Language Model Priming
for
Cross-Lingual Event Extraction

Steven Fincke, Shantanu Agarwal, Scott Miller, Elizabeth Boschee
AAAI 2021

Contextualized Representations

- **Incorporating a word's context into its representation is crucial**
  - *"They protested his arrest"*
  - *"She is a cardiac arrest survivor"*

- **Even that is not enough**
  - *"Activists protested his arrest"*
  - Activists representation remains the same for different events.

Approaches

- **Prompting**
  - Reformulate the problem, e.g. QA
    - *"Which is the agent in arrested?"*
  - Not suitable for cross-lingual setting

- **Priming**
  - Augment the input to the encoder
  - Provide additional task-specific information

IE-Baseline

- **Two components**
  - Trigger extraction
  - Argument extraction
  - Architecture
    - Sequence-labeling (BIO)
    - Classification layer
    - CRF layer
    - Average for multiple wordpieces

IE-Baseline: Argument extraction

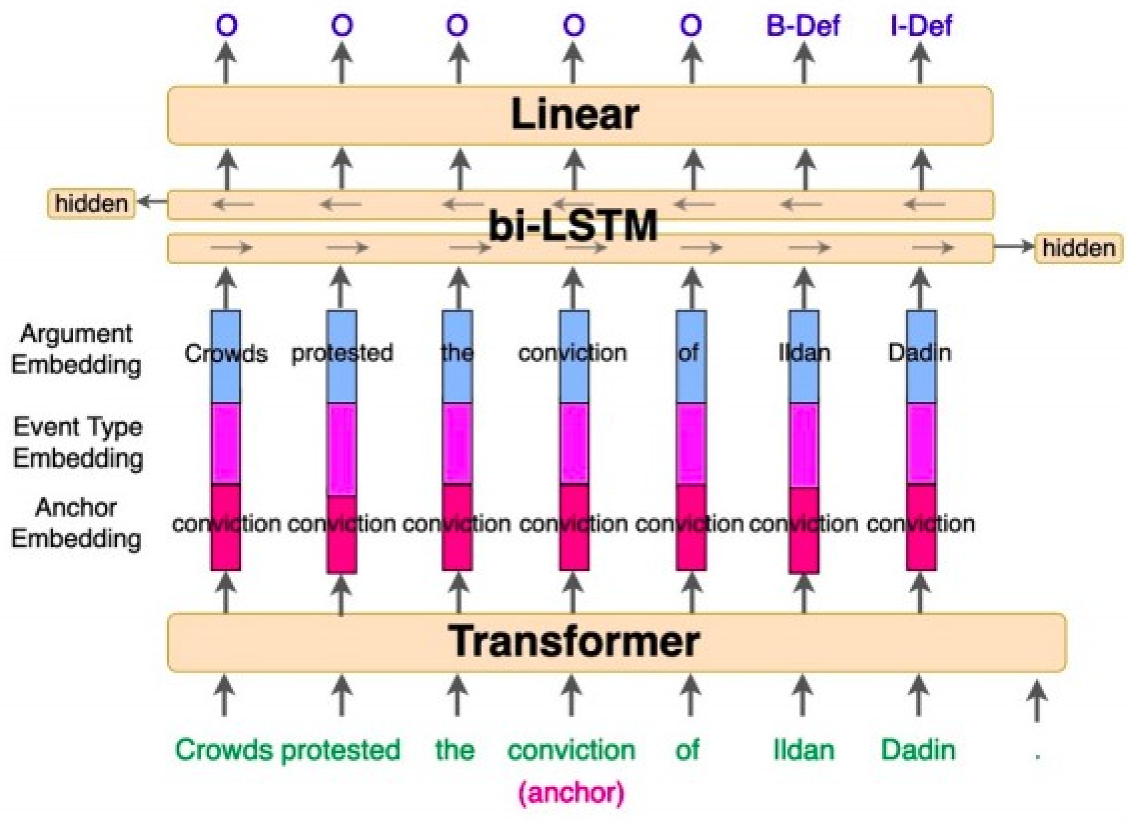

Figure 1: Baseline argument attachment architecture.

IE-PRIME: Argument extraction

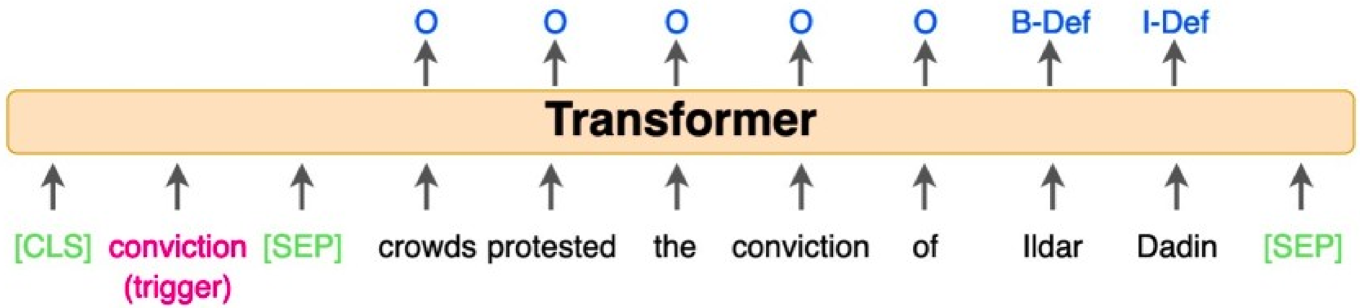

Figure 2: Priming a sentence for the trigger *conviction*. The span *Ildar Dadin* is identified as a DEFENDANT argument.

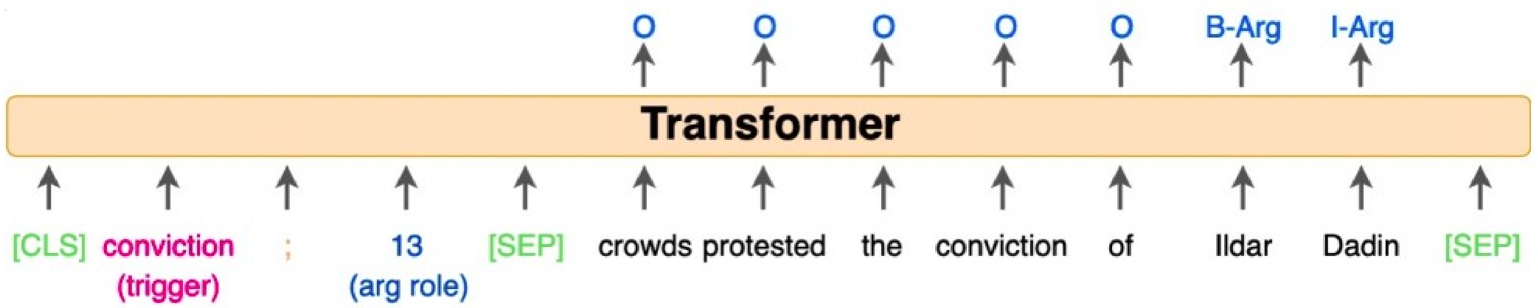

Figure 3: Priming a sentence for the trigger *conviction* and the argument role DEFENDANT. The span *Ildar Dadin* is identified as an argument and is therefore assigned the role being queried (DEFENDANT).

IE-PRIME: Event Extraction

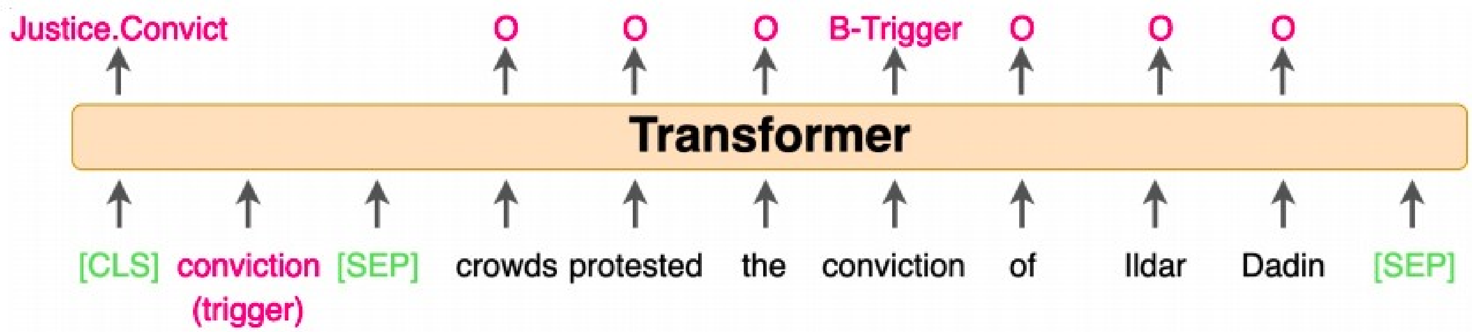

Figure 4: Priming a sentence to determine whether the span *conviction* is a trigger of type JUSTICE.CONVICT.

Experiment settings

- **Dataset**
  - ACE05
  - English and Arabic
- **Encoders**
  - Monolingual
    - Large, cased BERT
  - Cross-lingual
    - Large XLM-RoBERTa

Results

	Mono		Cross
	en→en	ar→ar	en→ar
Du and Cardie (2020)	65.4	–	–
Lin et al. (2020)	69.3	–	–
IE-BASELINE	63.2	53.3	44.7
IE-PRIME	**72.4**	**67.7**	**50.3**

Table 1: Gains in argument classification F1 score from priming for argument extraction (using gold triggers in our primary experimental setting). As for all reported results in this paper, IE-PRIME uses the trigger+role configuration for priming unless otherwise specified.

Results

	Priming method	Recall	Precision	F-Measure
Subburathinam et al. (2019)	–			61.8
Ahmad, Peng, and Chang (2021)	–			68.5
IE-BASELINE	–	67.1	79.6	72.8
IE-PRIME	trigger + role	66.5	82.9	73.8
IE-PRIME	trigger	**67.5**	**83.7**	**74.7**

Table 2: Secondary experimental setting: Argument classification F1 in the zero-shot cross-lingual condition (train on English, test on Arabic) with gold triggers and gold entity mentions, following splits from (Subburathinam et al. 2019).

	Monolingal				Cross-lingual	
	en→en		ar→ar		en→ar	
	trigger	argument	trigger	argument	trigger	argument
Wadden et al. (2019)	69.7	48.8	–	–	–	–
Wadden et al. (2019)†	70.4	52.2	**61.5**	44.4	41.6	22.0
Du and Cardie (2020)	72.4	53.1	–	–	–	–
Lin et al. (2020)	**74.7**	**56.8**	–	–	–	–
IE-PRIME (arguments only)	71.2	55.3	61.2	**48.9**	42.4	30.2
IE-PRIME (arguments + triggers)	68.1	52.9	60.2	48.7	**51.0**	**32.4**

Table 3: Trigger and argument classification F1 for end-to-end systems in our primary experimental setting. The first version of IE-PRIME includes the baseline trigger component and the primed argument extraction component. The second version includes both the primed trigger component and the primed argument extraction component. † indicates our local re-run of (Wadden et al. 2019).

Encoder Size Impact

	en→en		ar→ar		en→ar	
	base	large	base	large	base	large
IE-BASELINE	60.0	66.0	46.9	53.3	35.6	44.7
IE-PRIME	69.0	**74.1**	60.8	**67.7**	40.2	**50.3**

Table 4: Comparison of argument classification F1 (using gold triggers) based on size of pretrained language model.

Training Set Size Impact

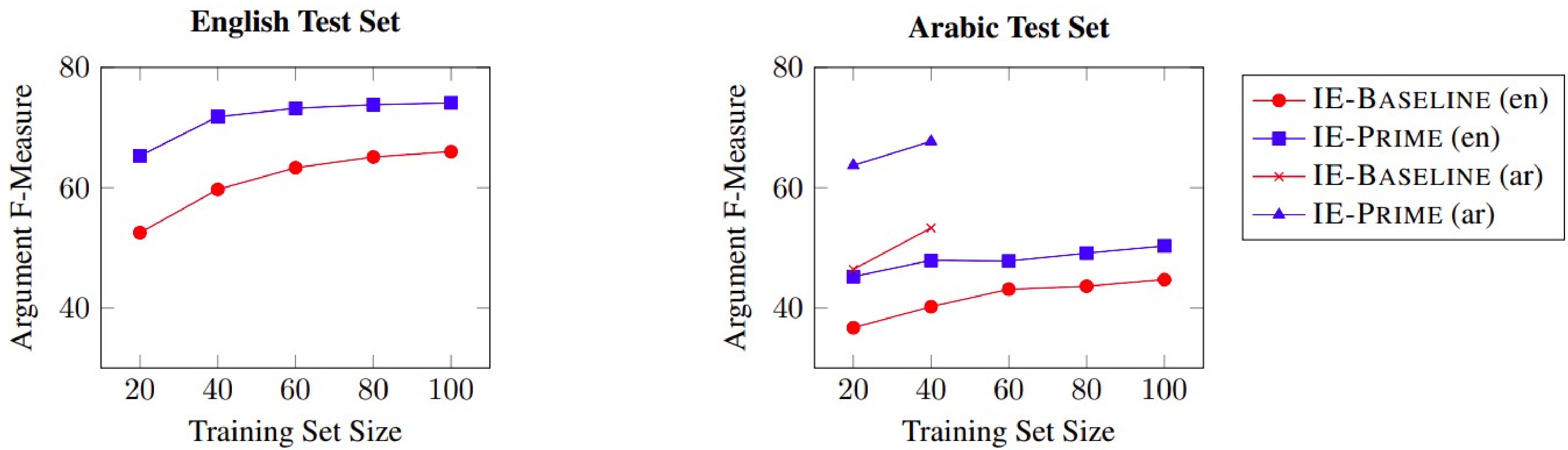

Figure 7: Comparison of IE-BASELINE and IE-PRIME by approximate training set size. Training size here is calculated as the number of events in a document set and is shown as a percentage of full English training set size. The language of the data in which the models were trained are denoted in parenthesis. Experiments in this figure use gold triggers.