Journal of Information Technology Management
ISSN-OnLine: 2980-7972
University of Tehran

Advancing Automatic Sarcasm Detection in the French Language

Yannick U. Tchantchou Samen*

Corresponding author, Assistant Prof., Senior Lecturer, Ph.D., Department of Mathematics and Computer Science, Faculty of Science, University of Maroua, P.O Box:814 Maroua, Cameroon; Laboratoire de Recherche en Sciences Informatiques et Applications (LRSIA), UAC, Abomey-Calavi, Benin. E-mail: yannick.samen@imsp-uac.org

Abou A. Djafar

Ph.D. Candidate, Department of Mathematics and computer Science, Faculty of Science, University of Maroua, Maroua, Cameroon. E-mail: bigdjaf@gmail.com

Kaladzavi Guidedi

Associate Prof., Department of Computer Science and telecommunications, National Advanced School of Engineering of Maroua, University of Maroua, Maroua, Cameroon. E-mail: kaladzavi@univ-maroua.cm

Fréjus A. A. Laleye

Opscidia, 9 Rue des colonnes, Paris, France. E-mail: frejus.laleye@opscidia.com

Journal of Information Technology Management, 2026, Vol. 18, Issue 3, pp. 202-227
Published by the University of Tehran, College of Management
doi: https://doi.org/10.22059/jitm.2026.414898.4482
Article Type: Research Paper

Received: January 13, 2026
Received in revised form: March 08, 2026
Accepted: May 21, 2026
Published online: July 01, 2026

Abstract

The proliferation of user-generated content necessitates robust sentiment analysis and opinion mining. Sarcasm detection presents a significant challenge in natural language processing, as sarcastic statements often invert literal meaning. Accurate sarcasm detection is crucial for enhancing sentiment analysis, opinion mining, recommender systems, and public opinion monitoring. While transformer-based and deep learning models have advanced sarcasm detection in English and other languages, French remains underdeveloped due to limited annotated datasets and the absence of models tailored to its linguistic nuances. Consequently, existing methods struggle to generalize and fail to leverage complementary semantic information to improve sarcasm recognition. This paper introduces a novel hybrid multi-task learning framework for French sarcasm detection. This architecture integrates transfer learning with sentiment analysis by combining a fine-tuned CamemBERT encoder, an auxiliary sentiment-analysis module, and a Bi-LSTM classifier. This approach jointly models

contextual and affective information to better distinguish between sarcastic and non-sarcastic text. A significant contribution of this study is the development and release of a large-scale French sarcasm dataset comprising approximately 22,000 manually collected and curated annotated instances from diverse sources, providing a valuable benchmark for future research. Another key contribution is the proposed novel hybrid multi-task architecture that integrates transfer learning, sentiment analysis, and Bi-LSTM-based classification for improved sarcasm detection. Empirical evidence shows that sentiment-aware representations significantly improve sarcasm detection over standard transformer approaches. Experiments on the TransCasm dataset and the new corpus demonstrate the model's effectiveness, achieving F1-scores of 96% and 90%, respectively, surpassing prior results.

Keywords: Sentiment Analysis, Sarcasm, Multi-task Learning, Transfer Learning, French.

Introduction

The widespread adoption of social media platforms has led to an unprecedented growth in user-generated content, making sentiment analysis and opinion mining increasingly important for understanding public opinions, customer feedback, and online interactions. However, accurately identifying users' true intentions remains challenging because online communication frequently employs figurative language, particularly sarcasm. Sarcasm expresses an intended meaning that often contradicts the literal interpretation of the text, making it one of the most difficult Natural Language Processing (NLP) tasks. Failure to correctly identify sarcastic expressions can substantially degrade the performance of downstream applications, including sentiment analysis, opinion mining, recommender systems, and public opinion monitoring.

Over the past decade, remarkable progress has been achieved in automatic sarcasm detection through the development of deep learning and transformer-based approaches. Numerous studies have reported promising results for English (Lin, 2016; Bassem, 2024; Farabi, 2024; Sharma, 2026), which remains the most extensively investigated language, while increasing attention has also been devoted to Chinese (Simon, 2022; Zhang, 2022), Arabic (Naik, 2022; Obeidat, 2022; Ren, 2024; Hassan, 2026), Hindi (Hengle, 2021; Cakebread-Andrews, 2021), and multilingual environments. However, French, the fifth most spoken language globally in 2023 as indicated in Table 1, has received comparatively limited attention in sarcasm detection research, particularly when contrasted with other top-ranked languages. To date, the only documented study addressing sarcasm in French is the work proposed by Simon et al. (2022), which utilized the Transcasm dataset. This relative paucity of research may stem from the absence of a comprehensive, high-quality dataset specifically curated for sarcasm in the French language. Consequently, this situation raises two primary research considerations: first, determining the optimal methodology for constructing a large-scale dataset designed to capture nuances of sarcasm in French; and second, identifying the most effective strategies for leveraging such a dataset to develop a robust model capable of

accurately identifying sarcastic sentences in French, while also accommodating the specific colloquialisms inherent in the language.

Table 1. The 12 Most Spoken Languages on Earth

Rank	Language	Number of Speakers (2023)
1	English	1.500.000.000
2	Mandarin	1.100.000.000
3	Hindi	609.500.000
4	Spanish	59.100.000
5	French	309.800.000
6	Standard Arabic	274.000.000
7	Bengali	272.800.000
8	Portuguese	263.600.000
9	Russian	255.000.000
10	Urdu	231.700.000
11	Indonesian	199.100.000
12	German	133.200.000

This paper introduces a novel hybrid multi-task learning framework designed for automated sarcasm detection in French. This model integrates transfer learning with sentiment-aware representations through three key components: a fine-tuned CamemBERT encoder for contextual representation, a sentiment analysis module to identify affective cues, and a Bi-LSTM classifier to model sequential dependencies and classify sarcasm. This integrated approach allows the model to leverage both semantic context and sentiment information, thereby enhancing its capacity to differentiate between sarcastic and non-sarcastic text. Furthermore, this research involved the creation of a new, large-scale French sarcasm corpus. This corpus comprises approximately 22,000 manually collected and curated text instances sourced from various online platforms. To our knowledge, this corpus represents one of the most extensive publicly available resources for French sarcasm detection, offering a valuable benchmark for subsequent research in this domain.

The remainder of this paper is organized as follows. Section 2 reviews related work on sarcasm detection. Section 3 presents the methodology and the construction of the French sarcasm corpus. Section 4 reports the experimental setup and results. Finally, Section 5 concludes the paper and outlines future research directions.

Literature Review

Sarcasm is something that people use in languages. Over the past 10 years, it has been studied a lot, especially when it comes to figuring out when someone is being sarcastic in different languages.

Sarcasm detection in English

English has received by far the greatest research attention in automatic sarcasm detection and has served as the primary benchmark for evaluating new methodologies. The availability of large annotated corpora has enabled the development of increasingly sophisticated approaches, ranging from conventional machine learning techniques to recent transformer-based architectures and Large Language Models.

Early deep learning approaches demonstrated the effectiveness of neural architectures for capturing contextual dependencies in sarcastic expressions. For example, Ayyasamy (2021) proposed a hybrid deep learning model based on distributed word representations to improve sarcasm detection in English texts. As research progressed, attention shifted toward multimodal learning, where textual information is combined with complementary sources such as facial expressions, visual content, and contextual cues. In this direction, Jose and Benedict (2023) introduced the DeepASD framework, illustrating the advantages of multimodal learning for understanding complex sarcastic expressions.

More recently, transformer-based language models have substantially advanced the state of the art. Heraldi and Ruskanda (2024) leveraged the effectiveness of fine-tuning Llama 2 for sarcasm detection, highlighting the potential of Large Language Models to capture subtle semantic relationships. Likewise, Farhan et al. (2024) developed a hybrid architecture combining GloVe embeddings and Bi-LSTM, achieving competitive performance across different social media platforms. Ibrahim et al. (2025) further improved sarcasm detection by integrating convolutional neural networks, enhanced LSTM architectures, and emoji representations, emphasizing the importance of combining contextual and affective information for figurative language understanding.

These studies clearly illustrate the evolution of English sarcasm detection from conventional deep learning architectures toward transformer-based and hybrid models capable of exploiting increasingly rich semantic representations. Despite these significant advances, sarcasm detection remains challenging because sarcastic meaning is strongly influenced by contextual, cultural, and emotional factors that are not always explicitly expressed in text. Consequently, improving the integration of contextual and affective information remains an active research direction.

Sarcasm detection in Chinese

Chinese is among the earliest non-English languages for which sarcasm detection has attracted considerable research attention. Initial efforts focused on the development of annotated corpora to facilitate supervised learning approaches (Lin and Hsieh, 2016; Gong et al., 2020). These resources have subsequently enabled the emergence of deep learning models capable of capturing contextual and semantic information more effectively. For instance, Jia

and Zan (2022) designed a hybrid BERT-BiGRU architecture that significantly improved sarcasm classification by combining contextual embeddings with sequential dependency modeling. Similarly, Zhang et al. (2022) introduced the Retrospective Reader model, which exploits contextual interactions to better interpret implicit sarcastic expressions. More recently, some works (Tian and Yang, 2024) integrated BERT, title attention, and Bi-LSTM mechanisms to jointly model textual content and contextual information, achieving competitive performance on Chinese sarcasm detection tasks. Collectively, these studies demonstrate the effectiveness of transformer-based architectures for Chinese sarcasm detection. Nevertheless, most approaches remain language-specific, and their ability to generalize across different linguistic and cultural contexts remains limited.

Encouraged by these promising results, similar approaches have been gradually and even simultaneously studied for other Asian languages, particularly Hindi.

Sarcasm detection in Hindi

Research on sarcasm detection in Hindi has gained increasing attention in recent years, primarily driven by the rapid growth of social media and the widespread use of code-mixed Hindi-English content. Early studies focused on the construction of annotated corpora and the development of baseline machine learning models capable of handling the linguistic complexity of code-mixed texts (Swami et al., 2018). Subsequently, researchers explored distributed word representations and bilingual embedding techniques to better capture semantic relationships between Hindi and English words occurring in the same sentence (Aggarwal et al., 2020).

With the emergence of deep learning, more sophisticated architectures have been introduced to improve sarcasm detection. Several studies investigated recurrent neural networks and attention-based mechanisms to model contextual dependencies in code-mixed social media texts, leading to noticeable improvements over conventional machine learning approaches. More recently, Kumar et al. (2023) defined a framework combining textual and emoji embeddings, demonstrating that incorporating non-textual semantic cues can substantially improve sarcasm recognition in informal online conversations. Pokhriyal and Jain applied the Zipf-Mandelbrot technique for text analysis. Their model, however, suffers from excessive complexity, hindering its scalability for large-scale applications (Pokhriyal and Jain, 2025).

Overall, these studies demonstrate the rapid evolution of Hindi sarcasm detection from traditional feature engineering toward deep contextual representation learning. Nevertheless, most existing approaches remain highly dependent on code-mixed datasets and language-specific resources, which limits their transferability to other linguistic settings and highlights the need for more generalized sarcasm detection frameworks.

Beyond Asian languages, sarcasm detection has also attracted growing attention in Arabic.

Sarcasm detection in Arabic

Research on sarcasm detection in Arabic has attracted growing attention in recent years owing to the increasing popularity of Arabic social media platforms and the linguistic complexity of the language. Early studies mainly focused on identifying lexical and contextual features capable of distinguishing sarcastic from literal expressions in social media texts (Al-Ghadhban et al., 2017). As annotated Arabic datasets became available, deep learning techniques gradually replaced traditional machine learning methods, enabling more effective modeling of contextual information.

Subsequent studies explored the use of multitask learning and transformer-based architectures to improve sarcasm detection by jointly exploiting semantic and sentiment-related information. For instance, Alharbi and Lee (2021) proposed a multitask learning framework, demonstrating that jointly learning sarcasm detection and sentiment analysis improves classification performance. Similarly, Obeidat et al. (2022) investigated the effectiveness of fine-tuned BERT models on several Arabic sarcasm datasets and reported significant performance gains over conventional approaches. More recently, Abuein et al. (2024) designed a deep learning model evaluated on the ArSa-Tweets dataset (Mihi, 2023), further confirming the effectiveness of contextual language models for Arabic sarcasm detection.

Overall, these studies demonstrate the progressive evolution of Arabic sarcasm detection from feature engineering approaches to transformer-based architectures capable of capturing complex contextual information. Nevertheless, most existing models are evaluated on language-specific datasets and remain primarily focused on textual information, limiting their ability to generalize across different communication contexts and multimodal social media environments.

Multilingual Sarcasm Detection

The growing diversity of multilingual social media content has stimulated increasing interest in multilingual sarcasm detection, aiming to develop models capable of recognizing sarcastic expressions across multiple languages while reducing dependence on language-specific resources. Recent advances in multilingual representation learning have enabled the development of models that leverage shared semantic knowledge to improve sarcasm detection in multilingual environments (Mihi, 2022; Pandey, 2025).

Derbala et al. (2024) investigated deep learning approaches for multilingual sarcasm detection and proved the effectiveness of contextual representation learning across different

languages. Their study highlighted the potential of multilingual models to learn language-independent semantic features while preserving language-specific contextual information. Similarly, Sukhavasi and Dondeti (2024) explored transformer-based architectures for multilingual sarcasm detection, reporting improved classification performance through the use of contextual embeddings capable of capturing semantic relationships across multiple languages. Yue et al. (2024) introduced the SarcNet dataset, a large-scale multilingual and multimodal benchmark designed to facilitate the development and evaluation of sarcasm detection models under more realistic communication settings. This resource represents an important step toward the standardization of multilingual sarcasm detection by providing diversified data collected from multiple languages and modalities.

Overall, these studies demonstrate that multilingual transformer-based models have considerably improved the generalization capability of sarcasm detection systems across languages. Nevertheless, sarcasm remains highly dependent on linguistic, cultural, and contextual factors that vary significantly from one language to another. Consequently, models trained on multilingual corpora often experience performance degradation when applied to low-resource languages such as French, highlighting the continuing need for language-specific datasets and architectures adapted to their linguistic characteristics. Table 2 presents a summary of several automatic sarcasm detection models in various languages.

Table 2. A comparative study of research into the detection of sarcasm in different languages

Studies	Method	Dataset	Key Contribution	Limitation
Lin and Hsieh (2016)	Crowdsourced corpus	Chinese user comments	First Chinese sarcasm corpus	Limited to conversational data
Gong et al. (2020)	Language-specific corpus	Chinese corpus	Developed a dedicated Chinese dataset	Limited generalization
Jia and Zan (2022)	BERT + Bi-GRU	Chinese social media	Captures contextual word relations	Weak on implicit and multimodal sarcasm
Zhang et al. (2022)	Retrospective Reader	Chinese dataset	Exploits temporal/contextual information	Difficulty with implicit sarcasm
Tian and Yang (2024)	BERT + Title Attention + Bi-LSTM	Chinese news/text	Joint modelling of title and content	Sensitive to ambiguous title-text relations
Swami et al. (2018)	English-Hindi corpus	Hinglish Tweets	First Hindi-English sarcasm dataset	Limited diversity
Aggarwal et al. (2020)	Bilingual Word Embeddings	Hinglish Tweets	Word-embedding representation	Difficult to transfer to pure Hindi
Kumar et al. (2023)	Word + Emoji Embeddings	Indian language social media	Combines textual and emoji information	Limited handling of subtle sarcasm
Al-Ghadhban et al. (2017)	Contextual word analysis	Arabic Twitter	Early Arabic sarcasm detector	Weak on implicit sarcasm
Alharbi and Lee (2021)	Multi-task Learning	Arabic corpus	Joint sarcasm and sentiment analysis	Scalability issues
Obeidat et al.	Fine-tuned BERT	Multiple Arabic	Demonstrated benefit of	Strong dependence on

(2022)		datasets	transfer learning	dataset quality
Abuein et al. (2024)	Deep Learning (ArSa-Tweets)	Arabic Tweets	Robust tweet classification	Lower performance on implicit sarcasm
Ayyasamy (2021)	Hybrid + Word Embeddings	English news headlines	Hybrid architecture	Less effective on informal text
Jose and Benedict (2023)	DeepASD (Multimodal)	Text + Facial emotions	Multimodal sarcasm detection	Requires high-quality visual data
Heraldi and Ruskanda (2024)	Fine-tuned Llama 2	English corpus	First LLM-based sarcasm detection	Weak on implicit sarcasm
Farhan et al. (2024)	GloVe + Bi-LSTM	Cross-platform social media	Effective hybrid deep learning model	Context-dependent optimization
Ibrahim et al. (2025)	CNN + ELSTM + Emoji	English Tweets	Integrates emoji semantics	High computational cost
Derbala et al. (2024)	Deep Learning	Multilingual corpus	Multilingual sarcasm detection	Performance decreases on complex multilingual data
Sukhavasi and Dondeti (2024)	Transformer	Multilingual datasets	Transformer-based multilingual detection	Cultural adaptation remains limited
Yue et al. (2024)	SarcNet Dataset	Multilingual multimodal corpus	Large multilingual multimodal benchmark	Dataset contribution only
Simon et al. (2022)	Naïve Bayes Multinomial	TransCasm (≈2,000 tweets)	First French bilingual sarcasm dataset	Small corpus; simple baseline; limited robustness

Overall, recent advances demonstrate a clear transition from traditional machine learning methods toward deep learning, transformer-based architectures, and, more recently, Large Language Models for sarcasm detection. Although remarkable progress has been achieved for English, Chinese, Hindi, Arabic, and multilingual environments, research on French sarcasm detection remains relatively limited. Existing studies mainly rely on contextual language representations and are often constrained by the scarcity of large annotated French corpora. These limitations motivate the present work, which introduces a new large-scale French sarcasm corpus together with a hybrid multi-task framework that jointly exploits contextual and sentiment-aware representations.

Methodology

Data acquisition

Acquiring a dataset for the study of sarcasm presents considerable challenges. While a substantial body of research addresses sarcasm in English, datasets about French sarcasm are notably scarce. Indeed, limited research has been conducted on the automatic detection of sarcasm within the French language. The sole publicly available dataset in the literature is the

TransCasm dataset (Simon et al., 2022). TransCasm represents the pioneering bilingual corpus of sarcastic tweets that includes French. In this work, we present the construction of a dataset comprising 22,000 instances, composed of sarcastic and non-sarcastic French utterances. The methodology for constructing this dataset is illustrated in Figure 1.

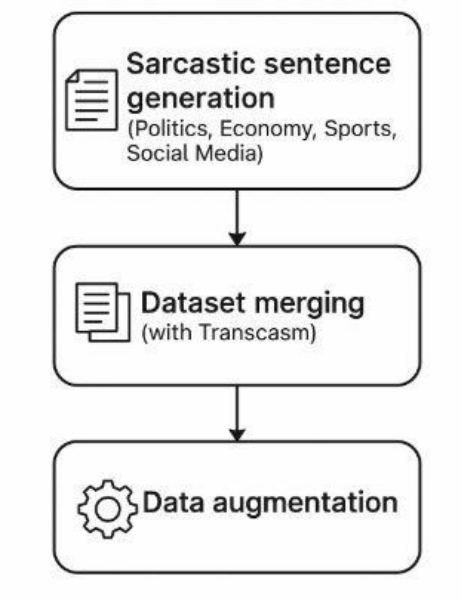


Figure 1. Dataset construction steps

TransCam dataset

The TransCasm dataset was derived from the SIGN corpus, introduced by Peled and Reichart (2017), and consists of 3,000 sarcastic tweets. The dataset has an average tweet length of 13.87 tokens and a vocabulary of 8,788 unique words. To build the SIGN corpus, the authors leveraged the X (formerly Twitter) API to collect tweets containing the hashtag #sarcasm between January 2016 and June 2016. Following collection, the authors commissioned five human annotators to provide sincere, non-sarcastic interpretations of these tweets, aiming to ascertain the original intended meaning. This process yielded a monolingual parallel corpus where sarcastic English tweets were paired with their non-sarcastic English counterparts. To create TransCasm, the SIGN corpus was translated from English into French using MateCat, an openly accessible web-based translation tool. Within the corpus, each tweet is presented alongside its corresponding interpretations. The original tweet and its interpretations are demarcated by a comma. The tweets encompass a diverse range of topics, including weather, traffic, employment, and education.

TransCasm pre-processing and analysis

To optimize the utility of the TransCasm corpus, pre-processing is required. This initial stage facilitates the restructuring and normalization of the data. Restructuring the data entailed forming a tabular structure comprising each sarcastic utterance and its non-sarcastic

interpretation. Subsequently, spaces, square brackets, double apostrophes, and other superfluous characters were eliminated. Following these operations, each element within the resulting array conformed to the pattern: ['sarcasm 1, interpretation 1'; 'sarcasm 1, interpretation 2'; ...; 'sarcasm 1, interpretation 5'; ...; 'sarcasm n, interpretation 5'].

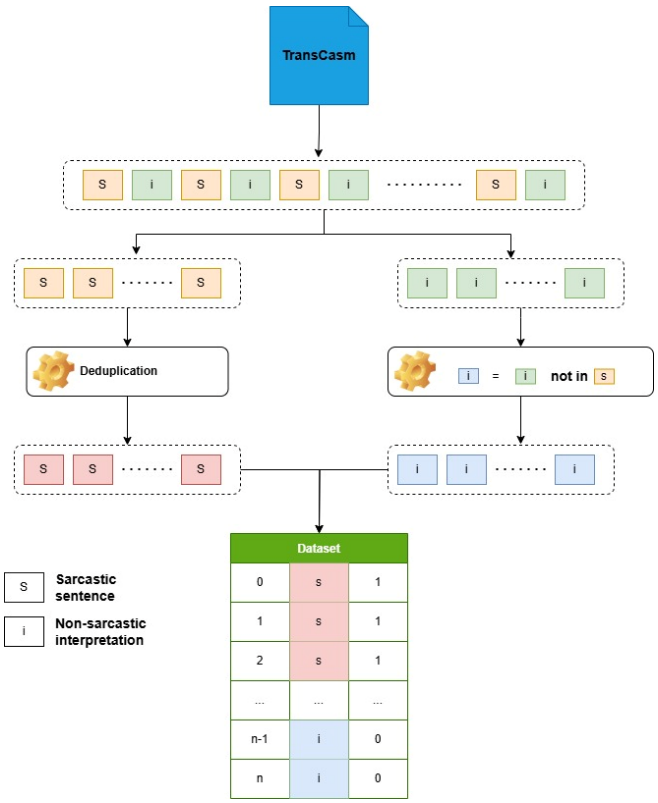


Figure 2. Pre-processing of the TransCasm dataset

The pre-processing and analysis workflow is summarized in Figure 2. Dataset normalization involved segregating sarcastic tweets from their corresponding non-sarcastic interpretations. This yielded two distinct lists: one of sarcastic comments, with 1,831 items, and another of non-sarcastic interpretations, each containing an equivalent number of items.

	Unnamed: 0	propos	sarcastique
0	0	ok ça va juste prendre quelques années de plu...	1
1	1	yelp jusqu'à 20 est-ce que einhorn l'a déjà v...	1
2	2	dressler sur le banc de touche problèmes à la...	1
3	3	j'adore comme plein de gens quittent clash po...	1
4	4	pcq je suis jeune il y a tellement d'autres f...	1
5	5	un autre chapitre d'un long livre sur la tota...	1
6	6	apparemment on a tous choisi fulmer pas vrai ?	1
7	7	j'aime totalement les appels téléphoniques de...	1
8	8	ah oui impossible que ce psychopathe ait comm...	1
9	9	excellent travail l'amérique vous avez élu do...	1

Figure 3. Pre-processed TransCasm DataFrame

Duplicate sarcastic comments and analogous non-sarcastic interpretations were subsequently removed, resulting in two datasets comprising 897 sarcastic comments and 1,647 non-sarcastic comments. Sarcastic instances were assigned a label of 1, while non-sarcastic instances received a label of 0, as depicted in Figure 3. Figure 4 presents the tag distribution within the pre-processed dataset. An imbalance between sarcastic and non-sarcastic comments is evident. Consequently, balancing this dataset necessitated reducing the proportion of non-sarcastic data to 897 instances, which meant reducing the sarcastic data too to match. This adjustment yielded a balanced dataset comprising 1,794 labeled instances. Given the inherently textual nature of the data examined in this study, the application of automatic natural language processing techniques is crucial for dataset analysis, particularly for identifying linguistic components frequently employed in sarcastic sentence construction. Accordingly, extracting grammatical characteristics of the dataset's elements facilitated the derivation of Part-of-Speech (PoS) tag distributions for both sarcastic data from TransCasm and non-sarcastic data.

The POS shows that both sarcastic and non-sarcastic sentences are predominantly composed of nouns, verbs, and adpositions (ADP), though sarcastic sentences exhibit a higher frequency of pronouns. Crucially, this analysis of PoS tag distribution provides insight into which tags might be effectively leveraged for subsequent dataset expansion. The subsequent sections will elaborate on how this information can be utilized to augment the dataset size.

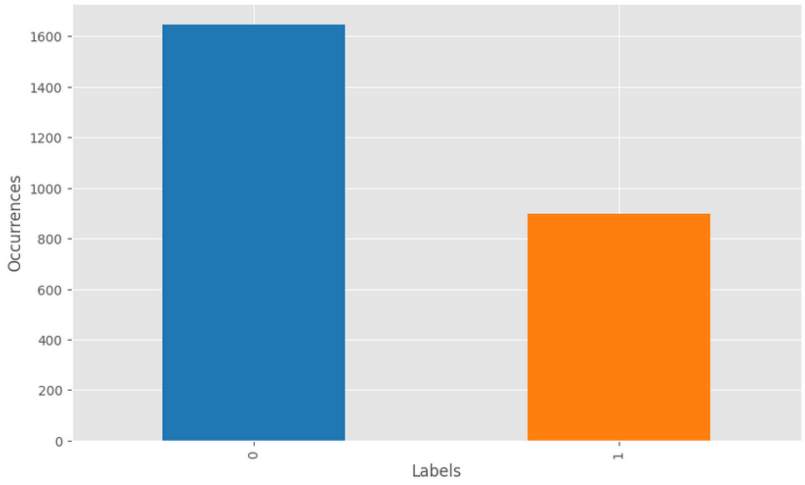


Figure 4. Label distribution

An attempt to construct a set of sarcastic data

The initial dataset of 1,800 instances proved inadequate for developing a robust sarcasm detection model in French. To address this limitation, a larger dataset comprising 22,000 instances was constructed, integrating Transcasm's preprocessed data with new data collected and annotated by French language experts.

Current restrictions on tweet extraction from X precluded direct acquisition of French sarcastic tweets. Consequently, an existing Twitter dataset primarily composed of English tweets2 from a prior sarcasm detection study was utilized. This dataset contained 15,000 sarcastic and 15,000 non-sarcastic tweets.

The sarcastic English data were extracted and pre-processed using the same methodology as the Transcasm dataset. Subsequently, these data were translated into French using a machine translation tool and categorized by domain (politics, economics, sport, social media). Each domain-specific data subset was then assigned to a corresponding expert. The experts' task was to ascertain whether the sarcastic intent was preserved in the French translation. If the sarcastic meaning was not retained, the sentence was manually revised to re-establish its sarcastic nature. This rigorous process yielded 5,000 sarcastic sentences. Non-sarcastic sentences were generated by direct translation of the non-sarcastic English tweets from the same source dataset.

This procedure resulted in a dataset of 10,000 labeled sentences, which was subsequently merged with the Transcasm dataset, culminating in a combined dataset of approximately 12,000 labeled sentences. To further expand the dataset size, data augmentation techniques were employed.

Data augmentation

Data augmentation typically leverages human understanding of invariances, rules, or heuristics; for instance, the class of an object in an image remains consistent even if the image is rotated.

In natural language processing, establishing universal rules that consistently guarantee the quality of augmented data across diverse domains presents a significant challenge. A conventional method involves replacing words with synonyms derived from curated ontologies, such as WordNet, or based on semantic similarity. However, the limited number of words possessing exact or near-exact synonyms restricts the applicability of synonym-based expansion to a small fraction of the vocabulary. As previously noted, sarcasm entails conveying positive sentiments to imply negative intent or ridicule. Given its inherent subtlety, linguistic manipulations such as back-translation or reverse-engineering of words are highly prone to altering the original sarcastic meaning of an utterance.

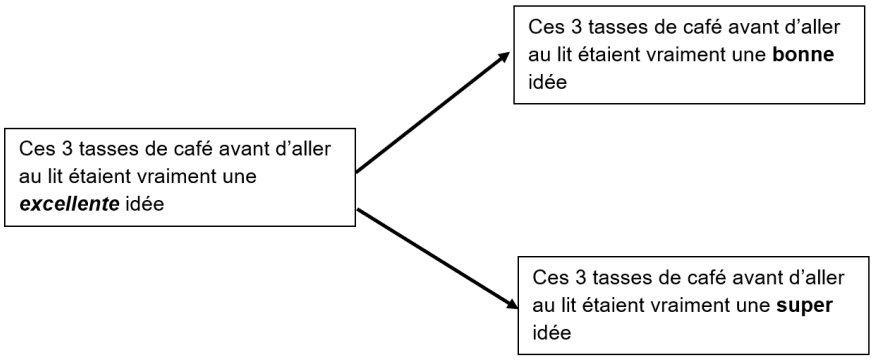


Figure 5. Illustration of the increase in data

Prior research on the grammatical structure of sarcasm indicates a strong correlation with specific grammatical elements, such as adverbs. Consequently, adjectives were identified as the sole grammatical category that could be manipulated without compromising the sarcastic semantics. Our approach involved identifying adjectives within each sarcastic comment. For each identified adjective, two synonyms were generated. This process was iteratively applied to other adjectives present in the sentence, if any (refer to Figure 5 for an illustrative example). This method yielded multiple sentences preserving the original semantics but with minor lexical variations.

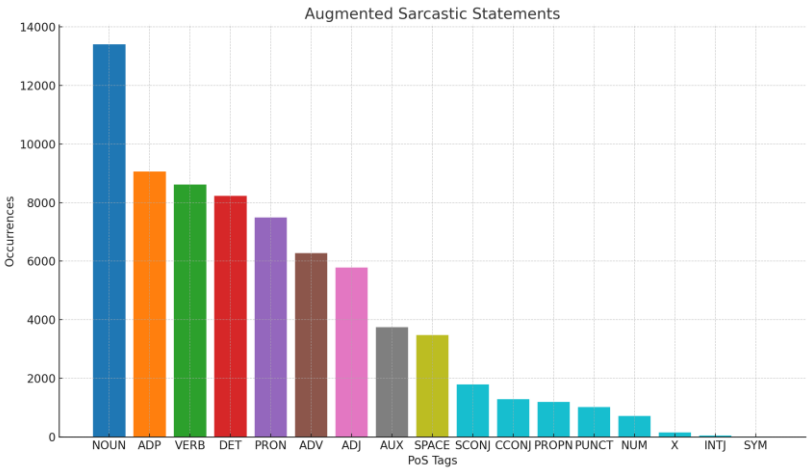


Figure 6. PoS of the dataset after augmentation

The pre-trained Camembert model (Martin, 2020) was utilized for synonym generation to facilitate this data augmentation. Camembert inherently possesses a word generation capability, enabling it to accurately predict missing words when strategically integrated into a sentence, thereby demonstrating its advanced linguistic proficiency. This methodology expanded the dataset from approximately 12,000 entries to around 22,000. As depicted in Figure 6, a substantial increase in data volume was achieved. Importantly, the structural integrity of the sentences, including the order and presence of grammatical elements, was preserved. Furthermore, a similarity assessment was conducted among the sarcastic sentences within the augmented dataset. Figure 7 illustrates that less than 8% of the sarcastic instances exhibited a similarity score exceeding 0.5. This low proportion of highly similar sentences indicates a broad coverage within the dataset and confirms that the generation process did not introduce a significant number of near-identical sentences.

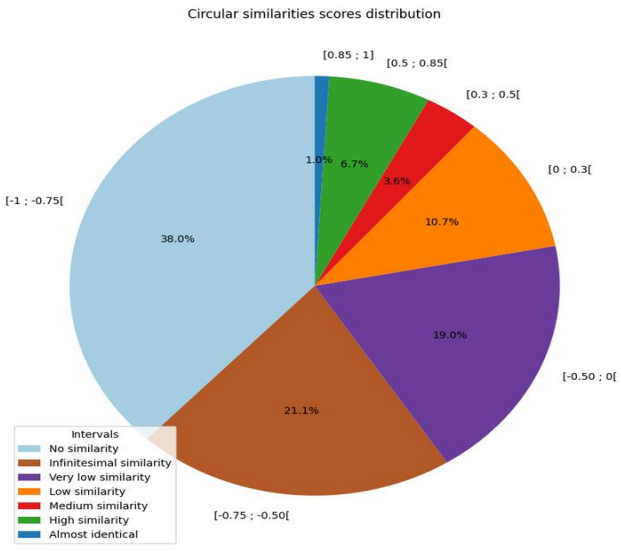


Figure 7. Similarity distribution score between sarcastic sentence

Tokenization step

This section details the tokenization process, wherein the sarcasm dataset is adapted for compatibility with the CamemBERT model's input requirements. Initially, the maximum sequence length, encompassing both sarcastic and non-sarcastic utterances, was empirically determined to be 221 tokens. Subsequently, CamemBERT's tokenizer is employed, using the predefined maximum sequence length as a parameter, to transform each input sequence into a 2 x 211 tensor. This tensor comprises a 211-dimensional vector for token IDs and a corresponding vector of the same dimension for the attention mask. Sequences shorter than 211 tokens are padded to this maximum length; padding tokens are assigned a value of 1, while their corresponding attention mask entries are set to 0.

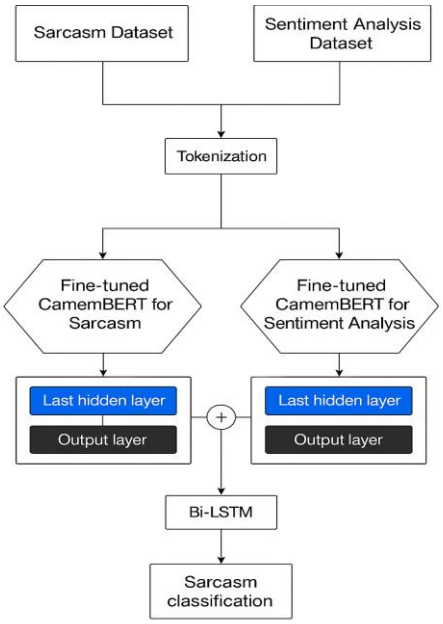


Figure 8. Architecture of the model

CamemBERT fine-tuning

In this work, the CamemBERT model is fine-tuned at two levels. CamemBERT represents an adaptation of the RoBERTa (Robustly Optimized BERT Approach) model, specifically tailored for the French language and built upon the Transformer architecture. Developed by Facebook AI in 2019, RoBERTa is a natural language processing model that improves upon BERT by optimizing training methodologies and leveraging a larger volume of data (Liu, 2019).

CamemBERT Fine-tuning for Sarcasm Detection

Following data tokenization, the CamemBERT model underwent training on the designated dataset to specialize in sarcasm detection. After defining the hyperparameters and completing the tokenization process, the model was trained on the dataset. Figure 9 illustrates the

progression of both loss and accuracy for the CamemBERT model, specifically fine-tuned for sarcasm detection.

CamemBERT Fine-tuning for sentiment analysis

The CamemBERT model was fine-tuned for sentiment analysis using the dataset introduced by Radford (2019). This dataset, originating from X, comprises French sentiment analysis data that were translated from English. The CamemBERT model was trained on this sentiment analysis dataset, employing the same hyperparameters utilized in the sarcasm detection task.

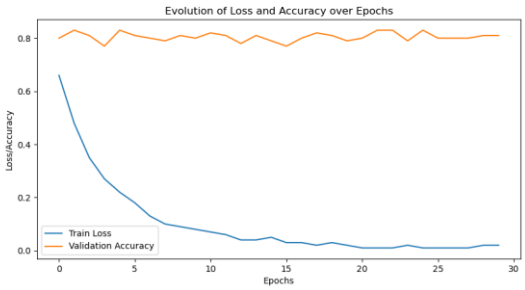


Figure 9. Loss and accuracy evolution curve of the CamemBERT fine-tuned model for sarcasm detection

Figures 10 presents the loss and accuracy values obtained from the CamemBERT model fine-tuned for sentiment analysis.

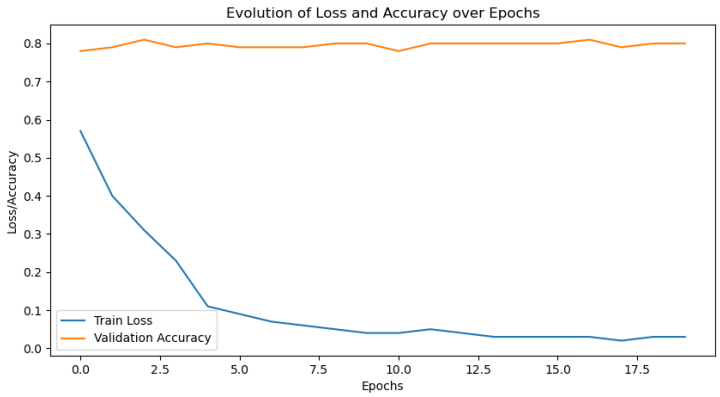


Figure 10. Loss and accuracy evolution curve of the CamemBERT fine-tuned model for sentiment analysis

Bi-LSTM Classifier for Sarcasm Detection

Deep recurrent neural networks (RNNs) have shown considerable efficacy in Natural Language Processing (NLP) tasks. This effectiveness stems from their recurrent layers, which incorporate feedback loops, enabling the retention of information as a form of temporal memory. However, training conventional RNNs to capture long-term temporal dependencies,

as required in tasks involving sentences or tweets, presents challenges due to the vanishing gradient problem, where the loss function's gradient diminishes exponentially over time. Consequently, Bidirectional Long Short-Term Memory (Bi-LSTM) networks are employed. Bi-LSTMs are a class of recurrent neural networks designed to process sequential data in both forward and backward directions.

As depicted in Figure 11, this bidirectional processing, combined with the capabilities of LSTMs, allows the model to capture both past and future contextual information from the input sequence.

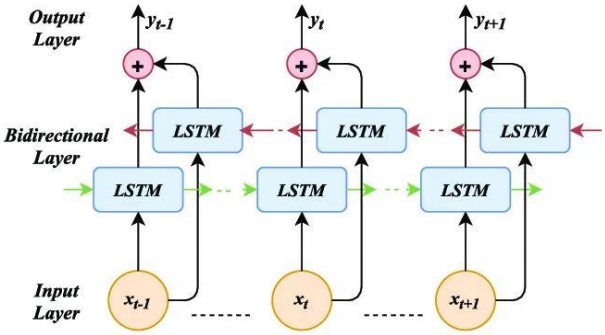


Figure 11. Bi-LSTM classifier architecture

After the fine-tuning of the two models—one dedicated to sarcasm detection and the other to sentiment analysis—their respective encoder outputs were extracted and subsequently concatenated. This concatenated output then serves as input to a Bi-LSTM model, which utilizes these encoded data to predict the class of each expression.

Results

Experimental validation of the model was conducted in multiple stages. Initially, the Transcasm dataset was utilized, followed by experiments on the newly constructed dataset.

Experiments on the Transcasm dataset

A series of evaluations were conducted using the Transcasm dataset, employing various models. Initially, the data were trained using a Naive Bayesian Multinomial Classifier (NBCM). This classifier was previously employed in (Simon et al., 2022) for French sarcasm detection with the Transcasm dataset.

Subsequently, the CamemBERT model was trained on Transcasm. This was followed by a configuration combining CamemBERT with a Random Forest classifier. Further experiments involved coupling the CamemBERT model with a BiLSTM classifier. Finally, our hybrid model, integrating sentiment analysis with a Bi-LSTM classifier, was trained.

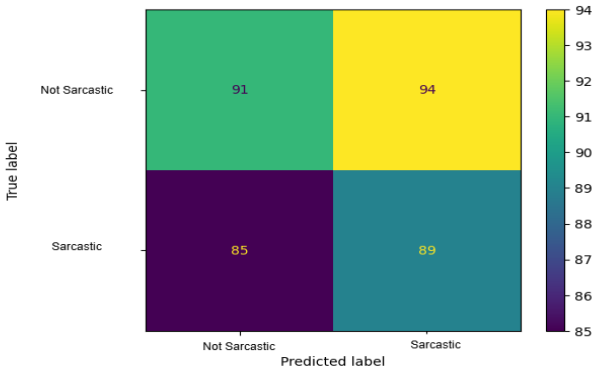


Figure 12. Confusion matrix for the CNBM-based model

Figure 12 presents the confusion matrix for the CNBM model. The matrix indicates an accuracy of 51% on the Transcasm dataset. This performance suggests the model's limited capacity to distinguish between sarcastic and non-sarcastic comments. Deployment of the CamemBERT model led to an improvement in sarcasm detection performance, with accuracy approaching 77% (Figure 13).

	precision	recall	f1-score	support
0	0.73	0.85	0.79	179
1	0.82	0.69	0.75	180
accuracy			0.77	359
macro avg	0.78	0.77	0.77	359
weighted avg	0.78	0.77	0.77	359

Figure 13. Performance of the fine-tuned CamemBERT sarcasm detection model

Combining the CamemBERT model with a Random Forest classifier yielded an accuracy of 88%, while its integration with a Bi-LSTM classifier achieved 89% accuracy.

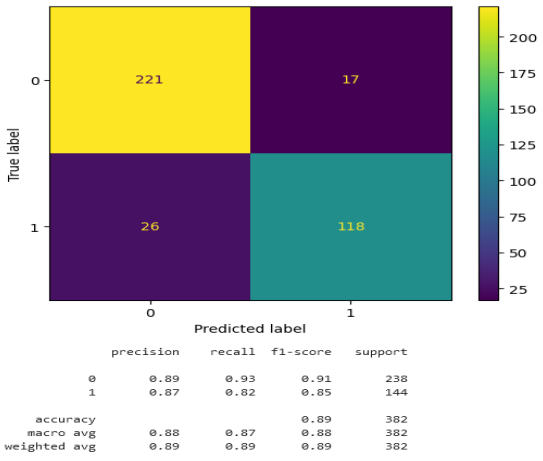


Figure 14. Confusion matrix for the hybrid model:
CamemBERT encoder and Bi-LSTM classifier

Figure 14 displays the confusion matrix and performance metrics for this hybrid configuration: for the CamemBERT encoder model coupled with BiLSTM.

The integration of a CamemBERT encoder with a classifier appears to be a promising approach for automated sarcasm detection in French. Given the potential negative implications of sarcasm, a sentiment analysis module was incorporated into this hybrid architecture, as detailed in the preceding section.

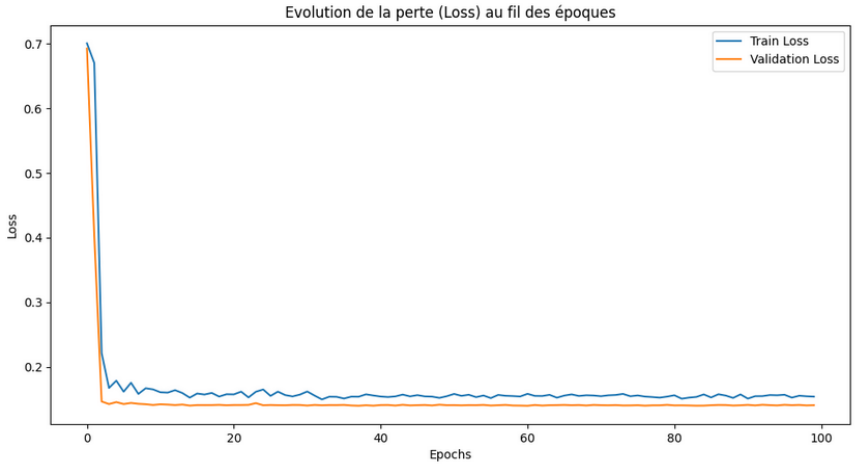


Figure 15. Evolution curve of the loss function of the model

Training the complete hybrid model on the Transcasm dataset yielded a substantial performance improvement over the preceding configuration. Figure 15 illustrates the progression of the loss function for the proposed model. The graph demonstrates that after 100 epochs, the loss function value decreased to below 0.1, indicating sufficient minimization of losses.

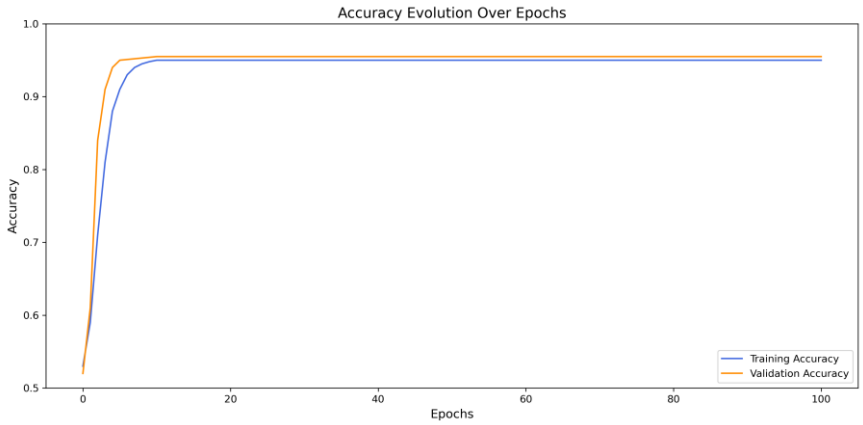


Figure 16. Evolution curve of the accuracy function of the model

Figure 16 depicts the evolution of model accuracy during the training process. During training, the model's accuracy approached 99%. Finally, the confusion matrix demonstrates an F1-score of 96% on the Transcasm dataset for the proposed model. However, it is important

to note that high performance on a relatively small dataset does not inherently guarantee robust generalization. To address this limitation, the model was subsequently trained on the custom-developed 22k dataset.

Experimentation with the new dataset

The model was trained on this dataset, as detailed in Section X. The F1-score during the training phase was 98% for the new dataset, marginally lower than the approximately 99% observed with the Transcasm dataset.

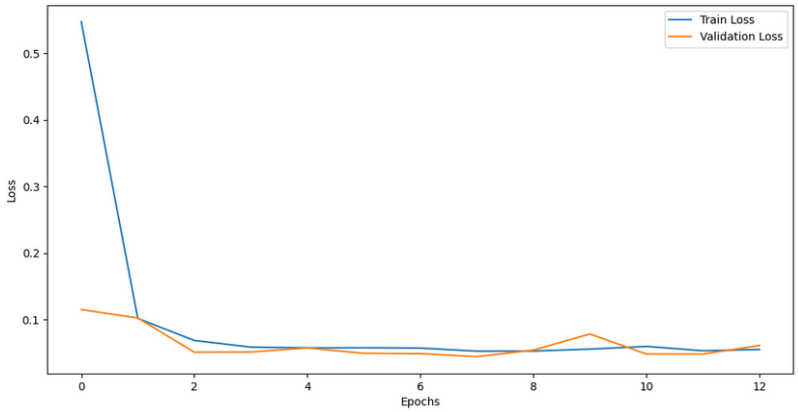
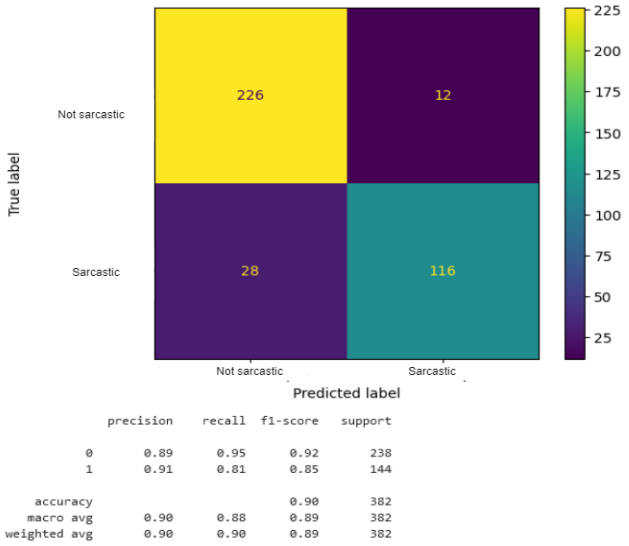


Figure 17. Evolution of lost function for the new dataset

Conversely, the loss function exhibited a sawtooth pattern, as depicted in Figure 17, closely mirroring the behavior observed when utilizing the Transcasm dataset. This characteristic is attributed to the larger scale of our dataset, which necessitated a significantly extended training duration and greater computational resources compared to the Transcasm data. The analysis of these curves indicates a proficient training process, characterized by a negligible error margin. As detailed in the confusion matrix, our model achieved an F1-score of 98%, representing a 2% improvement over the performance associated with the Transcasm dataset. A primary distinction, however, lies in the model's predictive performance.

Figure 18 shows that the model yielded an F1-score of 90% on the new dataset, in contrast to 96% achieved on the Transcasm dataset. It is noteworthy that the model demonstrated enhanced robustness when evaluated on our dataset compared to the Transcasm dataset, which comprises merely 2,000 data instances. The newly constructed dataset is not only larger but, crucially, encompasses diverse information sources, including data from multiple domains, tweets, syntactically complete sentences, and newspaper headlines. This broader scope is anticipated to improve the model's applicability for sarcasm detection in varied professional contexts. Furthermore, potential errors originating from the sentiment analysis phase may propagate and influence the subsequent sarcasm classification.



	precision	recall	f1-score	support
0	0.89	0.95	0.92	238
1	0.91	0.81	0.85	144
accuracy			0.90	382
macro avg	0.90	0.88	0.89	382
weighted avg	0.90	0.90	0.89	382

Figure 18. Confusion matrix for the new dataset

Discussion

The experimental results demonstrate the effectiveness of the hybrid multi-task framework for French sarcasm detection. Across both evaluation datasets, the adopted architecture consistently outperformed previously reported approaches, confirming that the combination of contextual semantic representations and sentiment-aware information provides a more comprehensive understanding of sarcastic expressions.

One of the main findings of this study is that incorporating sentiment analysis as an auxiliary task significantly improves sarcasm detection performance. Unlike literal statements, sarcastic utterances frequently express an implicit polarity that differs from their apparent lexical sentiment. Consequently, relying exclusively on contextual embeddings may not always be sufficient to capture the underlying intention of the speaker. By jointly learning contextual representations through CamemBERT and affective information through the sentiment analysis module, the designed architecture is better able to distinguish between literal and sarcastic expressions. This observation is consistent with previous studies suggesting that sentiment inconsistency constitutes one of the strongest indicators of sarcasm.

Another important observation concerns the contribution of transfer learning. The use of CamemBERT, which has been pre-trained on a large French corpus, enables the model to capture rich contextual and syntactic information that would be difficult to learn from a relatively limited sarcasm dataset alone. Fine-tuning this language model on the sarcasm detection task considerably improves feature representation compared with conventional machine learning approaches relying on handcrafted features or static word embeddings.

The integration of the Bi-LSTM classifier also contributes to the overall performance of the designed model. Although transformer models efficiently encode contextual information, sequential recurrent layers remain effective for modelling local sequential dependencies and refining contextual representations before the final classification stage. The complementary roles of CamemBERT and Bi-LSTM therefore contribute to the robustness of the hybrid architecture.

The model achieved an F1-score of 96% on the TransCasm dataset, representing a substantial improvement over the previously reported state-of-the-art results. This improvement demonstrates that combining contextual language modelling with sentiment-aware representations provides significant benefits for French sarcasm detection. Furthermore, the model maintained a high F1-score of 90% on the newly constructed corpus containing approximately 22,000 instances. The relatively small decrease in performance between the two datasets suggests that the model generalizes well across different writing styles and textual sources, including both informal social media messages and more structured French texts.

Beyond this, one of the major contributions of this work lies in the construction of a new large-scale French sarcasm corpus. The scarcity of annotated resources has long represented one of the principal obstacles to the development of French sarcasm detection systems. By providing a substantially larger dataset than existing French benchmarks, this work contributes to the evaluation of the adopted approach, and also to the future development and comparison of deep learning approaches for French figurative language understanding.

Overall, the findings demonstrate that combining transfer learning, sentiment-aware representation learning, and sequential modelling constitutes an effective strategy for sarcasm detection in low-resource languages. More broadly, this work contributes to reducing the resource gap in French Natural Language Processing by introducing both a new benchmark dataset and a robust hybrid architecture that can serve as a foundation for future research on figurative language understanding.

Conclusion

This paper presented a hybrid multi-task learning framework for automatic French sarcasm detection that combines transfer learning, sentiment-aware representation learning, and sequential deep learning. By integrating a fine-tuned CamemBERT encoder, a dedicated sentiment analysis module, and a Bi-LSTM classifier. The model provides an effective solution for modeling the semantic and affective characteristics of sarcastic expressions in French. Beyond the proposed architecture, this work contributes to addressing one of the major challenges in French Natural Language Processing, namely the scarcity of publicly available resources for sarcasm detection. The construction of a large-scale annotated French

sarcasm corpus containing approximately 22,000 instances constitutes a valuable contribution that can support future research on figurative language understanding, sentiment analysis, and related Natural Language Processing tasks. The experimental evaluation demonstrates the competitiveness of the adopted framework on both an existing benchmark and the newly developed corpus, confirming its suitability for French sarcasm detection across different textual sources. More importantly, the results suggest that combining contextual language representations with sentiment-aware information constitutes a promising direction for improving figurative language understanding in low-resource languages. Future work will focus on several research directions. First, the model will be extended to multimodal sarcasm detection by incorporating visual information, emojis, and conversational context commonly found on social media platforms. Second, the corpus will be further enriched with additional linguistic phenomena and contextual annotations in order to facilitate more comprehensive research on figurative language understanding in French.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

Abuein, Q., Ra'ed, M., Migdady, A., Jawarneh, M. S., & Al-Khateeb, A. (2024). ArSa-Tweets: A novel

Aggarwal, A., Wadhawan, A., Chaudhary, A., & Maurya, K. (2020). "Did you really mean what you said?": Sarcasm detection in Hindi-English code-mixed data using bilingual word embeddings. In *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)* (pp. 7-15).

Al-Ghadhban, D., Alnkhilan, E., Tatwany, L., & Alrazgan, M. (2017). Arabic sarcasm detection in Twitter. In *2017 International Conference on Engineering & MIS (ICEMIS)* (pp. 1-7). IEEE.

Alharbi, A. I., & Lee, M. (2021). Multi-task learning using a combination of contextualised and static word embeddings for Arabic sarcasm detection and sentiment analysis. In *Proceedings of the sixth Arabic natural language processing workshop* (pp. 318-322).

Arabic sarcasm detection system based on deep learning model. *Heliyon*, *10*(17).

Ayyasamy S. (2021). Construction of a hybrid model for English news headline sarcasm detection by word
embedding technique. Journal of Electrical Engineering and Automation, 3:184–198, 11.

Băroiu, A. C., & Trăușan-Matu, Ș. (2022). Automatic sarcasm detection: Systematic literature review. *Information*, *13*(8), 399.

Bharti, S. K., Sathya Babu, K., & Jena, S. K. (2017). Harnessing online news for sarcasm detection in Hindi tweets. In *International conference on pattern recognition and machine intelligence* (pp. 679-686). Cham: Springer International Publishing.

Derbala Yacoub, A., Elsayed Aboutabl, A., & O Slim, S. (2024). Multilingual sarcasm detection for enhancing sentiment analysis using deep learning algorithms. *Journal of Communications Software and Systems*, *20*(4), 278-289.

Dubey, A., Joshi, A., & Bhattacharyya, P. (2019). Computational sarcasm for different languages: A survey.

Farabi, S., Ranasinghe, T., Kanojia, D., Kong, Y., & Zampieri, M. (2024). A survey of multimodal sarcasm detection. *arXiv preprint arXiv:2410.18882*.

Farhan, S., Shoukat, R., & Aslam, A. (2024). Automatic Sarcasm Detection on Cross-Platform Social Media Datasets: A GLoVe and Bi-LSTM Based Approach. *Journal of Universal Computer Science (JUCS)*, *30*(5).

Frenda, S. (2023). Sarcasm and Implicitness in Abusive Language Detection: A Multilingual Perspective. *Procesamiento del Lenguaje Natural*, *70*, 239-242.

Ganganwar, V., Manvainder, Singh, M., Patil, P., & Joshi, S. (2024). Sarcasm and humor detection in code-mixed Hindi data: A survey. In *International Conference on Computing and Machine Learning* (pp. 453-469). Singapore: Springer Nature Singapore.

Gong, X., Zhao, Q., Zhang, J., Mao, R., & Xu, R. (2020). The design and construction of a Chinese sarcasm dataset. In *Proceedings of the twelfth language resources and evaluation conference* (pp. 5034-5039).

Hassan, M. A., García-Méndez, S., & de Arriba-Pérez, F. (2026). Contribution to Sarcasm Detection in Arabic Using Natural Language Processing Techniques. *Applied Sciences*, *16*(6), 2724. https://doi.org/10.3390/app16062724

Heraldi, F. D., & Ruskanda, Z. (2024). Effective intended sarcasm detection using fine-tuned llama 2 large language models. In *2024 11th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA)* (pp. 1-6). IEEE.

Ibrahim, S. A. M., Deshpande, M. M., & Pawar, V. N. (2025). Automated Sarcasm Detection in English Tweets Using CCNN and ELLSTM with Text and Emoji Embeddings. In *2025 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)* (pp. 1-8). IEEE.

Jia, C., & Zan, H. (2022). Context-Based Sarcasm Detection Model in Chinese Social Media Using BERT and Bi-GRU Models. In *CONF-SPML* (pp. 42-50).

Jose, J. M., & Benedict, S. (2023). Deepasd framework: A deep learning-assisted automatic sarcasm detection in facial emotions. In *2023 8th International Conference on Communication and Electronics Systems (ICCES)* (pp. 998-1004). IEEE.

Kumar, A., Sangwan, S. R., Singh, A. K., & Wadhwa, G. (2023). Hybrid deep learning model for sarcasm detection in Indian indigenous language using word-emoji embeddings. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(5), 1-20.

Lin, S. K., & Hsieh, S. K. (2016). Sarcasm detection in Chinese using a crowdsourced corpus. In *Proceedings of the 28th Conference on Computational Linguistics and Speech Processing (ROCLING 2016)* (pp. 299-310).

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized BERT pre-training approach. *arXiv preprint arXiv:1907.11692*.

Martin, L., Muller, B., Suarez, P. O., Dupont, Y., Romary, L., de La Clergerie, É. V., ... & Sagot, B. (2020). CamemBERT: a tasty French language model. In *Proceedings of the 58th annual meeting of the Association for Computational Linguistics* (pp. 7203-7219).

Mazumder, D., Kumar, A., & Patro, J. (2024). Revealing the impact of synthetic native samples and multi-tasking strategies in Hindi-English code-mixed humour and sarcasm detection. *arXiv preprint arXiv:2412.12761*.

Mihi, S., Ait Ben Ali, B., & Laachfoubi, N. (2022). A Comparative Review of Tweets Automatic Sarcasm Detection in Arabic and English. In *International conference on advanced intelligent systems for sustainable development* (pp. 841-849). Cham: Springer Nature Switzerland.

Mihi, S., Ait Benali, B., & Laachfoubi, N. (2023). Automatic sarcasm detection in Arabic tweets: resources and approaches. *Journal of Intelligent & Fuzzy Systems*, *45*(6), 9483-9497.

Naik, P. K., Chenjeri, S. S., Sruthy, S., & Mamatha, H. (2022). Sarcasm Detection in English Text using Tweets and Headlines. In *2022 International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)* (pp. 192-196). IEEE.

Obeidat, R., Bashayreh, A., & Younis, L. B. (2022). The impact of combining Arabic sarcasm detection datasets on the performance of a BERT-based model. In *2022 13th International Conference on Information and Communication Systems (ICICS)* (pp. 22-29). IEEE.

Pandey, R., Kumar, A., Singh, J. P., & Tripathi, S. (2025). A hybrid convolutional neural network for sarcasm detection from multilingual social media posts. *Multimedia Tools and Applications*, *84*(16), 15867-15895.

Peled, L., & Reichart, R. (2017). Sarcasm SIGN: Interpreting sarcasm with sentiment-based monolingual machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1690-1700).

Pokhriyal, H., & Jain, G. (2025). Supposititious sarcasm detection and sentiment analysis coping with Hindi language in social networks harnessing Zipf-Mandelbrot probabilistic optimisation and perplexity entropy learning. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *24*(2), 1-28.

Prajapati, A., Singh, A., Arora, V., & Bansal, N. (2024). Sarcasm detection in English-Hindi code-mixed text data. In *2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)* (Vol. 1, pp. 94-99). IEEE.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), 9.

Rašl, M., Žalik, M., & Keršič, V. (2021). Transformer-based Sarcasm Detection in English and Slovene Language. In *Proceedings of the 2021 7th Student Computer Science Research Conference* (p. 43).

Sharma, E., N. K. Gondhi, Chaahat, M. Murugappan, G. Naik, and L. Murugan. (2026). "Toward Effective Sarcasm Detection in Social Media: A Review of Artificial Intelligence-Based Automated Approaches." Expert Systems43, no. 3: e70203. https://doi.org/10.1111/exsy.70203.

Simon, D., Castilho, S., Lohar, P., & Afli, H. (2022). TransCasm: A Bilingual Corpus of Sarcastic Tweets. In *Proceedings of the LREC 2022 workshop on Natural Language Processing for Political Sciences* (pp. 98-103).

Sukhavasi, V., & Dondeti, V. (2024). Effective automated transformer model-based sarcasm detection using multilingual data. *Multimedia Tools and Applications*, *83*(16), 47531-47562.

Swami, S., Khandelwal, A., Singh, V., Akhtar, S. S., & Shrivastava, M. (2018). A corpus of English-Hindi code-mixed tweets for sarcasm detection. *arXiv preprint arXiv:1805.11869*.

Thorat, M., & Shaikh, N. F. (2024). Novel Deep Neural Network Approach for the Sarcasm Detection in Hindi Language. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(10s), 487–494.

Tian, Y., & Yang, S. (2024). Chinese Sarcasm Detection Using Title-Attention and Bidirectional LSTM Based on Bert. In *2024 13th International Conference of Information and Communication Technology (ICTech)* (pp. 34-38). IEEE.

Wu, B., Tian, H., Liu, X., Hu, W., Yang, C., & Li, S. (2024). Sarcasm detection in Chinese and English text with fine-tuned large language models. In *2024 IEEE 10th Conference on Big Data Security on Cloud (BigDataSecurity)* (pp. 47-51). Ieee.

Yue, T., Shi, X., Mao, R., Hu, Z., & Cambria, E. (2024). SarcNet: a multilingual multimodal sarcasm detection dataset. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 14325-14335).

Zhang, L., Zhao, X., Song, X., Fang, Y., Li, D., & Wang, H. (2022). A novel Chinese sarcasm detection model based on retrospective reader. In *International Conference on Multimedia Modeling* (pp. 267-278). Cham: Springer International Publishing.

Bibliographic information of this paper for citing:

Tchantchou Samen, Yannick U.; Djafar, Abou A.; Guidedi, Kaladzavi & Laleye, Fréjus A. A. (2026). Advancing Automatic Sarcasm Detection in the French Language. *Journal of Information Technology Management*, 18 (3), 202-227.
https://doi.org/10.22059/jitm.2026.414898.4482