

Bottom-Up Constituency Parsing and Nested Named Entity Recognition with Pointer Networks



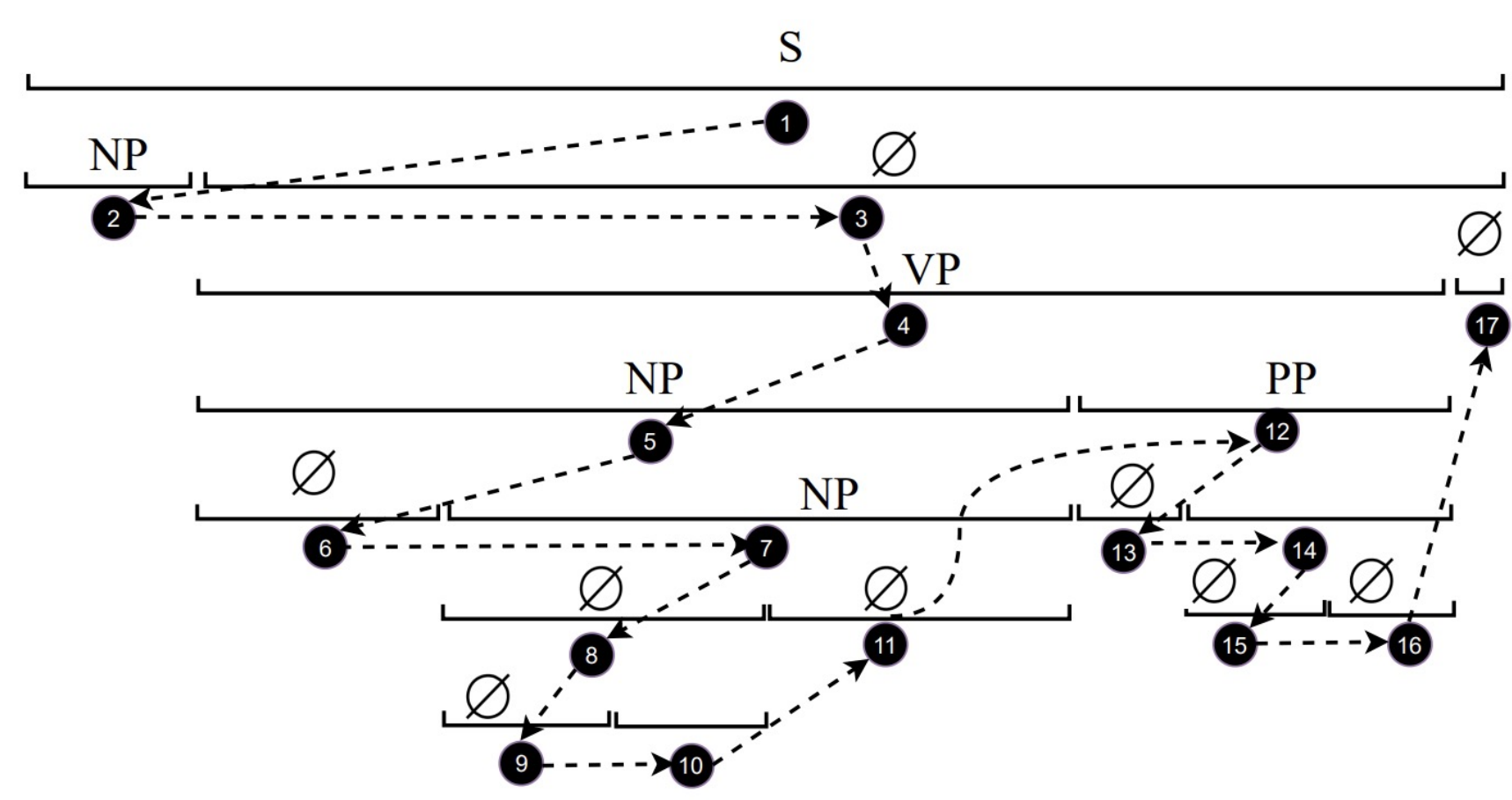
Songlin Yang, Kewei Tu

School of Information Science and Technology, ShanghaiTech University

Motivation

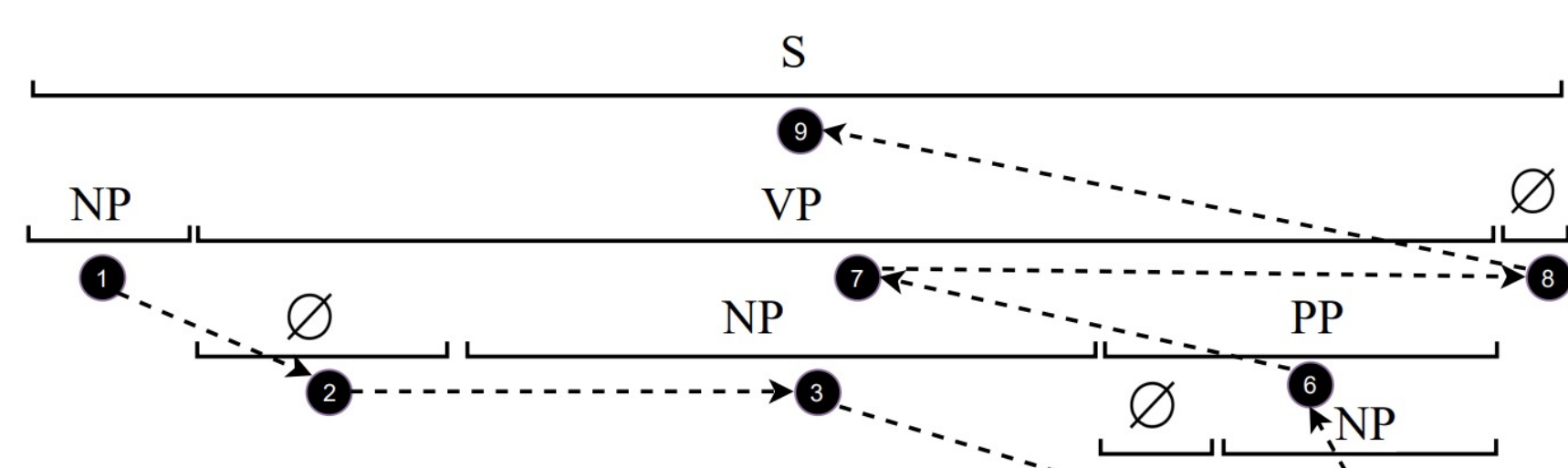
Constituency parsing and nested NER are similar tasks. They both aim to predict a collection of non-crossing and nested spans. If two spans overlap, one must be the subspan of the other. We propose a single model to tackle both tasks, casting nested NER to constituency parsing.

Seq2seq span generation



Mary acquired those three businesses in May 1988 .

(a)



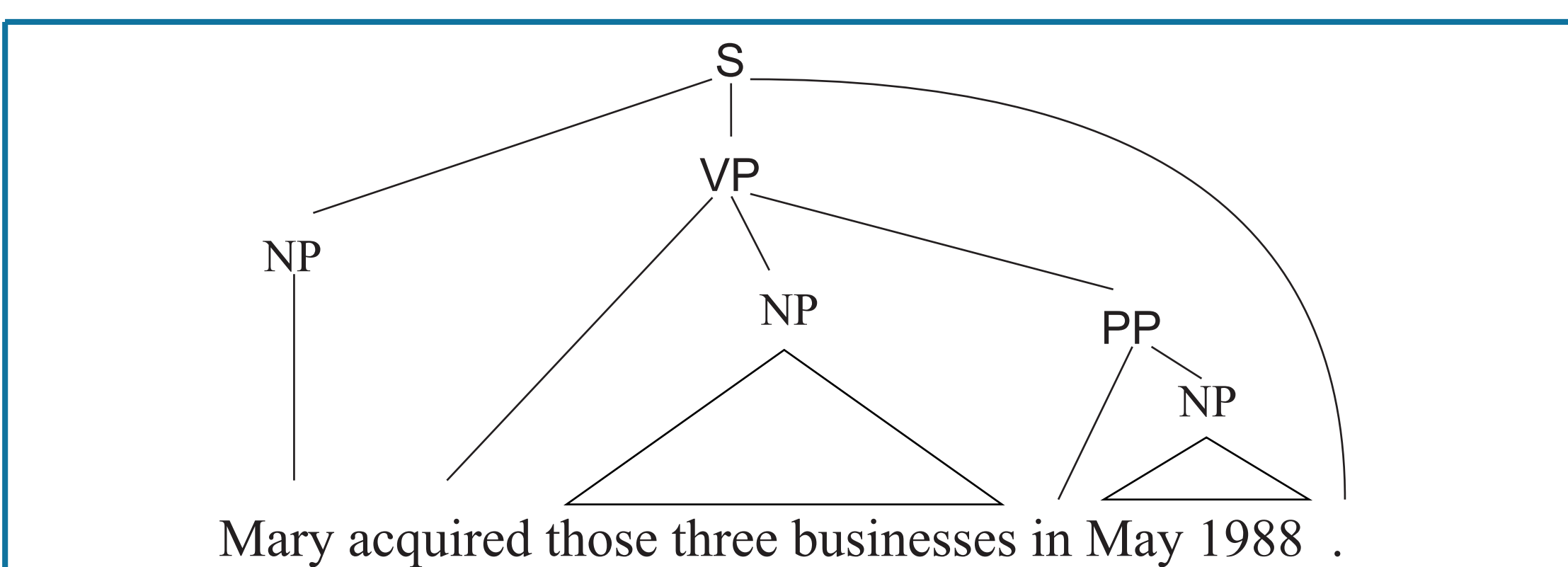
Mary acquired those three businesses in May 1988 .

(b)

(a) Nguyen et al. (2020) cast constituency parsing to top-down splitting, propose a seq2seq model with pointer networks to generate spans in pre-order.

(b) We propose to generate spans in post-order to leverage subspan information and to produce n-ary constituency trees directly.

Post-order pointing representation



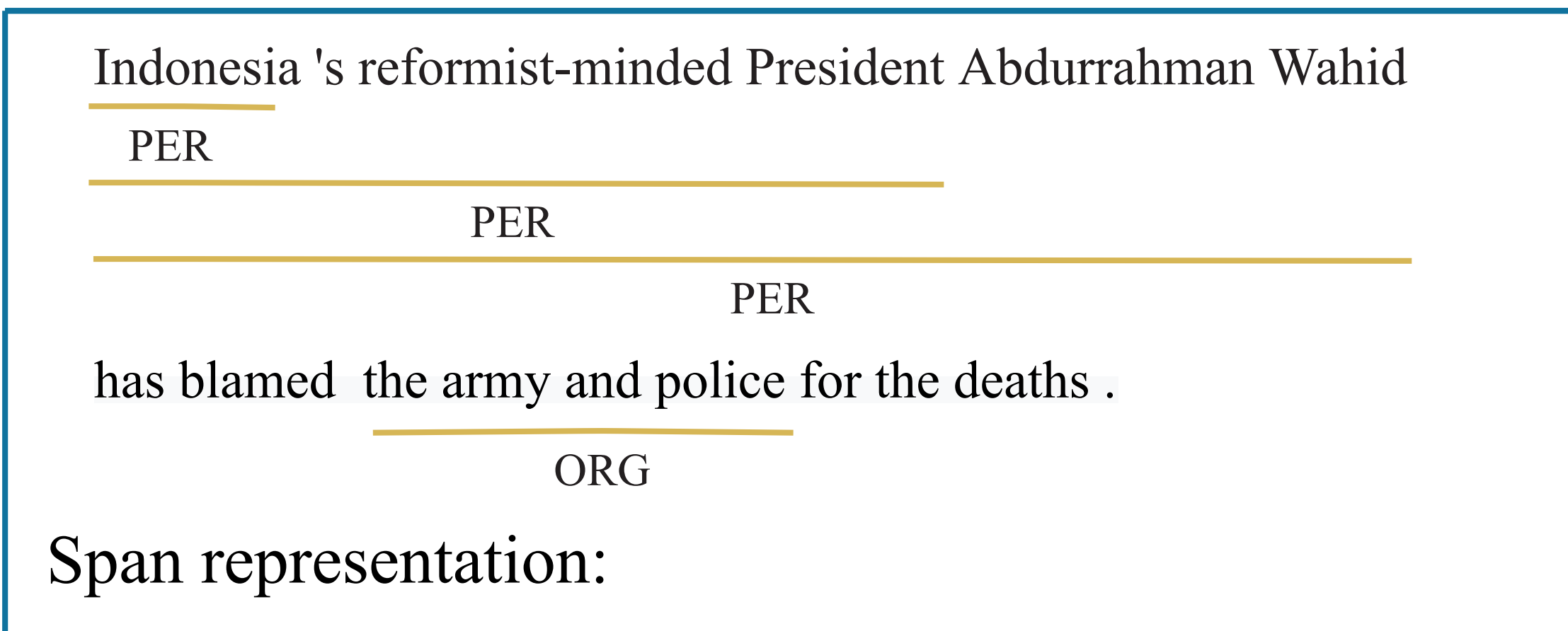
Span representation:

{ (0, 1, NP), (2, 5, NP), (6, 8, NP), (5, 8, PP), (1, 8, VP), (0, 9, S) }

Pointing representation:

{ (0->1, NP), (1->2, ∅), (2->5, NP), (5->6, ∅), (6->8, NP), (8->5, PP), (8->1, VP), (8->9, ∅), (9->0, S) }

(a)



Span representation:

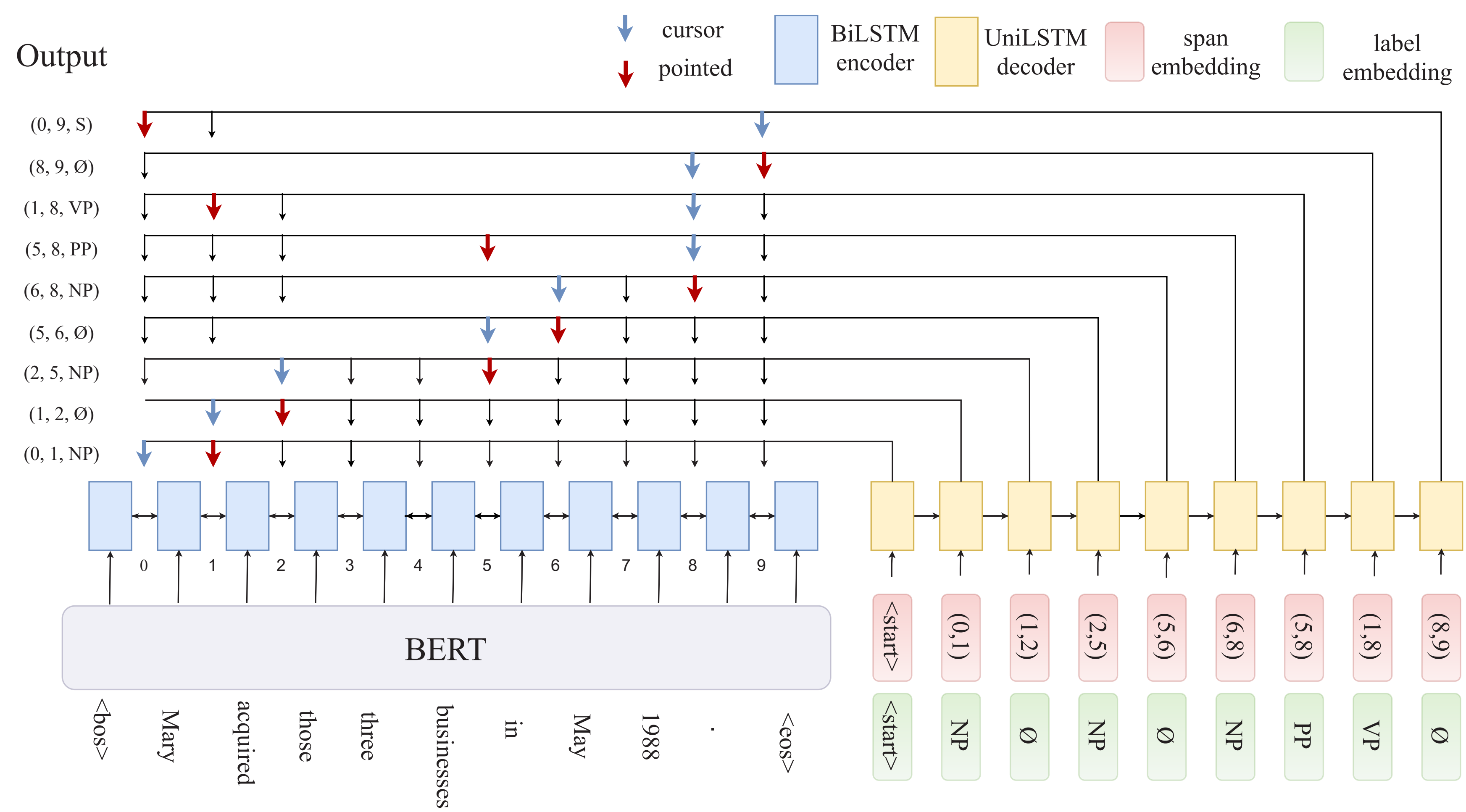
{ (0, 1, PER), (0, 4, PER), (0, 6, PER), (8, 12, ORG) }

Pointing representation:

{ (0->1, PER), (1->4, ∅), (4->0, PER), (4->6, ∅), (6->0, PER), (6->8, ∅), (8->12, ORG), (12->16, ∅), (16->0, ∅) }

(b)

Seq2seq architecture



Experiments

| Model | P | R | F |
|--------------------------|--------------|--------------|--------------|
| Kitaev et al. (2019) [S] | 95.46 | 95.73 | 95.59 |
| Zhou and Zhao (2019) [S] | 95.70 | 95.98 | 95.84 |
| Zhang et al. (2020) [S] | 95.85 | 95.53 | 95.69 |
| Yang and Deng (2020) [T] | 96.04 | 95.55 | 95.79 |
| Nguyen et al. (2020) [S] | - | - | 95.48 |
| Wei et al. (2020) [S] | 95.5 | 96.1 | 95.8 |
| Tian et al. (2020) [S] | 96.09 | 95.62 | 95.86 |
| Xin et al. (2020) [S] | 96.29 | 95.55 | 95.92 |
| Nguyen et al. (2021) [Q] | - | - | 95.7 |
| Cui et al. (2021) [S] | 95.70 | 96.14 | 95.92 |
| Ours [Q] | 96.19 | 95.83 | 96.01 |

Table 1. Results on PTB. All models use BERT as encoders. S: span-based methods. T: transition-based methods. Q: seq2seq-based methods.

| Model | ACE2004 | | | ACE2005 | | | GENIA | | |
|-------------------------|---------|-------|-------|--------------|--------------|--------------|--------------|--------------|--------------|
| | P | R | F | P | R | F | P | R | F |
| Shibuya and Hovy (2020) | 84.71 | 83.96 | 84.33 | 82.58 | 84.29 | 83.42 | 79.92 | 76.55 | 78.20 |
| Wang et al. (2020) | 86.08 | 86.48 | 86.26 | 83.95 | 85.39 | 84.66 | 79.45 | 78.94 | 79.19 |
| Wang et al. (2021) | 86.27 | 85.09 | 85.68 | 85.28 | 84.15 | 84.71 | 79.20 | 78.16 | 78.67 |
| Fu et al. (2021) | 86.7 | 86.5 | 86.6 | 84.5 | 86.4 | 85.4 | 78.2 | 78.2 | 78.2 |
| Xu et al. (2021) | 86.9 | 85.8 | 86.3 | 85.7 | 85.2 | 85.4 | 80.3 | 78.9 | 79.6 |
| Yan et al. (2021) | 87.27 | 86.41 | 86.84 | 83.16 | 86.38 | 84.74 | 78.57 | 79.3 | 78.93 |
| Shen et al. (2021) | 87.44 | 87.38 | 87.41 | 86.09 | 87.27 | 86.67 | 80.19 | 80.89 | 80.54 |
| Tan et al. (2021) | 88.46 | 86.10 | 87.26 | 87.48 | 86.64 | 87.05 | 82.31 | 78.66 | 80.44 |
| Ours | 86.60 | 87.28 | 86.94 | 84.61 | 86.43 | 85.53 | 78.08 | 78.26 | 78.16 |

Table 2. Results on ACE2004, ACE2005 and GENIA. All models use BERT as encoders. Our method beats previous best parsing-based method (Fu et al. 2021).

Analysis

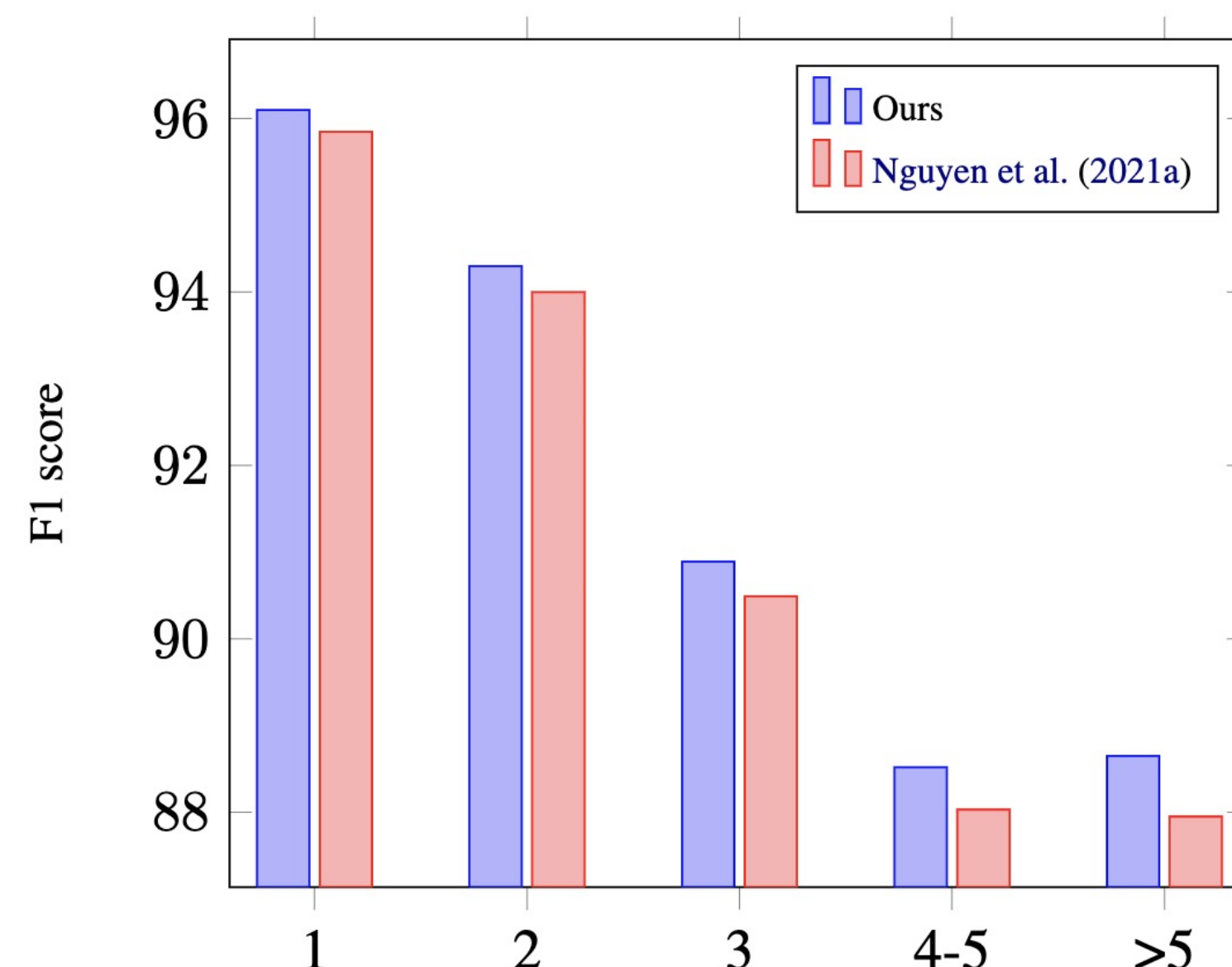


Figure 3. F1 scores on constituent nodes with different numbers of children on PTB test set.

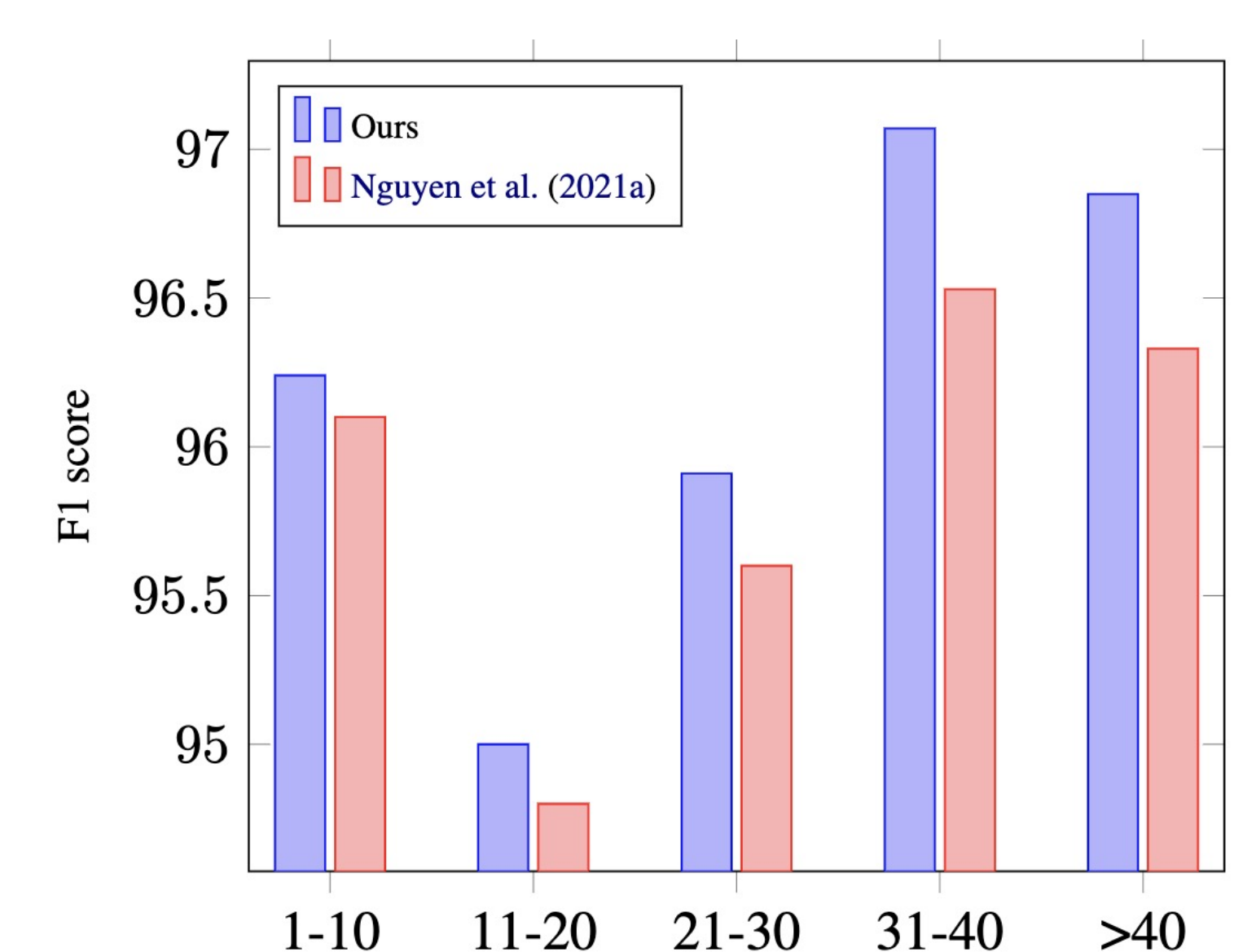


Figure 4. F1 scores against constituent span length on PTB test set.