LACELL at PoliticHeadlinES-IberLEF 2026: A Late Fusion System Combining Semantic, Linguistic, and Visual Signals for Headline Ranking

Desirée Martínez Carrión[1,*,†], Ángela Almela[1,†] and Pascual Cantos-Gómez[1,†]

[1]*Facultad de Letras, Universidad de Murcia, Campus de La Merced, 30001 Murcia, Spain*

Abstract

This paper describes the participation of the LACELL team in the PoliticHeadlinES shared task at IberLEF 2026, which focuses on headline ranking for Spanish political news. The primary challenge of this task lies in the presence of hard distractors that share overlapping entities, topics, and temporal frames with the target headline. To address this, we propose a late fusion architecture that integrates three independent but complementary signals: semantic similarity via multilingual E5 base sentence embeddings, fine-grained linguistic and stylistic features extracted via UMUTextStats, and visual-textual alignment using CLIP. In a strictly zero-shot setting, our unsupervised system achieved an official score of 0.66 on the CodaBench leaderboard. These results demonstrate that combining deep semantic embeddings with interpretable linguistic and visual signals offers a robust baseline that successfully complements global semantic similarity in political framing tasks.

Keywords

UMUTextStats, Linguistic Features, Sentence Embeddings, Headline Ranking, Late Fusion, Political Framing

1. Introduction

In the current digital era, political communication is predominantly consumed through digital platforms such as X (formerly Twitter), Facebook, and so on. Nevertheless, this content often take the form of sensationalist headlines that may incur clickbait and misinformation spreading, serving as the primary window users have into social media. This practice is the perfect gateway for attention and framing. Drawing on the foundational work of Entman [1], framing entails how information influences human consciousness, depending on how it is conveyed. In political journalism, selecting a specific headline serves an editorial purpose, highlighting particular aspects of reality to shape public opinion or manage public scrutiny (establishing urgency or neutrality, for instance). Evaluating stylistic and editorial fidelity of these headlines remains a critical challenge for Natural Language Processing (NLP) applied to political discourse [2].

This paper details the system developed by our team for the PoliticHeadlinES shared task at IberLEF 2026[3]. The task requires ranking ten candidate headlines for a given news article to place the ground truth at the top. To increase difficulty, the dataset introduces 'hard distractors' that share overlapping entities or topics, and temporal frames with the original headline. Under van Dijk's [4] theory of news as discourse, headlines function as macro-structures that require a deeper understanding of communicative intent beyond simple keyword matching. Dominant NLP methodologies frequently fall short of capturing complex and subtle linguistic cues, exposing a cross-disciplinary gap in purely automated tools [5]. Indeed, capturing fine-grained morphosyntactic and pragmatic nuances remains a significant challenge for Large Language Models (LLMs) [6, 7], while the crucial visual elements of editorial framing are often overlooked in text-only analyses [8].

To address this gap, we propose a late fusion architecture that integrates three independent but complementary signals:

IberLEF 2026, September 2026, León, Spain

*Corresponding author.

†These authors contributed equally.

desiree.martinezc@um.es (D. M. Carrión); angelalm@um.es (Á. Almela); pcantos@um.es (P. Cantos-Gómez)

0009-0000-0559-3974 (D. M. Carrión); 0000-0002-1327-8410 (Á. Almela); 0000-0001-6329-2352 (P. Cantos-Gómez)

- Semantic Similarity: Captures the propositional content using the multilingual E5 base model for weakly-supervised contrastive pre-training [9]. This model is specifically designed for retrieval-style tasks, where asymmetric inputs are compared (long document vs. short query, for instance, a headline). This leverages the bi-encoder foundations of Sentence-BERT [10], while ensuring interpretable signals in framing contexts.
- Linguistic Features: Analyses the morphosyntactic, pragmatic, and stylometric features of the headlines. We incorporate a large set of linguistic features covering: morphosyntax, lexical categories, discourse markers, psycholinguistic features (e.g., emotion, cognition), and stylometry (such as length, readability, or punctuation). We extract these variables by using UMUTextStats [11], an enhanced Spanish version of LIWC [12]. By incorporating this signal, our system captures fine-grained linguistic biases and the subtle framing differences that global embeddings often overlook.
- Visual Similarity: Models the alignment between the news image and the headline text using CLIP [13]. Crucially, as multimodal framing extends beyond text, visual elements are crucial to how a headline is portrayed and, therefore, transmitted to public opinion.

Then, by leveraging the Transformers library [14], our system prioritises interpretable signals to analyse the influence of each dimension on the final ranking without the need for extensive supervised fine-tuning.

2. Dataset

According to the organisers, the PoliticHeadlinES 2026 dataset is designed to evaluate the ability of systems to distinguish an original headline from a set of carefully selected distractors. These distractors share overlapping entities, topics, and temporal frames with the target text to increase difficulty, framing the task closer to information retrieval and semantic matching than standard text classification. Furthermore, each news article is assigned a unique MD5 hash to ensure consistent identification across all data splits and files.

This dataset entails two experimental settings corresponding to:

- Text-only subtask: Participants are provided only with the full news article text and candidate headlines.
- Multimodal subtask: Added to the textual content, an image associated with it is also provided to aid in semantic disambiguation.

The organizers partitioned the dataset into training, development, and testing splits. To guarantee consistent system training and evaluation, all partitions maintain the same data structure and schema. As an illustration, Table 1 outlines the format of the candidate headlines and the target ranking, where the ground truth (GT) headline must be positioned at the top.

3. System Overview

The core design rationale of our system is rooted in explicit compositional elements rather than model training. Instead of deploying a complex, non-interpretable supervised classifier, our architecture relies on a strategic combination of three complementary and independent signals via late fusion: semantic similarity, linguistic features, and visual similarity. This approach ensures that candidate selection remains strictly grounded on three verifiable pillars: macro-semantic contextualisation, explicit linguistic framing, and contextual grounding from images. Formally, each signal computes an independent score for every candidate headline, which is subsequently normalised and aggregated into a final ranking metric. Finally, to ensure reproducibility, the complete pipeline and implementation details have been released in a public repository[1].

[1]https://github.com/desireemcarrion/politicalheadline-2026-lacell

Table 1
Illustrative example of PoliticHeadlinES2026 dataset.

Rank	ID	Candidate headline
1	GT	El PP contraataca en el Senado y ve un avance de la amnistía en el Congreso: "Hoy es un día negro"
2	D1	Los negociadores reescriben la amnistía para salvar a Puigdemont
3	D2	La Justicia europea avala reclamar una amnistía si la ley que lo perdonó se deroga
4	D3	López Miras llama a la protesta «pacífica» contra la amnistía de Sánchez
5	D4	"Todos al Congreso. Ya": así se organizaron las dos protestas ultras de Ferraz contra la amnistía
6	D5	Las diferencias entre PSOE y Junts sobre el alcance de la amnistía alejan de esta semana la investidura...
7	D6	PP y Vox tumban en el Parlamento murciano la comisión de investigación sobre el incendio de las discotecas
8	D7	Los socios del Gobierno presionan para acelerar la tramitación de la amnistía
9	D8	"Vendía una bici de 7.600€ online y cuando quedé para cerrar el trato me la robaron"...
10	D9	Cruce de acusaciones entre partidos por el uso del Senado en la agenda legislativa

3.1. Semantic Similarity (Sentence Embeddings)

The main signal of our system is based on the multilingual E5 base model[2] [9], engineered specifically for retrieval-style tasks involving asymmetric inputs.

First, for each instance, we generate an embedding for the article body (E_body) and for each candidate headline (E_title_i). Then, the semantic score is computed using cosine similarity:

$$\text{score}_{\text{semantic}}(i) = \cos(E_{\text{body}}, E_{\text{title\_i}})$$

3.2. Linguistic Features (LF / UMUTextStats)

We employ UMUTextStats to capture the stylistic and framing choices of political discourse within each candidate headline. This tool extracts an extensive array of linguistic features designed to detect subtle editorial manipulation, including shifts in verbal mood (such as using the subjunctive to project doubt into a statement) [15], the strategic use of passive voice to avoid political responsibility [16], and the deliberate use of pragmatic markers that can drastically alter text framing [17].

Concretely, the extracted features set encompasses: (i) morphosyntax and lexical categories such as verb forms, gender, number, entities, and social processes; (ii) discourse and pragmatic markers that indicate stance and communicative intent; and (iii) psycholinguistics and stylometry to address emotional valence, cognition, length, and readability metrics.

Preprocessing for this signal entails the application of a log transformation:

$$X' = \log(1 + X)$$

Followed by a z-score standardisation. The similarity is computed as:

$$\text{score}_{\text{lf}}(i) = \cos(LF_{\text{body}}, LF_{\text{title\_i}})$$

3.3. Visual Similarity (CLIP)

To capture the visual context, the presence of entities or scenes, and the alignment between the image and the headline, we employ CLIP[3]. This model is able to map images and text into a shared embedding

[2] https://huggingface.co/intfloat/multilingual-e5-base
[3] https://huggingface.co/openai/clip-vit-base-patch32

space.

For this matter, we represent the image as: embedding V_image; and the headline i as: embedding V_title_i. Indeed, to compute the similarity and its score, we calculate it as follows:

$$\text{score}_{\text{clip}}(i) = \cos(V_{\text{image}}, V_{\text{title\_i}})$$

3.4. Score Normalisation and Fusion

To ensure comparability between signals with different scales, each score matrix is normalised per instance using z-scoring:

$$\text{score}_{\text{norm}} = \frac{\text{score} - \text{mean}}{\text{std}}$$

Then, the final scoring is obtained by a weighted linear combination of these normalised scores:

- **Textual Subtask:**

$$\text{score}_{\text{task1}}(i) = \alpha \cdot \text{score}_{\text{semantic}}(i) + \beta \cdot \text{score}_{\text{lf}}(i)$$

  Typical weights are: $\alpha = 0.98$ and $\beta = 0.02$.
- **Multimodal Subtask:**

$$\text{score}_{\text{task2}}(i) = \alpha \cdot \text{score}_{\text{semantic}}(i) + \beta \cdot \text{score}_{\text{lf}}(i) + \gamma \cdot \text{score}_{\text{clip}}(i)$$

  In here, typical weights being: $\alpha = 0.92$, $\beta = 0.03$, and $\gamma = 0.05$.

The asymmetric distribution of weights is motivated by the distinct scales and structural roles of each signal. The semantic component ($\text{score}_{\text{semantic}}$) serves as a macro-contextual filter to capture global thematic alignment. However, because hard distractors frequently share overlapping topics or entities, this semantic baseline tends to reach saturation. Thus, the explicit linguistic features (score_{lf}) and the visual information ($\text{score}_{\text{clip}}$), despite their smaller weight, function as critical auxiliary signals. To better illustrate the system architecture, see Figure 1.

Given the unsupervised and zero-shot nature of our system, standard ablation metrics would misrepresent the true utility of these auxiliary dimensions. Features derived from UMUTextStats or CLIP perform poorly if isolated, as they lack the macro-contextual filtering of semantic embeddings. Rather than driving global performance independently, their contribution is designed to be structural. Indeed, they operate as fine-grained tie-breakers for the final ranking decision, capturing subtle editorial framing that global embeddings overlook when reaching saturation.

4. Experimental Results

The performance of our late fusion system was evaluated on the official test set of the PoliticHeadlinES shared task. Evaluation metrics for this focus on ranking accuracy, specifically assessing the system's ability to place the original headline at the top position amidst highly competitive candidates.

Our unsupervised system achieved an official score of 0.66 on the CodaBench leaderboard. This result underscores the potential of our interpretable approach in a challenging zero-shot setting.

Given the complexity of the task at hand, relying solely on lexical overlap or traditional keyword matching typically results in near-random performance. In contrast, we explore how integrating deep semantic representations with explicit linguistic metrics and visual alignments can complement global textual matching. Rather than claiming a definitive solution, our findings suggest that this multi-disciplinary approach offers a nuanced baseline that can help mitigate the limitations of raw semantic alignment in competitive discursive contexts.

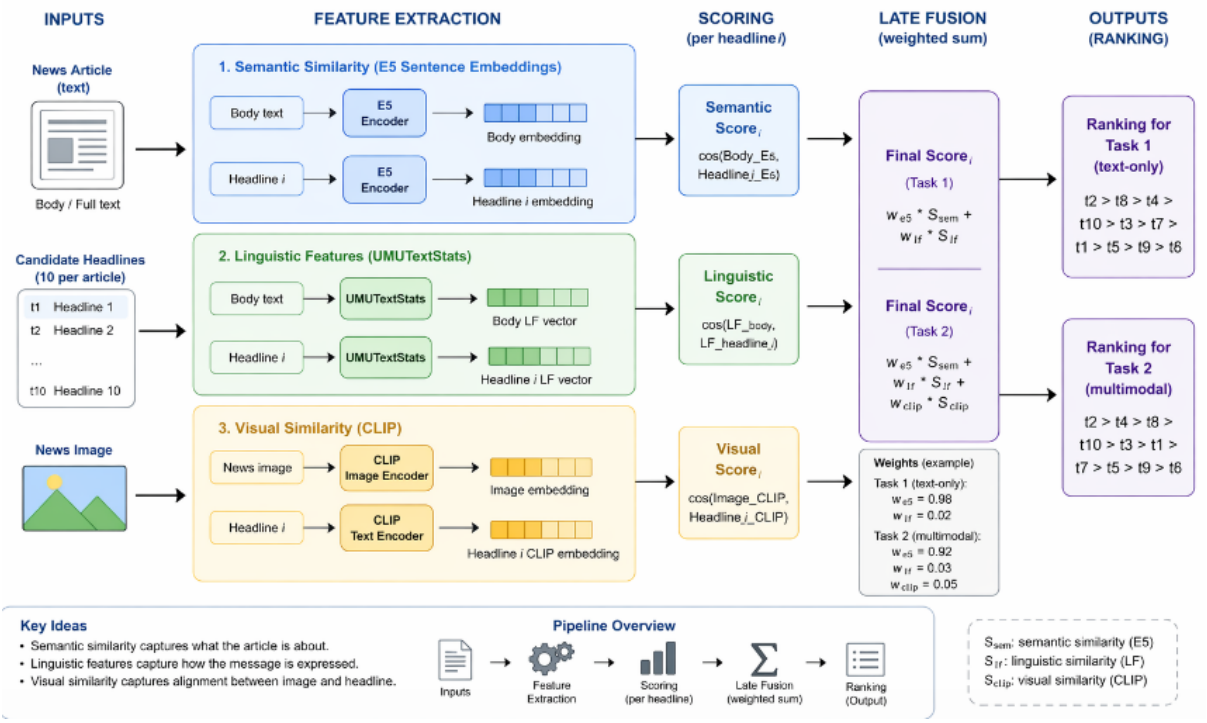

Figure 1: Approximation and experimental results of our system architecture.

4.1. Relation with Participating Systems

Table 2 contextualises our performance within the broader scope of the PoliticHeadlinES 2026 shared task. Our team secured the 14th position out of 20 participating systems in both subtasks, achieving scores of 65.98 and 65.86 for subtasks 1 and 2, respectively. These results contrast with the top-performing architectures on the leaderboard, such as Theron, LKE-BUAP, and VerbaNex AI Lab, which establish a highly competitive ceiling for this task (with scores reaching 95.43, 92.75, and 92.22, respectively).

When contextualising these figures, it is worth noting that top-ranking architectures typically rely on supervised fine-tuning and large language models adapted to the task's training partitions. Crucially, our unsupervised proposal significantly outperforms the official baselines (51.65 and 51.34 for both tasks, respectively).

Therefore, achieving a solid performance in a strictly zero-shot and unsupervised setting highlights the relevance of combining pre-trained text representations with explicit linguistic features. By prioritising explainability and computational efficiency, our architecture offers valuable insights into how certain pragmatic and stylometric attributes can act as critical decision criteria (nuances that are frequently lost within purely generative models).

5. Error Analysis and Limitations

An analysis of these results highlights the intrinsic challenges of the task. Based on empirical observations and Figure 1, sentence embeddings alone provide a strong baseline performance. Nonetheless, explicit linguistic features and visual information act as auxiliary signals since they perform poorly when isolated.

To provide a clearer understanding of how our system components interact, we analysed the final ranking behaviour across the 4,912 news articles included in the official test partition. The results show a visible interaction between the textual and multimodal dimensions: the final order of the candidate headlines changes in 2,377 cases (48.4% of the dataset) when auxiliary information is incorporated. Crucially, our experimental logs reveal that explicit linguistic metrics act as the primary drivers of

Table 2
Official results for PoliticHeadlinES 2026.

	Task 1		Task 2	
Team	Score	Rank	Score	Rank
Theron	95.43629	1	95.43629	1
LKE-BUAP	92.75108	2	92.75108	2
VerbaNex AI Lab	92.22656	3	92.62804	3
acho_gpt	90.26359	4	90.26359	4
GACOP	89.67746	5	89.67718	5
HEART-UJA-UA	89.45794	6	89.12555	6
franrodrigo	88.41059	7	88.41059	7
I2C-UHU	82.59159	8	82.59159	8
LYS-UDC	82.41000	9	80.82452	10
francordel	80.74166	10	81.28402	9
ITO	79.85917	11	79.61529	11
UTP	73.94271	12	77.04902	12
UC-UCO-Plenitas	70.49079	13	71.10051	13
LACELL	**65.98162**	**14**	**65.86319**	**14**
enrique_lr	65.56252	15	65.56252	15
UIE	64.47657	16	64.47657	16
I2C-UHU-Perseus	62.31144	17	62.31144	17
PUCE-UTP	58.06160	18	55.45929	18
baseline	51.65320	19	51.34444	19
MIMIC	34.79614	20	34.79614	20

this structural re-ranking. While the semantic embeddings establish the general thematic boundaries, this baseline frequently faces ambiguity when dealing with competitive distractors that share identical political topics and entities (such as the overlapping references to ”amnistía” shown in Table 1).

In these specific scenarios, features extracted by UMUTextStats appear to exert a stronger, more decisive influence on the decision boundaries than the visual signal, effectively functioning as tentative tie-breakers. Even with a small numerical weight in the final global fusion, the inclusion of these explicit linguistic metrics, which evaluate stylistic and pragmatic shifts, may successfully guide the architecture to mitigate semantic saturation and help prioritise the narrative structure of the original headline.

Consequently, a primary limitation stems from the independence of signals, which restricts deeper interaction between modalities. It is precisely in these intervals of semantic proximity where our interpretable, unsupervised architecture offers its main advantage. While the E5 encoder is blind to stylistic staging, the auxiliary linguistic signal from UMUTextStats captures variations in morphosyntax and discursive markers. This allows the model to evaluate the explicit rhetorical framing of the text, acting as a micro-stylistic tie-breaker that penalises generic distractors. Ultimately, this aids the final ranking without relying on computationally expensive black-box architectures. Furthermore, the visual signal heavily depends on image quality and relevance, and can be ambiguous if news images are generic, falling out when disambiguating highly similar headlines.

Another limitation entails the choice of linguistic markers. UMUTextStats is a solid and robust extractor that provides a strong baseline for linguistic study, but its results may be hindered when analysing hard distractors since the illocutionary force of the message is not fully captured [18]. Especially when these distractors share the same entities but differ in the pragmatic intent (i.e., whether a headline uses irony or rhetorical framing [19]). Lastly, the absolute performance of our current baseline may have been constrained by the absence of supervised learning or tuning on the development data splits.

6. Conclusions and Future Work

In this paper, we have described the participation of the LACELL team in the PoliticHeadlinES 2026 shared task. Rather than relying on a black-box architecture, our late fusion framework prioritises interpretability, offering a nuanced baseline that captures the multi-layered nature of editorial framing. Additionally, while global embeddings provide a robust foundation for general thematic matching, our findings suggest how explicit linguistic features can serve as subtle and essential regularisers, serving as a complementary view of raw textual alignment and communicative intent.

Building upon these insights, future work will explore cross-encoder architectures for re-ranking by incorporating explicit named entities and dynamic weighting strategies to improve the multimodal alignment of our system. Further, we intend to deepen our linguistic analysis by integrating pragmatic content, specifically focusing on Speech Act Theory [20], as it better captures the author's implicit stance [21]. In this way, the model will actively penalise distractors that alter the original communicative intention of the headline.

Acknowledgments

This work is part of the research project LaTe4PoliticES (PID2022-138099OB-I00) funded by MICIU/AEI/10.13039/501100011033 and the European Regional Development Fund (ERDF/EU - FEDER/UE)-a way of making Europe. Ms. Desirée Martínez Carrión is supported by University of Murcia through the predoctoral programme.

Declaration on Generative AI

During the preparation of this work, the author(s) used Gemini 3.1 Flash-Lite in order to: Grammar, technical terms, and spelling check. Further, the author(s) used ChatGPT 5.4 for figure 1 in order to: Generate images for illustrative purposes. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] R. M. Entman, Framing: Toward clarification of a fractured paradigm, Journal of Communication 43 (1993) 51–58. URL: https://academic.oup.com/joc/article/43/4/51-58/4160153. doi:10.1111/j.1460-2466.1993.tb01304.x.

[2] R. Pan, T. Bernal-Beltrán, J. Gómez-Navalon, J. A. García-Díaz, R. Valencia-García, Spanish PoliSUM-2025: Evaluating stylistic and editorial fidelity in abstractive summarization of political news, Procesamiento del Lenguaje Natural 76 (2026) 177–189.

[3] J. Gómez-Navalón, T. Bernal-Beltrán, R. Pan, G.-S. Francisco, J. A. García-Díaz, R. Valencia-García, Overview of PoliticHeadlinES at IberLEF 2026: Multimodal headline ranking in Spanish political news, Procesamiento del Lenguaje Natural 77 (2026).

[4] T. A. van Dijk, News As Discourse, Taylor & Francis, 1990.

[5] G. Vallejo, T. Baldwin, L. Frermann, Connecting the dots in news analysis: Bridging the cross-disciplinary disparities in media bias and framing, in: Proceedings of the Sixth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS 2024), Association for Computational Linguistics, Mexico City, Mexico, 2024, pp. 16–31. URL: https://aclanthology.org/2024.nlpcss-1.2. doi:10.18653/v1/2024.nlpcss-1.2.

[6] T. Spinde, S. Hinterreiter, F. Haak, T. Ruas, H. Giese, N. Meuschke, B. Gipp, The media bias taxonomy: A systematic literature review on the forms and automated detection of media bias, 2024. URL: http://arxiv.org/abs/2312.16148. doi:10.48550/arXiv.2312.16148, arXiv:2312.16148 [cs].

[7] H. Avetisyan, D. Broneske, Identifying and understanding game-framing in online news: BERT and fine-grained linguistic features, in: M. Abbas, A. A. Freihat (Eds.), Proceedings of the 4th International Conference on Natural Language and Speech Processing (ICNLSP 2021), Association for Computational Linguistics, Trento, Italy, 2021, pp. 95–107. URL: https://aclanthology.org/2021.icnlsp-1.12/.

[8] A. Arora, S. Yadav, M. Antoniak, S. Belongie, I. Augenstein, Multi-modal framing analysis of news, 2025. URL: https://arxiv.org/abs/2503.20960. arXiv:2503.20960.

[9] L. Wang, N. Yang, F. Wei, B. Chang, M. Zhou, Text embeddings by weakly-supervised contrastive pre-training, arXiv preprint arXiv:2212.03533 (2022).

[10] N. Reimers, I. Gurevych, Sentence-BERT: Sentence embeddings using siamese BERT-networks, in: K. Inui, J. Jiang, V. Ng, X. Wan (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 3982–3992. URL: https://aclanthology.org/D19-1410/. doi:10.18653/v1/D19-1410.

[11] J. A. García-Díaz, P. J. Vivancos-Vicente, Â. Almela, R. Valencia-García, UMUTextStats: A linguistic feature extraction tool for Spanish, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 6035–6044. URL: https://aclanthology.org/2022.lrec-1.649/.

[12] N. Ramírez-Esparza, J. W. Pennebaker, F. A. García, R. Suriá Martínez, La psicología del uso de las palabras: un programa de computadora que analiza textos en español, Revista Mexicana de Psicología 24 (2007) 85–99. URL: http://hdl.handle.net/10045/25918.

[13] A. Radford, J. W. Kim, C. Hallacy, et al., Learning transferable visual models from natural language supervision, in: Proceedings of ICML, 2021.

[14] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, A. Rush, Transformers: State-of-the-art natural language processing, in: Q. Liu, D. Schlangen (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Online, 2020, pp. 38–45. URL: https://aclanthology.org/2020.emnlp-demos.6/. doi:10.18653/v1/2020.emnlp-demos.6.

[15] F. Bermúdez, Evidencialidad. La codificación lingüística del punto de vista, Ph.D. thesis, University of Stockholm, Estocolmo, Suecia, 2005.

[16] N. Fairclough, Language and Power, Longman, 2001.

[17] J. Portolés Lázaro, M. A. Martín Zorraquino, Los marcadores del discurso, in: Gramática descriptiva de la lengua española, Vol. 3, Espasa Calpe España, 1999, pp. 4051–4214. URL: https://dialnet.unirioja.es/servlet/articulo?codigo=2152618.

[18] S. Borchmann, Headlines as illocutionary subacts: The genre-specificity of headlines, Journal of Pragmatics 220 (2024) 73–99. doi:10.1016/j.pragma.2023.12.007.

[19] A. Pluwak, Towards application of speech act theory to opinion mining, Cognitive Studies | Études cognitives (2016) 33–44. doi:10.11649/cs.2016.004.

[20] J. R. Searle, Expression and Meaning: Studies in the Theory of Speech Acts, Cambridge University Press, 1979.

[21] A. Upadhyaya, M. Fisichella, W. Nejdl, Toxicity, morality, and speech act guided stance detection, in: H. Bouamor, J. Pino, K. Bali (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2023, Association for Computational Linguistics, 2023, pp. 4464–4478. doi:10.18653/v1/2023.findings-emnlp.295.