

# **Cross-Arabic Dialectal POS Tagging: A Comparative Study**

Nir Shoham & Omer Kuriel  
NLP Course, Reichman University, 2025

## Motivation

- Arabic has rich dialectal variation, and MSA differs significantly from spoken dialects.
- Annotated dialectal data is rare, and most POS tag models are trained on MSA.
- POS tagging is a core NLP task and a good testbed for studying cross-dialect generalization.

## Research Questions

- How an MSA-trained model performs on dialects in a zero-shot setting
- What changes when models are fine-tuned on one dialect.
- Which systematic errors emerge across dialects in both zero shot and fine tuned models

## Data

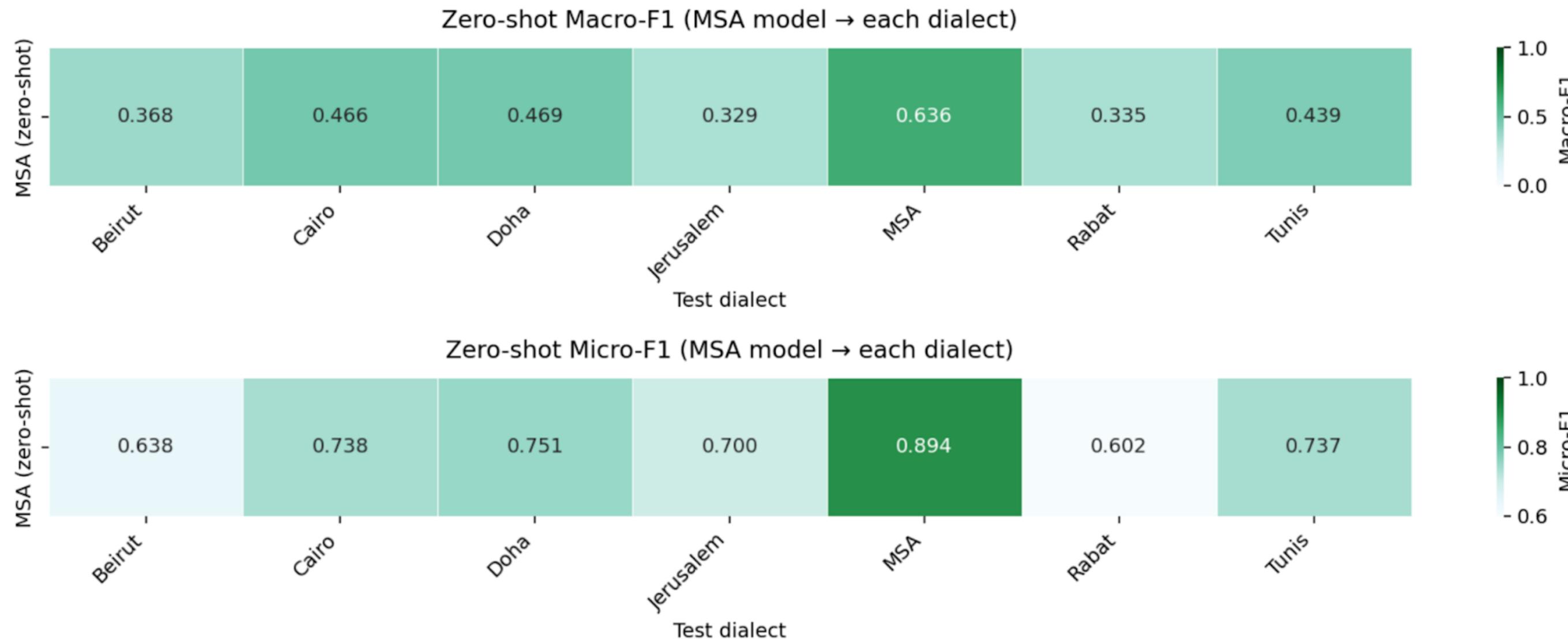
- We manually annotated 700 sentences across 7 dialects: Beirut, Cairo, Doha, Tunis, Rabat, Jerusalem, and MSA.
- Each subset included 100 sentences, tagged with a unified POS tagset.
- This dataset allowed for consistent cross-dialect evaluation.

## Methodology

- **Phase 1** - tested zero-shot performance using CAMeLBERT-CA trained on MSA.
- **Error analysis** for the zero shot performance with precision, recall, Macro-F1, Micro-F1, and Conservative F1 metrics, and most common mistakes
- **Phase 2** - fine-tuned the model on each dialect
- **Error analysis** across the fine tuned models with precision, recall, Macro-F1, Micro-F1, and Conservative F1 metrics and most common mistakes.

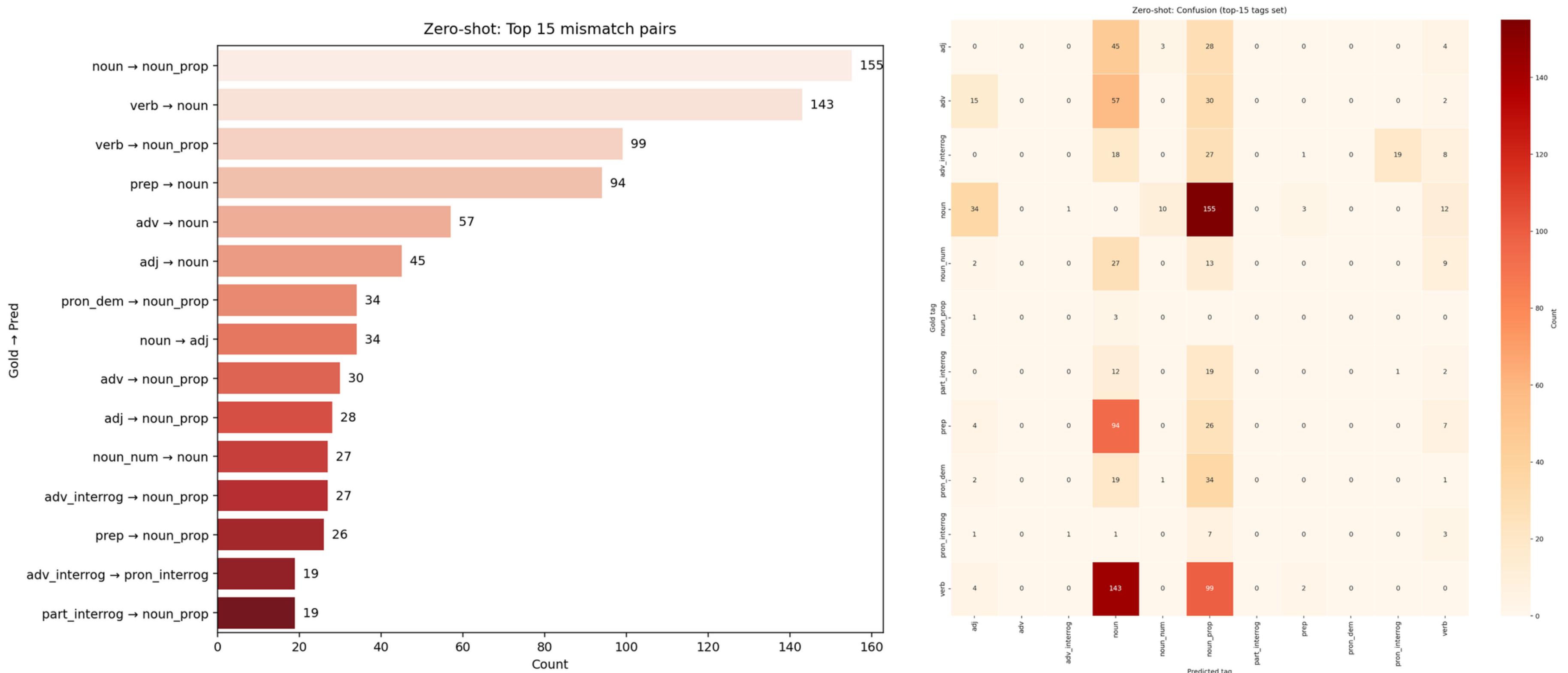
# Zero-Shot Results

- As expected, **MSA achieved the best performance** in both macro and micro scores.
- Across dialects, Micro-F1 was much higher than Macro-F1, showing the model relies on frequent tags.



# Zero-Shot Results

- Errors showed a strong bias toward predicting nouns and proper nouns.



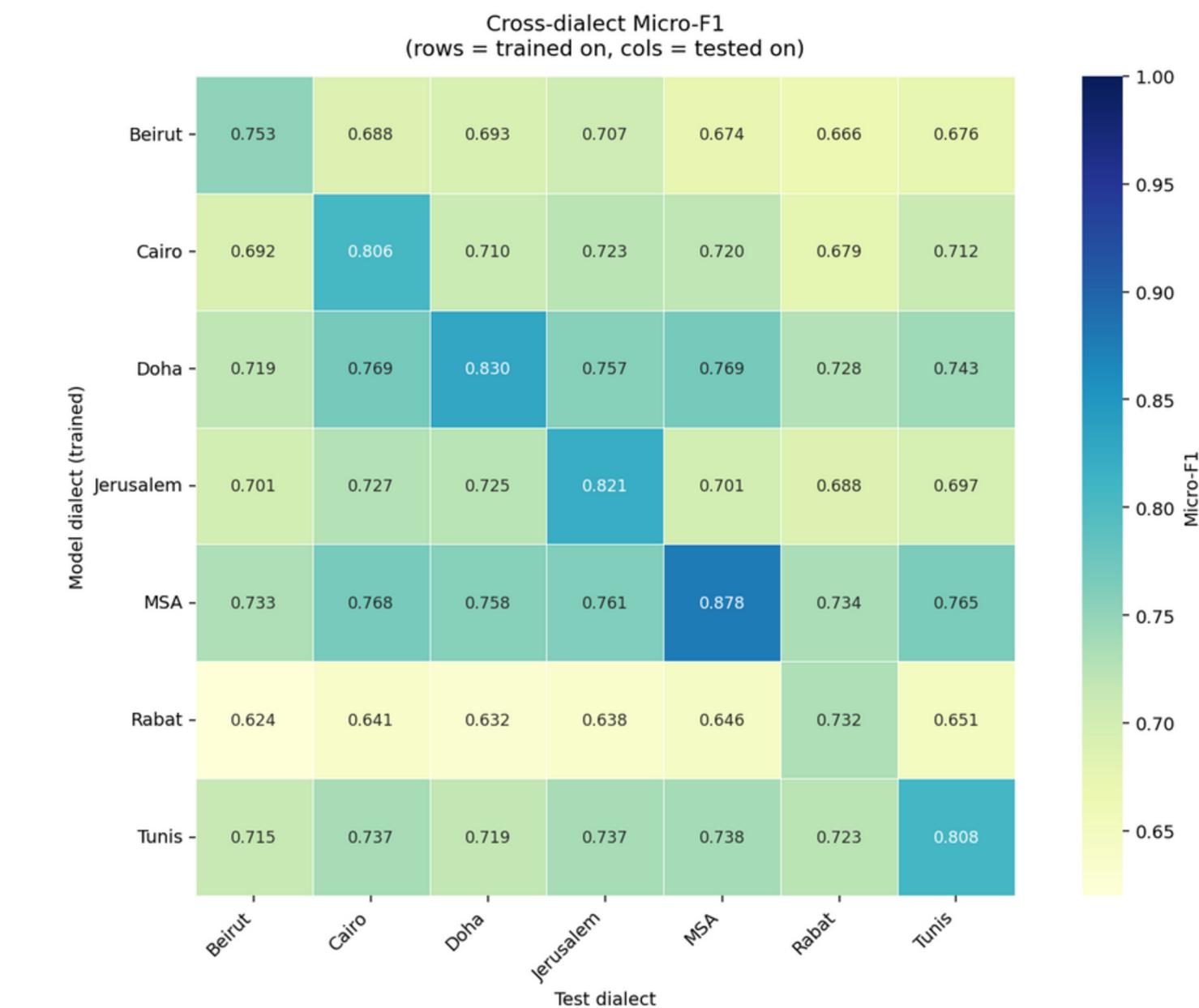
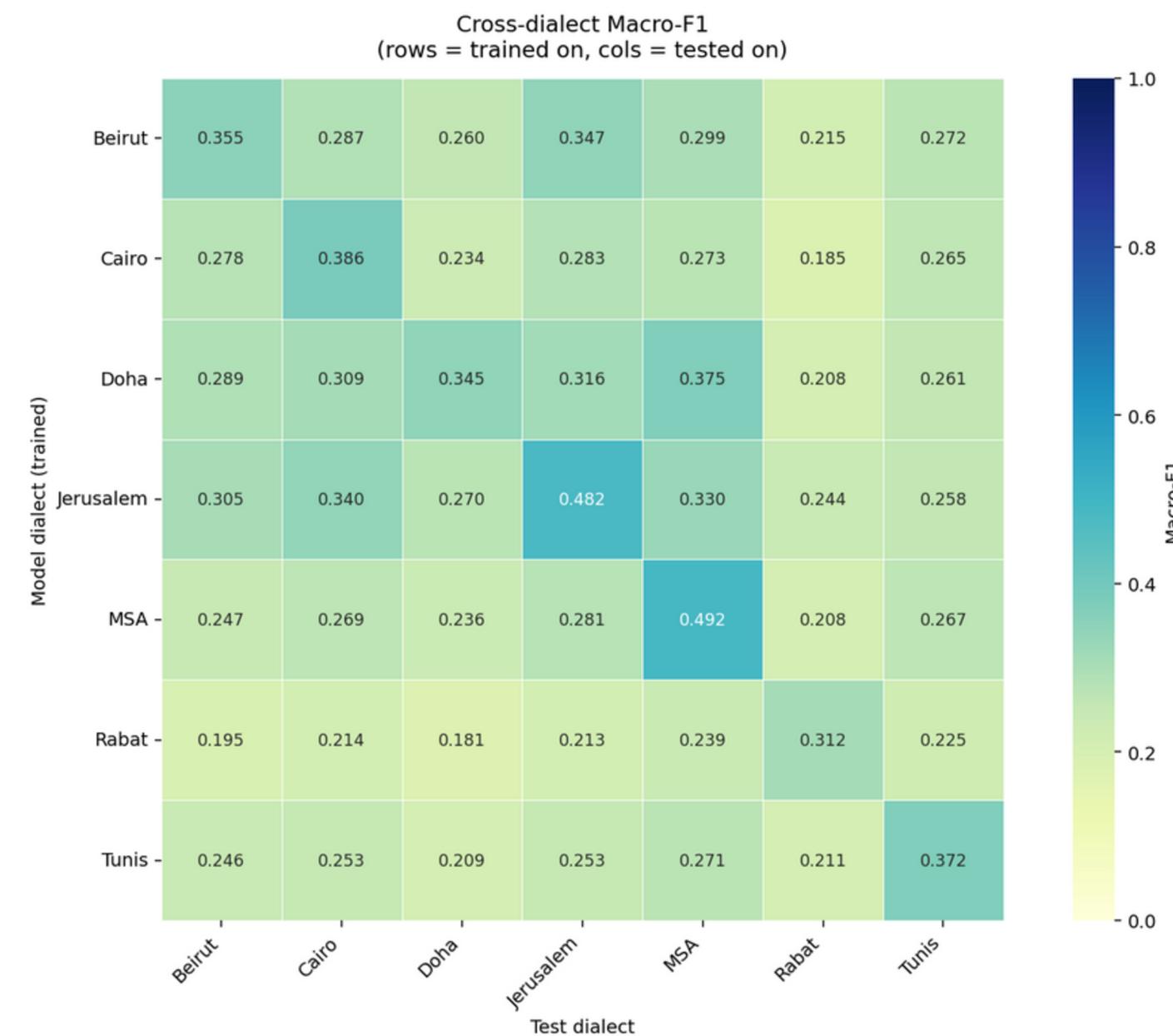
## Zero-Shot Results

Error patterns reveal:

- Difficulty to identify modal verbs - verbs carrying prefixes ("بـ") which represent an indicative/future meaning - the model tends to tag those tokens as nouns.
- Struggle with dialectal spelling and pronunciation variants - For instance, the Levantine form "هذا" (hāda, 'this') replaces the MSA "هذا" (hādhā)

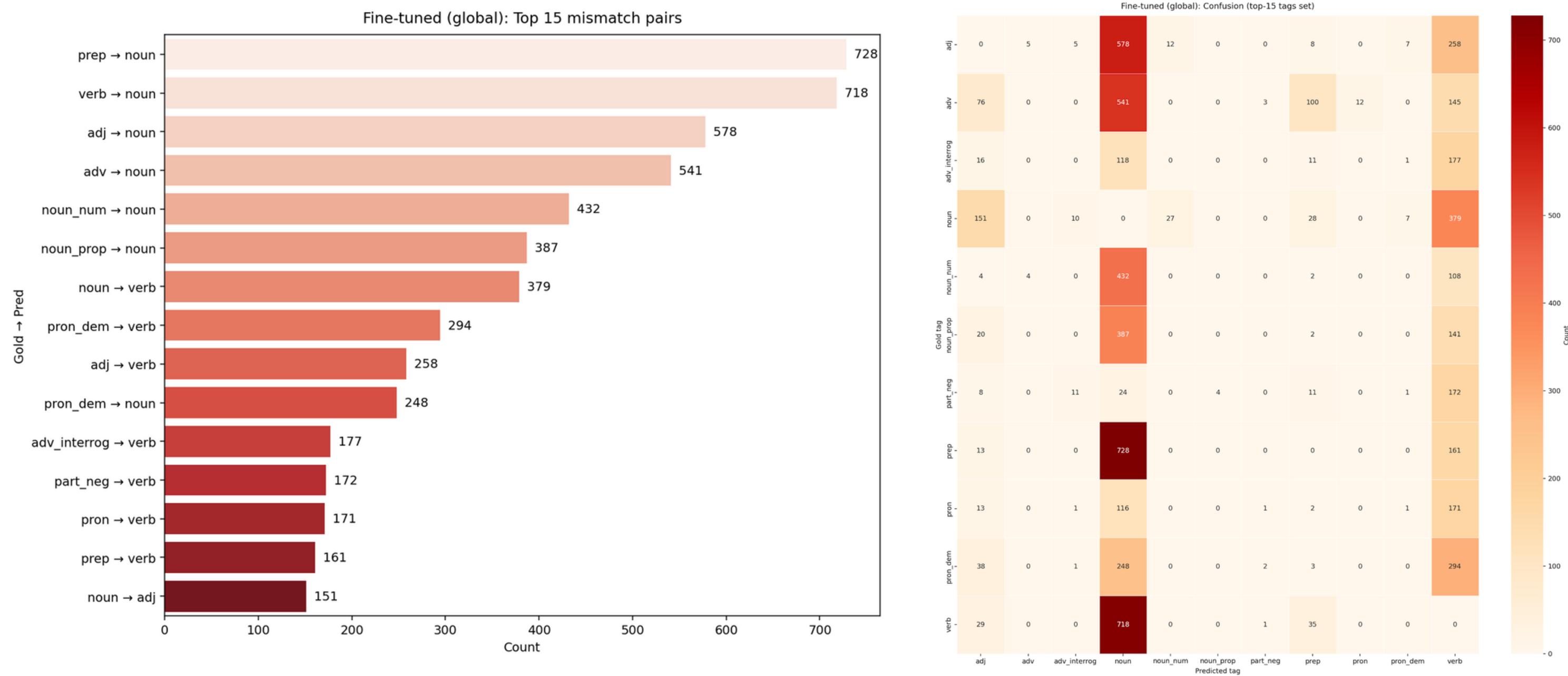
# Fine-Tuned Results

- Fine-tuning improved overall Micro-F1 performance and reduced proper-noun inflation.
- Levantine dialects (Beirut and Jerusalem) showed strong transitive connection.
- Rabat remained the hardest to bridge.



## Fine-Tuned Results

- After fine-tuning, the main bias shifted to noun–verb confusions.



## Key Insights

- Cross-dialect transfer may be influenced by geographic and linguistic distance: Levant dialects transfer best.
- Fine-tuning helps improving performance, but systematic biases remain.
- Rare and morphologically complex tags are especially difficult.
- MSA acts as a strong Micro-F1 “hub”, providing relatively stable performance across dialects.

## Discussion

- Overall, our findings suggest that cross-dialect transfer performance may be influenced by geographic and linguistic distance, with stronger transfer often occurring within regional clusters and weaker transfer across major dialect groups.
- Challenges remain in reducing the performance gap between divergent dialect families.
- These observations could inform future work on domain adaptation, dialect clustering, and the development of resources for underrepresented dialects.