Neural Metaphor Detection in Context

Ge Gao, Eunsol Choi, Yejin Choi, Luke Zettlemoyer

UWNLP

PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

Metaphors Are Pervasive

Metaphors Are Pervasive

The bank pumped $108 bn into the economy.

Metaphors Are Pervasive

The bank pumped $108 bn into the economy.

Metaphors Are Pervasive

- Some statistics (based on our study on benchmark corpus)
  - Sentence-level: 28%
  - Token-level: 11%

Metaphors Are Pervasive

- Some statistics (based on our study on benchmark corpus)
  - Sentence-level: 28%
  - Token-level: 11%

- Potential application of metaphor processing
  - Machine Translation   (Mao et al., 2018)
  - Sentiment Analysis (Mohammad et al., 2016)
  - Educational Applications (Kordoni, 2018)
  - Relation Extraction

  ……

Task

Task

1. Classification Task

✓

The bank pumped $108 bn into the economy .

2. Sequence Labeling Task

X ✓ ✓ X X X X X X

The bank pumped $108 bn into the economy .

Related Works

Feature-based approach

concreteness, abstractness, imaginability, feature norms, sensory features, bag-of-words features, WordNet features

Rei et al., 2017, Köper and im Walde, 2017, Bulat et al., 2017, Shutova et al., 2016, Tekiroglu et al., 2015, Tsvetkov et al., 2014, Broadwell et al., 2013, Strzalkowski et al., 2013, Hovy et al., 2013, Turney et al., 2011

Related Works

Limited Context

Feature-based approach

concreteness, abstractness, imaginability, feature norms, sensory features, bag-of-words features, WordNet features

Rei et al., 2017, Köper and im Walde, 2017, Bulat et al., 2017, Shutova et al., 2016, Tekiroglu et al., 2015, Tsvetkov et al., 2014, Broadwell et al., 2013, Strzalkowski et al., 2013, Hovy et al., 2013, Turney et al., 2011

Context is Important

Limited Context

[bank, pump, money]

Full Context

The bank pumped $108 bn into the economy.

Context is Important

Limited Context	Full Context
[bank, pump, money]	The bank pumped $108 bn into the economy.
[experts, examine, country]	The experts started examining the Soviet Union with a microscope to study perceived changes.

Related Works

Limited Context

Feature-based approach

concreteness, abstractness, imaginability, feature norms, sensory features, bag-of-words features, WordNet features

Rei et al., 2017, Köper and im Walde, 2017, Bulat et al., 2017, Shutova et al., 2016, Tekiroglu et al., 2015, Tsvetkov et al., 2014, Broadwell et al., 2013, Strzalkowski et al., 2013, Hovy et al., 2013, Turney et al., 2011

Full Context

Context-aware model

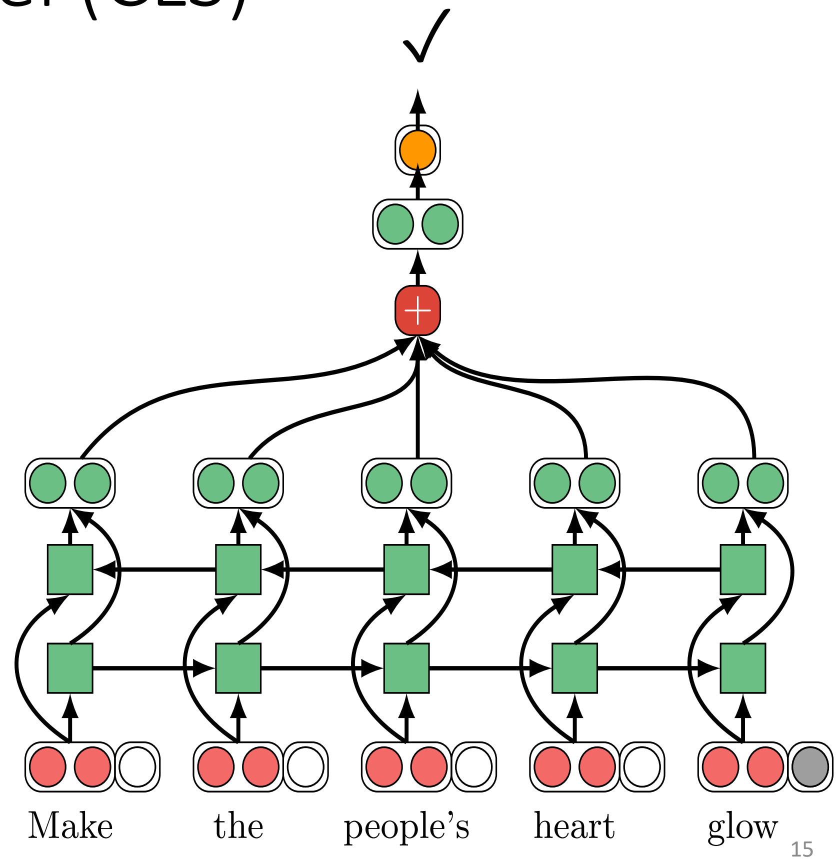
LSTM-based models and CRFs augmented by linguistic features: WordNet, POS tags, concreteness score, unigrams, lemmas, verb clusters

Swarnkar and Singh, 2018; Pramanick et al., 2018; Mosolova et al., 2018; Bizzoni and Ghanimifard, 2018; Wu et al., 2018

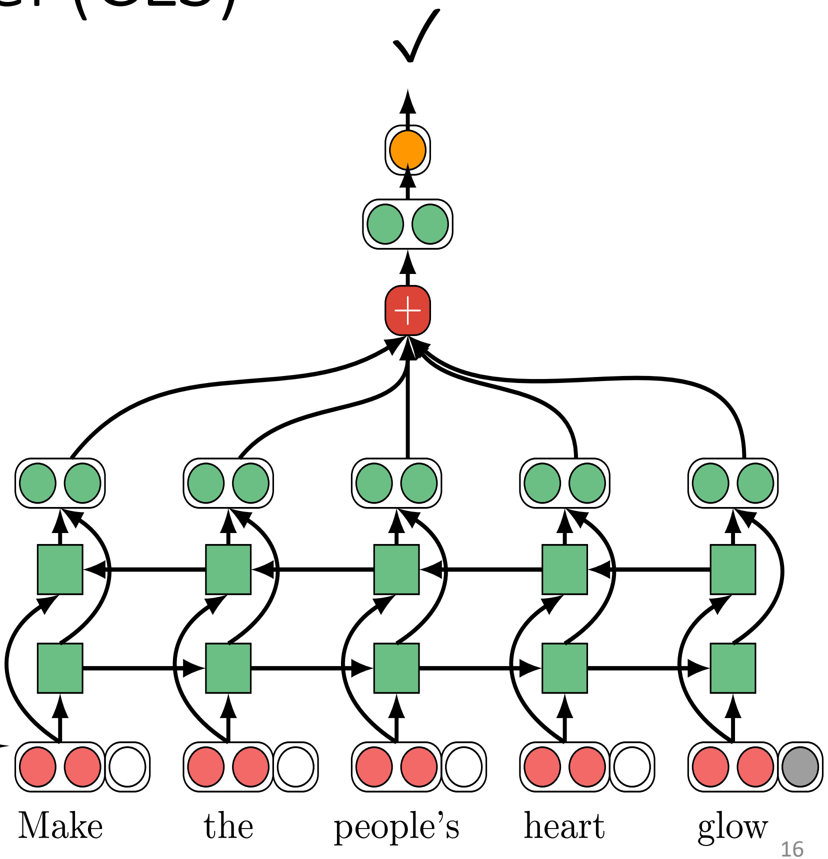
This Work

- Predicting the metaphoricity of all words in context
- BiLSTM models + contextualized word representation (ELMo)
- New state-of-the-arts on benchmark!

Our Classification Model (CLS)



Our Classification Model (CLS)



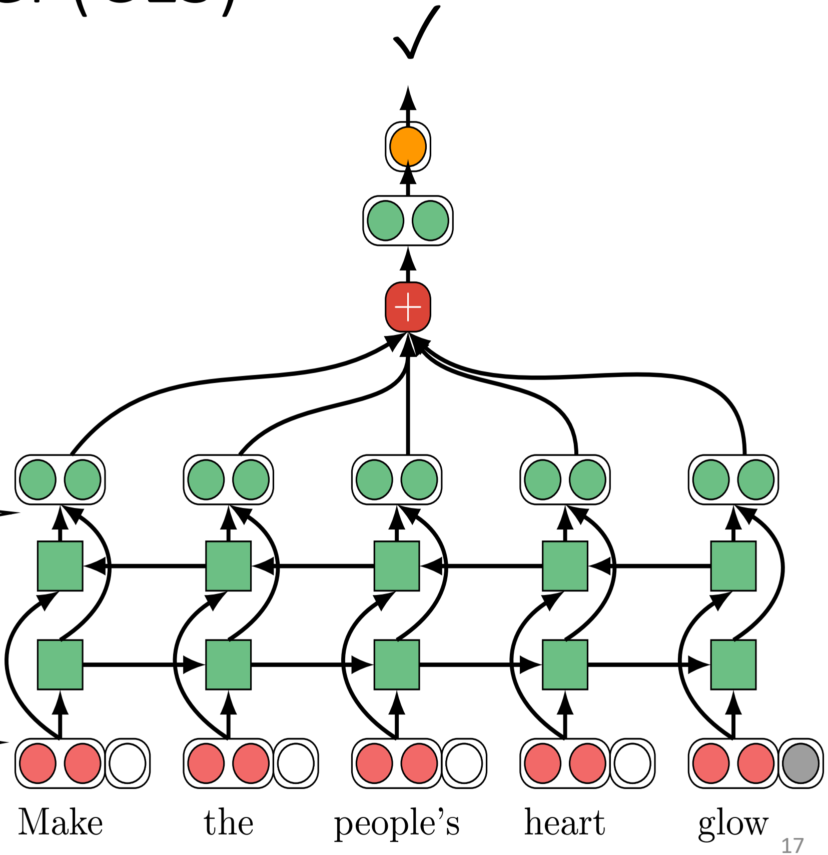
(1)Word Embeddings:
GloVe; ELMo; Binary embedding

GloVe: Pennington et al., (2014)
ELMo: Peters et al. (2018)

Our Classification Model (CLS)

(2) Contextualized representation encoded by BiLSTM

(1)Word Embeddings:
GloVe; ELMo; Binary embedding

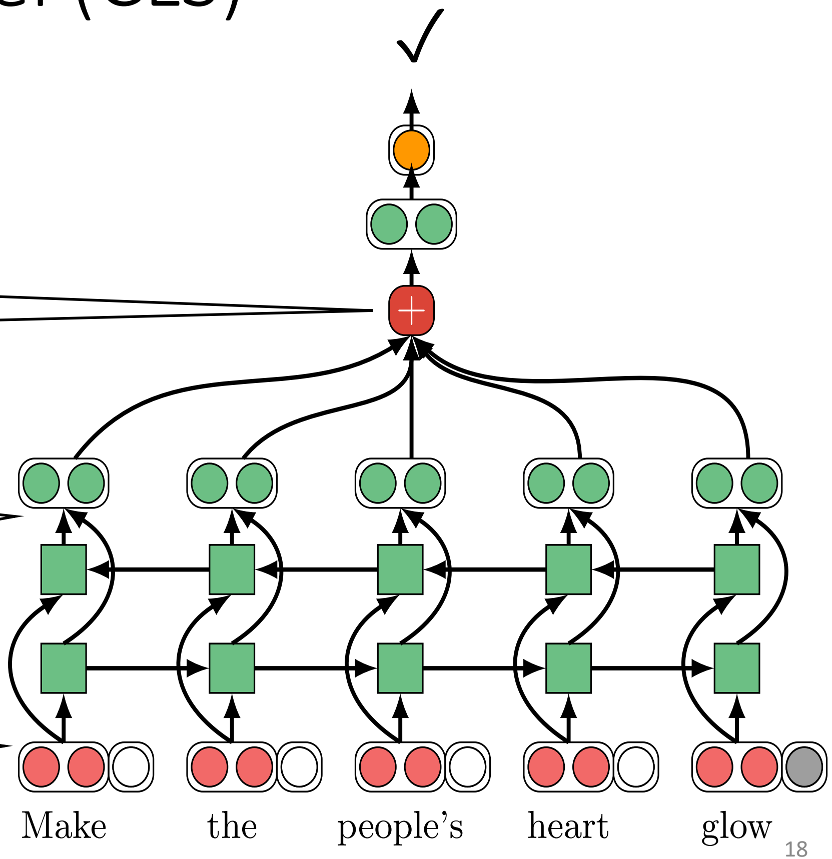


Our Classification Model (CLS)

(3) Attention layer

(2) Contextualized representation encoded by BiLSTM

(1)Word Embeddings:
GloVe; ELMo; Binary embedding



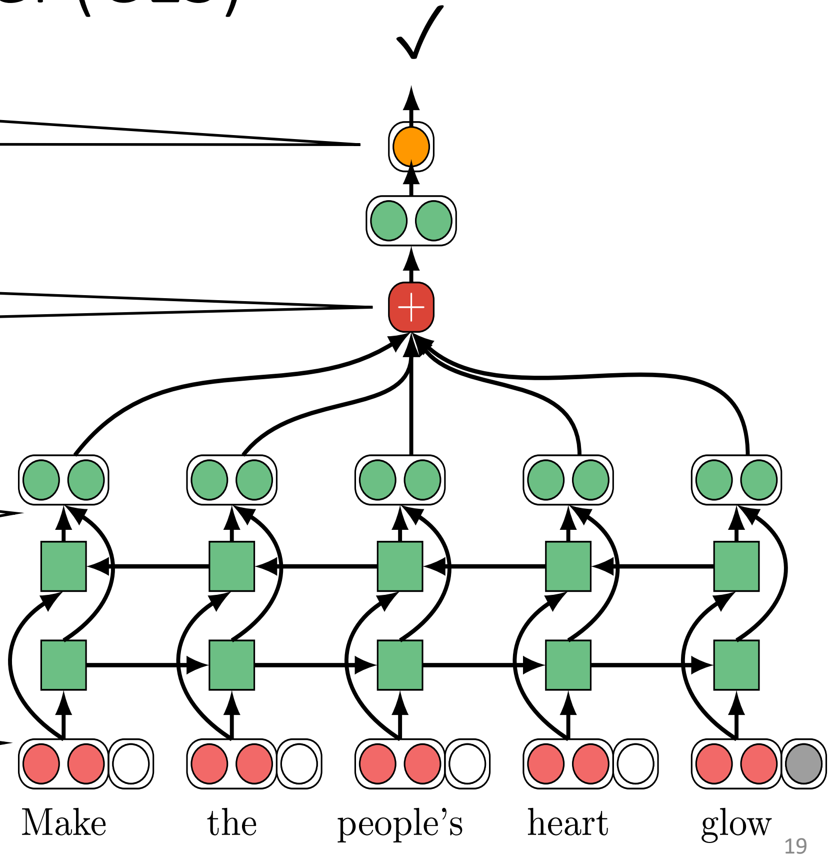
Our Classification Model (CLS)

(4) Feed forward layer

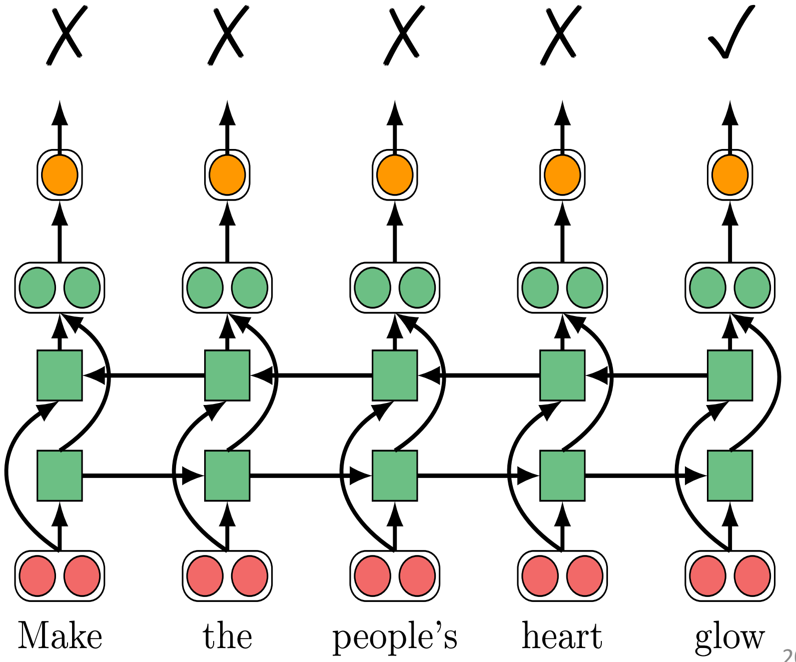
(3) Attention layer

(2) Contextualized representation encoded by BiLSTM

(1)Word Embeddings:
GloVe; ELMo; Binary embedding

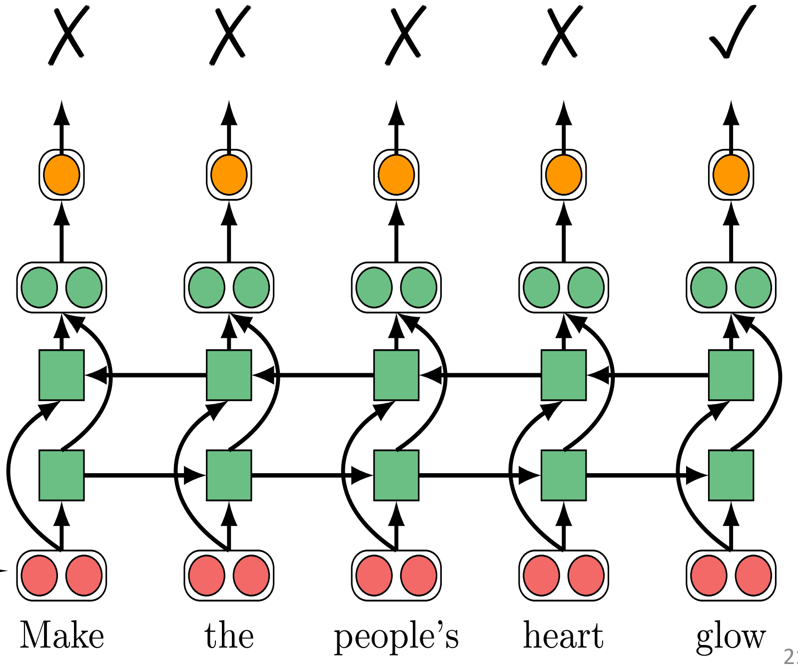


Our Sequence Labeling Model (SEQ)



Our Sequence Labeling Model (SEQ)

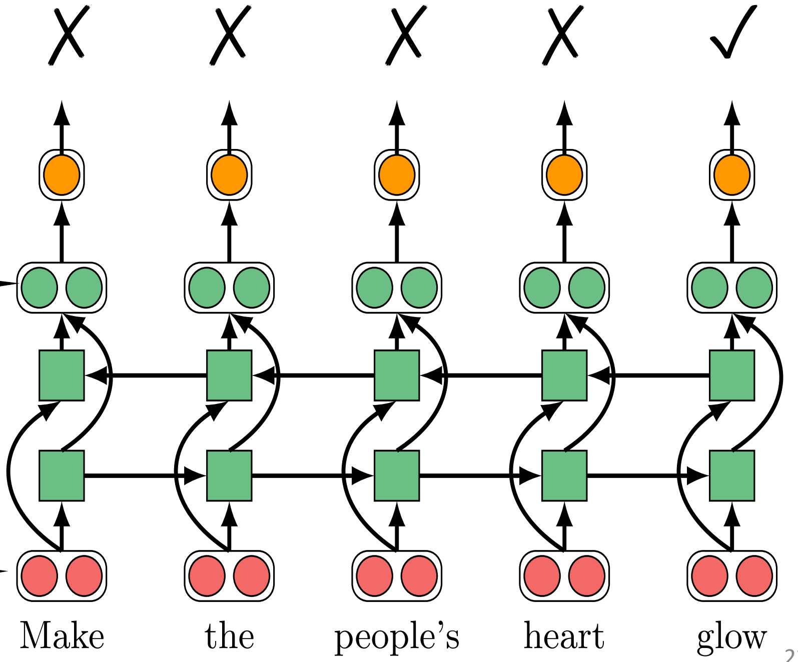
(1)Word Embeddings:
GloVe; ELMo



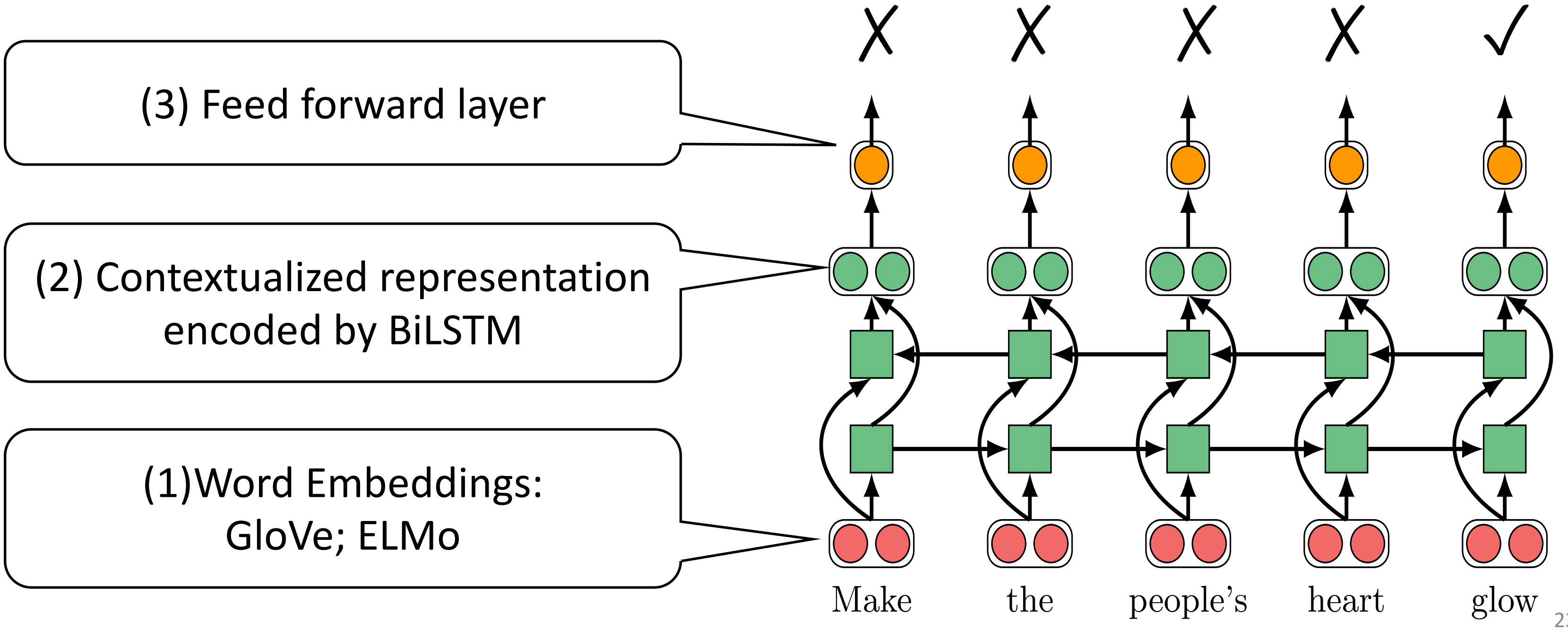
Our Sequence Labeling Model (SEQ)

(2) Contextualized representation encoded by BiLSTM

(1)Word Embeddings: GloVe; ELMo



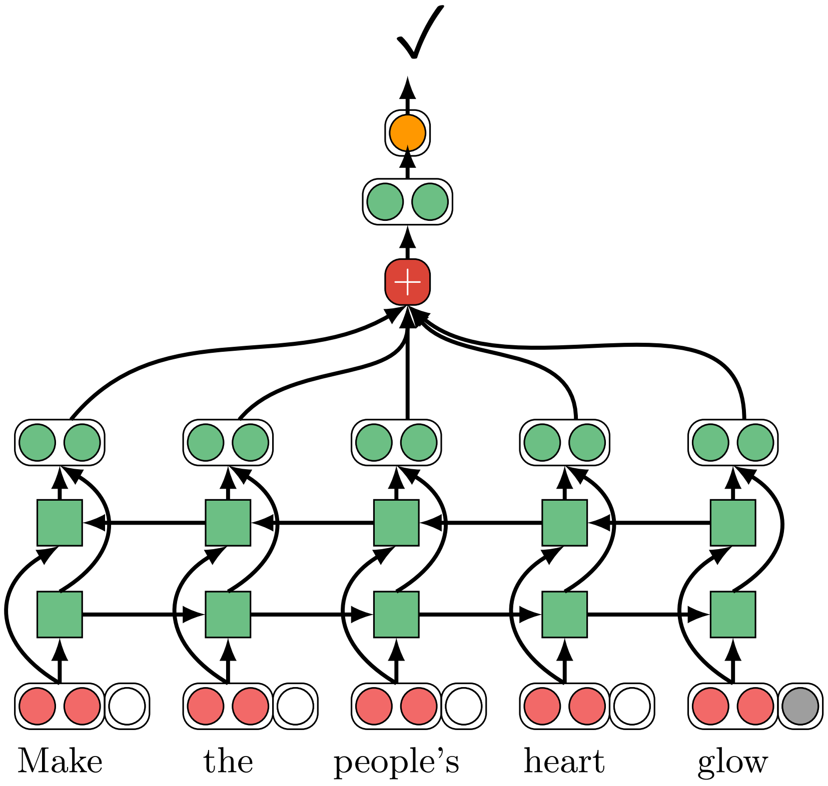
Our Sequence Labeling Model (SEQ)

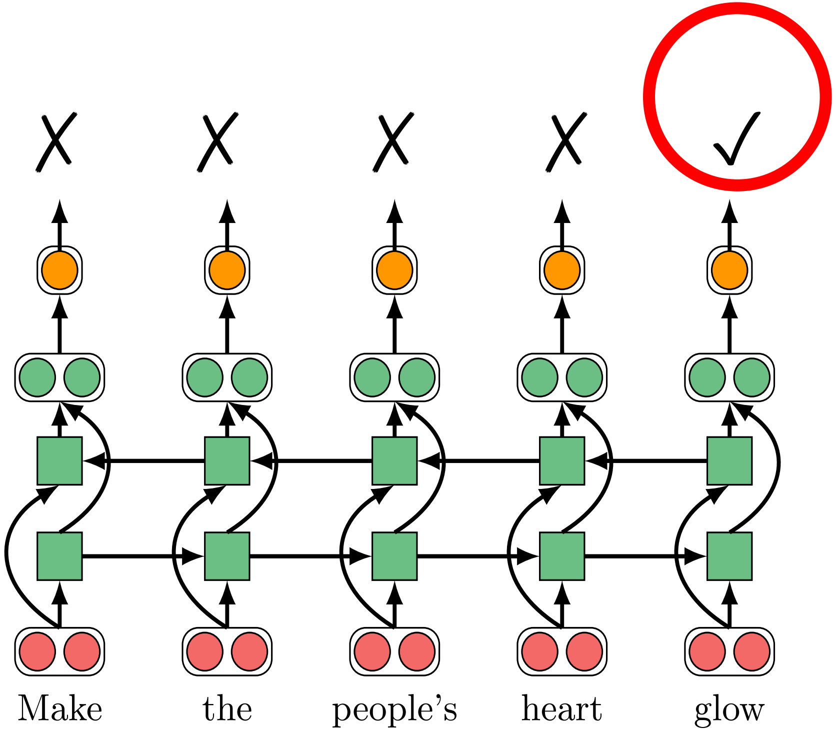


Dataset: Verb Classification

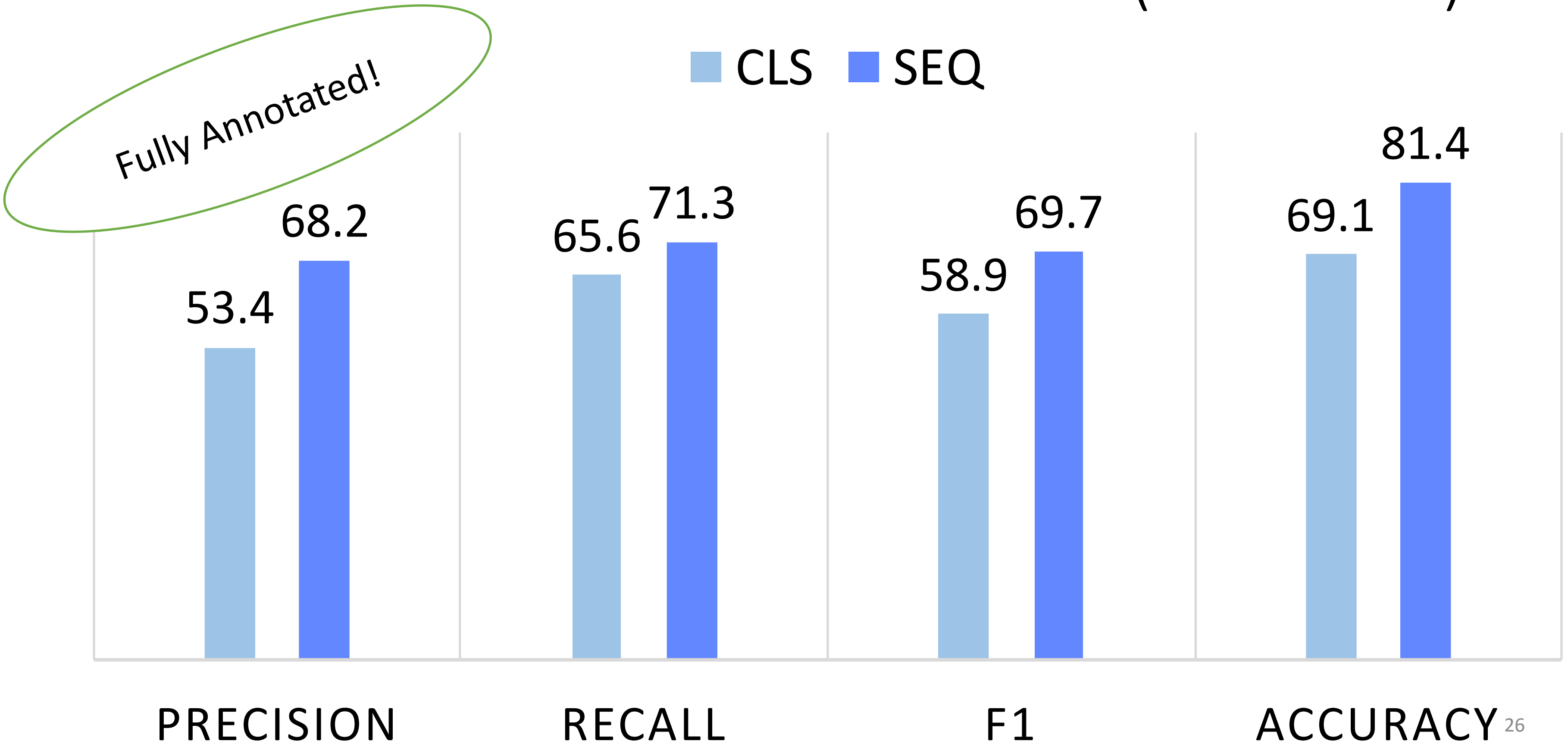
	# Examples	% Metaphor	# Uniq. Verb	Avg Sent. Len
MOH-X Mohammad et al., 2016	647	49%	214	8.0
TroFi Birke and Sarkar, 2006	3,737	43%	50	28.3
VUA Steen et al., 2010	23,113	28%	2047	24.5

Direct Comparison

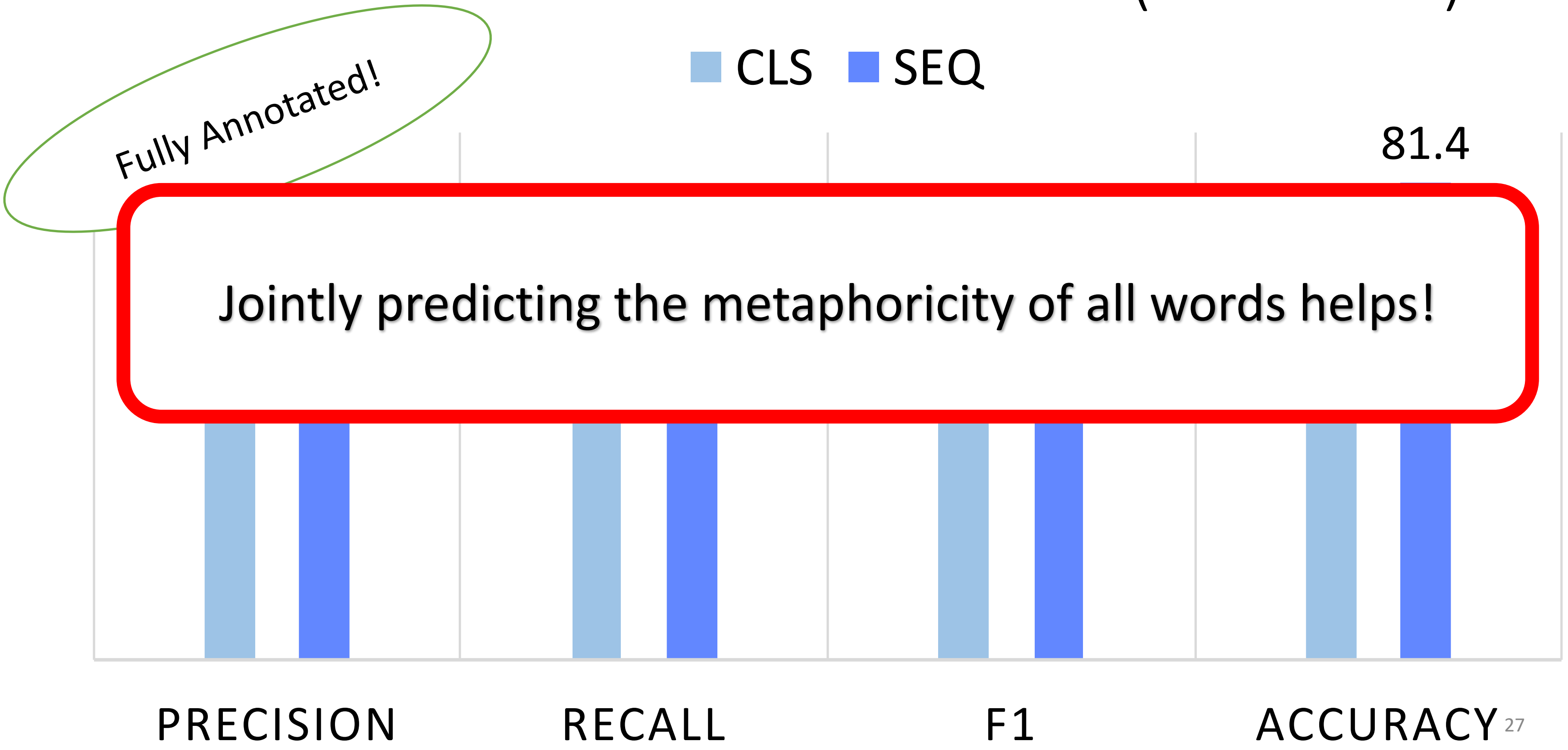




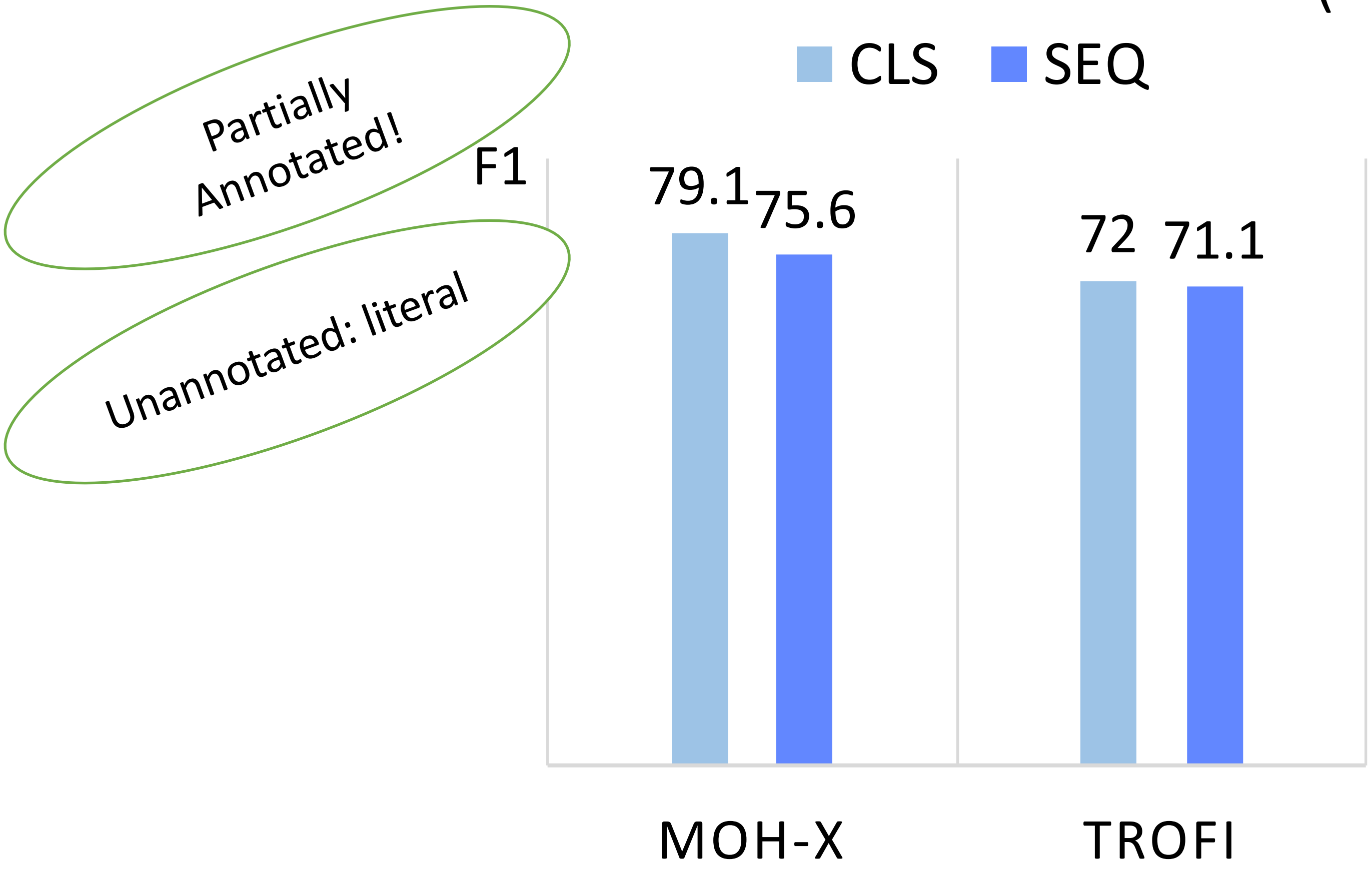
Performance: Verb Classification (VUA-verb)



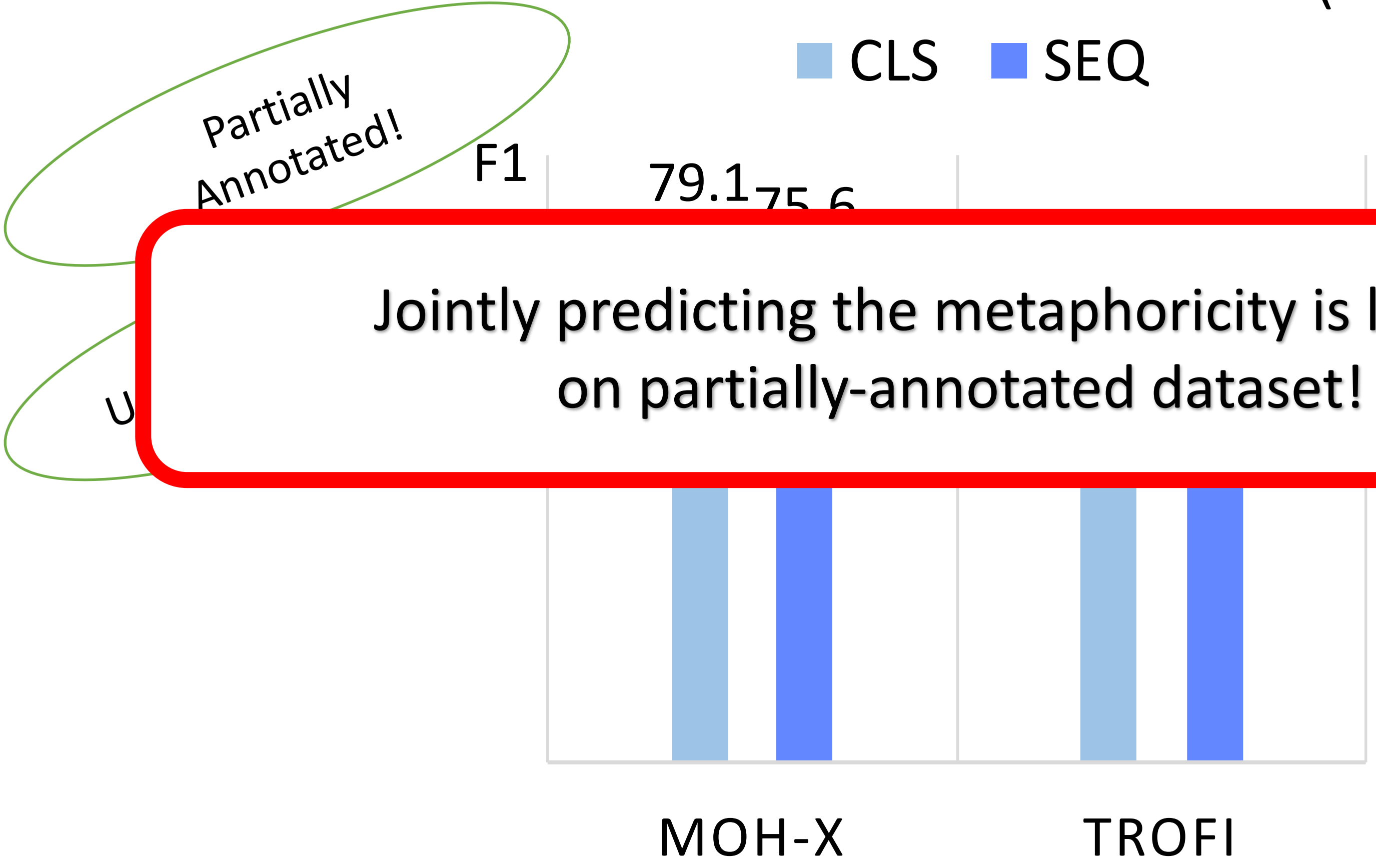
Performance: Verb Classification (VUA-verb)



Performance: Verb Classification (MOH, TroFi)



Performance: Verb Classification (MOH, TroFi)

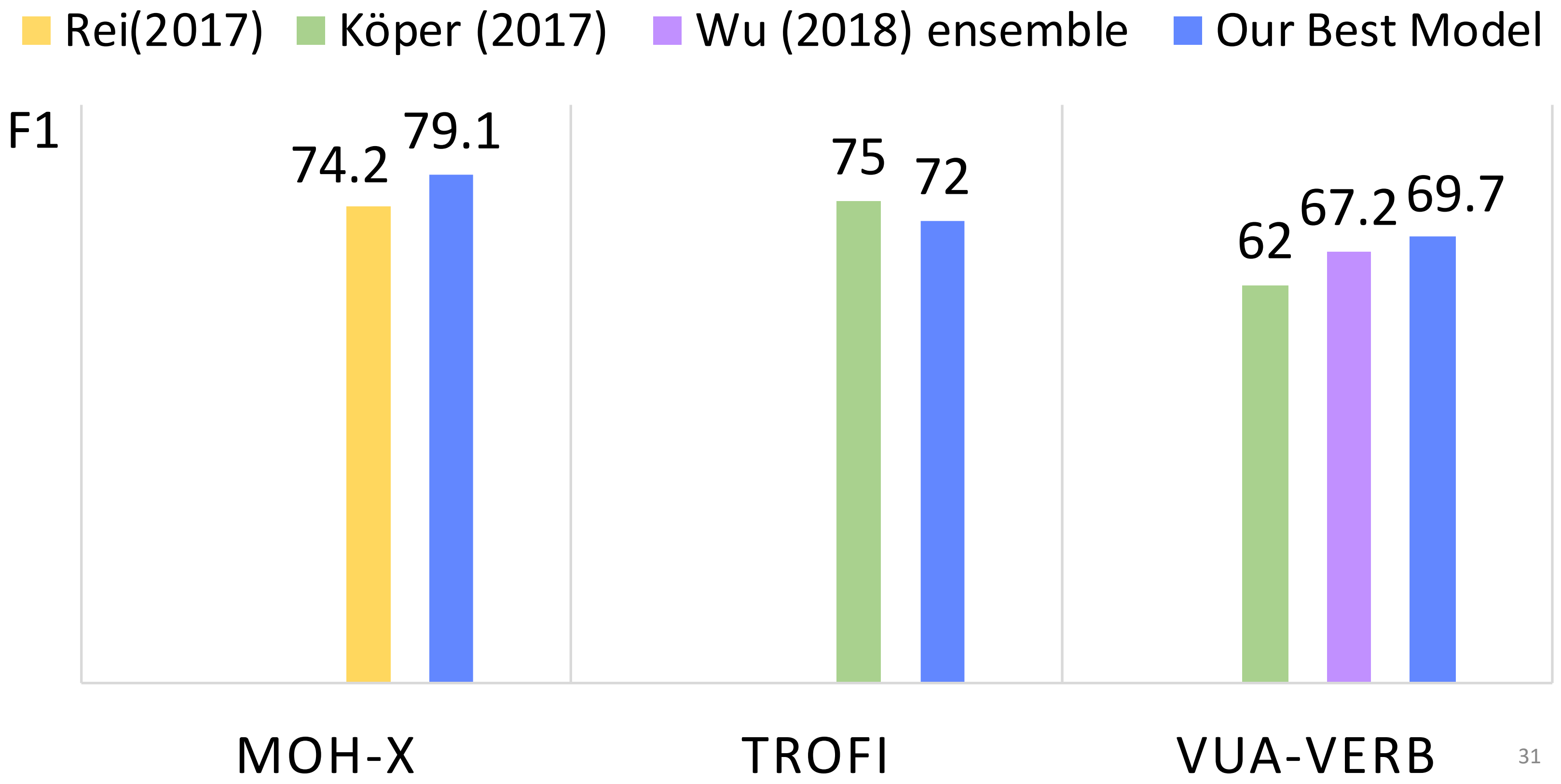


Jointly predicting the metaphoricity is limited on partially-annotated dataset!

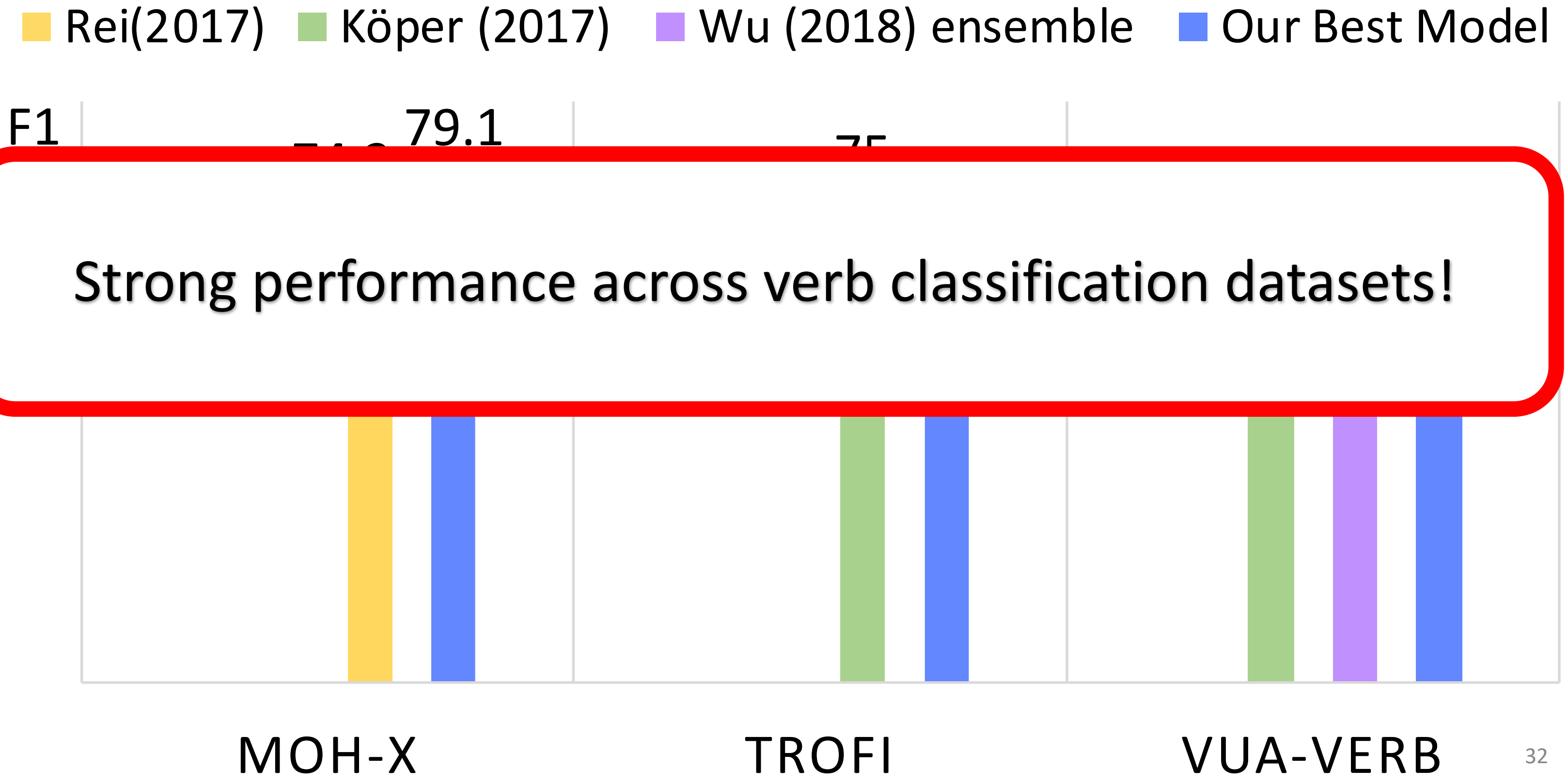
Comparison System

- Wu et al. (2018) Ensemble
  - a CNN-LSTM ensemble model with weighted-softmax classifier
  - pre-trained word2vec + POS tags + word cluster features
- Rei et al. (2017)
  - a neural similarity network on verb-noun pairs
  - pre-trained skip-gram word embeddings
- Köper et al. (2017)
  - a balanced logistic regression classifier
  - target verb lemma + 7 features based on abstractness rating

Performance: Verb Classification

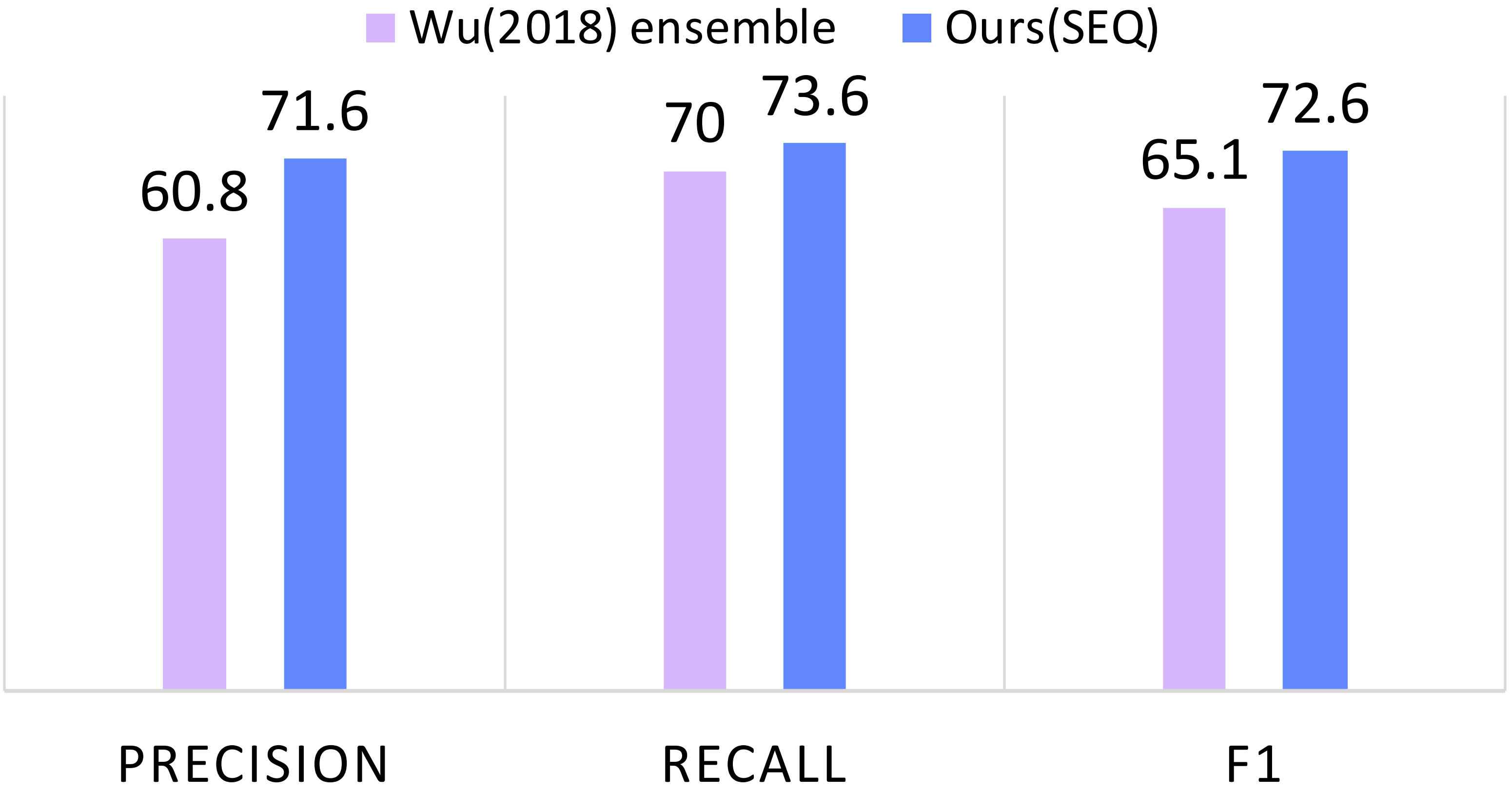


Performance: Verb Classification



Strong performance across verb classification datasets!

Performance: Sequence Labeling (VUA)

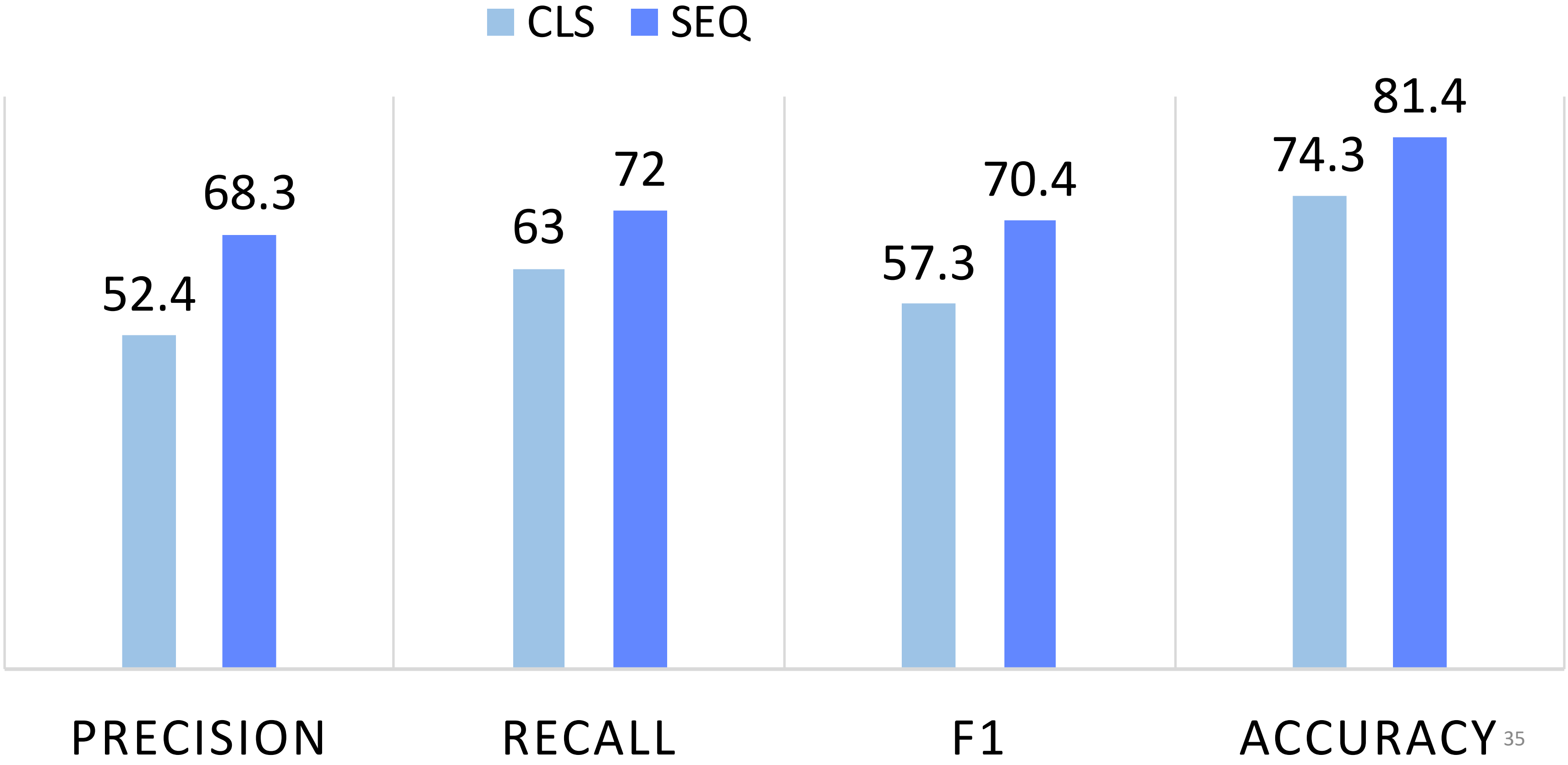


Analysis

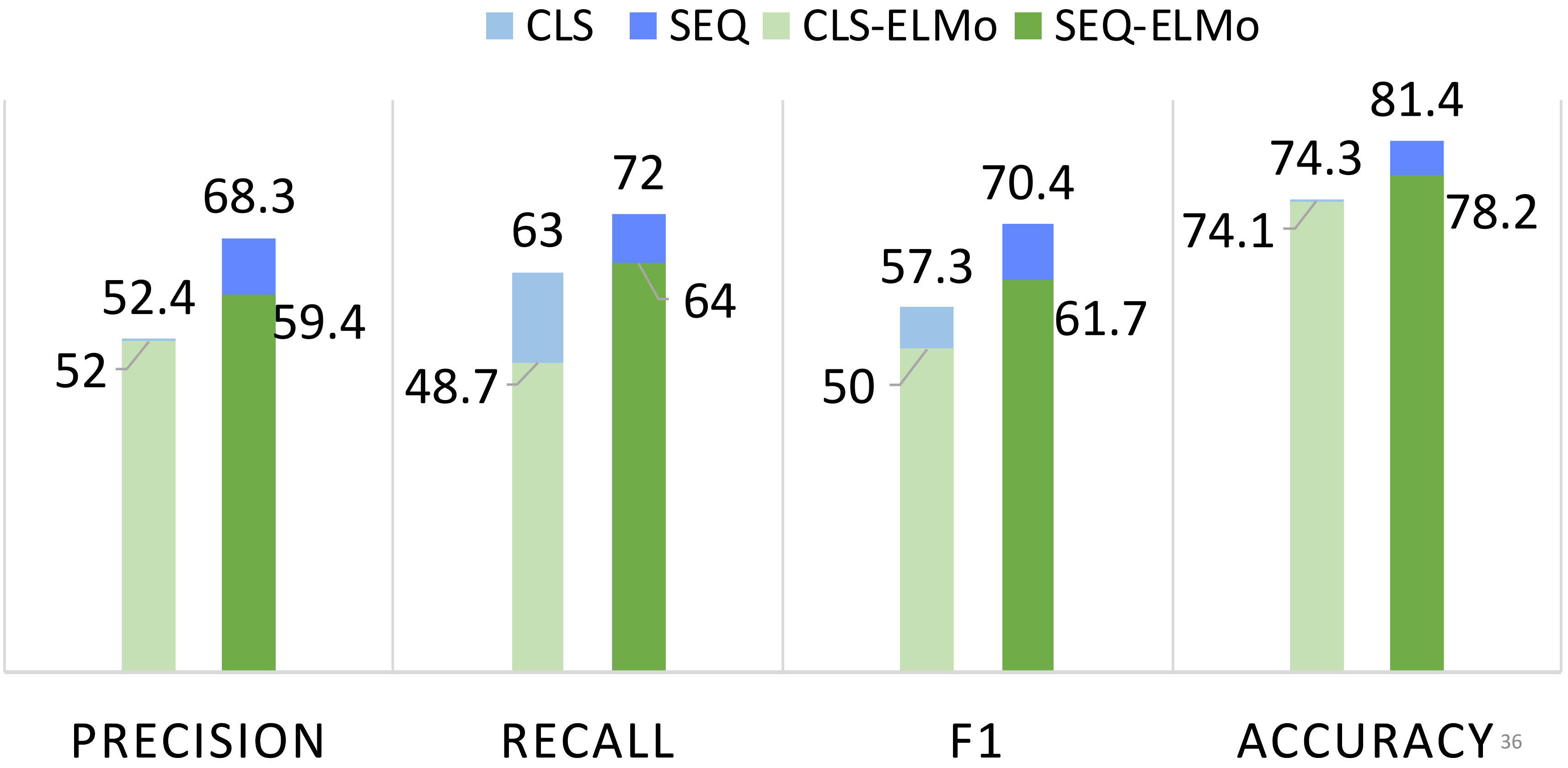
1. How helpful is the contextualized word representation?
2. How well does the model perform on non-verbal POS tags?
3. When does the sequence labeling model outperforms the classification model in general?
4. What kind of errors does the sequence labeling model make on verb classification?

……

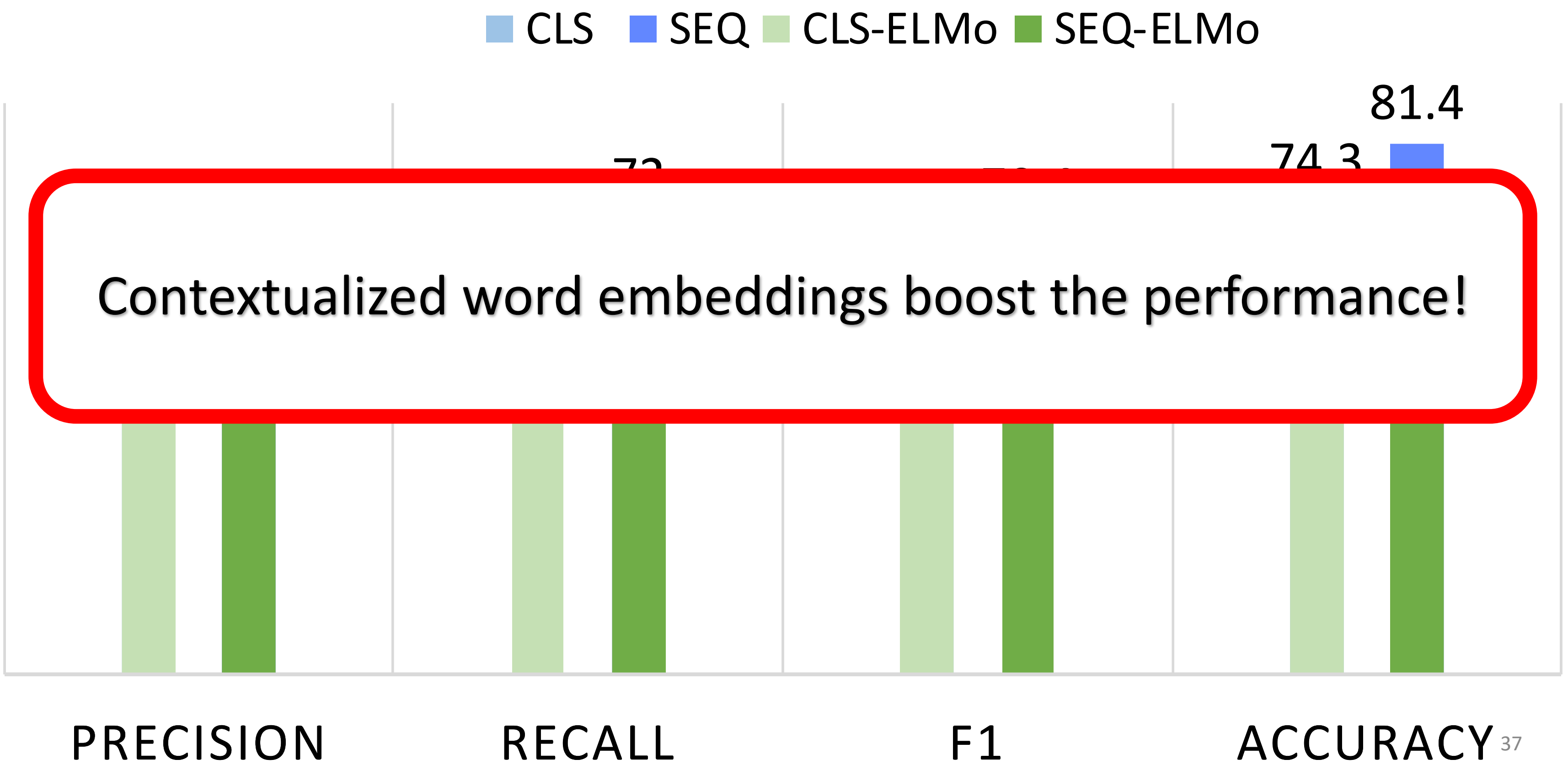
Ablation Study: Verb Classification (VUA dev)



Ablation Study: Verb Classification (VUA dev)

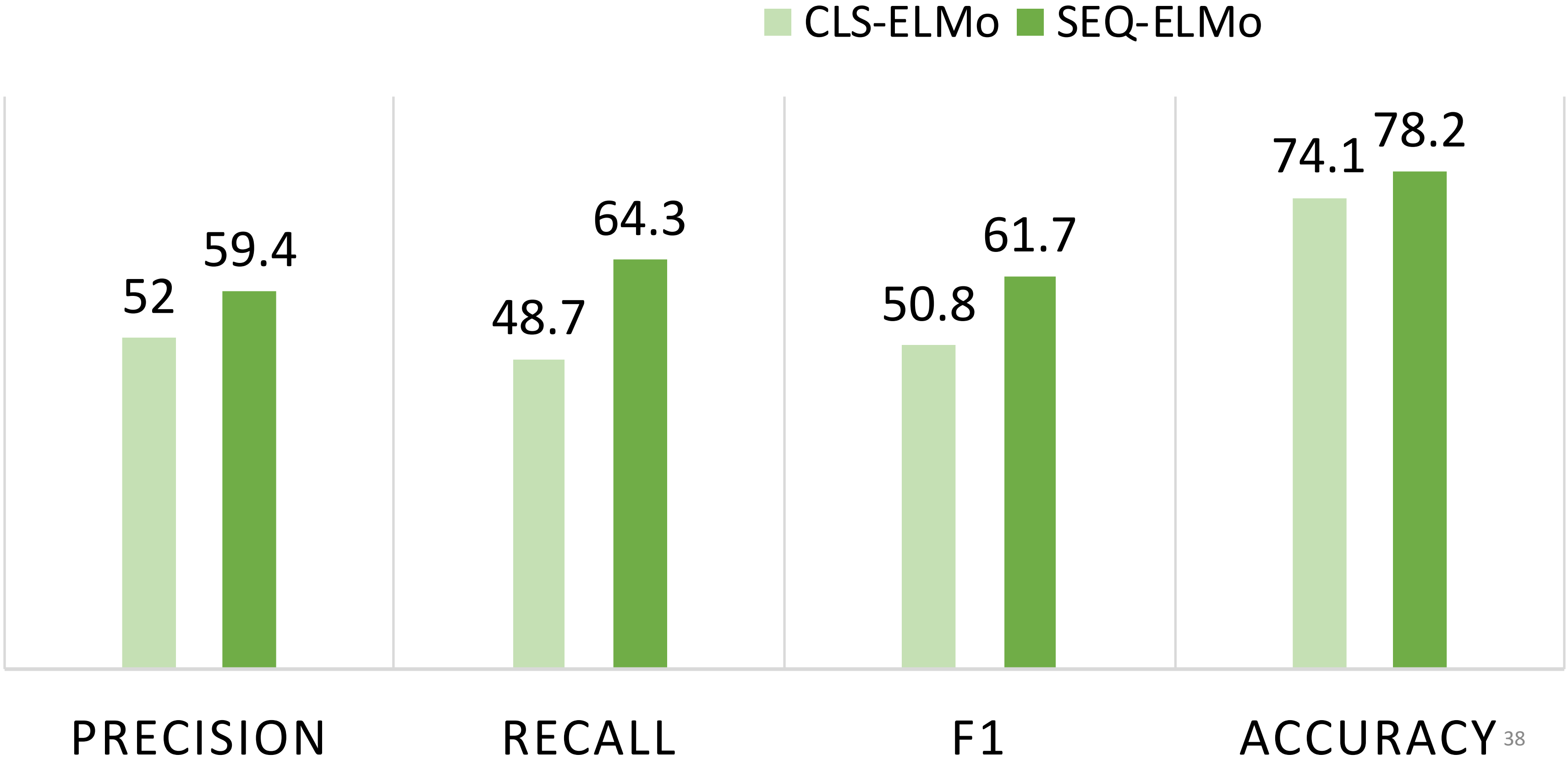


Ablation Study: Verb Classification (VUA dev)

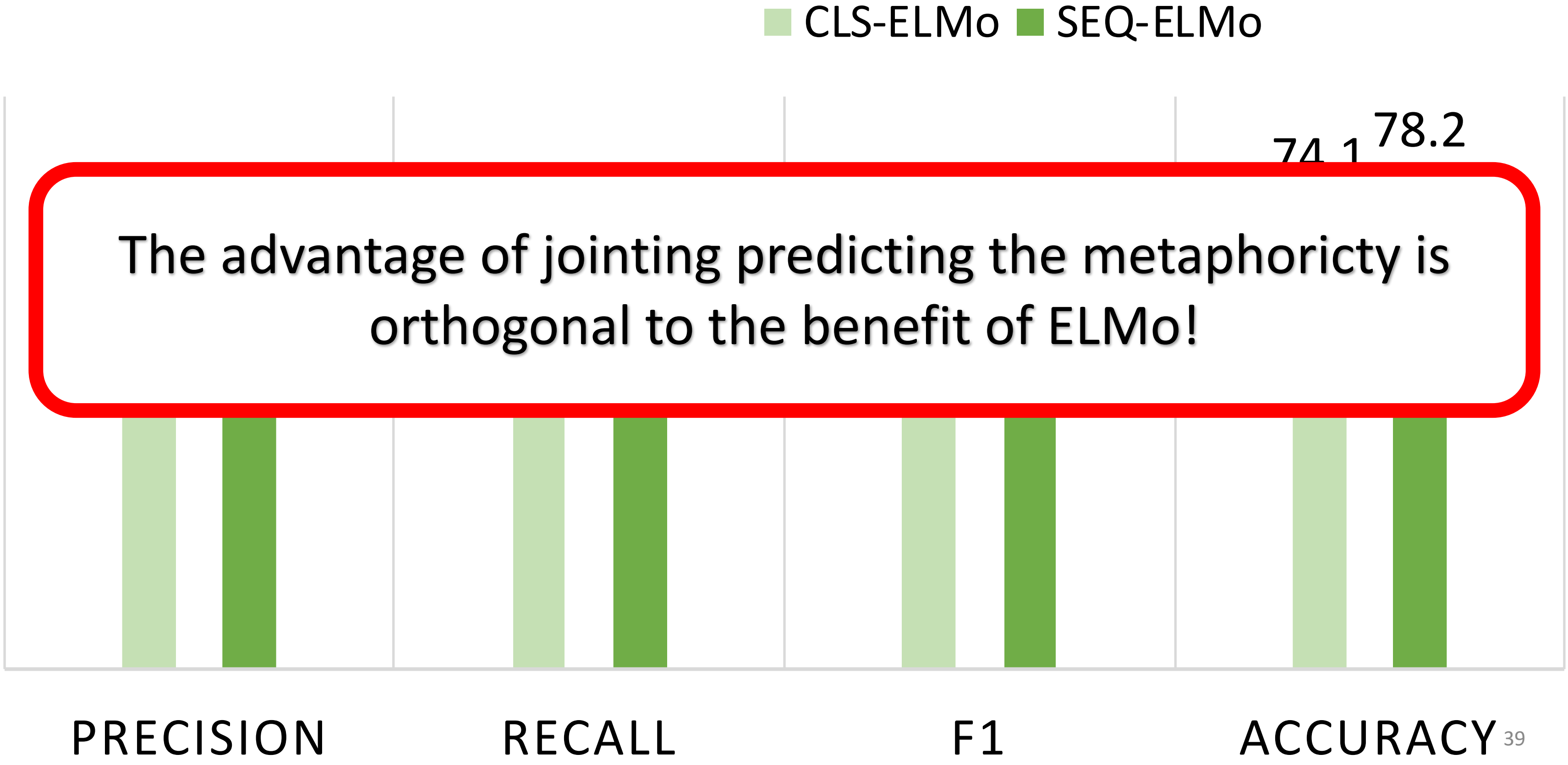


Contextualized word embeddings boost the performance!

Ablation Study: Verb Classification (VUA dev)



Ablation Study: Verb Classification (VUA dev)



The advantage of jointing predicting the metaphoricty is orthogonal to the benefit of ELMo!

Performance on Various POS tags

Performance on Various POS tags

- Verbs, nouns and adpositions are relatively easier to classify
  - Higher percentage of metaphors
  - More data

Performance on Various POS tags

- Verbs, nouns and adpositions are relatively easier to classify
  - Higher percentage of metaphors
  - More data
- Metaphorical particles are challenging to identify
  - Multiple meanings
  - Often associated with multi-word expressions

Performance on Various POS tags

- Verbs, nouns and adpositions are relatively easier to classify
  - Higher percentage of metaphors
  - More data

- Metaphorical particles are challenging to identify
  - Multiple meanings
  - Often associated with multi-word expressions

✓

  - Examples: put down the disturbance

Summary

- Jointly predicting the metaphoricity of all words in a sentence can enhance the classification performance.

- Standard BiLSTM models augmented with contextualized word representation perform strongly across various datasets.

Thank you!

Questions?