A neural joint model for Vietnamese word segmentation, POS tagging and dependency parsing

Dat Quoc Nguyen

04 December 2019

Motivation

- When written, white space is used in Vietnamese to separate syllables that constitute words
- 85% of Vietnamese word types are composed of at least two syllables
- Parsing real-world Vietnamese text where gold word segmentation is not available

  Pipeline: Vietnamese word segmenters and POS taggers have a non-trivial error rate, thus leading to error propagation

  A new multi-task learning model for joint word segmentation, POS tagging and dependency parsing

Tôi là sinh viên (I am student)

↓ Word segmenter

Tôi_{I} là_{am} $\text{sinh\_viên}_{\text{student}}$

↓ POS tagger

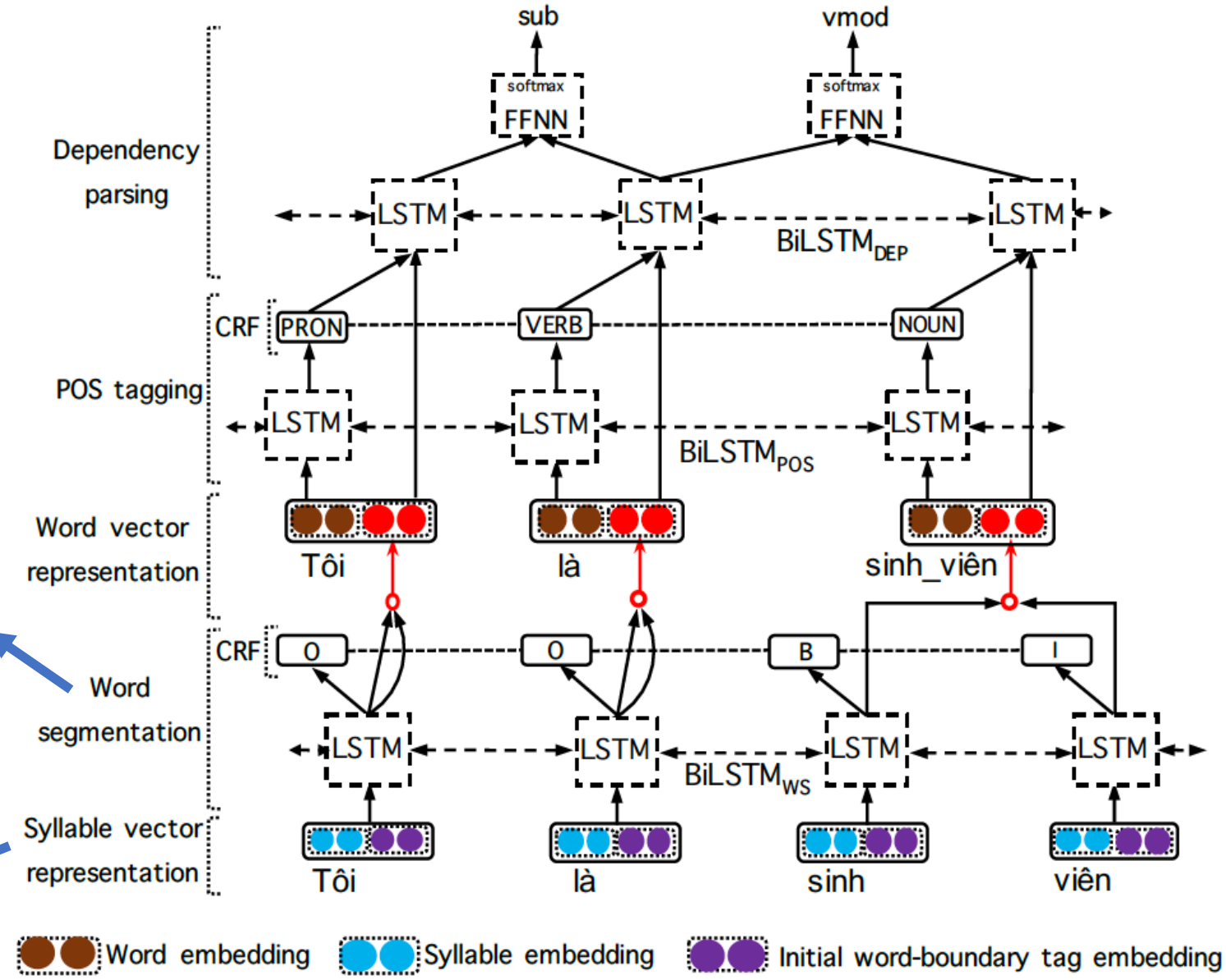
Tôi/PRON là/VERB sinh_viên/NOUN

↓ Dependency parser

ID	Form	POS	Head	DepRel
1	Tôi_{I}	PRON	2	sub
2	là_{am}	VERB	0	root
3	$\text{sinh\_viên}_{\text{student}}$	NOUN	2	vmod

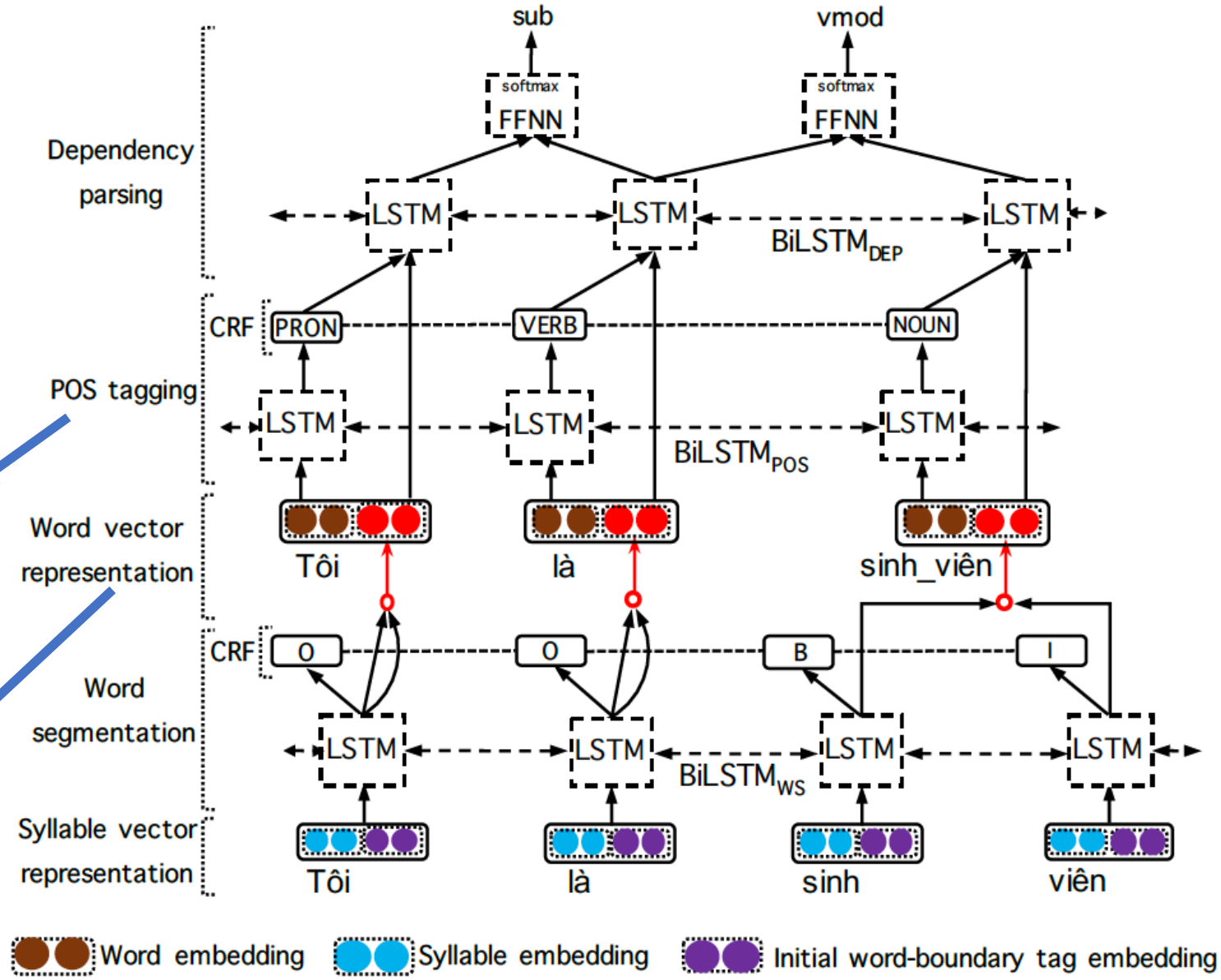
Our model

- A hierarchical structure of three main components: word segmentation, POS tagging and dependency parsing
  - *The word segmentation component uses a BiLSTM-CRF architecture to predict BIO word boundary tags from input syllables*, resulting in a word-segmented sequence
  - We apply a lexicon-based longest matching strategy to produce initial BIO word-boundary tags



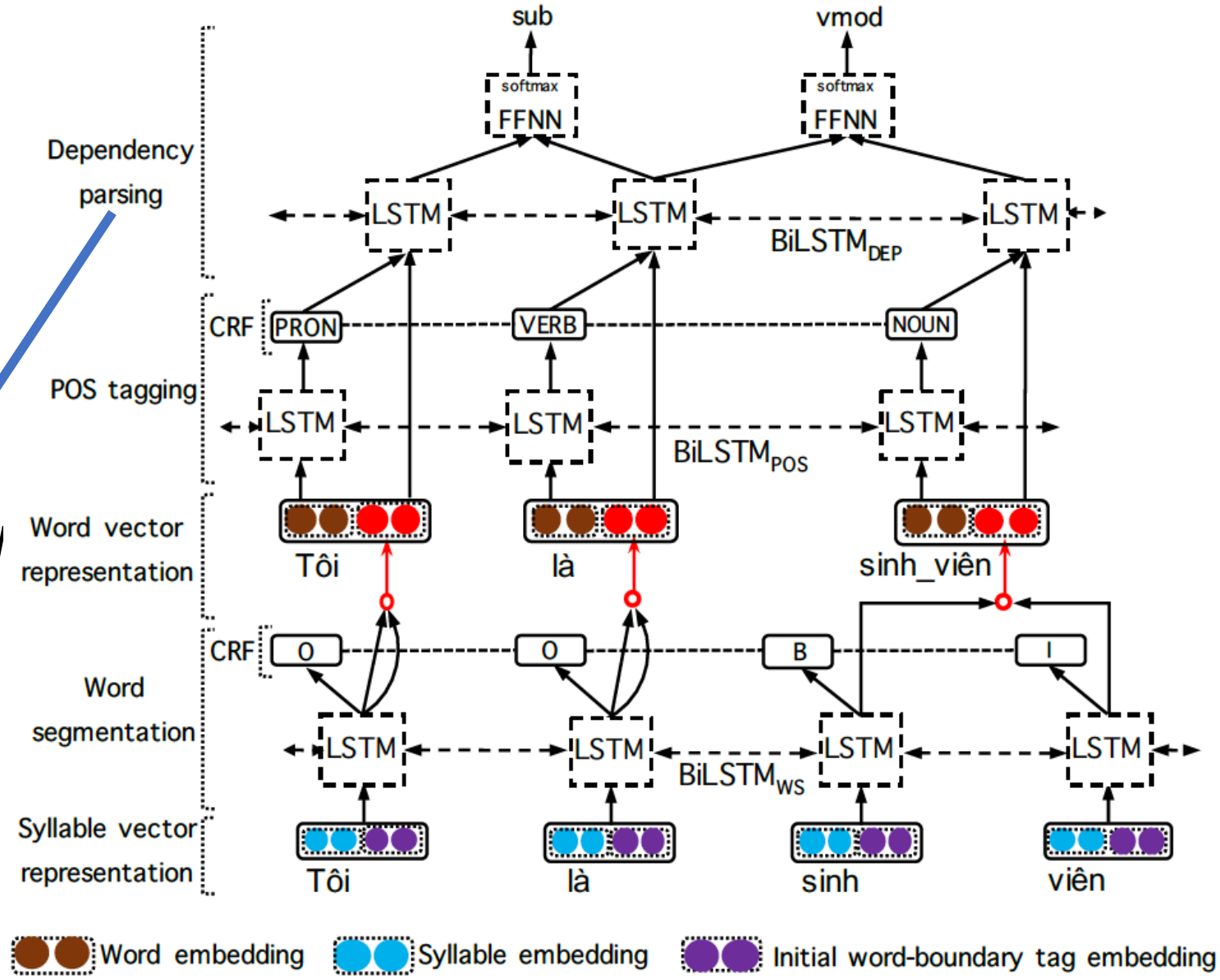
Our model

- A hierarchical structure of three main components: word segmentation, POS tagging and dependency parsing
  - *The POS tagging component also uses a BiLSTM-CRF to predict POS tags from the sequence of segmented words*
  - We create a vector to represent each word by concatenating its word embedding and its syllable-level word embedding



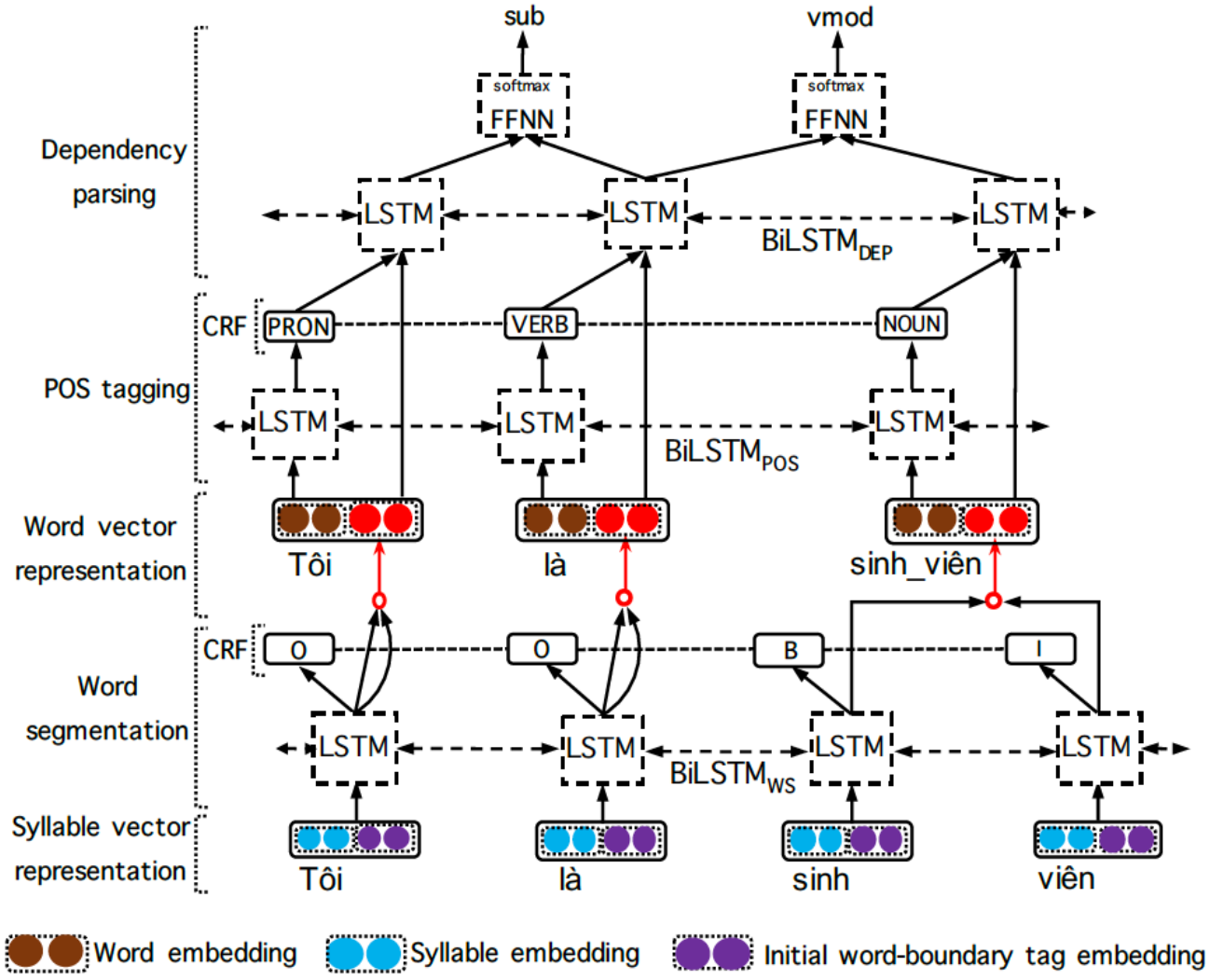
Our model

- A hierarchical structure of three main components: word segmentation, POS tagging and dependency parsing
  - *The dependency parsing component uses a graph-based architecture similarly to the one from Kiperwasser and Goldberg (2016) to decode dependency arcs and labels*
  - Nguyen et al. (2016) show that graph-based parsers perform better than transition-based parsers for Vietnamese



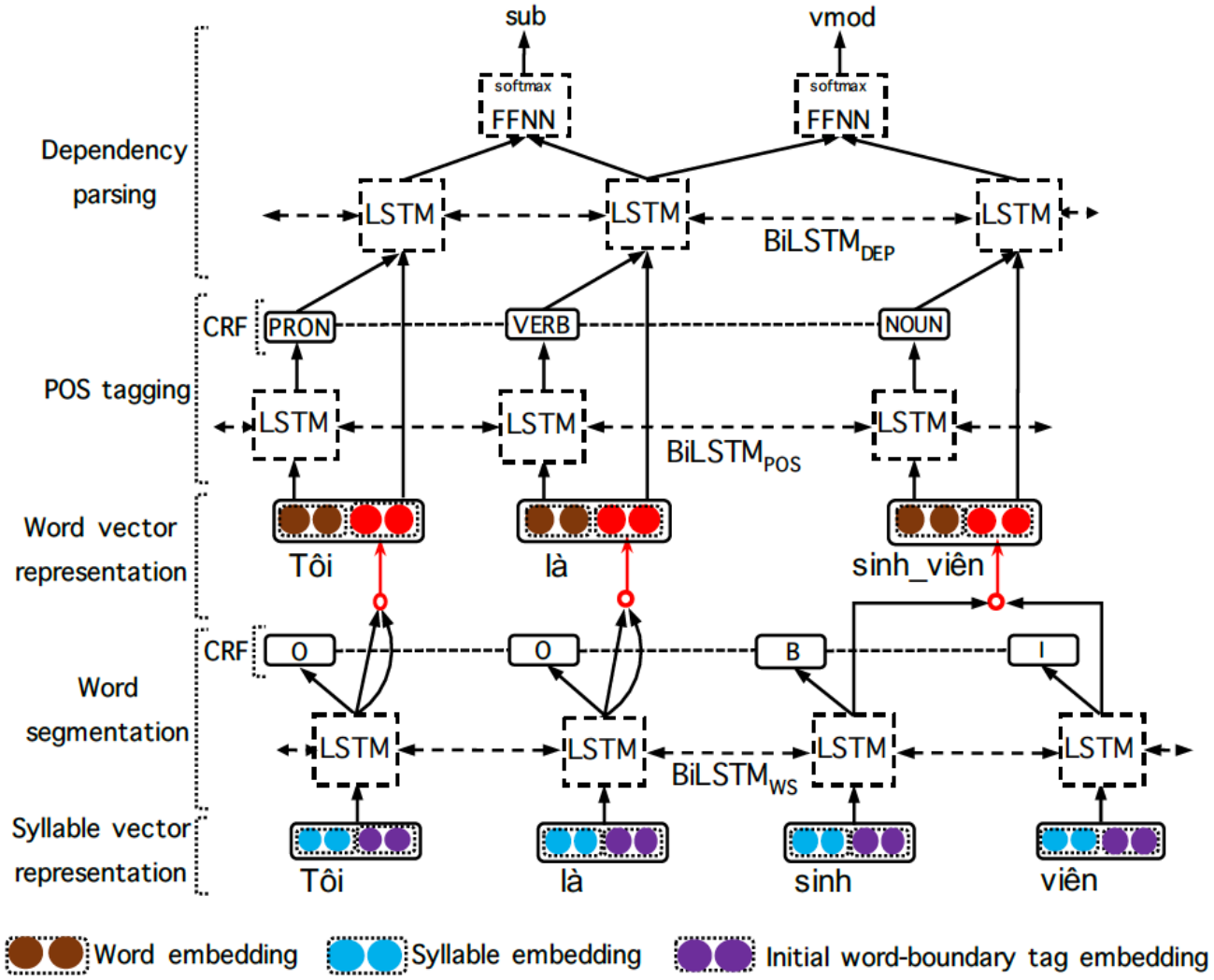
Our model

- Inspired by stack propagation based methods (Zhang and Weiss, 2016; Hashimoto et al., 2017) which are joint models for POS tagging and dependency parsing. For dependency parsing:
  - The Stack-propagation model (Zhang and Weiss, 2016) uses a transition-based approach
  - The joint multi-task model (Hashimoto et al., 2017) uses a head selection based approach (Zhang et al., 2017) which produces a probability distribution over possible heads for each word



Our model

- Compared to existing joint models for Chinese: they all use transition-based approaches
- Viewed as an extension of the joint POS tagging and dependency parsing model jPTDP-v2 (Nguyen and Verspoor, 2018), where we incorporate a BiLSTM-CRF for word boundary prediction
- See our paper for details of other improvements over jPTDP-v2



Test set results

- We follow the setup used in the Vietnamese NLP toolkit VnCoreNLP (Vu et al., 2018)

Model	WSeg	PTag	LAS	UAS
Our jointWPD	97.81	94.05	71.50	77.23
VnCoreNLP	97.90	94.06	68.84**	74.52**
jPTDP-v2	97.90	93.82*	70.78**	76.80*
Biaffine	97.90	94.06	72.59**	78.54**

- jointWPD obtains slightly lower word segmentation score than VnCoreNLP. A possible reason is because we used a bit smaller set of training data:

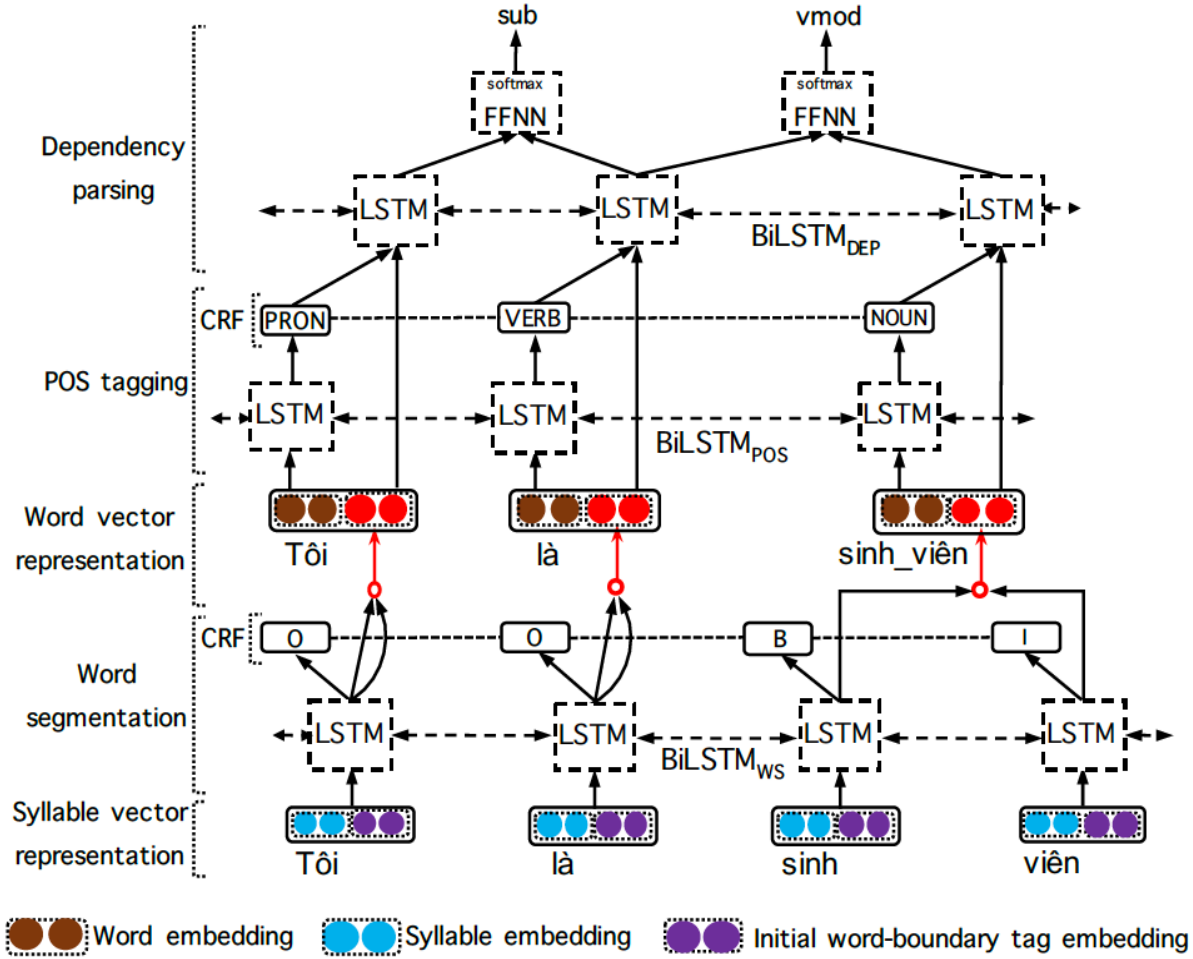
  VnCoreNLP used a set of 75K sentences for training, while we split this set into training (70K) and development (5K)

- jointWPD obtains 1.1+% lower LAS and UAS than Biaffine which uses a “biaffine” attention mechanism for predicting dependency arcs and labels

  Extend the parsing component with the biaffine attention mechanism in future work

Development set results

Model	WSeg	PTag	LAS	UAS
WS $\mapsto$ Pos $\mapsto$ Dep	98.48*	95.09*	70.68*	76.70*
Our jointWPD	98.66	95.35	71.13	77.01
(a) w/o $\text{Initial}_{\text{BIO}}$	98.25**	95.01*	70.34**	76.36**
(b) w/o CRF_{WSeg}	98.32**	95.06*	70.48**	76.47**
(c) w/o CRF_{PTag}	98.65	95.14*	71.00	76.94
(d) w/o PTag	98.63	95.10*	69.78**	76.03**



- WS → Pos → Dep: A pipeline approach
- (a): Without using initial word-boundary tag embedding
- (b): Using a softmax layer for word-boundary tag prediction instead of a CRF layer
- (c): Using a softmax layer for POS tag prediction instead of a CRF layer
- (d): Without using the POS tag embeddings for the parsing component

Conclusions

- We have presented the first multi-task learning model for joint word segmentation, POS tagging and dependency parsing in Vietnamese
- Experiments on Vietnamese benchmark datasets show that our joint multi-task model obtains results competitive with the state-of-the-art
- We will evaluate effects of the contextualized representations to our joint model
- A Vietnamese syllable is analogous to a character in other languages such as Chinese and Japanese
    - We will also evaluate the application of our model to those languages in future work

Acknowledgements: I would like to thank Karin Verspoor and Vu Cong Duy Hoang as well as the anonymous reviewers for their feedback!

Thanks for your attention!