The SlangTrack Dataset: Supporting the Detection of Words Used in Slang Senses

Afnan Aloraini, Goran Nenadic, Viktor Schlegel, Riza Batista-Navarro

The University of Manchester, UK | Qassim University, Saudi Arabia

LChange @ EACL 2026

afnan.aloraini@manchester.ac.uk

SlangTrack/SlangTrack-ST

The Problem: Slang as a Moving Target

Fluidity

Slang evolves rapidly, making it a persistent challenge for NLP systems.

Dual-Meaning Words

Many terms alternate between slang and non-slang senses.

Contextual Ambiguity

Simple lexical lookup fails to capture register-level distinctions.

Example: "Salty"

Context A: "The soup is too **salty**."
(Literal / Non-slang)

Context B: "She was **salty** after losing."
(Figurative / Slang)

Reframing the Task

Traditional Approach

Category-Level Detection

Focuses on identifying whether a word *belongs* to the category of slang.

Question: "Is this word slang?"

Method: Lexical lookup or dictionary-text mismatch.

Our Reframing

Instance-Level Disambiguation

Focuses on distinguishing slang from non-slang *senses* of the same lexical item in context.

Question: "Is this specific occurrence used in a slang sense?"

Method: Register-level contextual disambiguation.

Inspired by Word Sense Disambiguation (WSD)

SlangTrack Dataset — Design and Construction

10 Dual-Meaning Target Words

BMW	Brownie	Chronic	Climber	Cucumber	Eat	Germ	Mammy	Rodent	Salty

12,712

Labelled Instances

0.887

Cohen's Kappa

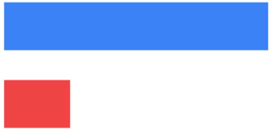
(High Reliability)

2

Distinct Registers

(Formal vs. Informal)

Class Distribution



Non-Slang 10,105 (80%)

Slang 2,607 (20%)

Diachronic Structure

CCOHA	Twitter
1980–2010	2010–2020

Captures both historical formal usage and contemporary informal usage.

Annotation Details & Quality

1. Pilot Phase

Initial testing with three annotators to refine guidelines and sense inventories.

2. Independent Labeling

Two annotators independently labelled all 12,712 instances using finalised guidelines.

3. Adjudication

Disagreements are resolved by a primary annotator with a linguistic background.

0.887

Cohen's Kappa

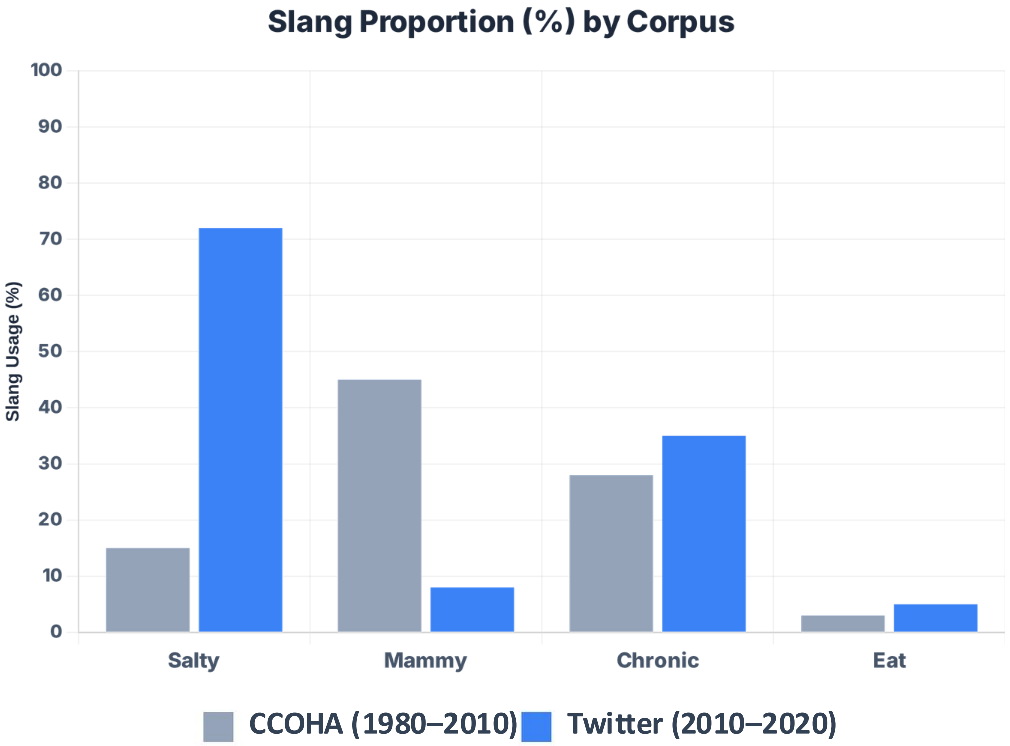
High Inter-Annotator Agreement (IAA)

SlangTrack vs. Existing Resources

Dataset	Source	Temp. Cov.	Polysemy	Annotation
Urban Dictionary	Crowdsourced	Contemp.		Word/Entry-level
Online Slang Dictionary (OSD)	Curated slang dictionary	Contemp.		Word/Entry-level
Green's Dictionary of Slang	Lexicographic dictionary	Hist.		Word/Entry-level
Reddit Glossaries	Community slang glossaries	Contemp.		Word/Glossary-level
OpenSub-Slang	Scripted dialogue	Contemp.		Sentence-level
SlangTrack (ST)	Naturally occurring text	Both	Dual-meaning	Instance-level

To our knowledge, SlangTrack is the first to provide gold-labelled, instance-level disambiguation for words that alternate between slang and standard senses across historical and contemporary registers.

Diachronic Variation: Slang Senses Shift



Innovation

Words like **'Salty'** show a massive shift from literal in CCOHA to overwhelmingly slang in Twitter (>70%).

Decline

'Mammy' slang senses were frequent in CCOHA but have become rare in modern Twitter data.

Stability

Items like **'Eat'** and **'Cucumber'** remain largely stable, with literal senses prevailing in both corpora.

Benchmarking & Methodology

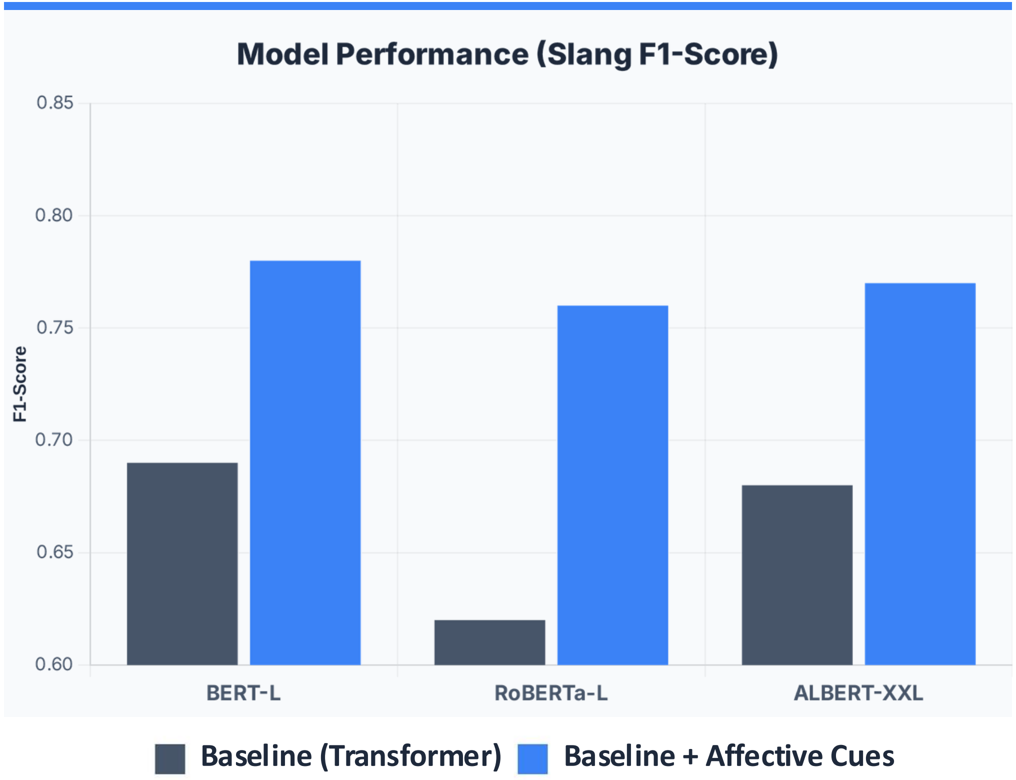
Model Benchmarking

- **Basic Neural Classifiers:** CNNs & BiLSTMs with embeddings (FastText, GloVe, BERT).
- **Transformer-based LMs:** Fine-tuned BERT-L, ALBERT-XXL, RoBERTa-L, XLNet-L.
- **Large Language Models (LLMs):** GPT-4o (Zero-/Few-shot, Fine-tuned) and LLaMA-3.1 (Fine-tuned).

Affective Cue Innovation

- **Sentiment:** HuggingFace sentiment model (nlptown/bert-base-multilingual-uncased-sentiment)
- **Emotion:** GoEmotions model (bhadresh-savani/bert-base-go-emotion)

Key Results: The Power of Affective Cues



Performance Boost

Incorporating **sentiment and emotion** features consistently improves F1-scores. For BERT-L, Slang F1 improves from 0.69 to 0.78 **(+0.09)**.

Transformers vs. LLMs

Fine-tuned models outperform prompting-only LLMs, highlighting the importance of supervised adaptation for slang disambiguation.

Per-Word Performance Analysis

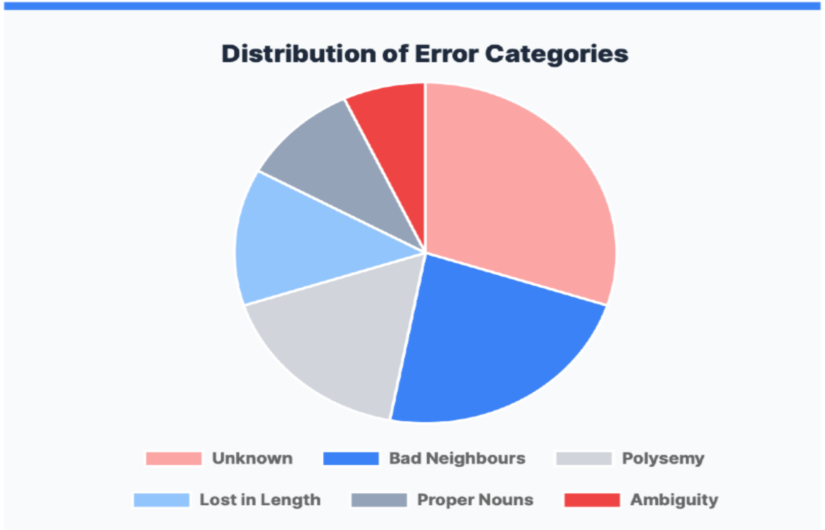
Performance varies significantly across target words, highlighting diverse levels of ambiguity and class balance for slang detection.

Word	Non-slang F1	Slang F1	Macro F1
BMW	0.995	**0.421**	0.708
Brownie	0.872	**0.800**	0.836
Chronic	0.931	**0.598**	0.765
Climber	0.964	**0.832**	0.898
Cucumber	0.960	**0.415**	0.688
Eat	0.905	**0.454**	0.680
Germ	0.848	**0.632**	0.740
Mammy	0.929	**0.458**	0.693
Rodent	0.844	**0.651**	0.748
Salty	0.870	**0.909**	0.890

Key Takeaway

Words like **'Salty'** achieve high slang F1, while ambiguous or infrequent slang senses (e.g., **'Cucumber', 'Mammy'**) yield lower scores, underscoring the challenge of fine-grained slang detection.

Error Taxonomy: Understanding Model Limitations



Manual analysis of 100 misclassified instances (Section 7).

Unknown (30%)
Unconventional abbreviations, rare slang, or novel expressions.

Bad Neighbours (23%)
Misleading local context or confounding lexical cues near the target word.

Polysemy (17%)
Target words with multiple meanings, both slang and non-slang.

Lost in Length (13%)
Very long or extremely short instances where pragmatic signals are diluted or truncated.

Proper Nouns (10%)
Proper nouns, especially bi-grams or tri-grams, misconstrued as slang.

Ambiguity (7%)
Broader sentence context creates uncertainty, even if the word's The meaning is clear.

Conclusion & Future Work

Key Contributions

Reframing: Shifted slang detection from lexical classification to instance-level register disambiguation.

Dataset: Released SlangTrack, a diachronic resource with 12.7k gold-labelled instances.

Innovation: Demonstrated the critical role of affective cues in resolving slang senses.

Future Directions

Expansion: Scaling to more target words, diverse languages, and varied social media platforms.

LLM Analysis: Investigating how diachronic shifts in slang are reflected in LLM internal knowledge.

Context: Integrating broader pragmatic and social context for more robust disambiguation.

Thank you! For questions or collaboration: afnan.aloraini@manchester.ac.uk