DUTH at SemEval-2026 Task 3: Multilingual Transformer Models for Dimensional Stance Prediction Across Tracks

Georgios Arampatzis **Avi Arampatzis**

Department of Electrical & Computer Engineering,
Democritus University of Thrace
{geoaramp,avi}@ee.duth.gr

Abstract

This paper presents **DUTH**, our system for Track A and Track B of SemEval-2026 Task 3 on Dimensional Sentiment Analysis, focusing on the Dimensional Aspect-Based Sentiment Regression (DimASR) subtask. DimASR requires predicting continuous Valence and Arousal (VA) scores for aspect terms in opinionated text and stance targets in public-issue discourse.

Our approach uses a multilingual Transformer encoder fine-tuned end-to-end to jointly encode the input text and its corresponding aspect or stance target, followed by a regression head for VA prediction. We evaluate DUTH on the official multilingual and multidomain datasets and compare it against the shared-task baselines. Results show competitive performance, with improvements over the strongest official baseline in Track A and over the mBERT baseline in Track B, while yielding consistently stronger predictions for Valence than for Arousal.

1 Introduction

Aspect-Based Sentiment Analysis (ABSA) has traditionally focused on discrete sentiment categories such as positive, negative, and neutral. More recent work in affective computing models sentiment as a continuous phenomenon along the Valence and Arousal (VA) dimensions, enabling finer-grained representations of emotional polarity and intensity. SemEval-2026 Task 3 introduces Dimensional Aspect-Based Sentiment Analysis (DimABSA) and Dimensional Stance Analysis (DimStance) as regression-based formulations of aspect-level sentiment and stance detection (Yu et al., 2026).

The shared task provides multilingual, multidomain datasets with continuous VA annotations across review and public-issue domains (Lee et al., 2026; Becker et al., 2026). We participate in both Track A and Track B, focusing on the Dimensional Aspect Sentiment Regression (DimASR) subtask. Track A predicts VA scores for given aspect terms in text, while Track B extends this setting to dimensional stance analysis, where stance targets are treated analogously to aspects in sociopolitical discourse. We adopt a multilingual encoder-based regression framework because the task spans multiple languages and domains, and a shared backbone offers a simple, reproducible, and effective means of transferring affective knowledge across settings.

Our system fine-tunes a multilingual pre-trained language model to jointly encode the input text and its aspect or stance target, followed by a regression head for continuous VA prediction. The aim is to establish a strong shared-task baseline rather than introduce task-specific architectural modifications.

Building on our prior work in multilingual shared tasks on affective modeling and classification (Arampatzis et al., 2023, 2025; Arampatzis and Arampatzis, 2025b,a), we evaluate this framework on both Track A and Track B of the official SemEval-2026 Task 3 benchmark. The results show competitive performance across both tracks, with consistently stronger prediction quality for Valence than for Arousal.

The remainder of this paper is organized as follows. Section 2 reviews related work, Section 3 describes the proposed system and datasets, Section 4 presents the experimental setup, Section 5 reports the results and analysis, and Section 6 concludes the paper.

2 Related Work

ABSA has evolved from feature-engineered methods (Pang and Lee, 2008; Liu, 2012) to Transformer-based models such as BERT and its variants (Devlin et al., 2019; Liu et al., 2019). Pretrained language models provide contextualized and aspect-aware representations, while joint encoding of the input sentence and target aspect has become a common strategy for fine-grained sen-

Proceedings of the The 20th International Workshop on Semantic Evaluation (2026), pages 592–598
July 3-4, 2026 ©2026 Association for Computational Linguistics

timent prediction across domains and languages (Sun et al., 2019; Xia et al., 2021).

Beyond categorical polarity classification, dimensional sentiment analysis models affect as continuous variables along the Valence and Arousal (VA) dimensions (Russell, 1980). Affective norm resources (Warriner et al., 2013), neural regression approaches (Buechel and Hahn, 2017), and shared tasks such as SemEval-2018 Task 1 (Mohammad et al., 2018) have highlighted both the usefulness and the challenges of continuous affect prediction.

SemEval-2026 Task 3 extends aspect-level sentiment and stance detection to continuous regression settings (Yu et al., 2026). Our work follows this direction by using a multilingual Transformer encoder for Dimensional Aspect Sentiment Regression (DimASR) and evaluating a unified framework across both tracks.

3 System Description

3.1 Dataset

We participate in both Track A and Track B of SemEval-2026 Task 3, focusing on the Dimensional Aspect Sentiment Regression (DimASR) subtask in each track. In both cases, the datasets provide continuous Valence and Arousal (VA) annotations built on top of existing aspect-based sentiment resources, following the shared-task specifications and the dimensional affect modeling paradigm.

3.2 Track A – DimABSA

The Track A dataset extends the ACOS dataset (Cai et al., 2021) by assigning real-valued VA scores to annotated aspect–opinion instances. Each example consists of a short opinionated text, one or more aspect terms, and the corresponding gold VA scores. VA values range from 1.00 to 9.00 and are rounded to two decimal places.

In this work, we focus on the English Laptop and Restaurant domains. The training sets retain implicit aspects and opinions annotated as `NULL`, following the original ACOS annotation scheme, whereas the official test sets exclude such implicit annotations.

Domain	Train	Dev	Test
Laptop	4,076	200	1,000
Restaurant	2,284	200	1,000
Total	**6,360**	**400**	**2,000**

Table 1: Dataset statistics for Track A – Subtask 1 (DimASR).

3.3 Track B – DimStance

Track B reformulates stance detection as a dimensional sentiment regression task by treating stance targets analogously to aspects and assigning continuous VA scores to each target. This formulation extends dimensional affect modeling beyond product reviews to public-issue discourse.

The dataset spans multiple languages and sociopolitical domains, including environmental protection and political discussions, enabling evaluation under multilingual and cross-domain conditions. Its multilingual composition also makes Track B suitable for examining how a shared encoder behaves in lower-resource settings.

Language	Train	Dev	Test
English	1,200	150	500
German	1,000	150	500
Nigerian Pidgin	800	100	400
Swahili	800	100	400
Total	**3,800**	**500**	**1,800**

Table 2: Dataset statistics for Track B – Subtask 1 (DimASR for dimensional stance analysis) across languages.

We use the official training and development splits for model training and hyperparameter selection, and evaluate the final models on the corresponding test sets. Gold labels for all test sets in both tracks were released after the evaluation phase and are used exclusively for final system reporting. We do not use external datasets or additional labeled resources.

3.4 Pre-processing

For each input instance, we flatten the aspect annotations by pairing each aspect term with the corresponding sentence, resulting in one training example per *(text, aspect)* pair. This formulation follows common aspect-aware sentiment modeling practice, where jointly encoding the target aspect with its surrounding context allows the model to focus on sentiment cues relevant to each specific opinion target (Sun et al., 2019; Xia et al., 2021).

Aspect terms are explicitly included in the input sequence to guide contextual representation learning and to disambiguate sentiment expressed toward multiple aspects within the same text.

Text is tokenized using the tokenizer associated with the underlying pre-trained language model. All inputs are truncated or padded to a fixed maximum sequence length to ensure computational efficiency and uniform batch processing.

No additional normalization, filtering, or data augmentation is applied. This design choice preserves the original linguistic signal and is consistent with prior findings showing that Transformer-based models are robust to surface-level noise and often benefit from minimal preprocessing in sentiment analysis tasks (Devlin et al., 2019; Liu et al., 2019).

3.5 Model

Our system uses `XLM-RoBERTa-base` (Conneau et al., 2020) as a multilingual Transformer backbone for both Track A (DimABSA) and Track B (DimStance). Transformer-based encoders have shown strong performance in aspect-based sentiment analysis because they can capture long-range contextual dependencies and produce aspect-aware representations (Devlin et al., 2019; Liu et al., 2019; Sun et al., 2019).

We selected XLM-RoBERTa-base because it offers broad multilingual coverage and a well-established pre-trained initialization for cross-lingual transfer, while keeping the overall system simple and reproducible. This choice is particularly suitable for SemEval-2026 Task 3, where the same modeling framework must generalize across both review and stance domains.

For each input instance, the text is concatenated with the corresponding aspect term or stance target and jointly encoded to obtain contextualized representations. This joint encoding strategy helps the model focus on sentiment cues relevant to each target and has been widely adopted in recent ABSA and opinion mining frameworks (Sun et al., 2019; Xia et al., 2021).

The pooled Transformer output is passed to a regression head consisting of a single linear layer that jointly predicts continuous Valence and Arousal (VA) scores. The model is trained in a multi-output regression setting using mean squared error loss over both affective dimensions, following common practice in dimensional sentiment modeling (Buechel and Hahn, 2017; Mohammad et al., 2018).

All parameters of the Transformer encoder and regression head are fine-tuned end-to-end on the task-specific training data. For Track A, we train separate domain-specific models for Laptop and Restaurant reviews to capture domain-dependent affective patterns. For Track B, we train independent models for each language-specific dataset. We adopt language-specific models in Track B to match the evaluation setup and to keep training stable across datasets of different sizes; joint multilingual training is left for future work.

Overall, this unified regression framework supports effective modeling of dimensional sentiment at both the aspect and stance levels across multiple domains and languages, while maintaining architectural simplicity and strong generalization capacity.

4 Experimental Setup

All experiments are conducted using the official training and development splits provided by the SemEval-2026 Task 3 organizers for both Track A (DimABSA) and Track B (DimStance). For Track A, we train separate domain-specific models for Laptop and Restaurant reviews in order to account for domain-dependent sentiment distributions. For Track B, we train independent models for each language-specific dataset, following the shared-task evaluation protocol. Model selection is performed on the development split using RMSE, and final test predictions are generated from the checkpoint that achieves the best development-set RMSE.

We use `XLM-RoBERTa-base` as a unified multilingual backbone and fine-tune all model parameters on the task-specific data. The regression head jointly predicts continuous Valence and Arousal scores for each input instance. Training is performed with the AdamW optimizer, and the same input formulation and network architecture are maintained across datasets to support consistent comparison across domains and languages.

All experiments are conducted with fixed random seeds to ensure reproducibility. Unless otherwise stated, the same optimization strategy and hyperparameter configuration are used across all datasets and tracks.

4.1 Evaluation Measures

System performance is evaluated using the official metrics defined by the SemEval-2026 Task 3 organizers. For the Dimensional Aspect Sentiment Regression (DimASR) subtask, the primary ranking metric is Root Mean Squared Error (RMSE), as defined by the official shared-task evaluation protocol.

In addition to RMSE, we report the Pearson Correlation Coefficient (PCC) separately for Valence and Arousal in order to quantify the linear association between predicted scores and gold affective

annotations (Benesty et al., 2009). Lower RMSE values indicate more accurate regression performance, whereas higher PCC scores reflect stronger agreement with human-provided affective ratings.

All reported results are computed using the official evaluation scripts released by the shared-task organizers to ensure direct comparability with other participating systems (Yu et al., 2026). To better contextualize our findings, we also compare our results against the corresponding official baseline scores released for the shared task.

4.2 Implementation Details

All models are implemented in PyTorch using the HuggingFace Transformers library. Training and inference are executed on NVIDIA GPUs with mixed-precision computation where supported, in order to improve memory efficiency and training speed. We use the `XLM-RoBERTa-base` checkpoint as the pre-trained encoder and fine-tune it end-to-end for the target regression task.

4.3 Training Configuration

Models are fine-tuned using mini-batch gradient descent, with evaluation and checkpoint saving performed at the end of each training epoch. The final model is selected automatically based on the lowest development-set RMSE. Dropout regularization is retained from the pre-trained backbone configuration.

For the reported experiments, we use AdamW optimization with a learning rate of $2e{-}5$, batch size 16, maximum sequence length 256, and 3 training epochs. The random seed is fixed to 42.

5 Results and Analysis

We evaluate our system on the official test sets using the metrics defined by the shared task. For DimASR, performance is measured using Root Mean Squared Error (RMSE) over the joint Valence–Arousal predictions, together with the Pearson Correlation Coefficient (PCC) computed separately for each affective dimension. In addition to the official task metrics, we compare our system against the corresponding official baselines in order to better contextualize the performance of fine-tuning `XLM-RoBERTa-base`.

5.1 Track A Analysis

The model achieves lower RMSE on the Laptop domain (1.59) than on Restaurant reviews (1.82),

Dataset	Best BL	DUTH	Δ	PCC$_V$	PCC$_A$
Laptop	2.1893	**1.5900**	+0.5993	0.63	0.47
Restaurant	2.1461	**1.8200**	+0.3261	0.62	0.46

Table 3: Track A results and comparison with the strongest official baseline (Kimi-K2 Thinking). Lower RMSE indicates better performance; positive Δ indicates improvement over the baseline.

indicating more stable regression behavior in technical product descriptions. Laptop reviews typically contain explicit evaluative language (e.g., *fast*, *slow*, *durable*), which provides strong and direct valence cues. In contrast, restaurant reviews often include more experiential, contextual, or figurative expressions, which increase semantic ambiguity and regression variance.

Compared with the strongest official Track A baseline (Kimi-K2 Thinking), our model achieves lower RMSE on both English domains, confirming that a simple multilingual encoder with target-aware input formatting provides a strong baseline for DimASR.

Across both domains, PCC$_V$ (0.63 / 0.62) consistently exceeds PCC$_A$ (0.47 / 0.46), indicating that the model captures sentiment polarity more reliably than emotional activation intensity. While valence predictions follow the gold trends relatively closely, arousal estimation exhibits greater dispersion, suggesting that cues related to emotional intensity are less directly encoded in surface lexical patterns.

5.2 Track B Analysis

Performance variability increases substantially in Track B. English achieves the lowest RMSE$_{VA}$ (2.1964), reflecting stronger regression stability in a relatively higher-resource setting. Lower-resource languages such as Swahili (SWA) and Nigerian Pidgin (PCM) exhibit higher RMSE values, suggesting greater sensitivity to domain shift and limited supervision.

Lang.	M3	mBERT	DUTH	Δ_M	PCC$_V$	PCC$_A$
ENG	**1.6430**	2.6985	2.1964	-0.5534	0.3303	0.0852
SWA	**2.2990**	2.7835	2.7267	-0.4277	0.1157	-0.0127
PCM	**1.7390**	3.2152	2.6718	-0.9328	0.5183	-0.3290
DEU	**1.5910**	2.3294	2.2921	-0.7011	0.0321	-0.0249

Table 4: Track B results. Lower RMSE$_{VA}$ is better. Δ_M = RMSE(Mistral-3) $-$ RMSE(DUTH); positive values indicate improvement over Mistral-3.

Compared with the strongest official Track B baseline (Mistral-3 14B), our model remains weaker across all languages. However, it consistently outperforms the mBERT baseline on every Track B dataset, indicating that fine-tuned multilin-

gual encoders provide a competitive conventional baseline in this setting.

Interestingly, PCM achieves the highest PCC_V (0.5183) despite a comparatively high $RMSE_{VA}$, suggesting that the predicted valence trends align with the gold annotations but remain affected by scale miscalibration. Negative or near-zero PCC_A values do not necessarily indicate overfitting to valence. Rather, they suggest that the model fails to preserve the relative ordering of arousal scores, often because predictions become compressed around the mean and are poorly calibrated under low-resource conditions. German (DEU) shows moderate RMSE but near-zero correlations, which points to conservative predictions clustered around the mean, reducing squared error while limiting variance-based alignment.

Across all Track B languages, PCC_A remains near zero or negative, highlighting the intrinsic difficulty of modeling arousal in short stance-oriented utterances. Unlike valence, which often depends on explicit evaluative polarity, arousal requires richer contextual and pragmatic cues that may not be strongly lexicalized.

Overall, these results suggest that multilingual Transformer models generalize effectively for valence regression across domains, whereas arousal estimation and low-resource cross-lingual transfer remain challenging aspects of dimensional sentiment modeling. A useful next step would be to compare these language-specific models with a jointly trained multilingual Track B model, which may improve stability for lower-resource languages.

5.3 Error Patterns

A qualitative inspection of development and test predictions reveals several recurring error patterns, consistent with prior findings in dimensional sentiment modeling. Representative examples are shown below.

Tr.	Tgt	Text	V	A	Error
A	donuts	fresh with each order	7.25	6.75	implicit sent.
B	EVs	not economically viable	3.25	5.88	mean-centered

Table 5: Representative difficult DimASR examples. V and A denote gold Valence and Arousal.

Implicit sentiment expressions often lead to predictions close to the dataset mean, as the model defaults to conservative VA estimates when lexical affective cues are weak or ambiguous. Although this behavior may reduce large RMSE errors, it also lowers output variance and harms correlation-based metrics, especially for arousal (Buechel and Hahn, 2017; Warriner et al., 2013).

Arousal estimation is less stable than valence prediction. Subtle preference statements often lack strong lexical markers of emotional activation, making intensity inference more difficult and leading to compressed output variance and lower PCC_A scores (Russell, 1980; Mohammad et al., 2018).

Low-resource and cross-domain settings further exacerbate these effects by weakening calibration and increasing central-tendency bias. Overall, the model appears more robust for valence than for arousal, particularly under domain shift and limited supervision.

6 Conclusion

This paper presented DUTH, a unified Transformer-based regression framework for Dimensional Aspect Sentiment Regression (DimASR) in SemEval-2026 Task 3. By fine-tuning a multilingual pre-trained encoder, our system predicts continuous Valence and Arousal (VA) scores for aspect terms and stance targets across multiple domains and languages.

Experimental results show competitive performance across both tracks. In Track A, our approach improves over the strongest official baseline on both English domains, while in Track B it consistently surpasses the mBERT baseline but remains below the stronger Mistral-3 14B baseline. Across settings, the model produces more reliable predictions for Valence than for Arousal, and our analysis highlights recurring challenges related to implicit sentiment, weak lexical intensity cues, and lower-resource conditions.

Overall, these findings indicate that a simple multilingual Transformer with target-aware input formatting provides a strong and reproducible baseline for continuous sentiment regression. Future work will examine joint multilingual training, alternative multilingual encoders, and calibration-oriented objectives to improve robustness, especially for arousal prediction.

References

Georgios Arampatzis and Avi Arampatzis. 2025a. DUTH at CLEF 2025 SimpleText track: Tackling scientific text simplification and hallucination detection. In *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2025)*, volume 4038 of *CEUR Workshop Proceedings*, pages 4211–4224. CEUR-WS.org.

Georgios Arampatzis and Avi Arampatzis. 2025b. DUTH at CLEF JOKER 2025 tasks 2 and 3: Translating puns and proper names with neural approaches. In *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2025)*, volume 4038 of *CEUR Workshop Proceedings*, pages 2791–2802. CEUR-WS.org.

Georgios Arampatzis, Vasileios Perifanis, Symeon Symeonidis, and Avi Arampatzis. 2025. DUTH at EXIST 2025: Multilingual sexism detection with soft labels and transformers. In *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2025)*, volume 4038 of *CEUR Workshop Proceedings*, pages 1793–1800. CEUR-WS.org.

Giorgos Arampatzis, Vasileios Perifanis, Symeon Symeonidis, and Avi Arampatzis. 2023. DUTH at SemEval-2023 task 9: An ensemble approach for Twitter intimacy analysis. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 1225–1230, Toronto, Canada. Association for Computational Linguistics.

Jonas Becker, Liang-Chih Yu, Shamsuddeen Hassan Muhammad, Jan Philip Wahle, Terry Ruas, Idris Abdulmumin, Lung-Hao Lee, Nelson Odhiambo, Lilian Wanzare, Wen-Ni Liu, Tzu-Mi Lin, Zhe-Yu Xu, Ying-Lung Lin, Jin Wang, Maryam Ibrahim Mukhtar, Bela Gipp, and Saif M. Mohammad. 2026. DimStance: Multilingual datasets for dimensional stance analysis. *Preprint*, arXiv:2601.21483.

Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*. Springer.

Sven Buechel and Udo Hahn. 2017. EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 578–585, Valencia, Spain. Association for Computational Linguistics.

Hongjie Cai, Rui Xia, and Jianfei Yu. 2021. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 340–350, Online. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Lung-Hao Lee, Liang-Chih Yu, Natalia Loukashevich, Ilseyar Alimova, Alexander Panchenko, Tzu-Mi Lin, Zhe-Yu Xu, Jian-Yu Zhou, Guangmin Zheng, Jin Wang, Sharanya Awasthi, Jonas Becker, Jan Philip Wahle, Terry Ruas, Shamsuddeen Hassan Muhammad, and Saif M. Mohammad. 2026. DimABSA: Building multilingual and multidomain datasets for dimensional aspect-based sentiment analysis. *Preprint*, arXiv:2601.23022.

Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.

Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. SemEval-2018 task 1: Affect in tweets. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, pages 1–17, New Orleans, Louisiana. Association for Computational Linguistics.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135.

James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.

Chi Sun, Luyao Huang, and Xipeng Qiu. 2019. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 380–385, Minneapolis, Minnesota. Association for Computational Linguistics.

Amy B. Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior Research Methods*, 45(4):1191–1207.

Qingrong Xia, Bo Zhang, Rui Wang, Zhenghua Li, Yue Zhang, Fei Huang, Luo Si, and Min Zhang. 2021. A unified span-based approach for opinion mining with syntactic constituents. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1795–1804. Association for Computational Linguistics.

Liang-Chih Yu, Jonas Becker, Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Lung-Hao Lee, Ying-Lung Lin, Jin Wang, Jan Philip Wahle, Terry Ruas, Alexander Panchenko, Ilseyar Alimova, Kai-Wei Chang, Lilian Wanzare, Nelson Odhiambo, Bela Gipp, and Saif M. Mohammad. 2026. SemEval-2026 task 3: Dimensional aspect-based sentiment analysis (DimABSA). In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*. Association for Computational Linguistics.