وحدة البحث اللساني وقضايا اللغة العربية في الجزائر

LINGUISTIC RESEARCH UNIT AND ARABIC LANGUAGE ISSUES IN ALGERIA

https://urclaa.dz · contact@urclaa.dz · (+213) 029.60.82.82 · (+213) 029.60.89.82

مركز البحث العلمي والتقني لتطوير اللغة العربية

Scientific and Technical Research Center for the Development of the Arabic Language

+213 23 180 095 · +213 23 180 097

www.crstdla.dz · contact@crstdla.dz

CRSTDLA

A Qualitative Error Analysis of Whisper-Based Sentiment Recognition in **Algerian Spoken Arabic:** Prosodic and Pragmatic Failures in a Low-Resource Dialect

Dr. Abdelouahab Baflah

Scientific and Technical Research Center for the Development of the Arabic Language (CRSTDLA)

Ouargla Research Unit, Algeria

a.baflah@crstdla.dz

1 ABSTRACT

While pre-trained sentiment analysis models have achieved high accuracy in standardized languages, they remain largely **"semantically blind"** to the intricate nuances of **Algerian spoken Arabic** (Darja). This paper challenges the localization of current AI models by presenting a rigorous error analysis based on real-world emotional narratives. Using a multi-layered framework—mapping Valence-Arousal-**Dominance** (VAD) against human-verified sentiment labels—we demonstrate how automated systems misinterpret the "silent data" within the dialect. **Our analysis focuses on two neglected dimensions: prosodic** (vowel **elongation, stuttering, and breath noises) and pragmatics (the use of cultural idioms and syntactic ellipses like "7 packs a day"). The findings reveal that dialectal markers, specifically /g/ phoneme in contexts of distress, are often dismissed by AI as linguistic noise rather than high-arousal indicators. By exposing the gap between machine-generated outputs and the "Golden Standard" of human linguistic intuition, this study argues that recognizing emotion in Algerian Arabic requires moving beyond the word level. We propose a hybrid model where prosodic cues and pragmatic context are not mere supplements but essential components for any reliable sentiment recognition system in North African dialects.**

Keywords: Algerian Arabic, sentiment analysis, prosody, pragmatics, error analysis, VAD model, spoken language processing.

2 INTRODUCTION

Sentiment analysis is experiencing remarkable progress with the emergence of modern artificial intelligence models. However, these tools are primarily designed for standardized languages or high-resource dialects, such as Modern Standard Arabic (MSA). Consquenlty, their results are often inaccurate when appplied to low-resource, morphologically complex dialects like Algerian Darja (spoken Arabic in Algeria). The characteristics of Algerian Darja include heavy code-swtiching, lexical borrowing, and a strong reliance on implicit contextual meaning, making it a major challenge for natural language processing (NLP) (Saadane and Babaheti, 2015).

Current automated systems predominantly rely on a "text-centric" approach, converting speech to text using Automatic Speech Recognition (ASR) systems—such as the Whisper model—and subsequently anylyzing the resulting text to extract sentiment (text classification or regression). This pipeline strips the speaker's language of crucial non-verbal cues. The expression of emotion is not conveyed through vocabulary alone; rather, emotional charge is densely encoded in "silent data": changes in stress, unintelligible words, and mumbles, in addition to culturally bound pragmatic structures. When automated models "clean" the speech to conform to MSA standards, they delete these essential emotional signals, causing the system to fail in recognizing high-arousal states such as anger or distress.

3 METHODOLOGY AND DATASET

3.1 Overall Methodology

We adopt a qualitative analytical approach to evaluate current AI systems in sentiment analysis for Algerian Darja. Evaluation is conducted by noubping system outputs and human baselines within the Valence–Arousal–Dominance (VAD) space. **Discrepancies between automated evaluations and the Golden Standard** are analyzed qualitialvelly to identily the **loss of prosodic and pragmatic information.**

Spoken Utterance (Algerian Darja)
↓
Automatic Speech Recognition (Whisper)
↓
Text Output
↓
Sentiment Analysis Model
↓
VAD Mapping
↓
Comparison with Human Golden Standard
↓
Error Analysis (Prosody & Pragmatics)

3.2 Dataset

The dataset comprises 57 utterance-level examples drawn from seven distinct sources:

- 9 utterances extracted from phone-in radio calls archive.
- 12 utterances from the personal narrative of a 26-year-old male speaker from Bordj R Kiffan.
- 4 utterances from the narrative of a female speaker.
- 22 utterances extracted from four segments of a radio program broodcast on YouTube.

Each utterance was manually segmented, transcribed, and annotated by three native speakers based on the VAD model.

VAD MODEL (Annotation Dimensions)

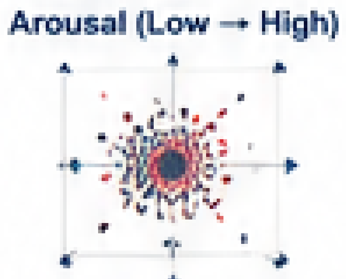
Valence	: Negative – Neutral – Positive
Arousal	: Low – Medium – High
Dominance	: Low – Medium – High

5 TEXT-CENTRIC PIPELINE: THE INFORMATION LOSS PROBLEM

Current automated systems follow a linear, text-centric pipeline that leads to the loss of crucial emotional cues. before annotating the data in its radio material:

1 Manual Transcription + Initial Segmentation (by three experts) → 2 ASR with Whisper (Large-v3 model) → 3 Sentiment Analysis (Text-based models) → 4 Mapping outputs to VAD space and error categorization

The texts and audio were processed using three models: Whisper (ASR), AraBERT (text classification), and DziriBERT (dialect-specific).

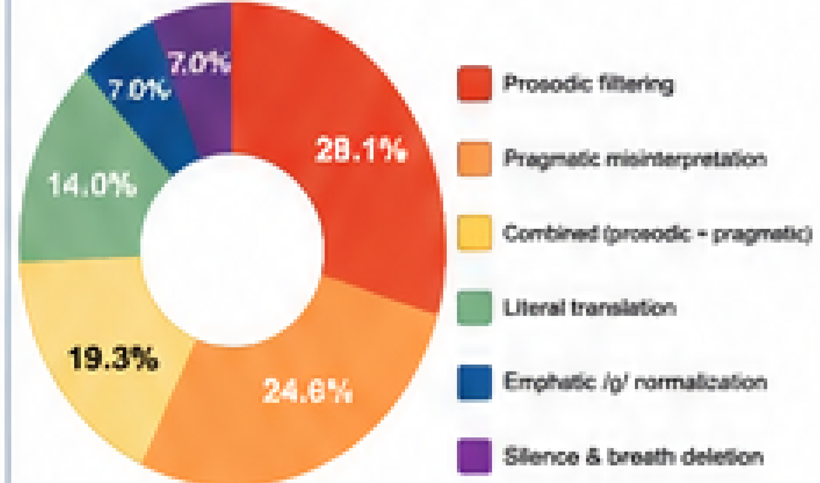


4 LINGUISTIC ERROR ANALYSIS

4.1 Distribution of Error Types Across the Dataset (N=57)

	Error Category	Description	Count	Percentage
1	**Prosodic filtering** (vowel elongation, stuttering, pitch, and mumbles)	Deletion of vowel elongation, stuttering, pitch variations, and mumbles as "noise"	16	28.1%
2	**Pragmatic misinterpretation** (idioms, metaphors, ellipses, and sarcasm)	Failure to grasp cultural idioms, metaphors, ellipsis, and sarcasm	14	24.6%
3	**Combined (prosodic + pragmatic)**	Cases where both prosodic and contextual cues are lost	11	19.3%
4	**Literal translation of cultural expressions**	Reduction of culturally bound expressions to their literal lexical meaning	8	14.0%
5	**Emphatic /g/ normalization (dialectal /g/ under stress)**	Flattening of the dialectal /g/ (realized as /g/ under stress) to standard "g" or "j"	4	7.0%
6	**Silence & breath deletion**	Removal of sighs, pauses, and breathiness as non-linguistic noise	4	7.0%

4.2 Error Distribution (N=57)



4.3 Key Linguistic Error Dimensions

1. **Prosodic Filtering (Audio Information Loss)**
   Vowel elongation in Distress, Emphatic dialectal /g/ produced under stress (anger, frustration, or high-arousal) is normalized to "g" or "j", removing an acoustic marker of emotion.
2. **Mumbles and Stuttering Deletion**
   Whisper filters out mumbles, hesitations, repetitions, and disfluencies, which are strong indicators of emotional intensity.
3. **Silences and Breath Sounds**
   Pauses, breathiness, and sighs are removed as "noise", depriving the system of intensity and rhythm markers.
4. **Pragmatic Errors (Contextual Information Loss)**
   Idioms and metaphors are misinterpreted; cultural metaphors and idioms are interpreted literally or misunderstood due to lack of cultural grounding.
5. **Syntactic Ellipses Ignored**
   Elliptical structures ("7 packs a day") are misread as nonsensical or neutral due to missing pragmatic cues.
6. **Sarcasm Not Detected**
   Sarcasm tone at the discourse level is lost after ASR normalization and literal interpretation.

4.4 Illustrative Examples

Example 1: Syntactic Ellipsis – "7 packs a day"	
Utterance (Darja):	كان يسور 7 علب في النهار (The used to do 7 packs a day)
ASR Output:	kan ydir 7 3olb fi nhaar
Literal Meaning (Model):	He does 7 boxes a day
Correct Pragmatic Meaning:	He smoked 7 packs of cigarettes a day (seeere addiction)
Error Type:	Pragmatic misinterpretation + Literal translation
Impact on VAD:	Underestimated negative arousal and dominance

Example 2: Prosodic Loss – High-Arousal Anger	
Utterance (Darja):	[illegible] !!! ("By God, I caaaaan't even bear them!")
ASR Output:	wallah ma n9adar hatta nasma3 hadha elkelam
What was lost:	Elongation of "[illegible]"., raised pitch, breathiness, pause
Error Type:	Prosodic filtering + Silence deletion
Impact on VAD:	Converted intense anger (high arousal) to medium or low arousal

4.5 Golden Standard (Human Annotation) vs. Model Outputs

Human Annotation (Golden Standard)
3 native speakers independently annotated S7 utterances using the VAD model. Majority vote was used as the final label.

Golden Standard Outcome
The human baseline captures high-arousal negative states (e.g., intense anger, severe stress) more accurately.

Model Outputs
Text-based models (AraBERT, DziriBERT) underesimante arousal and dominance due to prosodic/pragmatic loss.

6 CONCLUSION

This paper highlights the linguistic limits of current sentiment analysis models when applied to Algerian spoken Arabic (Darja). By conducting a qualitative error analysis within the VAD framework on S7 utterances, we ppragmatic dimeestons. The resulcsebnisiterrpret gonistion by dallent negative emotions—such as intense anger or severe stress—arr fnequently underesttnaatle omr misclassified.
severe stress—are frequentld.
Recognizing emotion in Algerian Arabic requires moving beyond the lexical level. We propose a hybrid approach where prosodic cues and pragmatic context are not mere supplements but essential components for any reliable sentiment recognition system in North African dialects.

7 REFERENCES

1. Abdulhail, M., Tetrault, J., A. Rached, E. Gatoui, ARABERT & MARBERT: Deep bidirectional transformers for Arabic, arXiv preprint arXiv:2003.00104, 2020.
2. Alomari, M. et al. Deep learning models for Arabic sentiment analysis: a review, IEEE Access, 10, 2022.
3. Babaheti, R., Saadane, R., When Darja meets NLP: an analytical study of sentiment analysis in Algerian dialect, Revue TAL, 56(1), 45-63, 2015.
4. Cambria, E. et al., practical guide to sentiment analysis. Springer, 2nd ed., 2020.
5. Mahammadi, K., et al., Emotion detection in low-resource languages: a review, ACM Comput. Surv., 55(4), 1–42, 2022.
6. Poria, S. et al., Deep learning for affective computing. Text-based emotion recognition in complex scenarios, [illegible], 2022.
7. Saadane, R., Babaheti, R., Sentiment analysis in Algerian dialects. Challenges and perspectives. International Journal of Advanced Computer Science and Applications, [illegible], 2018.
8. Seddah, D., & Rabah, I., L'Analyse des sentiments en langue arabe: état de l'art, TALN, 27(1), 2016.