a more elaborate or playful language style.* Conversely, non-satirical texts had longer average sentence lengths and more consistent punctuation. Personal pronouns and question marks were more frequent in satirical content, suggesting a rhetorical strategy to engage the audience. SINAI also emphasized the value of stylometric and linguistic features (e.g., num_words, num_chars, exclamations, uppercase_ratio, polarity, irony_score, PoS tags, rhetorical_questions, metaphors, n-grams) for capturing subtle stylistic signals.

Recognizing that satire often relies on intonation, pitch variation, stress patterns and rhythm, teams explored various methods for extracting acoustic features. Ferrara and UMU-Ev leveraged HuBERT, a self-supervised model that learns acoustic representations by predicting masked audio segments. HuBERT computes 1024-dimensional embedding that captures prosodic, rhythmic, and timbral features. ITST and UAE integrated Wav2Vec 2.0, another self-supervised speech representation model, for high-level acoustic features. UKR also utilized Wav2Vec2 embeddings. Several teams, including FlowIn, UMU-Ev, and UKR, used MFCCs, which approximate human sound perception and capture essential aspects of speech prosody and articulation that are often indicative of expressive intentions, such as emphasis, exaggeration, or irony. UMU-Ev also incorporated prosodic features and their deltas (first and second-order temporal derivatives of MFCCs) to capture dynamic changes over time.

A crucial aspect of the multimodal task was combining textual and acoustic features. Many teams, including ITST, UAE, and UKR, adopted a straightforward approach of concatenating textual embeddings (from BETO or RoBERTa) with audio embeddings (from Wav2Vec 2.0 or MFCCs). This "simple early fusion" demonstrated robust performance despite its simplicity. While concatenation was common, future work proposed by Ferrara and others included exploring deeper multimodal fusion strategies beyond simple feature concatenation, including attention-based or tensor fusion methods. UMU-Ev specifically evaluated weighted sum and attention-based fusion, finding attention to be effective in their best multimodal model.

Various machine learning and deep learning models were employed for classification. SVMs were a popular choice for both textual and multimodal tasks due to their effectiveness with high-dimensional feature vectors. ITST, UAE, and UMU-Ev successfully used SVMs. Ferrara used a feed-forward neural network for audio classification. FlowIn and UMU-Ev employed MLP or Fully Connected Dense Neural Networks (DNN) for their classification tasks, particularly in multimodal settings. Besides, FlowIn explored shallow models such as Logistic Regression, SVM, and XGBoost for text classification, highlighting the effectiveness of combining shallow and deep learning techniques. Finally, FlowIn applied a voting ensemble to combine model predictions, contributing to their improved post-evaluation ranking. SINAI also used a voting classifier, achieving high performance on their validation set.

The competition results highlight the effectiveness of multimodal approaches, though improvements over text-only models were sometimes modest.

## 6 Conclusions

This paper presents the first edition of the SatiSPeech shared task in IberLEF 2025, which consists of two subtasks. The first subtask concerns to the extraction of satiric con-

tent from text, which involves the identification of salient features and the determination of the most representative ones of each emotion within a dataset derived from real-life scenarios. The second subtask deals with satire classification from a multimodal perspective, using aligned text and audio segments. This requires the construction of a more complex architecture to solve the classification problem. The dataset for this task was compiled by collecting audio segments from different Spanish YouTube channels and was annotated as either satirical or non-satirical.

As this is the first time we have organized this event, we are very pleased with the response, with the registration of 30 users and the participation of 11 teams, who sent promising approaches for solving both tasks. The participants demonstrated significant progress in multimodal satire detection in Spanish. Transformer models like BETO and HuBERT proved highly effective for both text and audio feature extraction. While textual cues remain dominant, the integration of prosodic features offers valuable complementary information, even with simple fusion strategies. The shared task is still accessible in the post-evaluation phase `https://codalab.lisn.upsaclay.fr/competitions/21501#learn_the_details`. The Codalab page contains two notebooks for the preparation of the baseline and the submission file to be sent to the competition, as well as the full dataset excluding the golden labels for test set. It is our hope that these resources will prove beneficial to the Spanish NLP community.

One of the insights gained from the competition is that there are still many challenges to developing effective satire detection systems, particularly within multimodal and cross-cultural contexts. One critical limitation is the scarcity of rich multimodal satire datasets, which hampers the capacity to capture the full diversity and authenticity of real-world satirical expression. Compounding this issue is the complexity of integrating heterogeneous modalities; while current approaches often rely on simple feature concatenation, future research is directed toward more sophisticated fusion strategies, such as attention-based mechanisms or tensor fusion methods, to improve robustness. Additionally, the inherently cultural and regional nature of humor, especially in Spanish, introduces further complexity, as models must account for nuanced linguistic and sociocultural variations. Generalization also remains a concern, with some systems exhibiting performance degradation on more diverse or previously unseen data in official test settings compared to internal validation scenarios. To address these challenges, future work appears to be aimed at developing more sophisticated fusion techniques, addressing data scarcity, and improving models' ability to generalize across diverse satirical expressions, particularly considering the subtle cultural and linguistic complexities of Spanish satire. It is our intention to organize a second edition of this task. In this context, the dataset will be expanded to include additional channels and modalities, such as video.

## *Acknowledgments*

This work is part of the research project LaTe4PoliticES (PID2022-138099OB-I00) funded by MCIN/AEI/10.13039/501100011033 and the European Regional Development Fund (ERDF)-a way to make Europe. Mr. Tomás Bernal-Beltrán is supported by University of Murcia through the predoctoral programme.

## *References*

Almela, A., P. Cantos-Gómez, D. Granados-Meroño, and G. Alcaraz-Mármol. 2025. LACELL at SatiSPeech-IberLEF 2025: Multimodal Linguistic Features plus Embeddings for Satire Identification from YouTube. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41th Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS.org.*

Bao, N. M., T. T. Tran, N. T. Bao, and D. V. Thin. 2025. FlowInTeam at SatiSPeech-IberLEF 2025: Multimodal Speech-text Satire Recognition in Spanish. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41th Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS.org.*

Barceló-Milkova, A. J., A. Casamayor-Segarra, V. Ahuir, and M. J. Castro-Bleda. 2025. ELiRF-UPV at SatiSPeech-

IberLEF 2025: Multimodal Speech-text Satire Recognition in Spanish. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41th Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS.org*.

Bortolotti, M. 2025. Ferrara at SatiSPeech-IberLEF 2025: Leveraging BETO and HuBERT for Multimodal Speech-Text Satire Recognition. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41th Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS.org*.

Bredin, H. and A. Laurent. 2021. End-to-end speaker segmentation for overlap-aware resegmentation. In *Proc. Interspeech 2021*, Brno, Czech Republic, August.

Cañete, J., G. Chaperon, R. Fuentes, J. Ho, H. Kang, and J. Pérez. 2023. Spanish Pre-trained BERT Model and Evaluation Data. *CoRR*, abs/2308.02976.

Cedeño-Moreno, D., M. Vargas-Lombardo, and A. Delgado-Herrera. 2025. UTP at SatiSpeech–IberLEF 2025: FastText and-Wav2Vec2 for Satire Detection in Spanis. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41th Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS.org*.

Conneau, A., A. Baevski, R. Collobert, A. Mohamed, and M. Auli. 2021. Unsupervised Cross-Lingual Representation Learning for Speech Recognition. In H. Hermansky, H. Cernocký, L. Burget, L. Lamel, O. Scharenborg, and P. Motlícek, editors, *22nd Annual Conference of the International Speech Communication Association, Interspeech 2021, Brno, Czechia, August 30 - September 3, 2021*, pages 2426–2430. ISCA.

de la Rosa, J., E. G. Ponferrada, M. Romero, P. Villegas, P. G. de Prado Salas, and M. Grandury. 2022. BERTIN: Efficient Pre-Training of a Spanish Language Model using Perplexity Sampling. *Proces. del Leng. Natural*, 68:13–23.

del Pilar Salas-Zárate, M., G. Alor-Hernández, J. L. Sánchez-Cervantes, M. A. Paredes-Valverde, J. L. García-Alcaraz, and R. Valencia-García. 2020. Review of English literature on figurative language applied to social networks. *Knowledge and Information Systems*, 62(6):2105–2137.

Espin-Riofrio, C., J. Ortiz-Zambrano, and A. Montejo-Ráez. 2025. SINAI at SatiSPeech in IberLEF 2025: Detection of Satire in Spanish Texts Using Stylometric and Linguistic Features. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41th Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS.org*.

García-Díaz, J. A. and R. Valencia-García. 2022. Compilation and evaluation of the Spanish SatiCorpus 2021 for satire identification using linguistic features and transformers. *Complex & Intelligent Systems*, 8(2):1723–1736.

García-Díaz, J. A., P. J. Vivancos-Vicente, Á. Almela, and R. Valencia-García. 2022. UMUTextStats: A linguistic feature extraction tool for Spanish. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, and S. Piperidis, editors, *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 6035–6044, Marseille, France, June. European Language Resources Association.

Gladun, A., J. Rogushina, and R. Martínez-Béjar. 2025. UKR at SatiSPeech-IberLEF 2025: Multimodal Satire Detection in Spanish with a BETO-Based Text Encoder and MFCC-Derived Audio Features. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41th Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS.org*.

González-Barba, J. Á., L. Chiruzzo, and S. M. Jiménez-Zafra. 2025. Overview of IberLEF 2025: Natural Language Processing Challenges for Spanish and other Iberian Languages. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41st Conference of the Spanish Society for Nat-*

*ural Language Processing (SEPLN 2025), CEUR-WS. org*.

Gutiérrez-Fandiño, A., J. Armengol-Estapé, M. Pàmies, J. Llop-Palao, J. Silveira-Ocampo, C. P. Carrino, C. Armentano-Oller, C. Rodriguez-Penagos, A. Gonzalez-Agirre, and M. Villegas. 2022. Maria: Spanish language models. *Procesamiento del Lenguaje Natural*, 68(0):39–60.

Hsu, W.-N., B. Bolte, Y.-H. H. Tsai, K. Lakhotia, R. Salakhutdinov, and A. Mohamed. 2021. HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, 29:3451–3460, October.

Hu, E. J., Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. 2022. LoRA: Low-Rank Adaptation of Large Language Models. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net.

Jiang, T., H. Li, and Y. Hou. 2019. Cultural differences in humor perception, usage, and implications. *Frontiers in psychology*, 10:123.

Joulin, A., E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov. 2016. FastText.zip: Compressing text classification models. *CoRR*, abs/1612.03651.

Lagos-Ortiz, K., J. Medina-Moreira, and O. Apolinario-Arzube. 2025. UAE at SatiSpeech–IberLEF 2025: Multimodal Satire Detection Using BETO and-Wav2Vec2 Embeddings with SVM. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41th Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS.org*.

Li, L., O. Levi, P. Hosseini, and D. Broniatowski. 2020. A multi-modal method for satire detection using textual and visual cues. In *Proceedings of the 3rd NLP4IF Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 33–38.

Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR*, abs/1907.11692.

Ortega-Bueno, R., P. Rosso, and J. E. M. Pagola. 2022. Multi-view informed attention-based model for Irony and Satire detection in Spanish variants. *Knowledge-Based Systems*, 235:107597.

Paredes-Valverde, M. A. and M. del Pilar Salas-Zárate. 2025. ITST at SatiSPeech–IberLEF 2025: Leveraging Transformers for Textual and Multimodal Satire Detection in Spanish. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41th Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS.org*.

Radford, A., J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International conference on machine learning*, pages 28492–28518. PMLR.

Valero-Vilella, E. 2025. UMU-Ev at SatiSPeech-IberLEF 2025: Exploring Multimodal Satire Detection with Efficient Audio-Text Representations. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41th Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS.org*.

Wick-Pedro, G., C. F. da Silva, M. L. Inácio, O. A. Vale, and H. de Medeiros Caseli. 2024. Using Large Language Models for Identifying Satirical News in Brazilian Portuguese. In *Proceedings of the 16th International Conference on Computational Processing of Portuguese*, pages 156–167.

Zheng, T. F., G. Zhang, and Z. Song. 2001. Comparison of Different Implementations of MFCC. *J. Comput. Sci. Technol.*, 16(6):582–589.