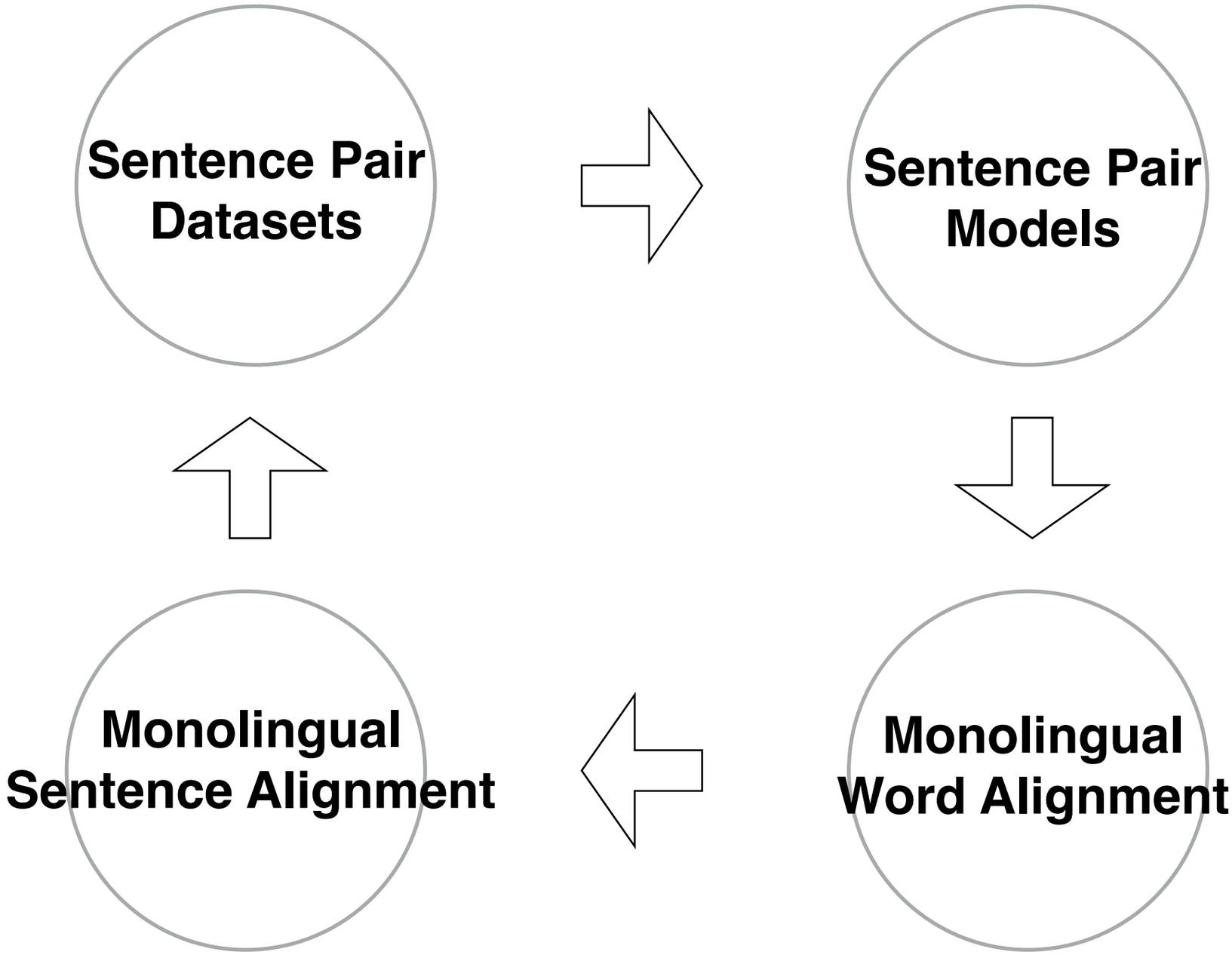
Sentence Pair Modeling in Natural Language Processing

Wuwei Lan

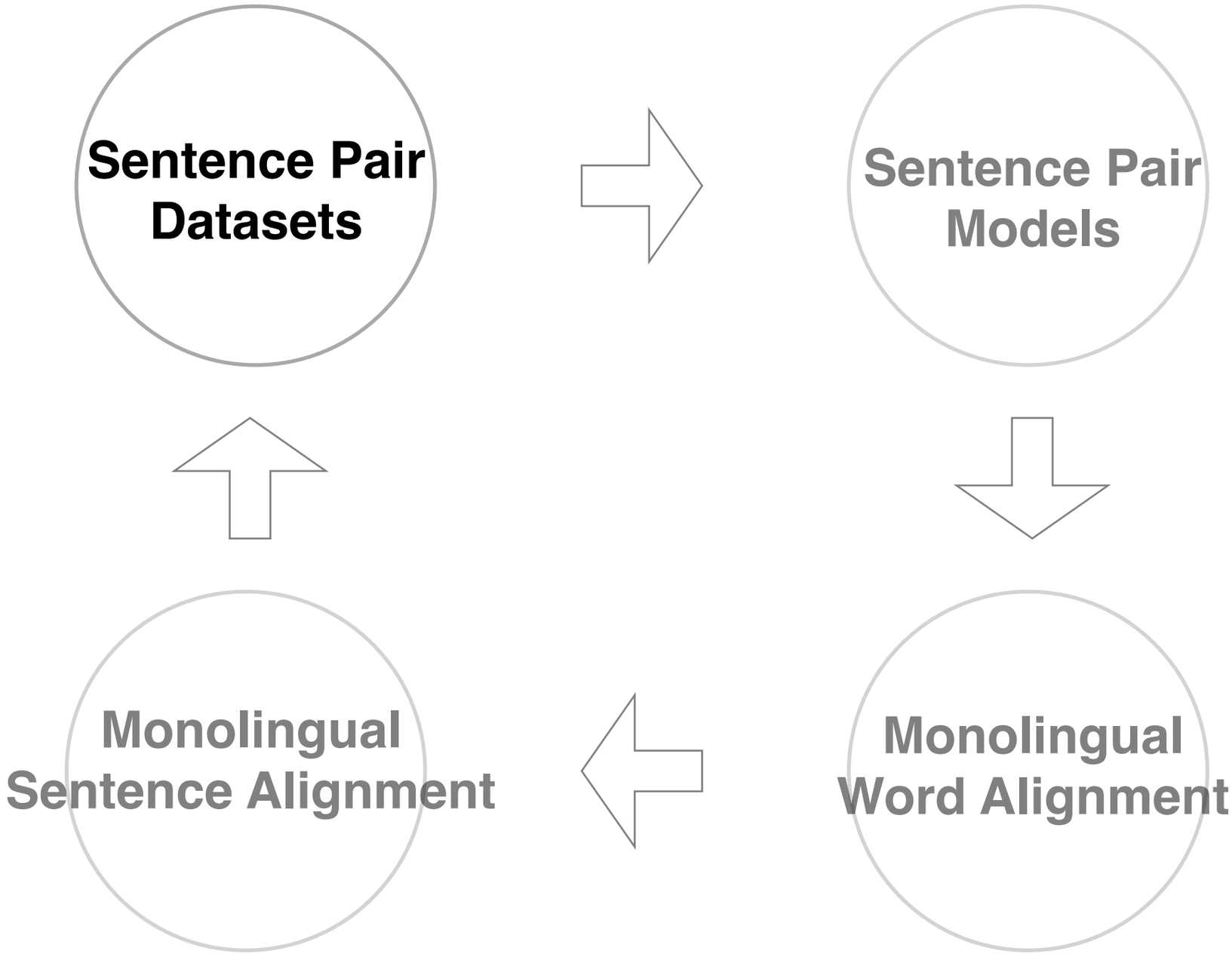
THE OHIO STATE UNIVERSITY

Department of Computer Science and Engineering

Outline


Sentence Pair Datasets
Sentence Pair Models
Monolingual Word Alignment
Monolingual Sentence Alignment

Outline


Sentence Pair Datasets
Sentence Pair Models
Monolingual Word Alignment
Monolingual Sentence Alignment

Sentence Pair Modeling Tasks

Paraphrase Identification

paraphrase | non-paraphrase

Dataset: Quora (400k), URL (51k), PIT (16k)

Semantic Textual Similarity

score[0,5]

Dataset: STS14 (11k), STS-B (8.6k)

Natural Language Inference

entailment | neutral | contradiction

Dataset: SNLI (570k), MNLI (432k)

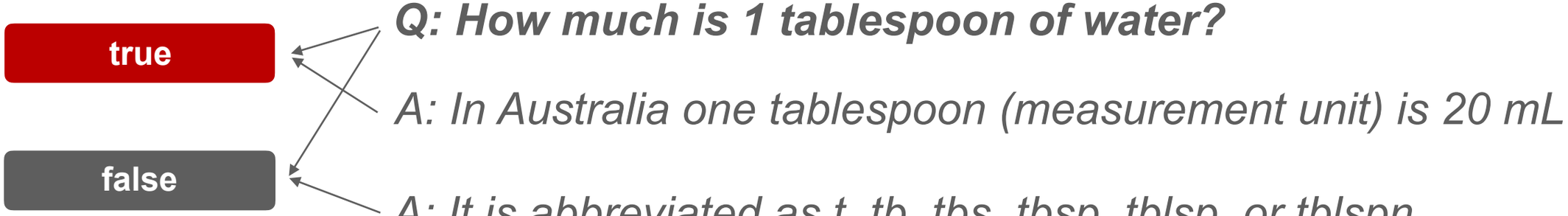
Question Answering

true | false

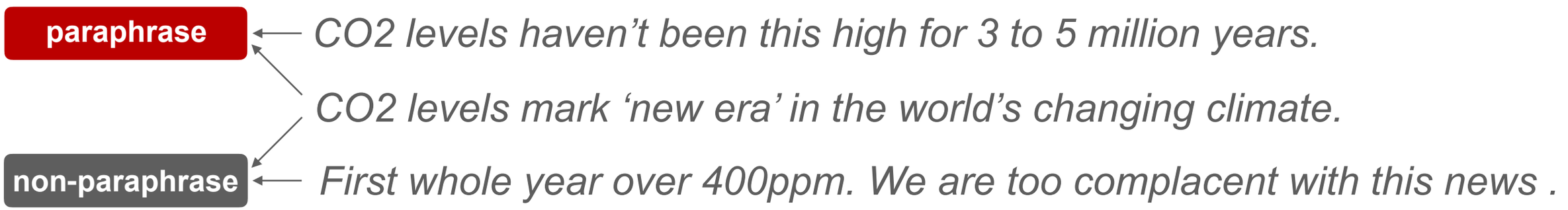
Dataset: WikiQA (12k), TrecQA (56k)

Examples

Question Answering [1]



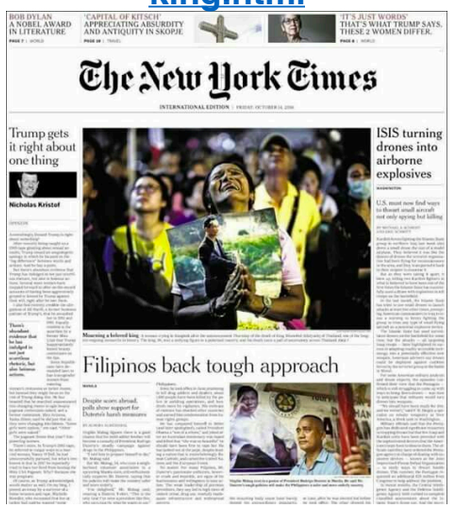
Paraphrase Identification [2]



[1] Yi Yang, Wen-tau Yih, and Christopher Meek. WikiQA: A challenge dataset for open-domain question answering. (EMNLP 2015).
[2] **Wuwei Lan**, Siyu Qiu, Hua He, and Wei Xu. A Continuously Growing Dataset of Sentential Paraphrases (EMNLP 2017).

Paraphrase Collection in Twitter[1]

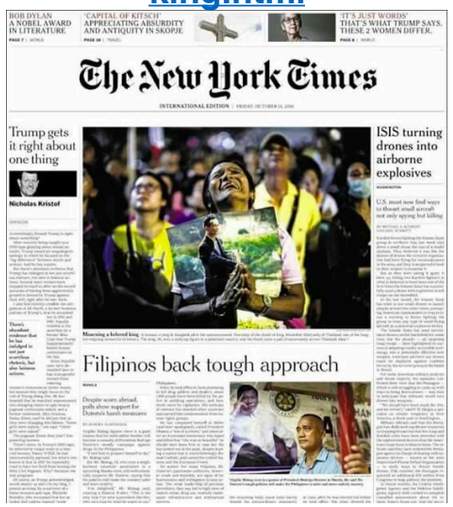
https://www.nytimes.com/2016/10/13/world/asia/thailand-king.html



[1] **Wuwei Lan**, Siyu Qiu, Hua He, and Wei Xu. A Continuously Growing Dataset of Sentential Paraphrases (EMNLP 2017).

Paraphrase Collection in Twitter[1]

https://www.nytimes.com/2016/10/13/world/asia/thailand-king.html


The New York Times

INTERNATIONAL EDITION

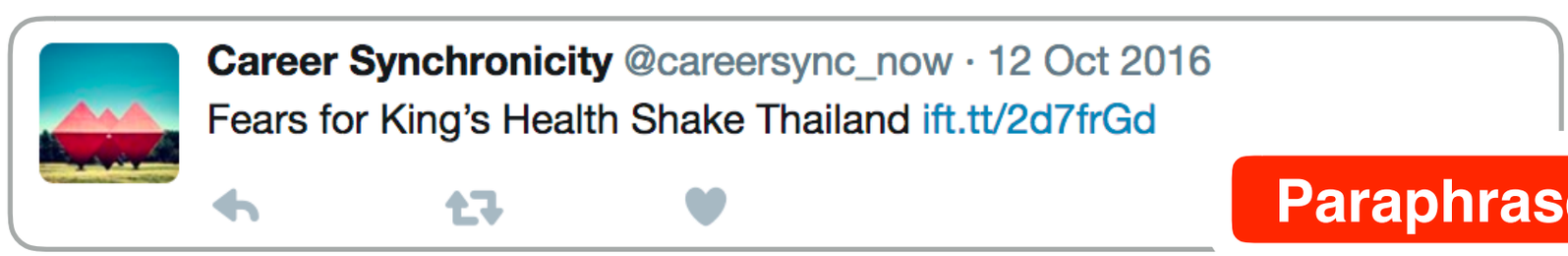
Trump gets it right about one thing

Nicholas Kristof

ISIS turning drones into airborne explosives

Filipinos back tough approach

Despite scorn abroad, polls show support for Duterte's harsh measures

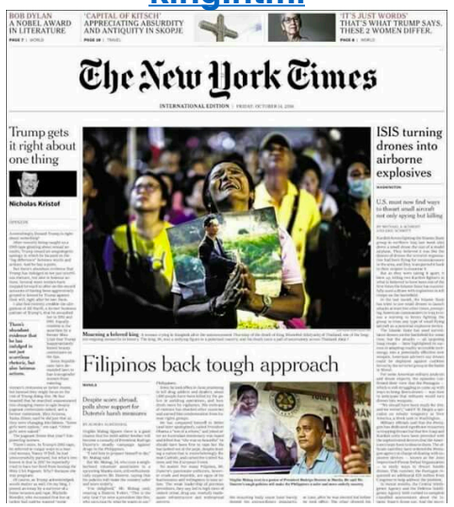


Paraphrase

[1] **Wuwei Lan**, Siyu Qiu, Hua He, and Wei Xu. A Continuously Growing Dataset of Sentential Paraphrases (EMNLP 2017).

Paraphrase Collection in Twitter[1]

https://www.nytimes.com/2016/10/13/world/asia/thailand-king.html



The New York Times @nytimes · 12 Oct 2016
Worries over the health of King Bhumibol Adulyadej are shaking Thailand
nyti.ms/2dRzPcr
5 261 144

Career Synchronicity @careersync_now · 12 Oct 2016
Fears for King's Health Shake Thailand ift.tt/2d7frGd

Paraphrase

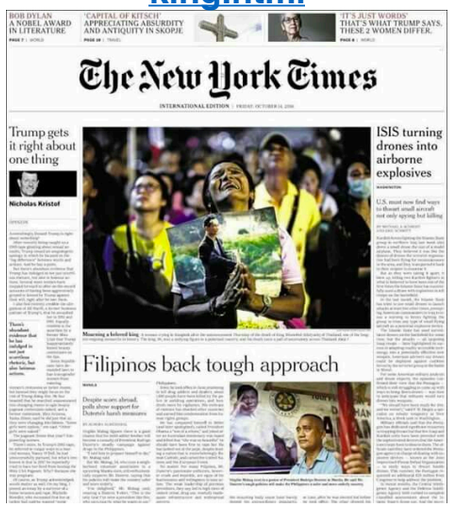
Herbert Buchsbaum @herbertnyt · 12 Oct 2016
New bulletin from Thai palace: King is still on a ventilator and in unstable condition. nyti.ms/2dW1A37

Non-Paraphrase

[1] **Wuwei Lan**, Siyu Qiu, Hua He, and Wei Xu. A Continuously Growing Dataset of Sentential Paraphrases (EMNLP 2017).

Paraphrase Collection in Twitter[1]

https://www.nytimes.com/2016/10/13/world/asia/thailand-king.html


The New York Times

Trump gets it right about one thing

ISIS turning drones into airborne explosives

Filipinos back tough approach

The New York Times @nytimes · 12 Oct 2016
Worries over the health of King Bhumibol Adulyadej are shaking Thailand
nyti.ms/2dRzPcr
5 261 144

Career Synchronicity @careersync_now · 12 Oct 2016
Fears for King's Health Shake Thailand ift.tt/2d7frGd

Herbert Buchsbaum @herbertnyt · 12 Oct 2016
New bulletin from Thai palace: King is still on a ventilator and in unstable condition. nyti.ms/2dW1A37

[1] **Wuwei Lan**, Siyu Qiu, Hua He, and Wei Xu. A Continuously Growing Dataset of Sentential Paraphrases (EMNLP 2017).

Paraphrase Collection in Twitter[1]

https://www.nytimes.com/2016/10/13/world/asia/thailand-king.html

same URL

The New York Times @nytimes · 12 Oct 2016
Worries over the health of King Bhumibol Adulyadej are shaking Thailand
nyti.ms/2dRzPcr
5 261 144

Career Synchronicity @careersync_now · 12 Oct 2016
Fears for King's Health Shake Thailand ift.tt/2d7frGd

Herbert Buchsbaum @herbertnyt · 12 Oct 2016
New bulletin from Thai palace: King is still on a ventilator and in unstable condition. nyti.ms/2dW1A37

[1] **Wuwei Lan,** Siyu Qiu, Hua He, and Wei Xu. A Continuously Growing Dataset of Sentential Paraphrases (EMNLP 2017).

Twitter-URL Corpus Construction

Step 1:Collecting tweets that refer to the same URL.

```
for account in {CNN, BBC, New York Times et al.}:
    for tweet in {account.tweets_of_yesterday}:
        if tweet contains URL:
            search_tweet = Twitter.search(key=URL)
            search_tweet.clean()
            if NotOverlap(tweet, search_tweet):
                candidate.append(search_tweet)
    end
end
```

Step 2: Labelling for gold standard corpus

Sampling ~51k candidate sentence pairs

Labeled by Amazon Mechanical Turk

Step 3: Training

Splitting 42k for training and 9k for testing

Run various paraphrase identification models

Get best classifier M

Step 4: Predicting

M.predict(unlabeled data)

Existing Paraphrase Corpora

Key for success:

- narrow the search space
- ensure diversity among sentences

Also Pitfalls …

[MSRP[1]]

clustered news articles

5,801 annotated pairs

needed a SVM classifier to select sentences before data annotation

[PIT-2015[2]]

Twitter trending topics

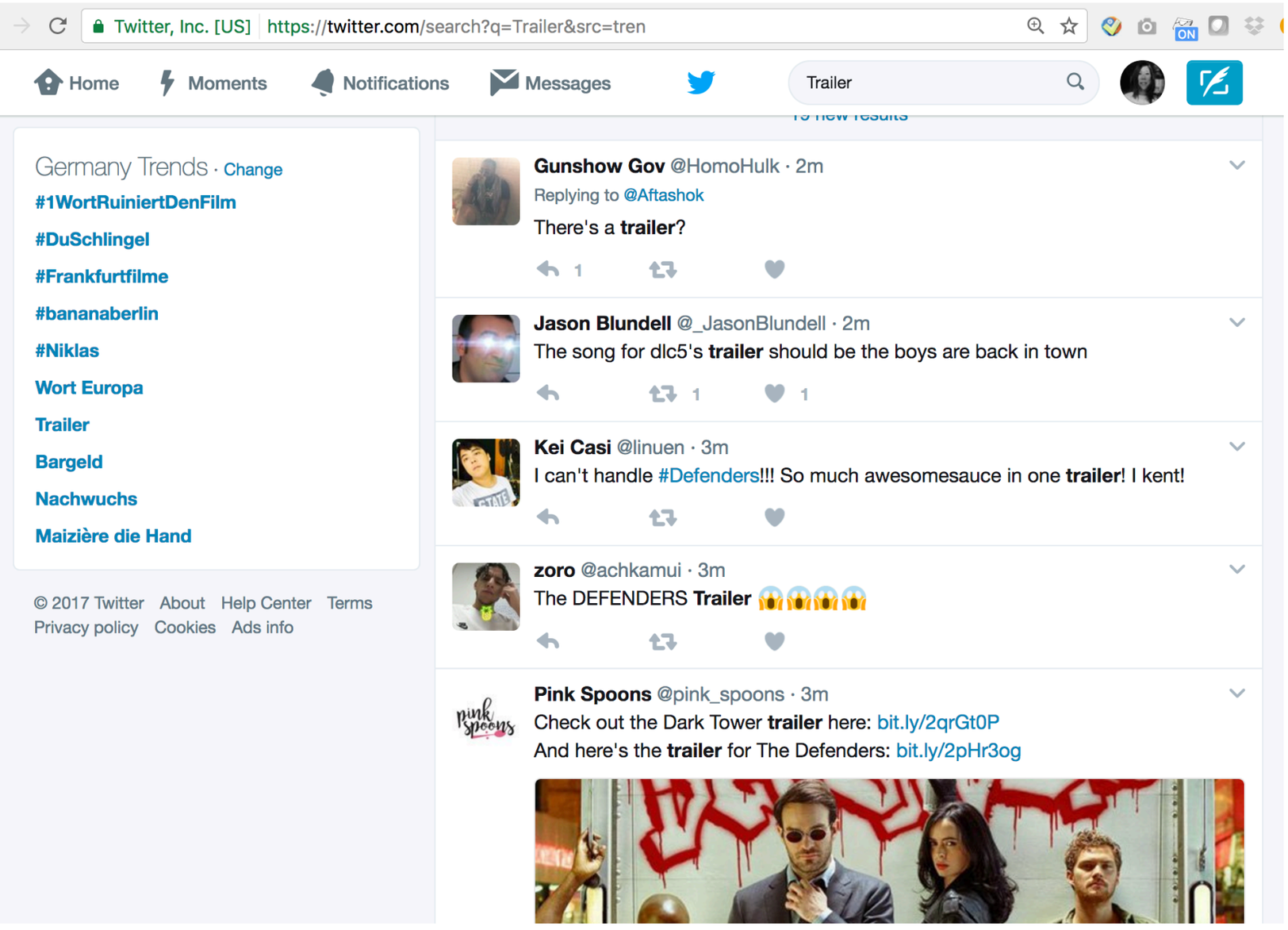
14,035 annotated pairs

needed human-in-the-loop to avoid “bad” topics

[1] William B. Dolan and et al., Unsupervised construction of large paraphrase corpora: Exploiting massively parallel news sources. (COLING 2004)

[2] Wei Xu and et al., Extracting lexically divergent paraphrases from Twitter. (TACL 2014)

Existing Paraphrase Corpora



[PIT-2015[2]]

Twitter trending topics

14,035 annotated pairs

needed human-in-the-loop to avoid “bad” topics

Existing Paraphrase Corpora

Key for success:

- narrow the search space
- ensure diversity among sentences
- **the simpler the better!**

[MSRP[1]]

clustered news articles

5,801 annotated pairs

[Twitter-URL]

URL-linked tweets

51,524 annotated pairs

[PIT-2015[2]]

Twitter trending topics

14,035 annotated pairs

no clustering or topic detection needed
no data selection steps needed

[1] William B. Dolan and et al., Unsupervised construction of large paraphrase corpora: Exploiting massively parallel news sources. (COLING 2004)

[2] Wei Xu and et al., Extracting lexically divergent paraphrases from Twitter. (TACL 2014)

Existing Paraphrase Corpora

Key for success:

- narrow the search space
- ensure diversity among sentences
- **the simpler the better! more effective automatic paraphrase identification**

[MSRP[1]]

clustered news articles

5,801 annotated pairs

[Twitter-URL]

URL-linked tweets

51,524 annotated pairs

[PIT-2015[2]]

Twitter trending topics

14,035 annotated pairs

30,000 new sentential paraphrases every month

[1] William B. Dolan and et al., Unsupervised construction of large paraphrase corpora: Exploiting massively parallel news sources. (COLING 2004)

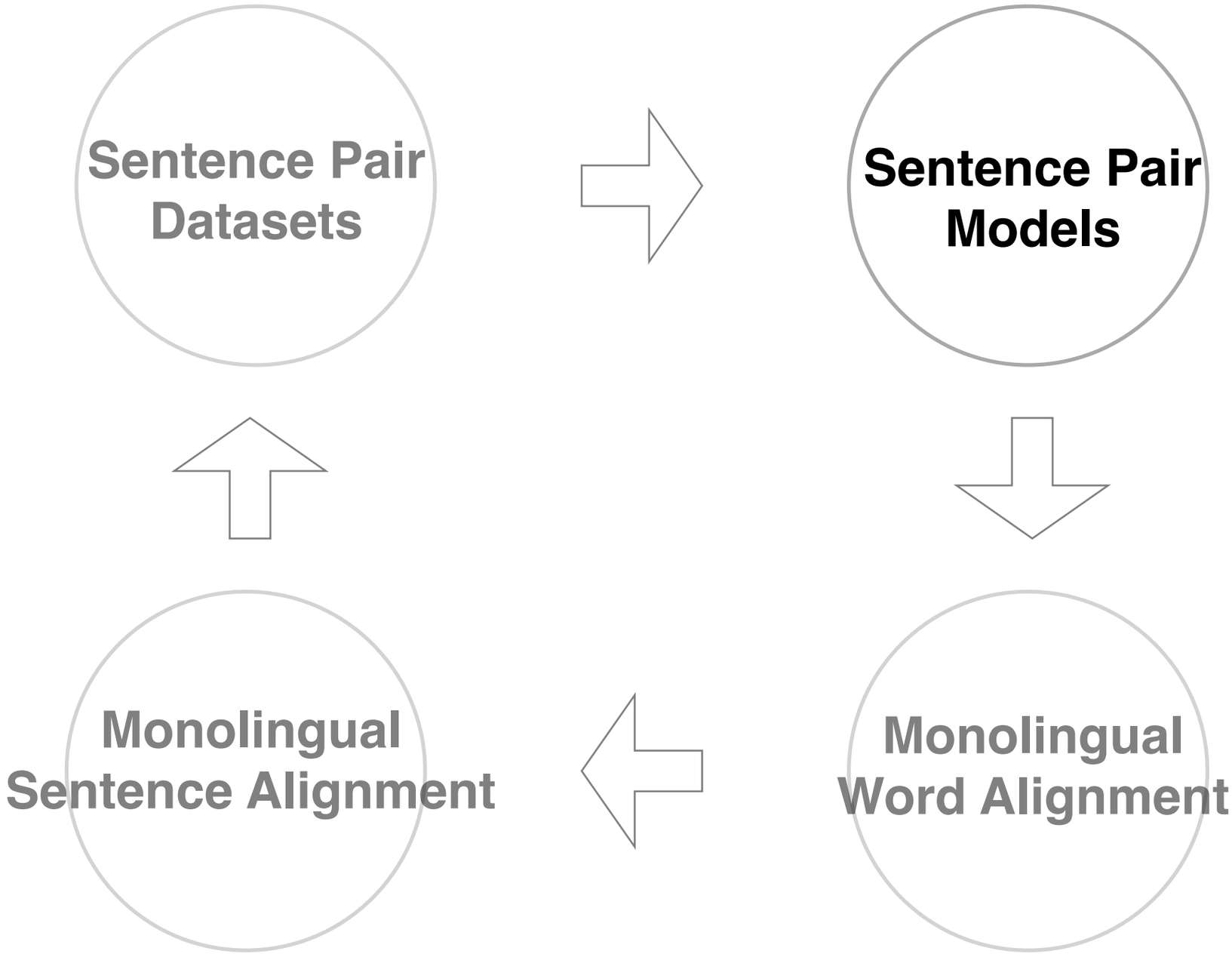
[2] Wei Xu and et al., Extracting lexically divergent paraphrases from Twitter. (TACL 2014)

Once we have a lot of sentential paraphrases

(we can learn name variations fully automatically with word alignment models)

Donald Trump, DJT, Drumpf, Mr Trump, Idiot Trump, Chump, Evil Donald, #OrangeHitler, Donald @realTrump, D*nald Tr*mp, Comrade #Trump, Crooked #Trump, CryBaby Trump, Daffy Trump, Donald KKKrump, Dumb Trump, GOPTrump, Incompetent Trump, He-Who-Must-Not-Be-Named, Pres-elect Trump, President-Elect Trump, President-elect Donald J . Trump, PEOTUS Trump, Emperor Trump

Outline


Sentence Pair Datasets
Sentence Pair Models
Monolingual Word Alignment
Monolingual Sentence Alignment

Out-of-vocabulary Problem in Social Media

Dataset	Training Size	Test Size	# INV	# OOV	OOV Ratio	Source
PIT-2015	11530	838	7771	1238	13.7%	Twitter trends
Twitter-URL	42200	9324	24905	11440	31.5%	Twitter/news
MSRP	4076	1725	16226	1614	9.0%	news

Table 2.7: Statistics of three benchmark datasets for paraphrase identification. The training and testing sizes are in numbers of sentence pairs. The number of unique in-vocabulary (INV) and out-of-vocabulary (OOV) words are calculated based on the publicly available GloVe embeddings

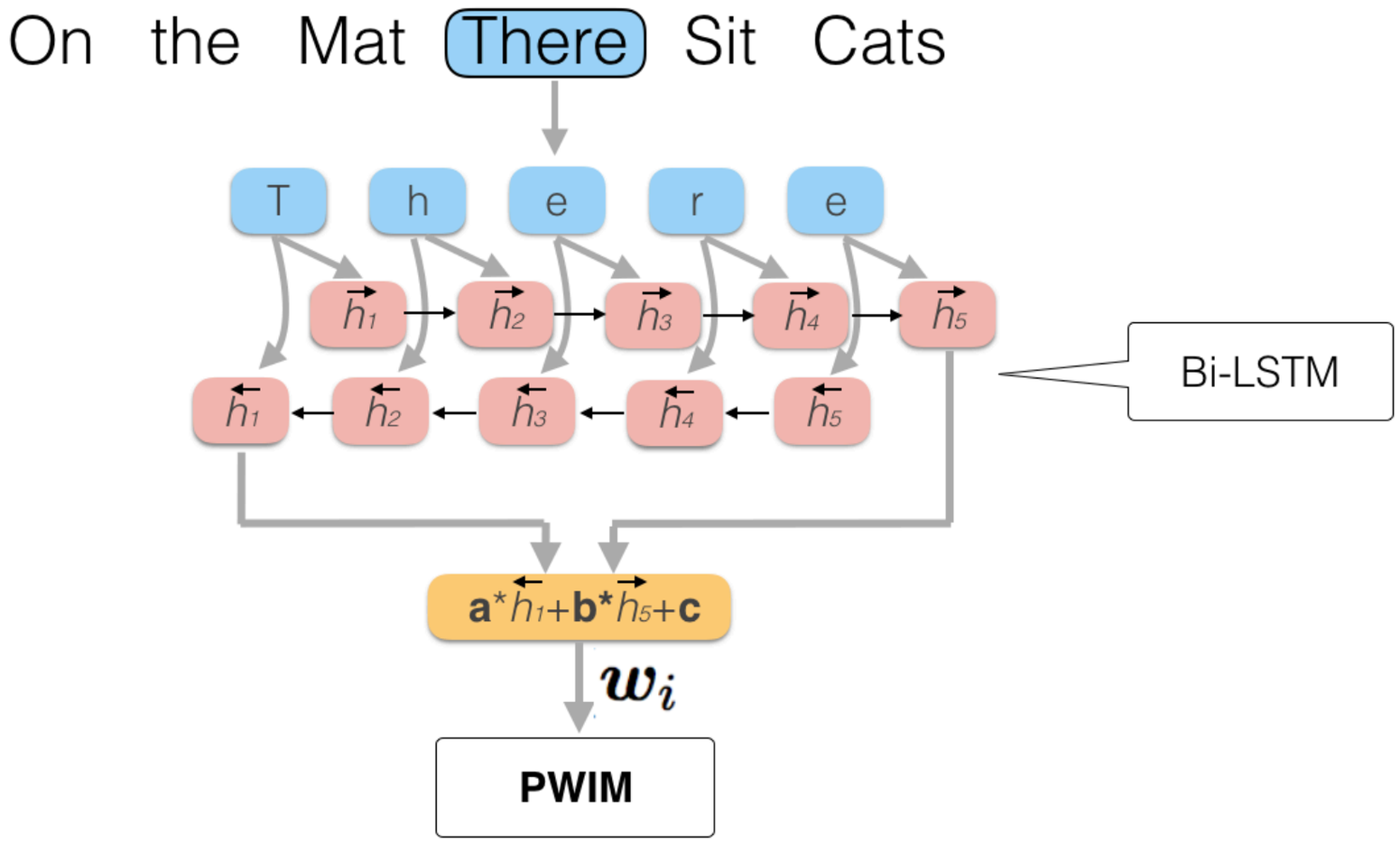
Representing Word with Smaller Units[1]

Unit	Output of σ(brexit)
unigram	b, r, e, x, i, t
bigram w overlap	br, re, ex, xi, it
bigram w/o overlap	br, ex, it
trigram w overlap	bre, rex, exi, xit
trigram w/o overlap	bre, xit
whole word	brexit

Table 1: Ngram examples for word ***brexit***.

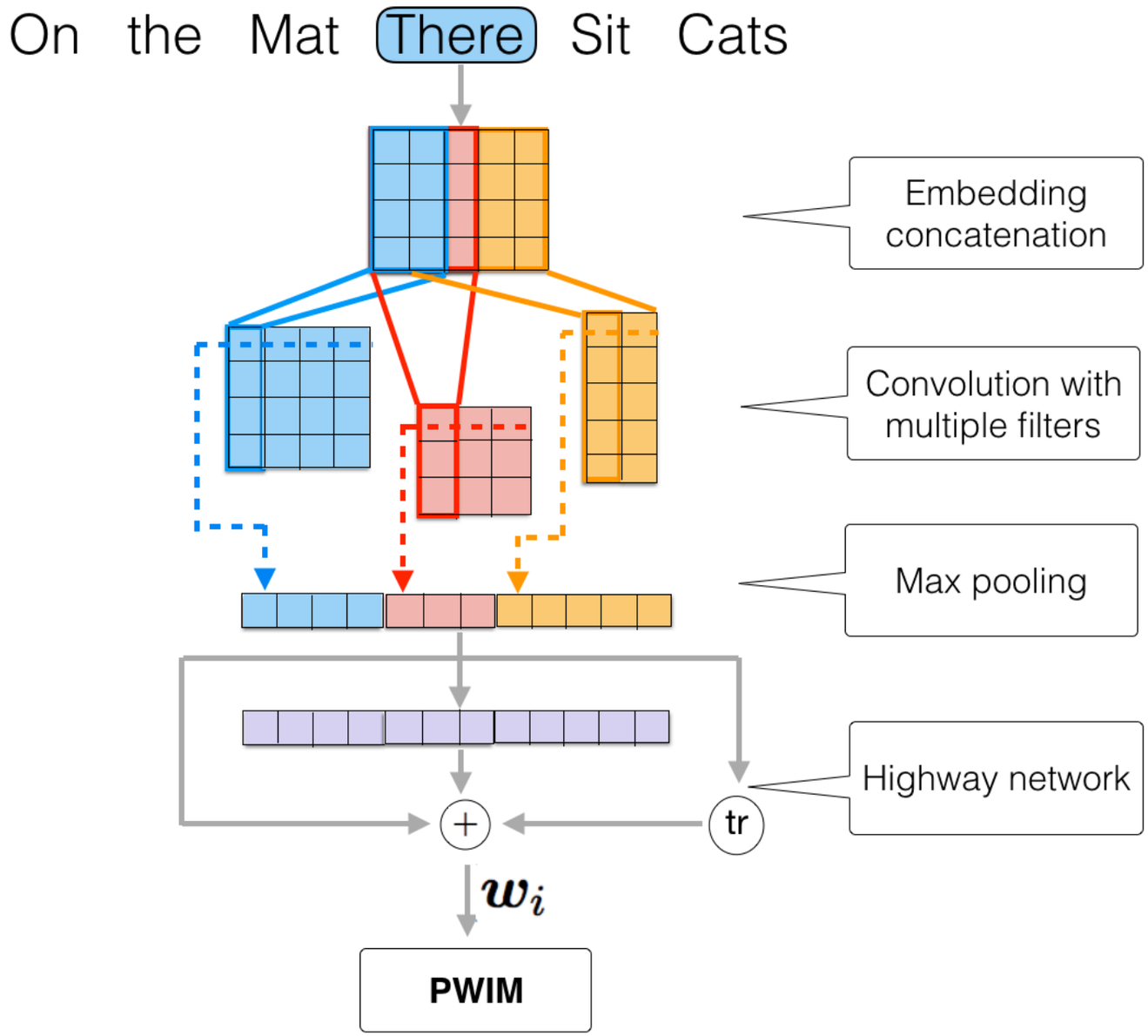
[1] **Wuwei Lan** and Wei Xu, The importance of subword embeddings in sentence pair modeling. (NAACL 2018)

RNN Based Character Embedding[1]



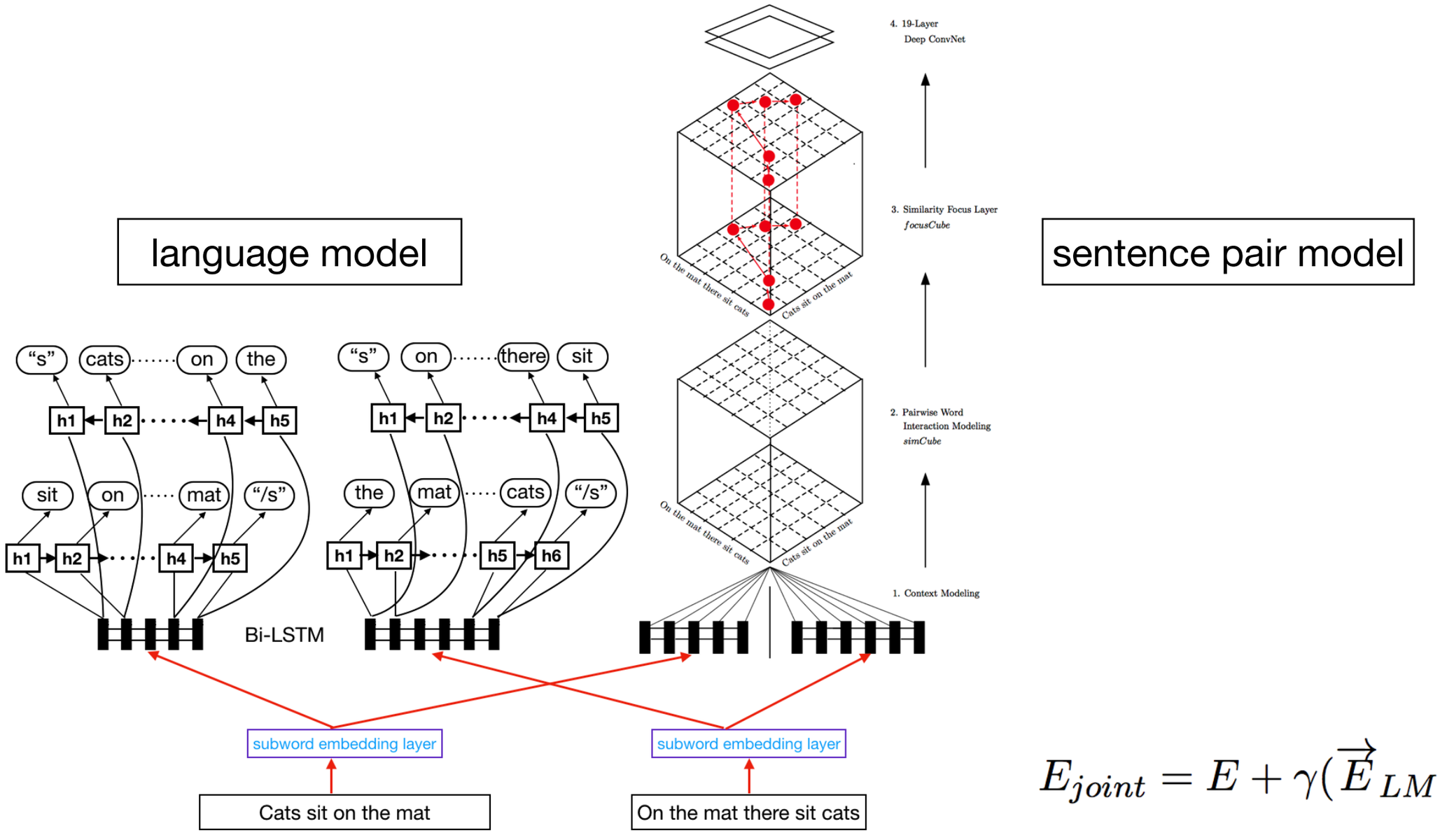
[1] Wang Ling, and et al., Finding function in form: Compositional character models for open vocabulary word representation. (EMNLP 2015)

CNN Based Character Embedding [1]



[1] Yoon Kim, and et al., Character-aware neural language models. (AAAI 2016)

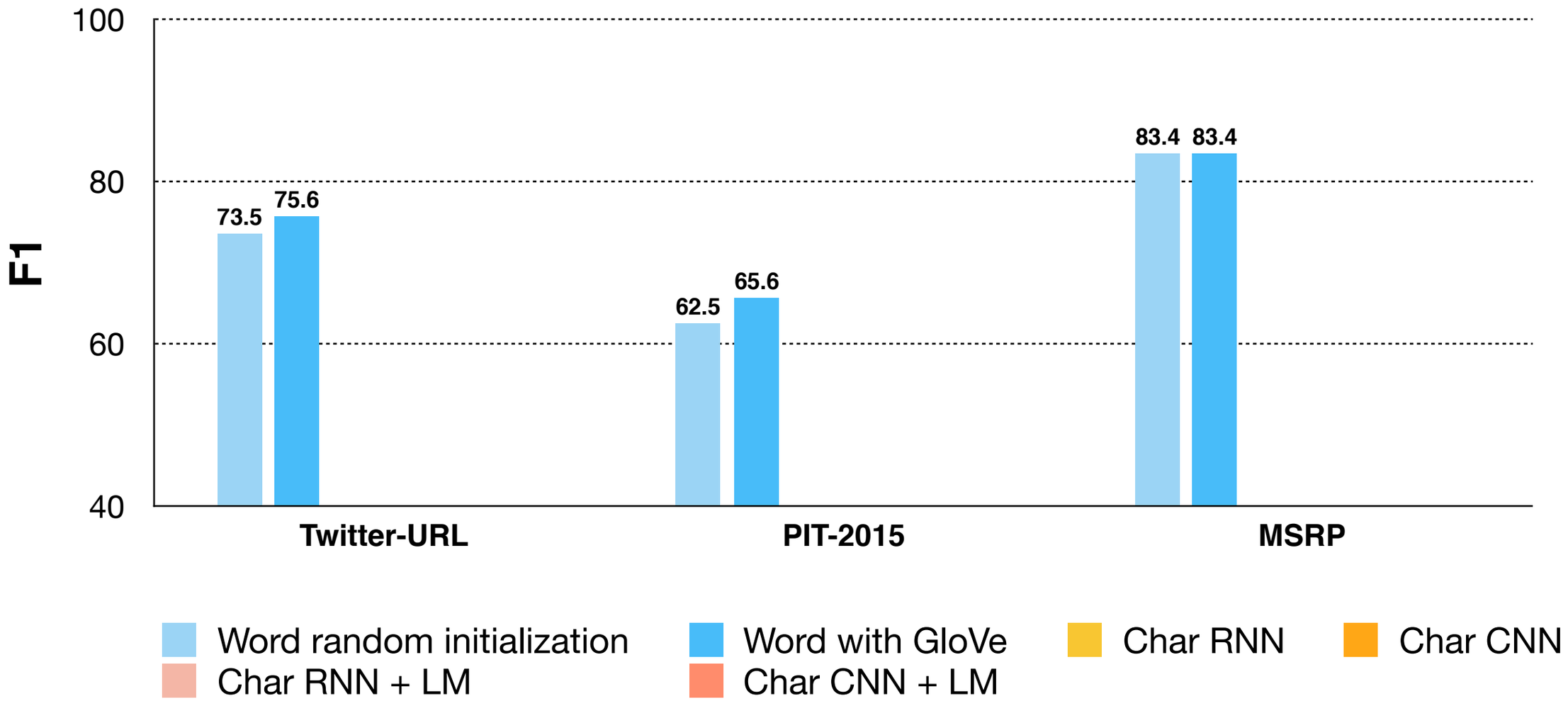
Multi-task Language Model[1]



$$E_{joint} = E + \gamma(\overrightarrow{E}_{LM} + \overleftarrow{E}_{LM})$$

[1] **Wuwei Lan** and Wei Xu, The importance of subword embeddings in sentence pair modeling. (NAACL 2018)

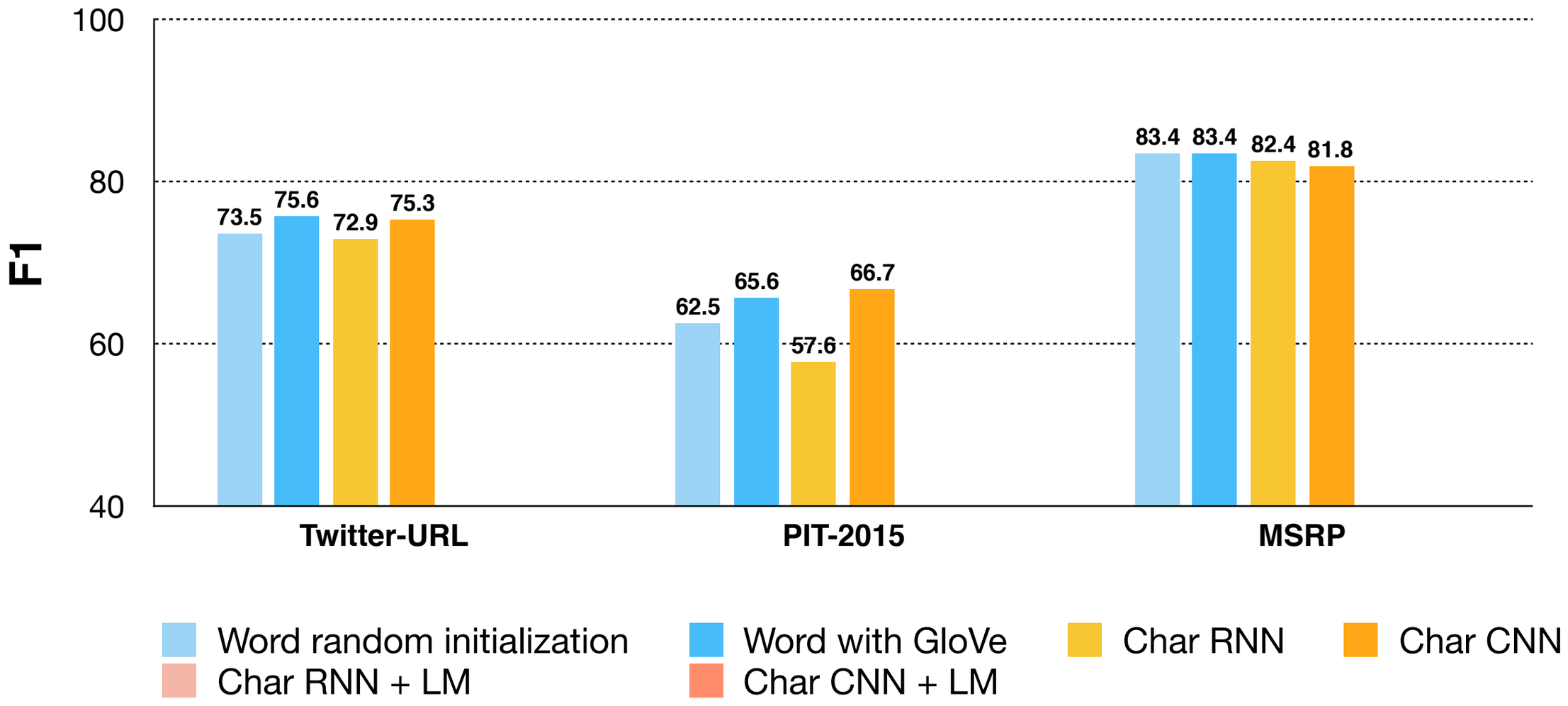
Experiment for Subword Models [1]



- Subword models have comparable performance with word models.
- Subword models present new SOTA with multi-task language model.

[1] **Wuwei Lan** and Wei Xu, The importance of subword embeddings in sentence pair modeling. (NAACL 2018)

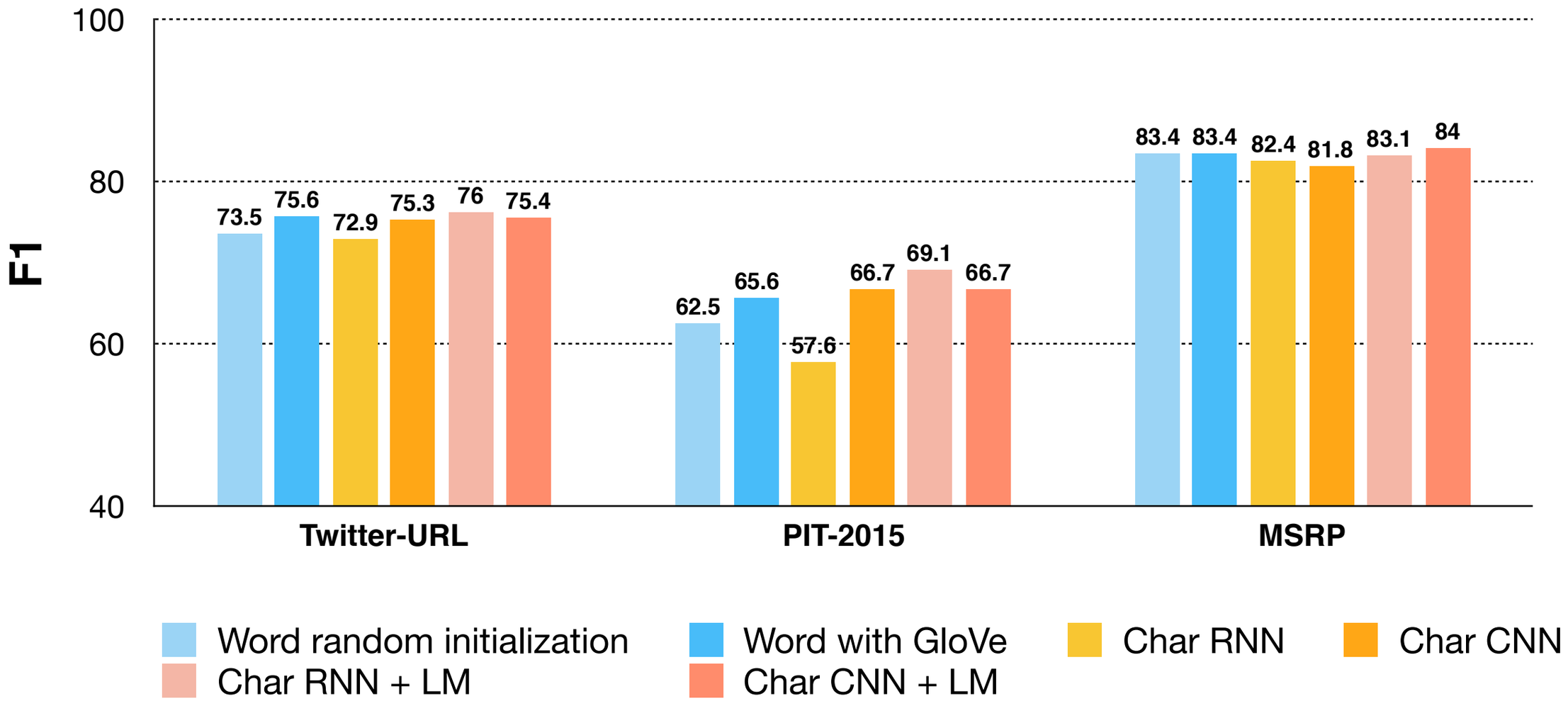
Experiment for Subword Models [1]



- Subword models have comparable performance with word models.
- Subword models present new SOTA with multi-task language model.

[1] **Wuwei Lan** and Wei Xu, The importance of subword embeddings in sentence pair modeling. (NAACL 2018)

Experiment for Subword Models [1]



- Subword models have comparable performance with word models.
- Subword models present new SOTA with multi-task language model.

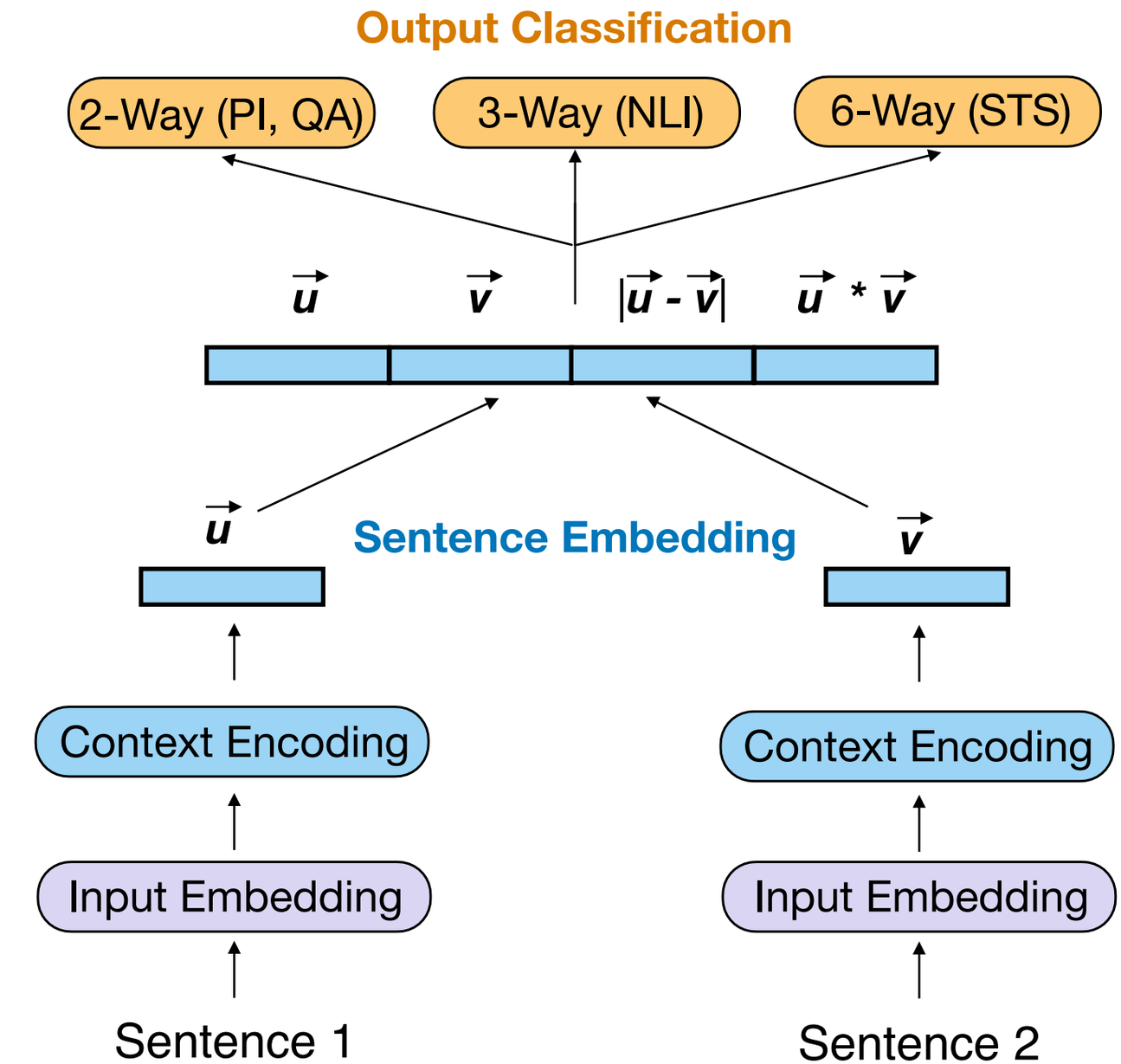
[1] **Wuwei Lan** and Wei Xu, The importance of subword embeddings in sentence pair modeling. (NAACL 2018)

General neural framework for SPM tasks

Type I: Sentence Encoding-based Models

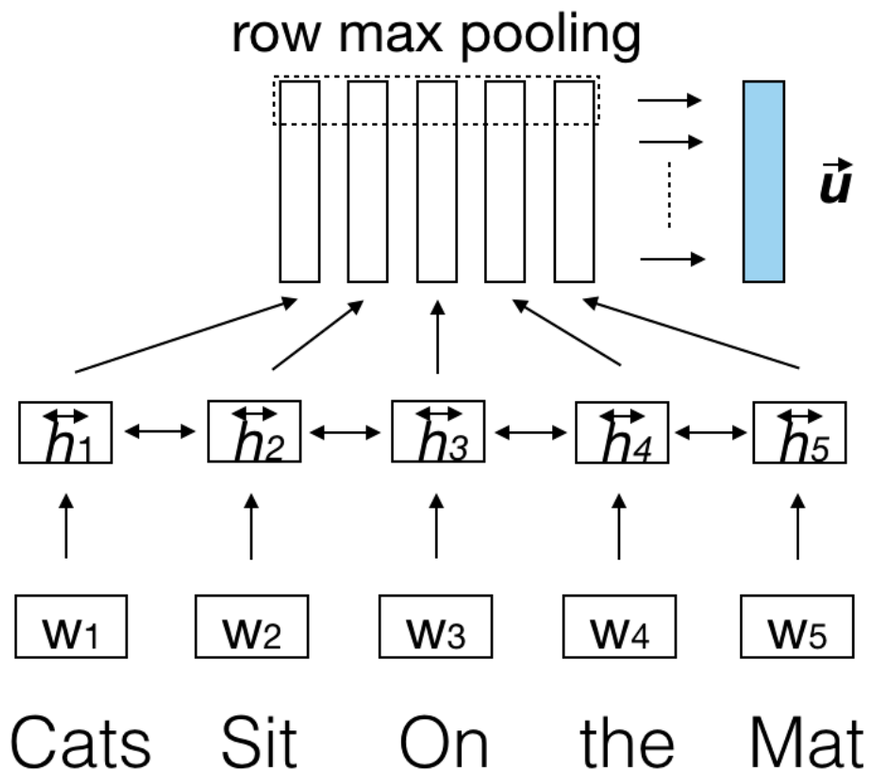
Type II: Word Interaction-based Models

Type I: Sentence Encoding-based Models[1]



[1] **Wuwei Lan** and Wei Xu, Neural Network Models for Paraphrase Identification, Semantic Textual Similarity, Natural Language Inference, and Question Answering. (COLING 2018)

Type I: Sentence Encoding-based Models[1]

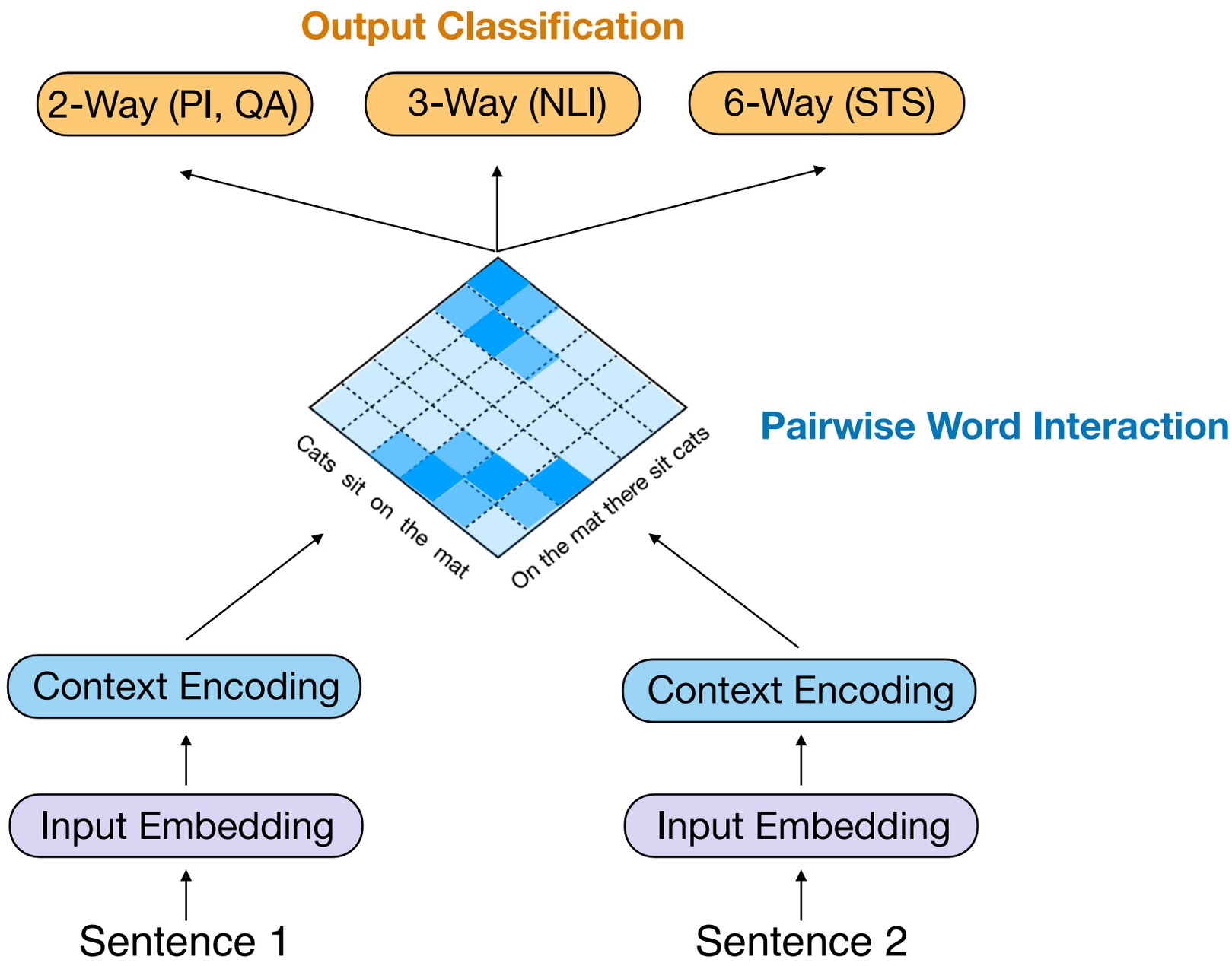


Sentence Embedding

InferSent: 1-layer Bi-LSTM.[2]
SSE: 3-layer Bi-LSTM with skip connection.[3]

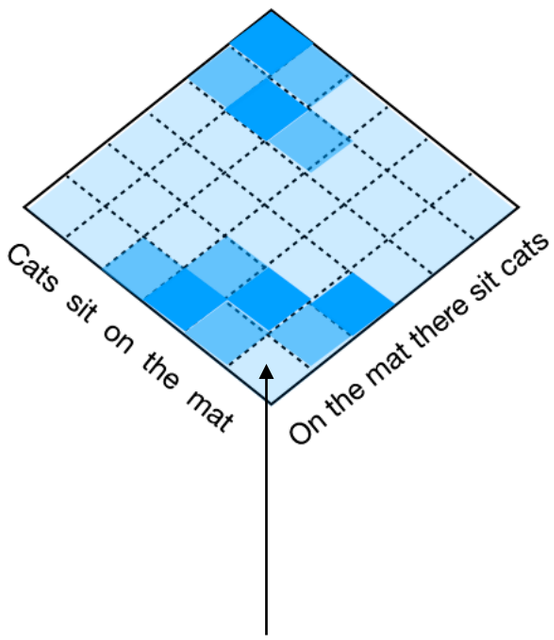
[1] **Wuwei Lan** and Wei Xu, Neural Network Models for Paraphrase Identification, Semantic Textual Similarity, Natural Language Inference, and Question Answering. (COLING 2018)
[2] Jihun Choi, Kang Min Yoo, and Sang-goo Lee: Unsupervised learning of task-specific tree structures with tree-LSTMs. (EMNLP 2017).
[3] Yixin Nie and Mohit Bansal. Shortcut-stacked sentence encoders for multi-domain inference. (RepEval 2017)

Type II: Word Interaction-based Models[1]



- semantic relation between two sentences depends largely on aligned words/phrases

[1] **Wuwei Lan** and Wei Xu, Neural Network Models for Paraphrase Identification, Semantic Textual Similarity, Natural Language Inference, and Question Answering. (COLING 2018)



F(,)

Pairwise Word Interaction

=G(,)

=G(,)

.....

=G(,)

Aggregate

y=H(+ + … +)

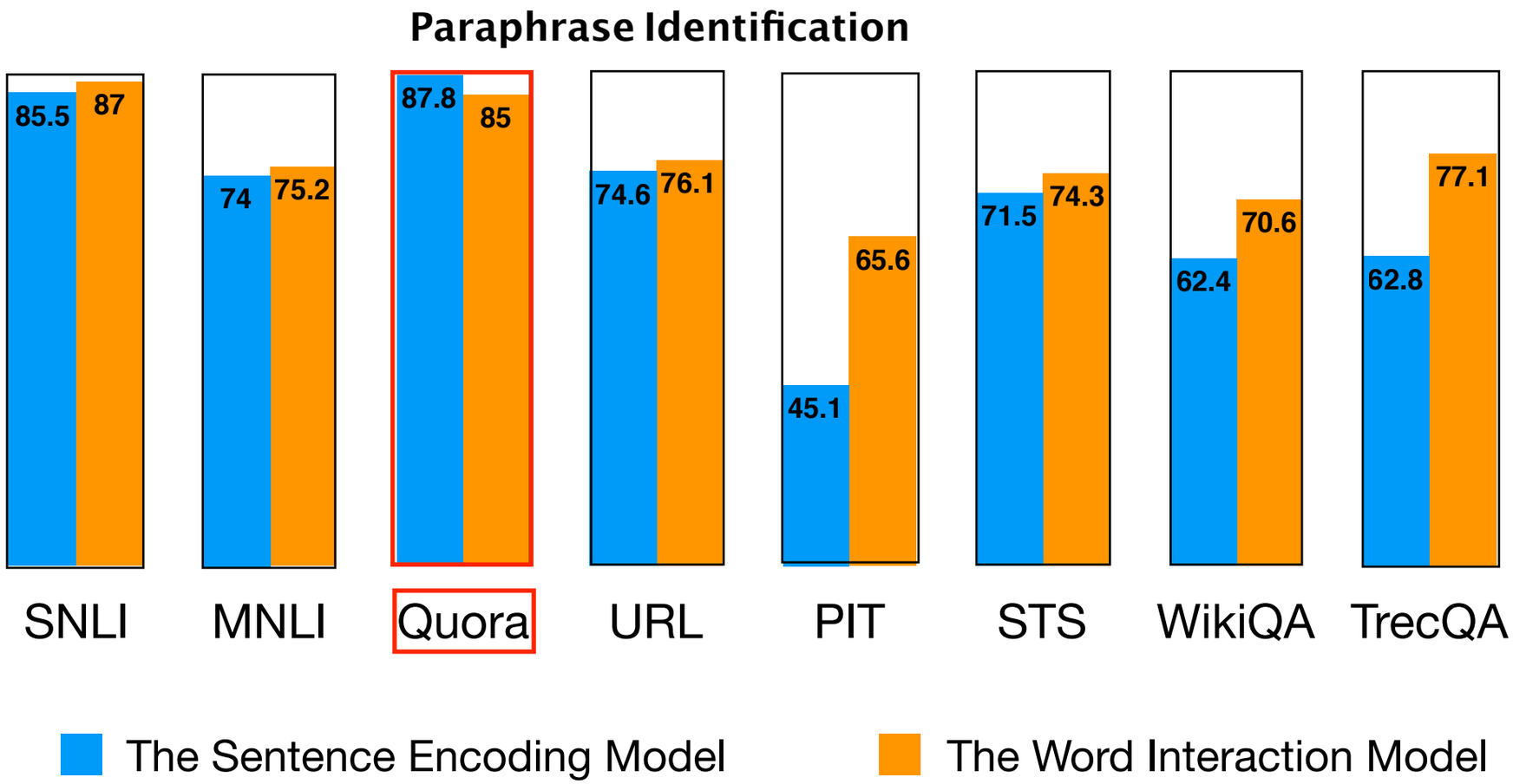
DecAtt[5]: **F** is dot product; **G, H** are feedforward networks.

ESIM[6]: more features in **G**(, , ,), and **G** is replaced with Bi-LSTM/Tree-LSTM.

PWIM[7]: **F** uses cosine, L2 and dot product; **G** (,) is "hard" attention; **H** is deep CNN.

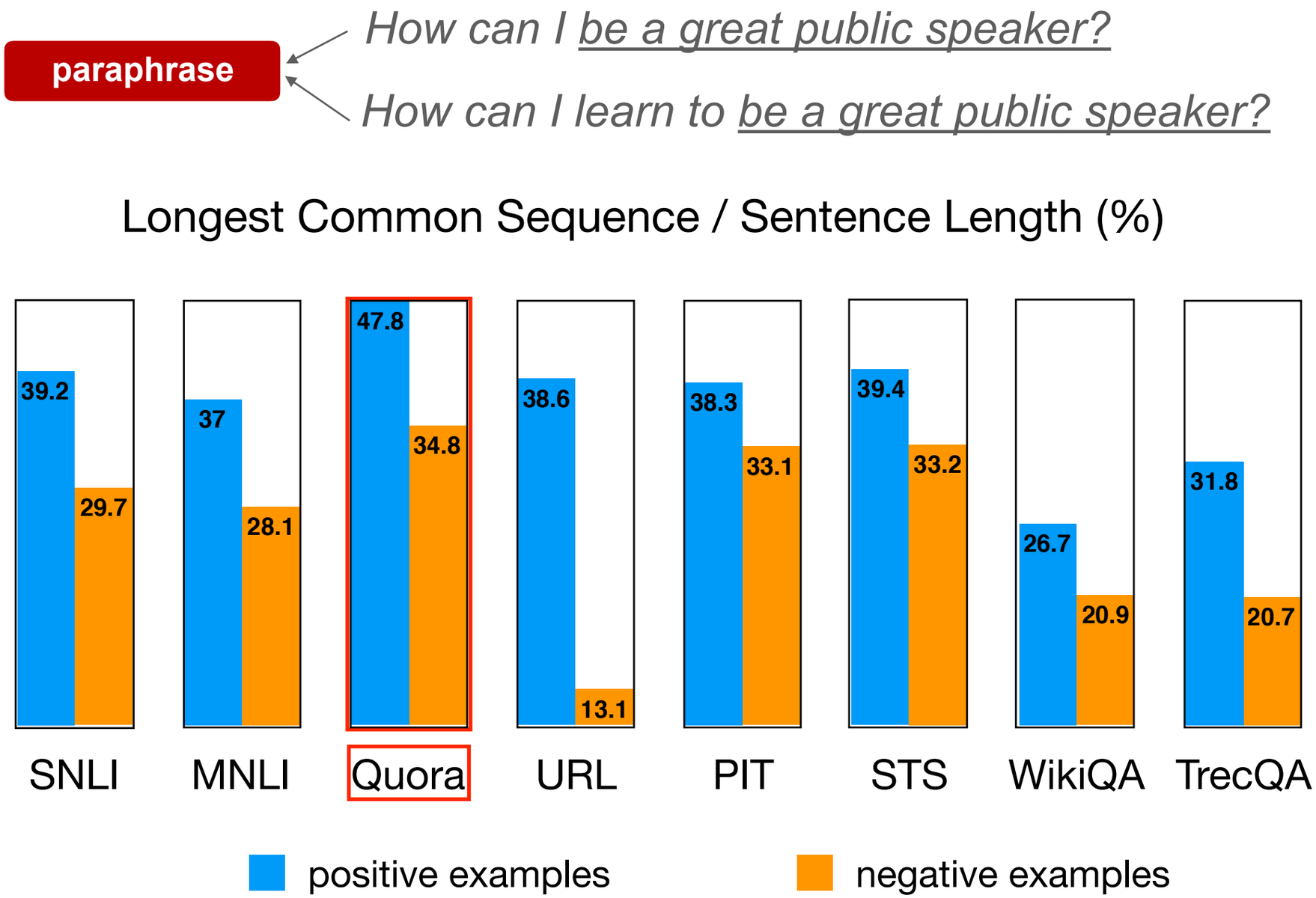
[5] Ankur Parikh, Oscar Tackstrom, Dipanjan Das, and Jakob Uszkorei. A decomposable ¨ attention model for natural language inference. (EMNLP 2016)
[6] Qian Chen, Xiaodan Zhu, Zhenhua Ling, Si Wei, Hui Jiang, and Diana Inkpen. Enhanced LSTM for natural language inference. (ACL 2017)
[7] Hua He and Jimmy Lin. Pairwise word interaction modeling with deep neural networks for semantic similarity measurement. (NAACL 2016)

What Type of Model performs better ?



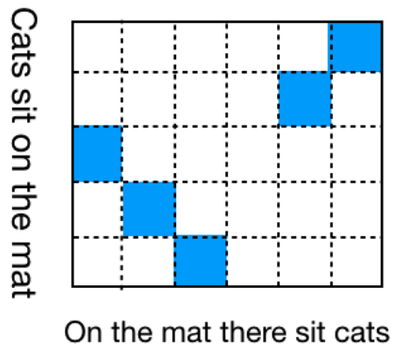
- Type II model outperforms Type I model significantly (except Quora).

Why is Quora an exception ?

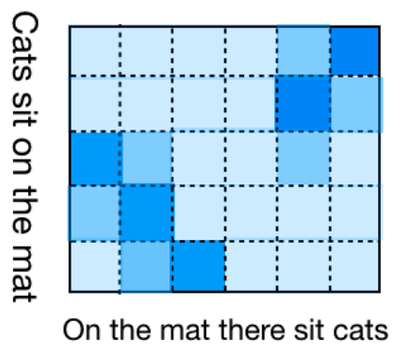


- More subsequences shared, less interaction involved.

In general, Type II model performs better



PWIM[8]**:** hard alignment.

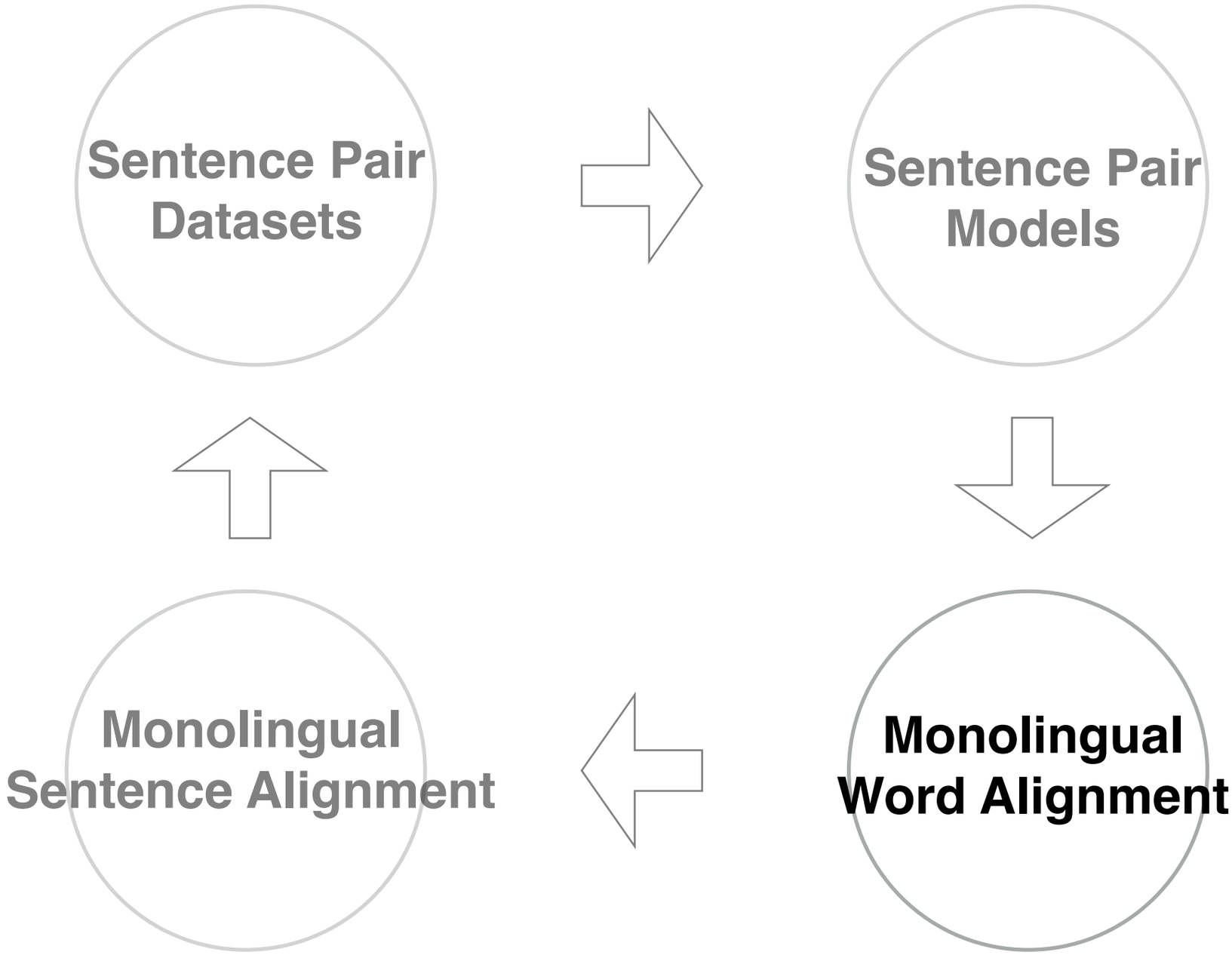


DecAtt[9] **ESIM**[10]**:** soft alignment.

Word alignment plays a key role!

[8] Hua He and Jimmy Lin. Pairwise word interaction modeling with deep neural networks for semantic similarity measurement. (NAACL 2016)
[9] Ankur Parikh, Oscar Tackstrom, Dipanjan Das, and Jakob Uszkorei. A decomposable ¨ attention model for natural language inference. (EMNLP 2016)
[10] Qian Chen, Xiaodan Zhu, Zhenhua Ling, Si Wei, Hui Jiang, and Diana Inkpen. Enhanced LSTM for natural language inference. (ACL 2017)

Outline


Sentence Pair Datasets
Sentence Pair Models
Monolingual Word Alignment
Monolingual Sentence Alignment

Example: Monolingual Word Alignment

Figure 1: Example from MTReference dataset. Black is *sure* alignment and grey is *possible* alignment.

Word alignment has a long history (20+ years)

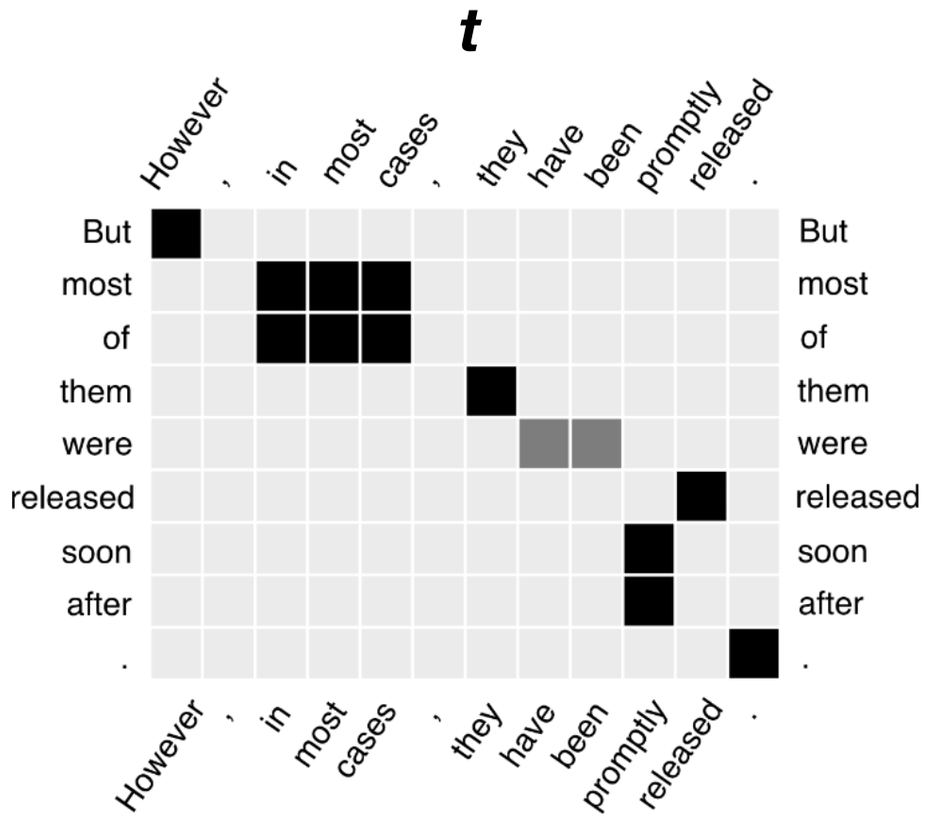
- Tools on IBM models [Brown+, CL-1993]: GIZA++ [Och and Ney, CL-2003] and FastAlign [Dyer+, NAACL 2013].
    - bilingual, SMT, unsupervised, low quality
- Some monolingual word aligners: Jacana [Yao+, ACL 2013] and Sultan's aligner [Sultan+, TACL 2014].
    - supervised, lots of rules and features, but not satisfying
- Neural based bilingual word aligners: [Garg+, EMNLP-2019], [Stengel-Eskin+, EMNLP-2019] and [Zenkel+ ACL-2020].
    - unsupervised, outperform GIZA++ but rely on NMT.

Our focus — monolingual, supervised, high quality !

Problem Formulation [1]

$|\boldsymbol{a}| = |\boldsymbol{s}|$

$\boldsymbol{a} = [1,3,3,7,0,11,10,10,12]$



discriminative classifier

$$p(\boldsymbol{a}|\boldsymbol{s},\boldsymbol{t}) = \frac{e^{\psi(\boldsymbol{a},\boldsymbol{s},\boldsymbol{t})}}{\sum_{\boldsymbol{a}' \in \mathcal{A}} e^{\psi(\boldsymbol{a}',\boldsymbol{s},\boldsymbol{t})}}$$

linear-chain CRF

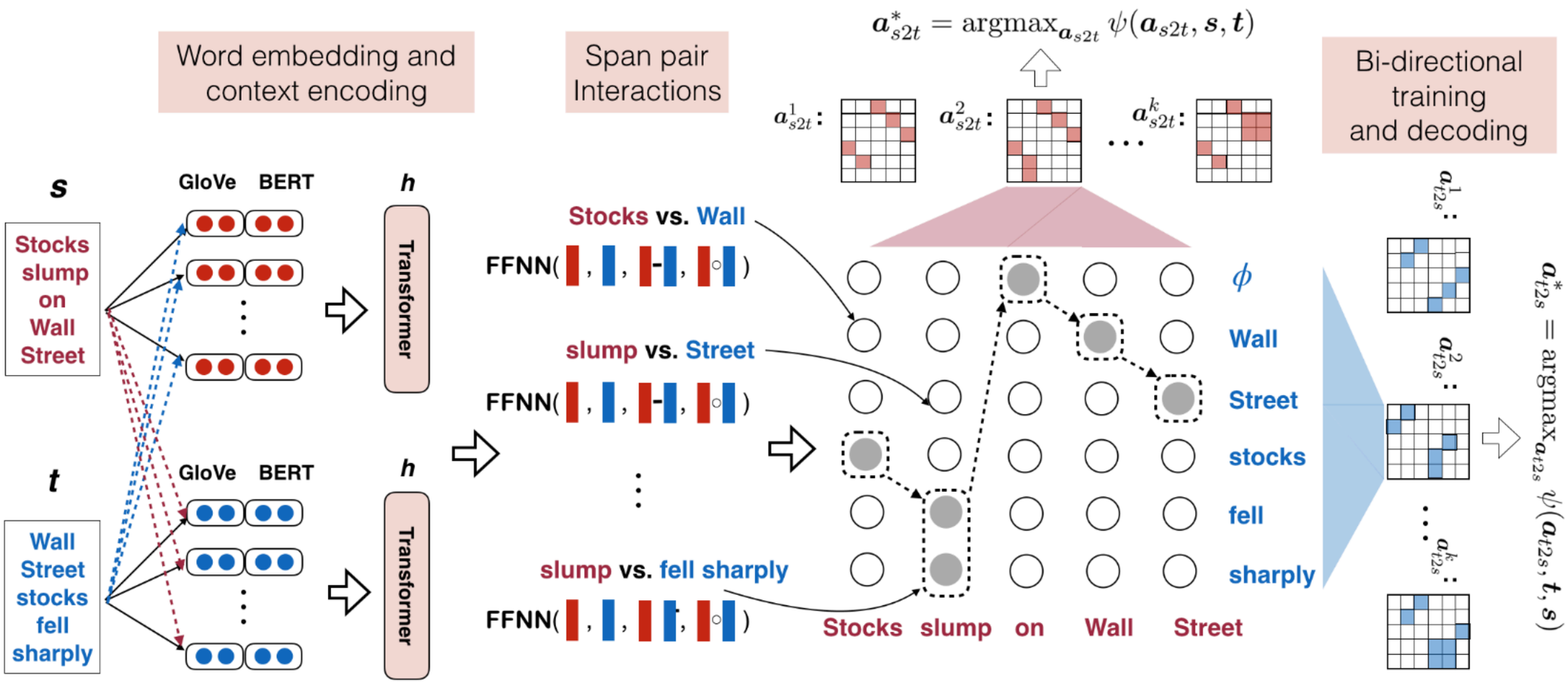
$$\psi(\boldsymbol{a},\boldsymbol{s},\boldsymbol{t}) = \sum_{i=1}^{|\boldsymbol{s}|} \upsilon(s_i, t_{a_i}) + \tau(a_{i-1}, a_i)$$

linear-chain semi-CRF

$$\psi(\boldsymbol{a},\boldsymbol{s},\boldsymbol{t}) = \sum_{i,d} \upsilon(s_{[i-d+1:i]}, t'_{a_i}) + \tau(a_{i-d}, a_i)$$

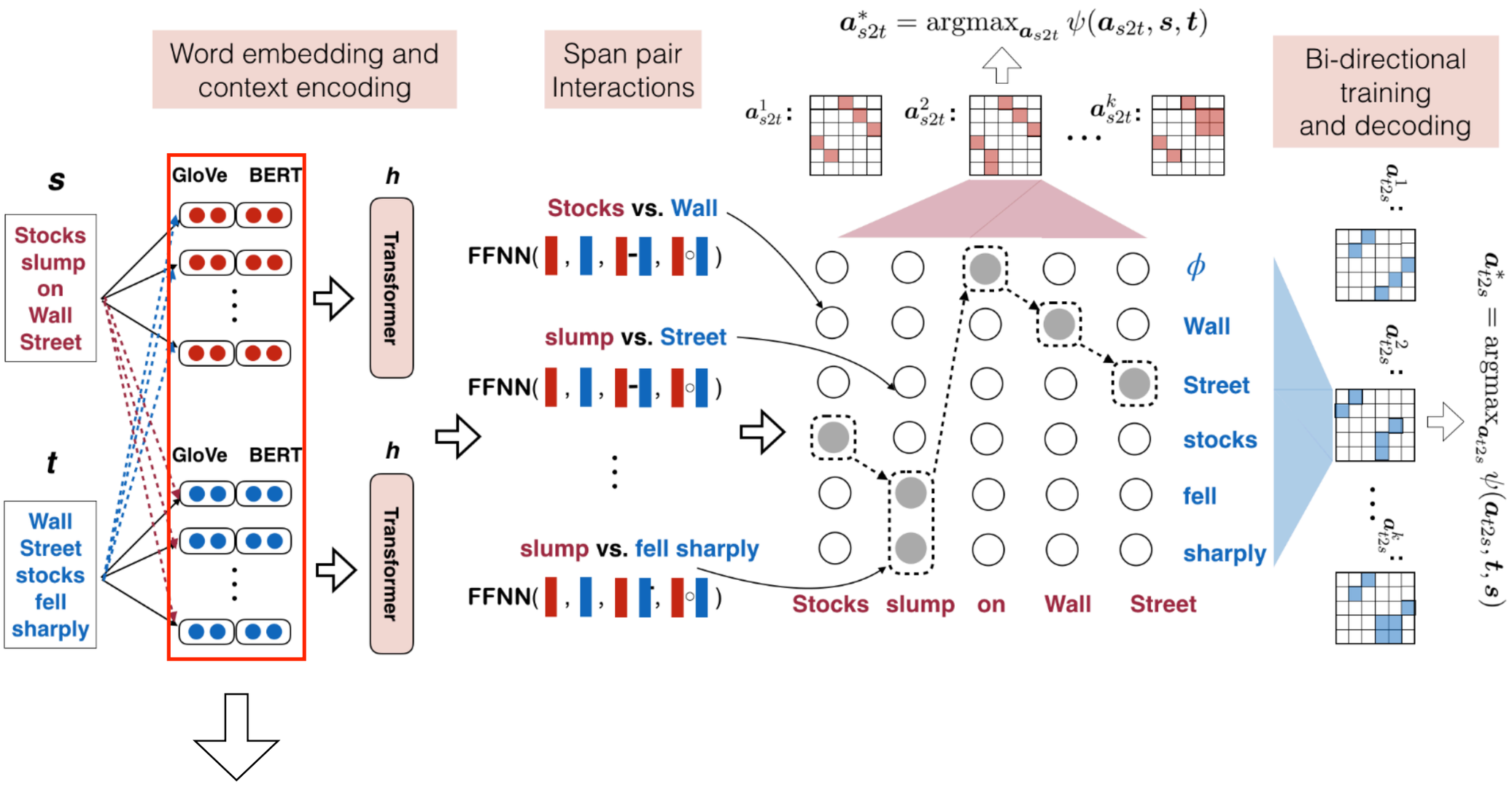
[1] Phil Blunsom and Trevor Cohn. Discriminative Word Alignment with Conditional Random Fields. (ACL 2006)

Neural Semi-CRF Word Aligner [1]



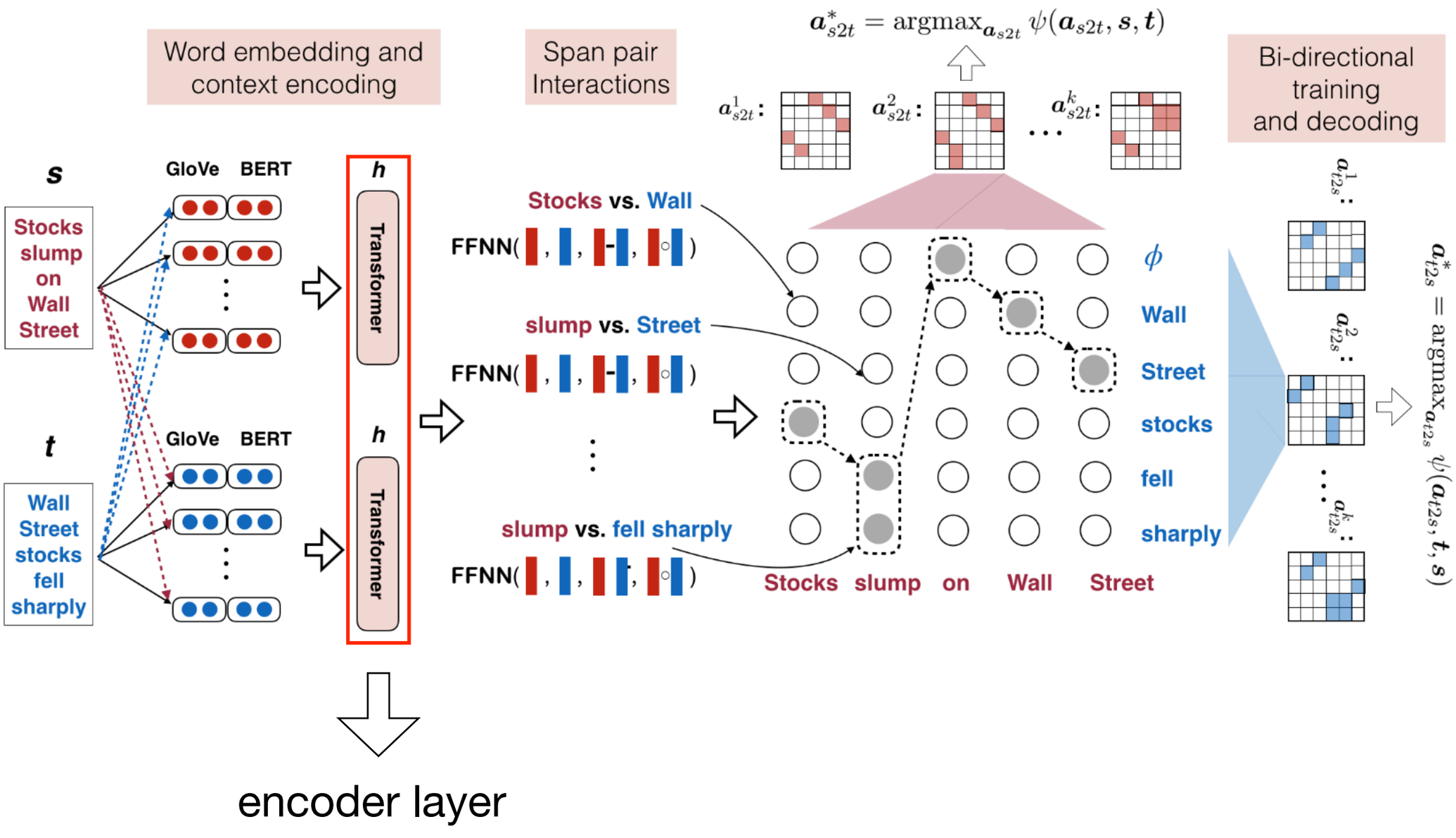
[1] **Wuwei Lan** and Wei Xu, Neural semi-Markov CRF for Monolingual Word Alignment. (in submission)

Neural Semi-CRF Word Aligner [1]



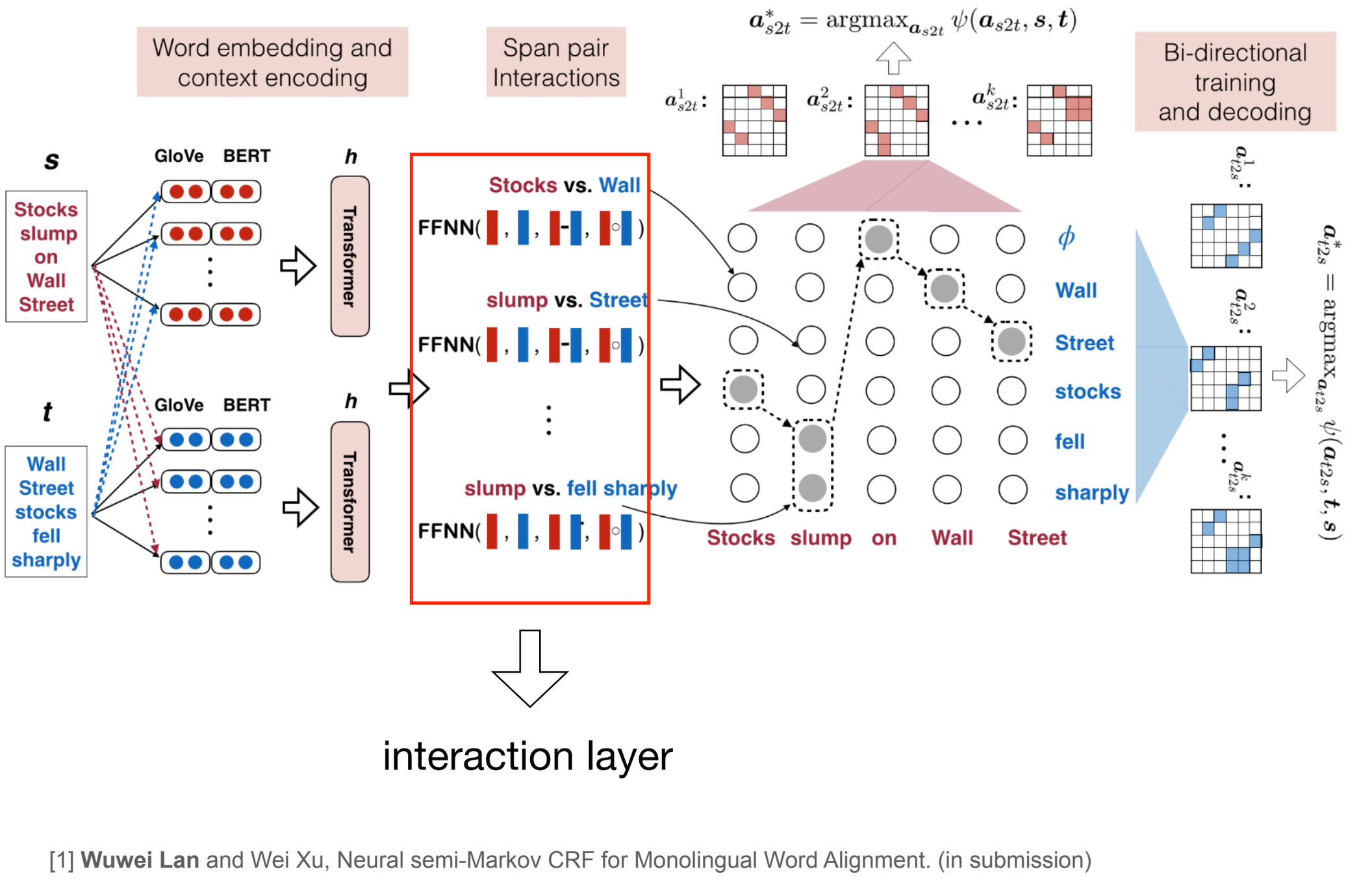
[1] **Wuwei Lan** and Wei Xu, Neural semi-Markov CRF for Monolingual Word Alignment. (in submission)

Neural Semi-CRF Word Aligner [1]



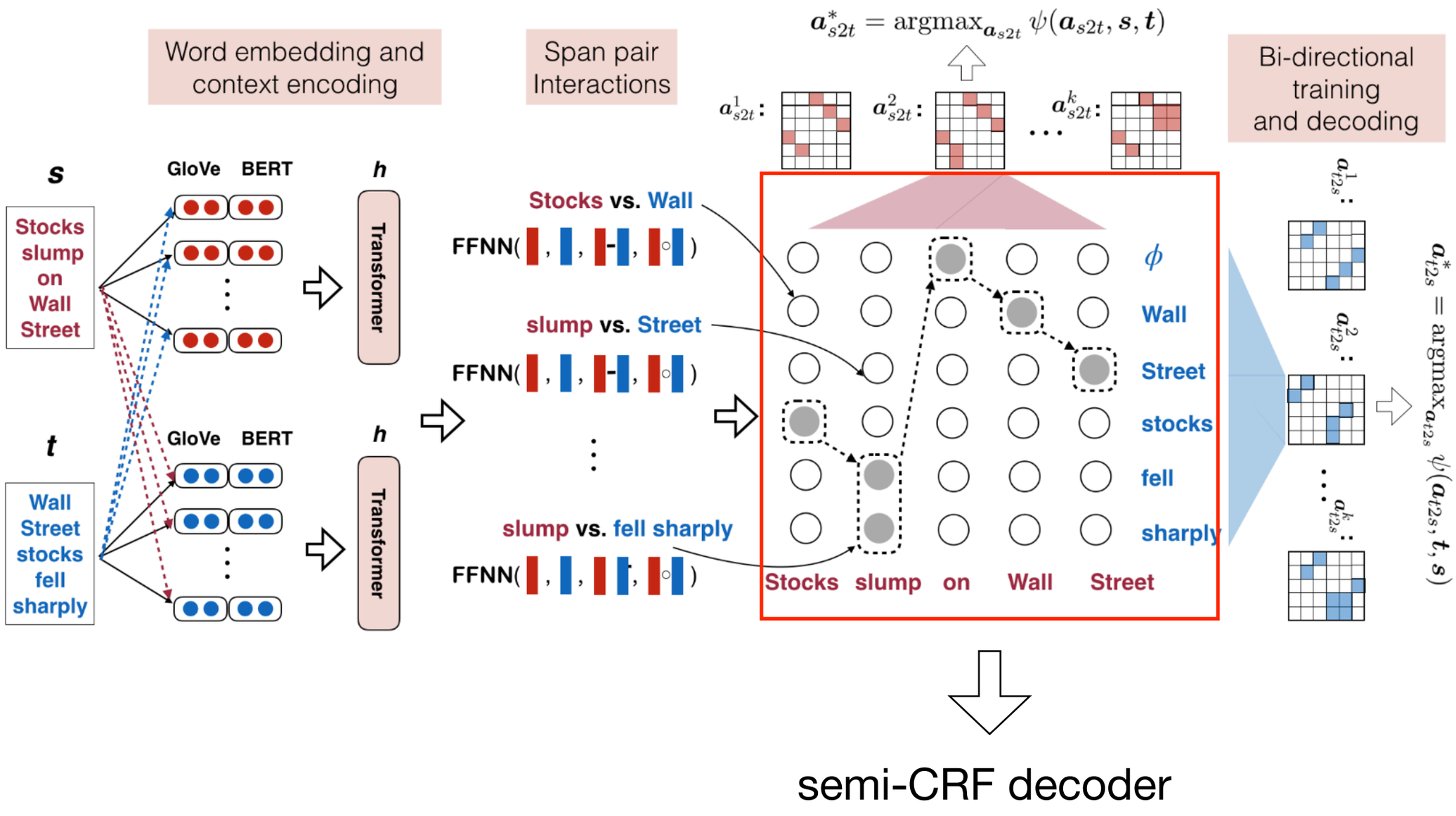
[1] **Wuwei Lan** and Wei Xu, Neural semi-Markov CRF for Monolingual Word Alignment. (in submission)

Neural Semi-CRF Word Aligner [1]



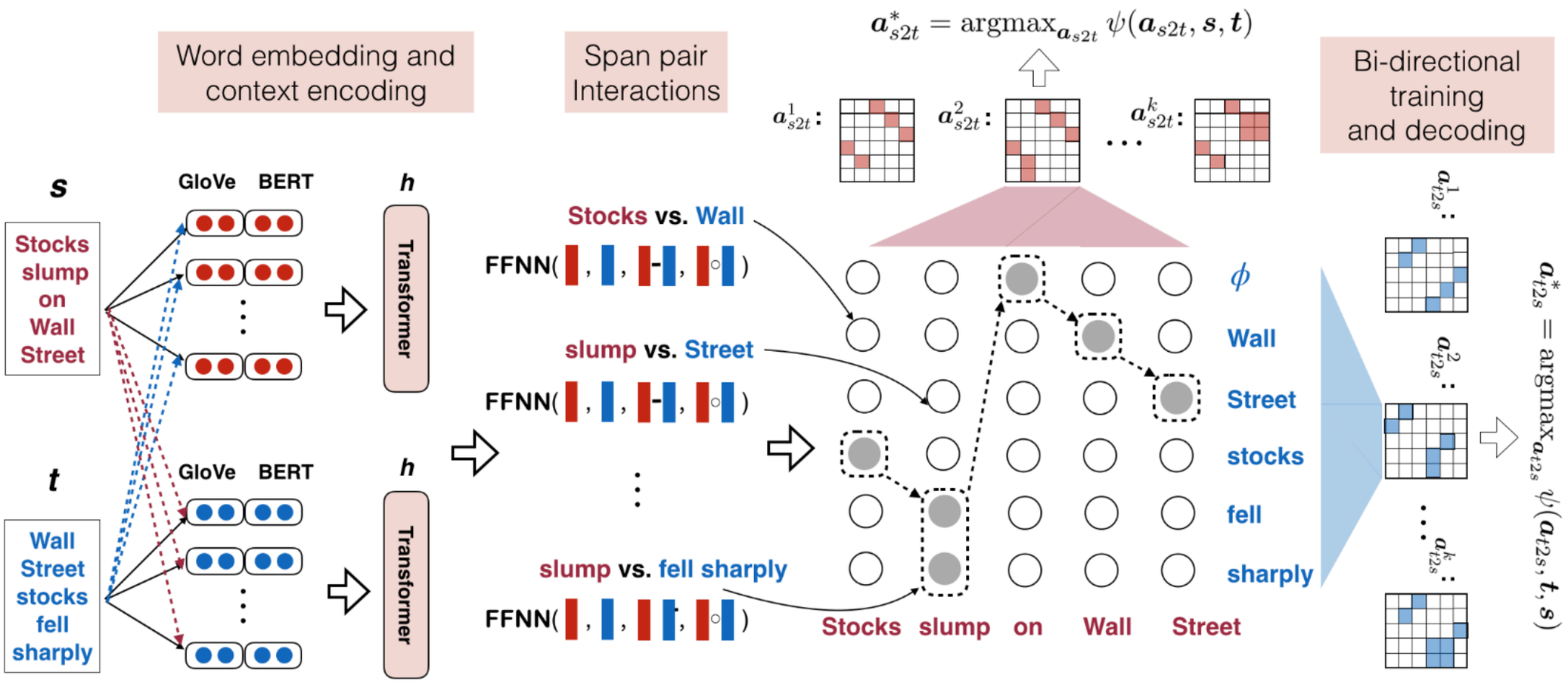
[1] **Wuwei Lan** and Wei Xu, Neural semi-Markov CRF for Monolingual Word Alignment. (in submission)

Neural Semi-CRF Word Aligner [1]



[1] **Wuwei Lan** and Wei Xu, Neural semi-Markov CRF for Monolingual Word Alignment. (in submission)

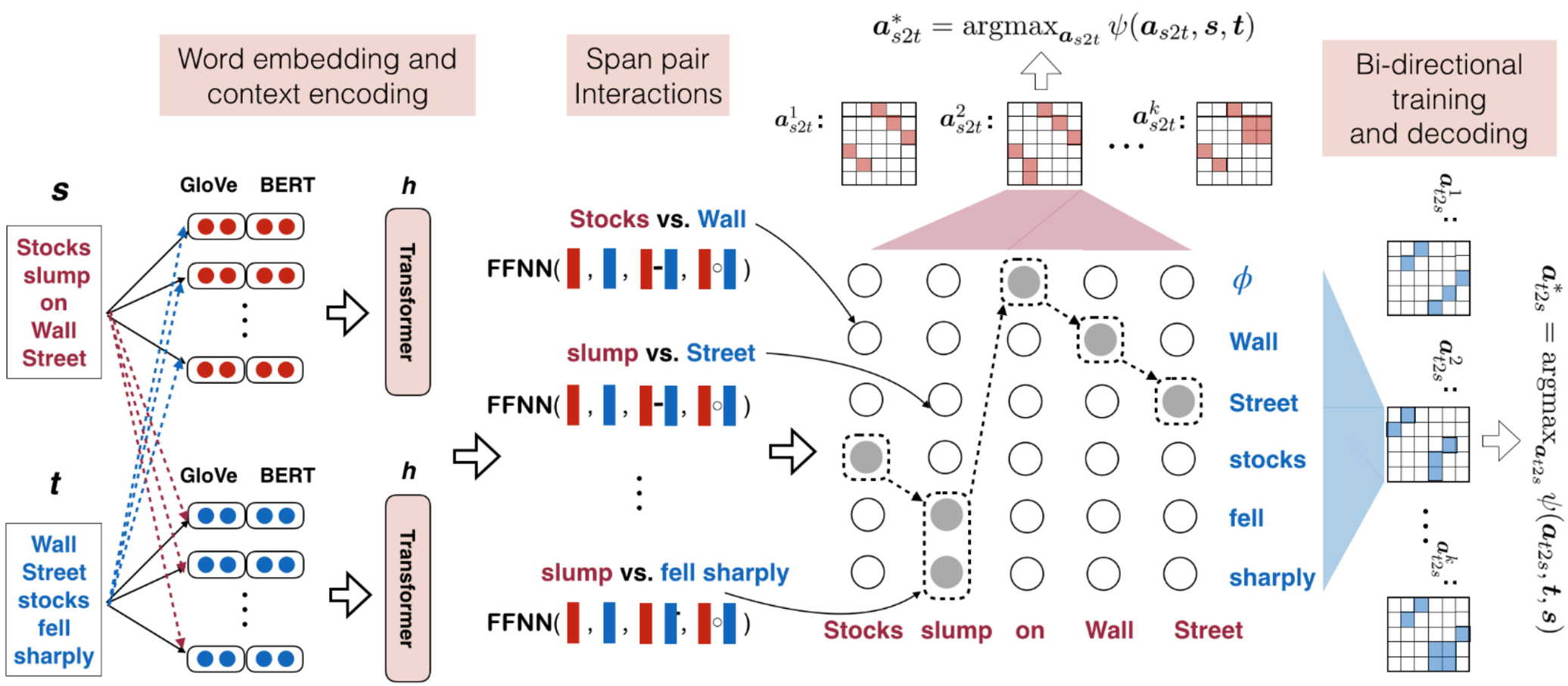
Neural Semi-CRF Word Aligner



- Learning: conditional log-likelihood $\sum_{(s,t,a^*) \in D} -\log p(a^*|s,t)$

- Inference: Viterbi semi-CRF decoder $\beta_i(a|s,t) = \max_{a'} \max_{d=1...L} [\beta_{i-d}(a'|s,t) + \upsilon(s_{[i-d+1:i]}, t'_a) + \tau(a', a)]$

Span Representation



$$\alpha_{i,t} = \frac{e^{\boldsymbol{w}\cdot \text{FF}_p(\boldsymbol{h_t})}}{\sum_{k=start(i)}^{end(i)} e^{\boldsymbol{w}\cdot \text{FF}_p(\boldsymbol{h_k})}}$$

$$\bar{\boldsymbol{g}}_i = \sum_{t=start(i)}^{end(i)} \alpha_{i,t} \cdot \boldsymbol{g_t}$$

Attention-based weighted average

$$\boldsymbol{w}_{1:T} = \text{BERT}(w_{1:T})$$

$$\boldsymbol{h}_{1:T} = \text{BiLSTM}(\boldsymbol{w}_{1:T})$$

$$\boldsymbol{g}_{i:j} = [\overrightarrow{\boldsymbol{h}}_j - \overrightarrow{\boldsymbol{h}}_{i-1}; \overleftarrow{\boldsymbol{h}}_i - \overleftarrow{\boldsymbol{h}}_{j+1}; \overrightarrow{\boldsymbol{h}}_{i-1}; \overleftarrow{\boldsymbol{h}}_{j+1}]$$

LSTM based relative representation

A Comprehensive Evaluation Benchmark

[MTReference]

NIST MT corpus

train/dev/test (2398/800/800)

In-domain evaluation

Out-of-domain evaluation

[arXiv]

paper revision

test only (200)

[Newsela]

text simplification

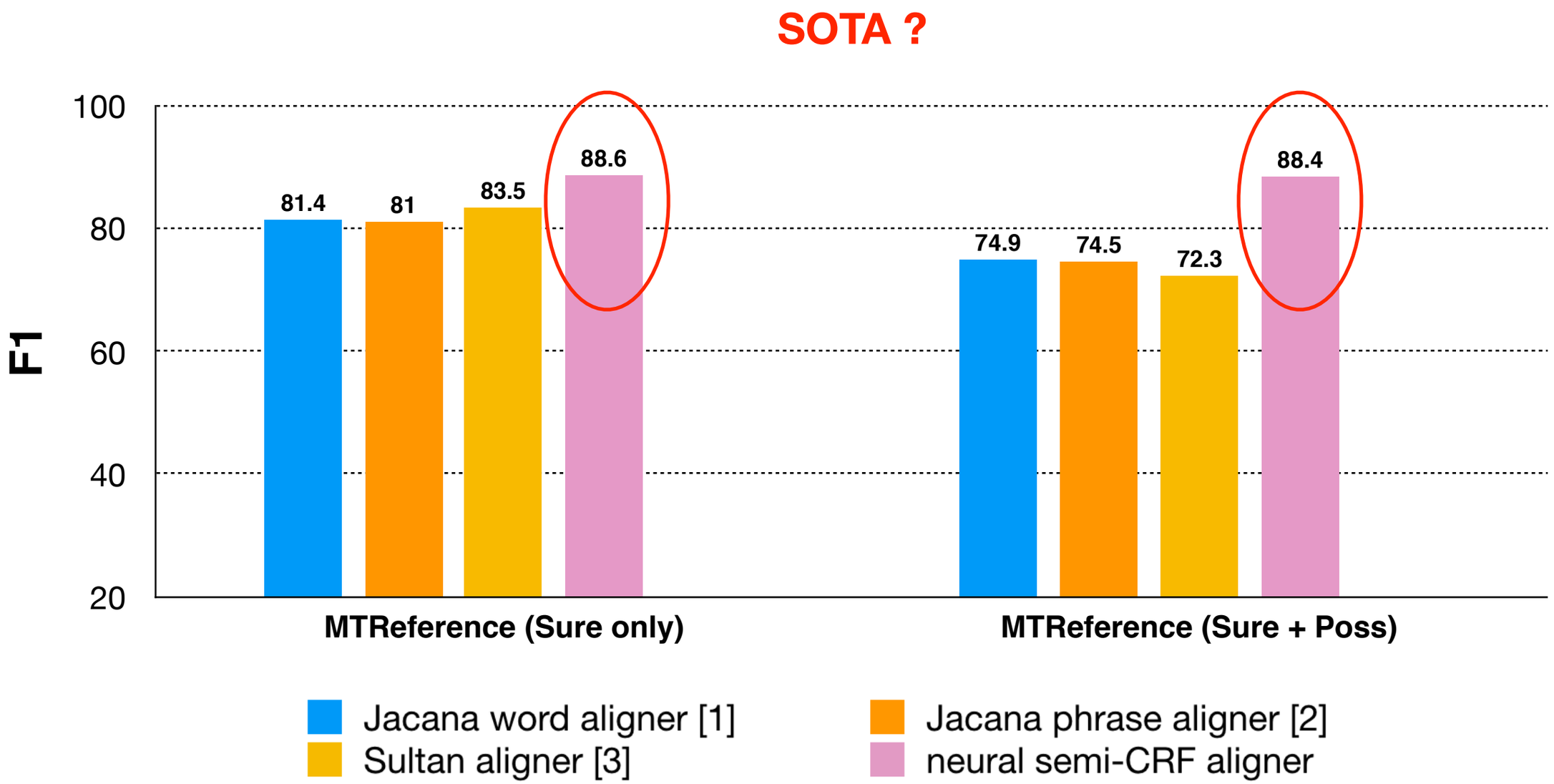
test only (500)

[Wikipedia]

edit history

test only (1052)

In-domain Evaluation



[1] Xuchen Yao, et al., A lightweight and high performance monolingual word aligner. (ACL 2013)

[2] Xuchen Yao, et al., Semi-markov phrase-based monolingual alignment. (EMNLP 2013)

[3] Md Arafat Sultan, et al., Back to basics for monolingual alignment: Exploiting word similarity and contextual evidence. (TACL 2014)

State-of-the-art Bilingual Word Aligner

Word alignment is formulated as QA and solved with BERT fine-tuning system. [1]

We used multilingual BERT for the cross-language span prediction

- Although it is made from monolingual texts, it works surprisingly well !!

[CLS] 言語 [SEP] language is a means of communication [SEP] ⇒ language
[CLS] は [SEP] language is a means of communication [SEP] ⇒ No_Answer
…

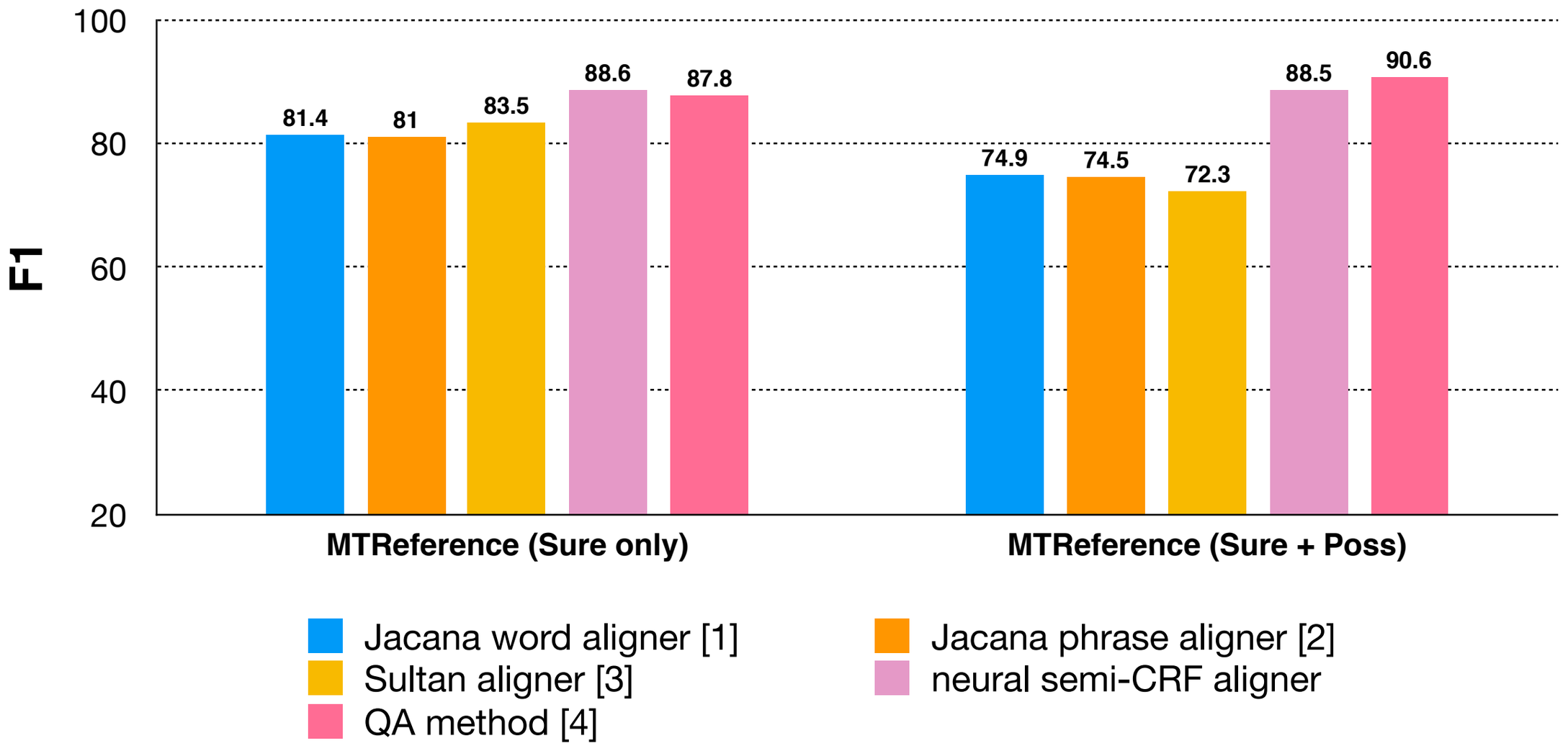
We added the question's context by enclosing the word with a boundary marker, pilcrow '¶' (paragraph mark)

- It works even better !!

[CLS] ¶ 言語 ¶ は コミュニケーション の 道具 で ある [SEP] language is a means of communication [SEP] ⇒ Language
[CLS] 言語 ¶ は ¶ コミュニケーション の 道具 で ある [SEP] language is a means of communication [SEP] ⇒ No_Answer
…

[1] Masaaki Nagata, Katsuki Chousa and Masaaki Nishino, A lightweight and high performance monolingual word aligner. (EMNLP 2020)

In-domain Evaluation



[1] Xuchen Yao, et al., A lightweight and high performance monolingual word aligner. (ACL 2013)

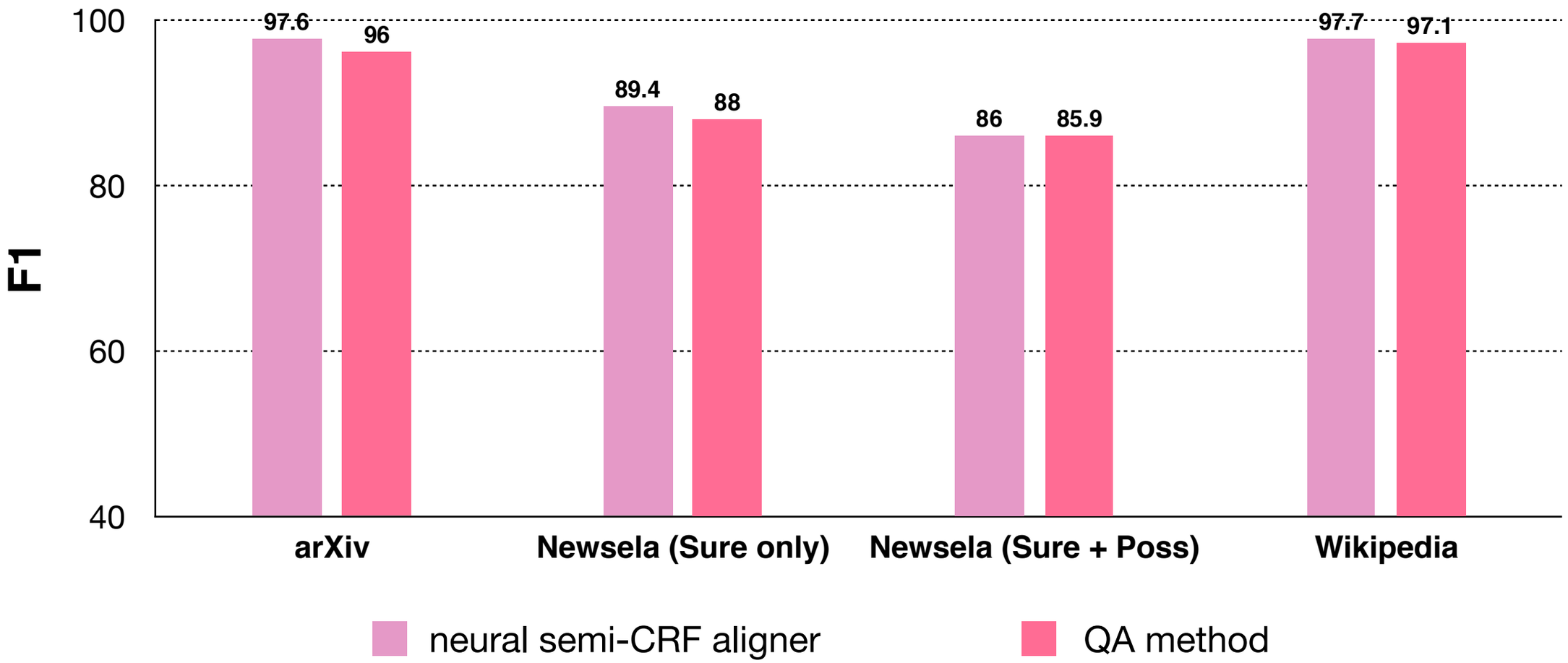
[2] Xuchen Yao, et al., Semi-markov phrase-based monolingual alignment. (EMNLP 2013)

[3] Md Arafat Sultan, et al., Back to basics for monolingual alignment: Exploiting word similarity and contextual evidence. (TACL 2014)

[4] Masaaki Nagata, Katsuki Chousa and Masaaki Nishino, A lightweight and high performance monolingual word aligner. (EMNLP 2020)

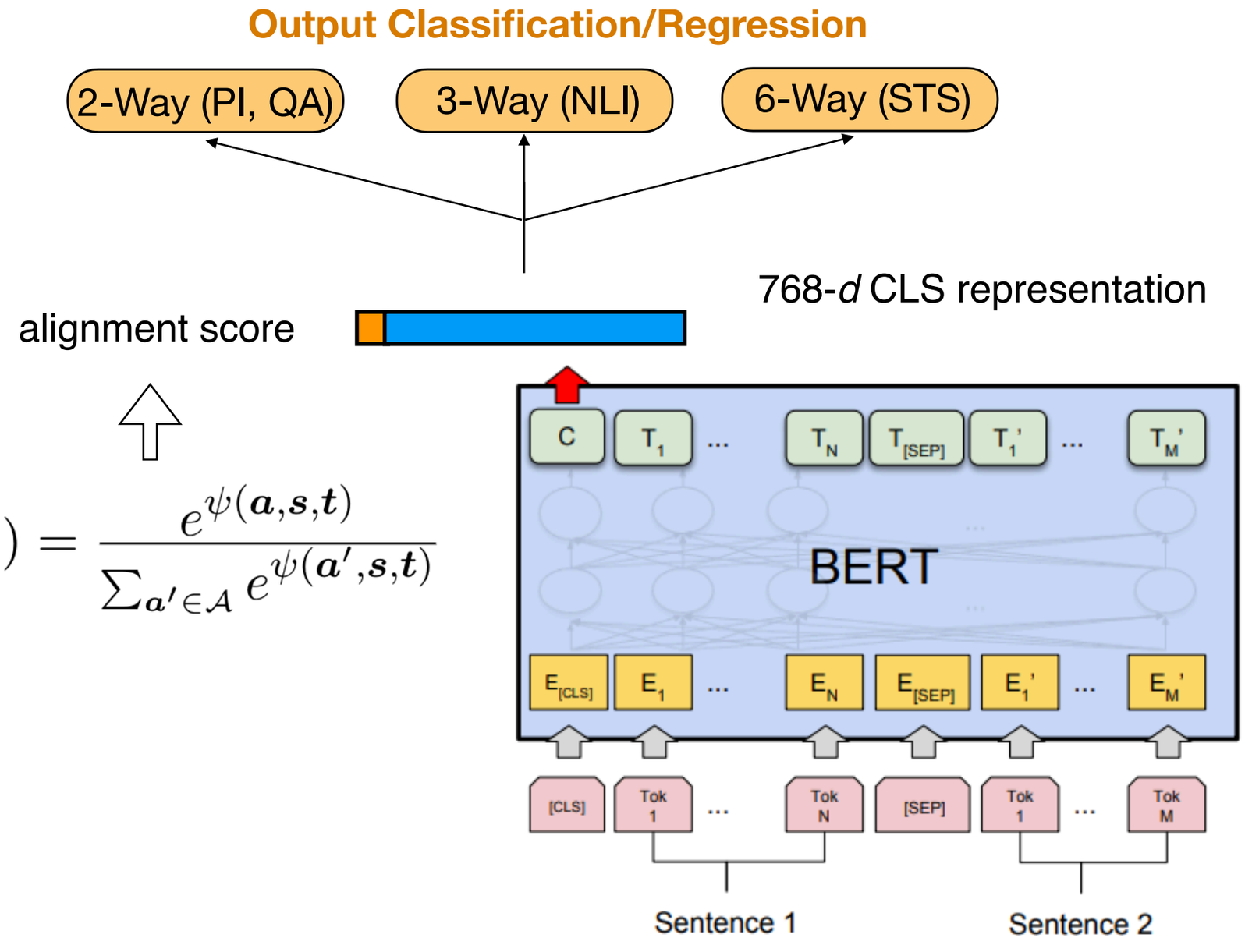
Out-of-domain Evaluation

(both models are trained with MTReference Sure + Poss setting)



Our neural semi-CRF has better generalization capability !

Downstream Applications



$$p(\boldsymbol{a}|\boldsymbol{s},\boldsymbol{t}) = \frac{e^{\psi(\boldsymbol{a},\boldsymbol{s},\boldsymbol{t})}}{\sum_{\boldsymbol{a}' \in \mathcal{A}} e^{\psi(\boldsymbol{a}',\boldsymbol{s},\boldsymbol{t})}}$$

(a) Sentence Pair Classification Tasks: MNLI, QQP, QNLI, STS-B, MRPC, RTE, SWAG

Downstream Applications

large-scale: 100k-500k

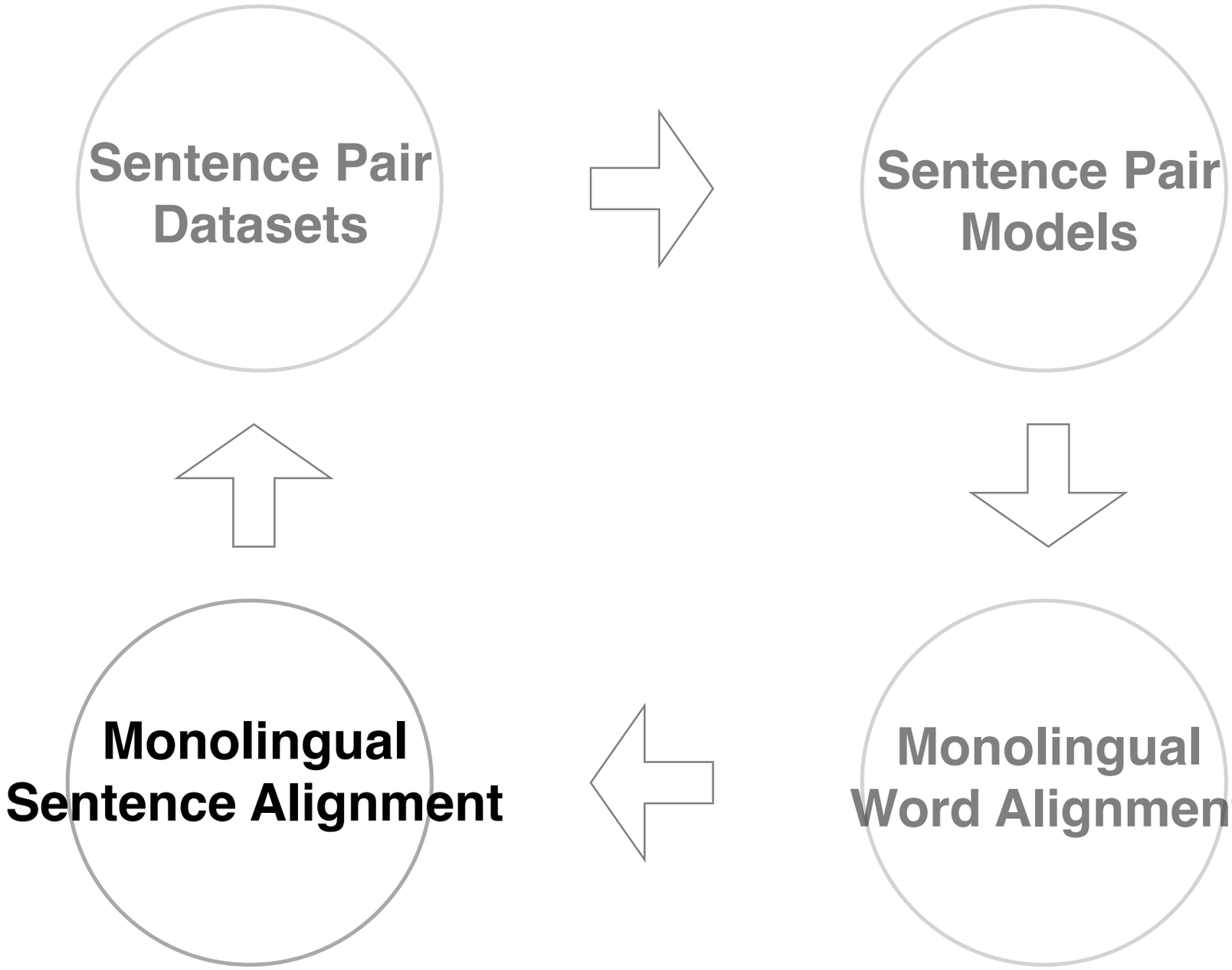
small-scale: 5k-50k

Models	SNLI	MNLI	QQP	QNLI	STS-B	MRPC	RTE	WikiQA	TrecQA	URL	PIT	STS14
	Acc	Acc	F1/Acc	Acc	r/ρ	F1/Acc	Acc	MAP/MRR	MAP/MRR	max_F_1	max_F_1	r
$BERT_{BASE}$	**90.8**	**84.4**/83.5	71.3/89.1	**91.1**	85.2/**83.7**	87.8/83.1	67.8	82.1/83.6	83.7/87.2	77.4	75.9	82.3
$BERT_{BASE}$ + Aligner	90.7	**84.4/83.6**	**71.6/89.2**	**91.1**	**85.3/83.7**	**88.7/84.5**	**68.2**	**82.9/84.3**	**86.5/89.5**	**77.9**	**76.3**	**83.0**

Table 4: Downstream evaluations on natural language inference (SNLI, MNLI, QNLI), paraphrase identification (QQP, MRPC, URL, PIT), question answering (WikiQA, TrecQA), recognizing textual entailment (RTE) and semantic textual similarity (STS-B, STS14).

- Performance improvement for PI and QA tasks, but not for NLI task [1].
- BERT-finetuning can potentially learn the word alignment.

Outline


Sentence Pair Datasets
Sentence Pair Models
Monolingual Word Alignment
Monolingual Sentence Alignment

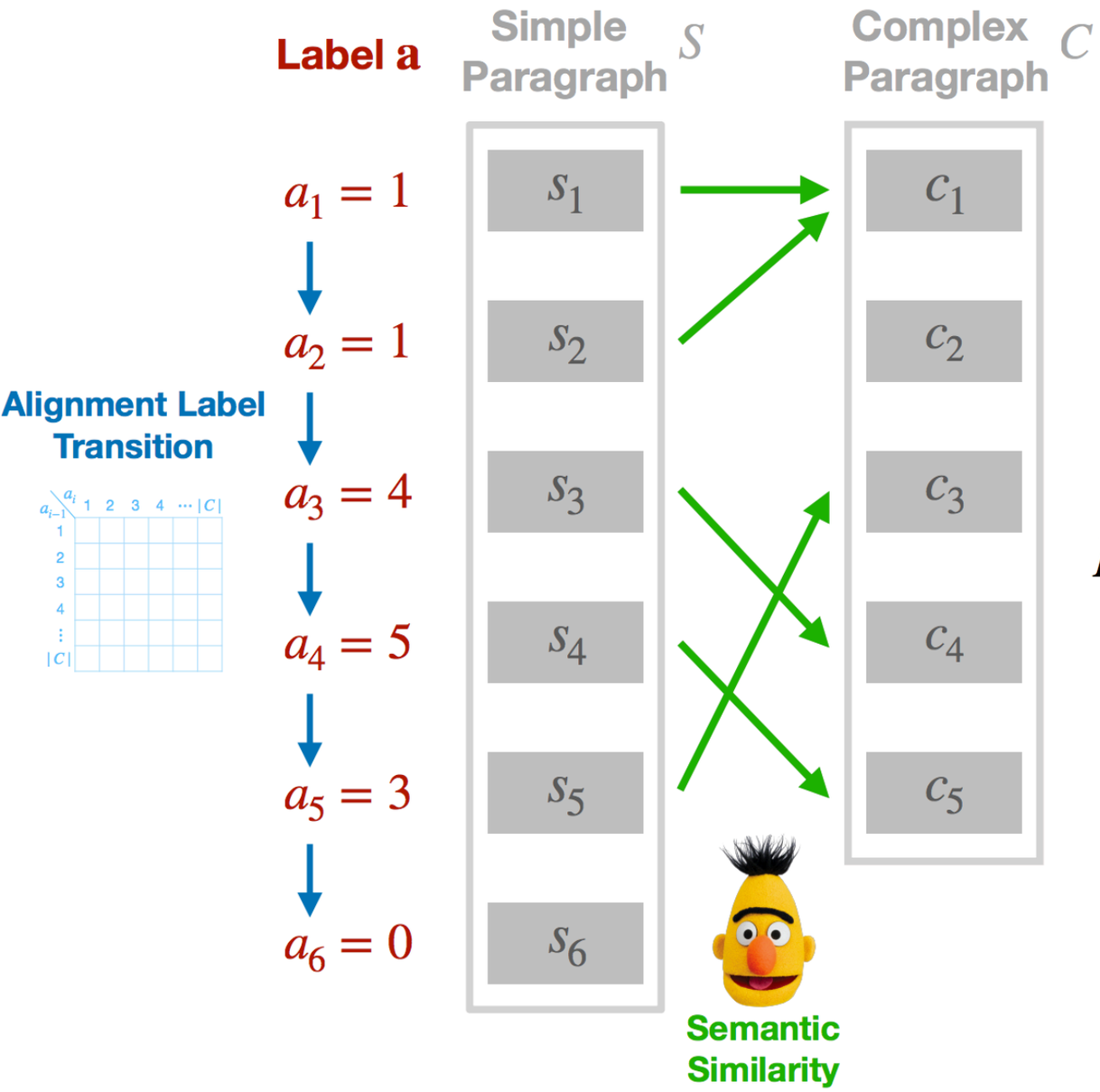
Word Alignment vs. Sentence Alignment

Simple article S	Label	Operation	Complex article C
s_1 The buildup of plaque can trap the bacteria that live in our mouths.	$a_1 = 1$	Splitting	c_1 The layers of plaque trap the bacteria that also live in our mouths and turns them into small fossils.
s_2 It turns them into tiny fossils.	$a_2 = 1$	Simplification	c_2 And when we die, these micro-fossils stay whole, even as most of the rest of us breaks down.
s_3 Even after death, these micro-fossils don't break down.	$a_3 = 2$	Deletion	c_3 Throughout most of the history of archaeology, researchers have seen the tooth plaque as waste.
s_4 They last for thousands of years.	$a_4 = 0$	Insertion	

The same problem with different granularity (word vs. sentence) !

Neural CRF sentence aligner [1]

(reuse the same neural CRF word aligner with some variations)



Semantic Similarity — Alignment Label Transition

$$\Psi(\mathbf{a}, S, C) = \sum_{i=0}^{|S|} sim(s_i, c_{a_i}) + T(a_i, a_{i-1})$$

$$P(\mathbf{a} \mid S, C) = \frac{\exp(\Psi(\mathbf{a}, S, C))}{\sum_{\mathbf{a} \in \mathbf{A}} \exp(\Psi(\mathbf{a}, S, C))}$$

Linear-chain CRF

all possible alignments (dynamic programming)

[1] Chao Jiang, Mounica Maddela, **Wuwei Lan**, Yang Zhong and Wei Xu, Neural CRF Model for Sentence Alignment in Text Simplification. (ACL 2020)

Sentence alignment evaluation on Newsela [1]

		aligned + partial vs. others		
		Precision	Recall	F1
Greedy	JaccardAlign (Xu et al., 2015)	98.66	67.58	80.22
Dynamic Programming	MASSAlign (Paetzold et al., 2017)	95.49	82.27	88.39
Greedy	CATS (Štajner et al., 2018)	88.56	91.31	89.92
Threshold	$\text{BERT}_{finetune}$	94.99	89.62	92.22
Threshold	$\text{BERT}_{finetune}$ + paragraph alignment	98.05	88.63	93.10
CRF	Ours CRF aligner	97.86	91.31	95.59

+5.7

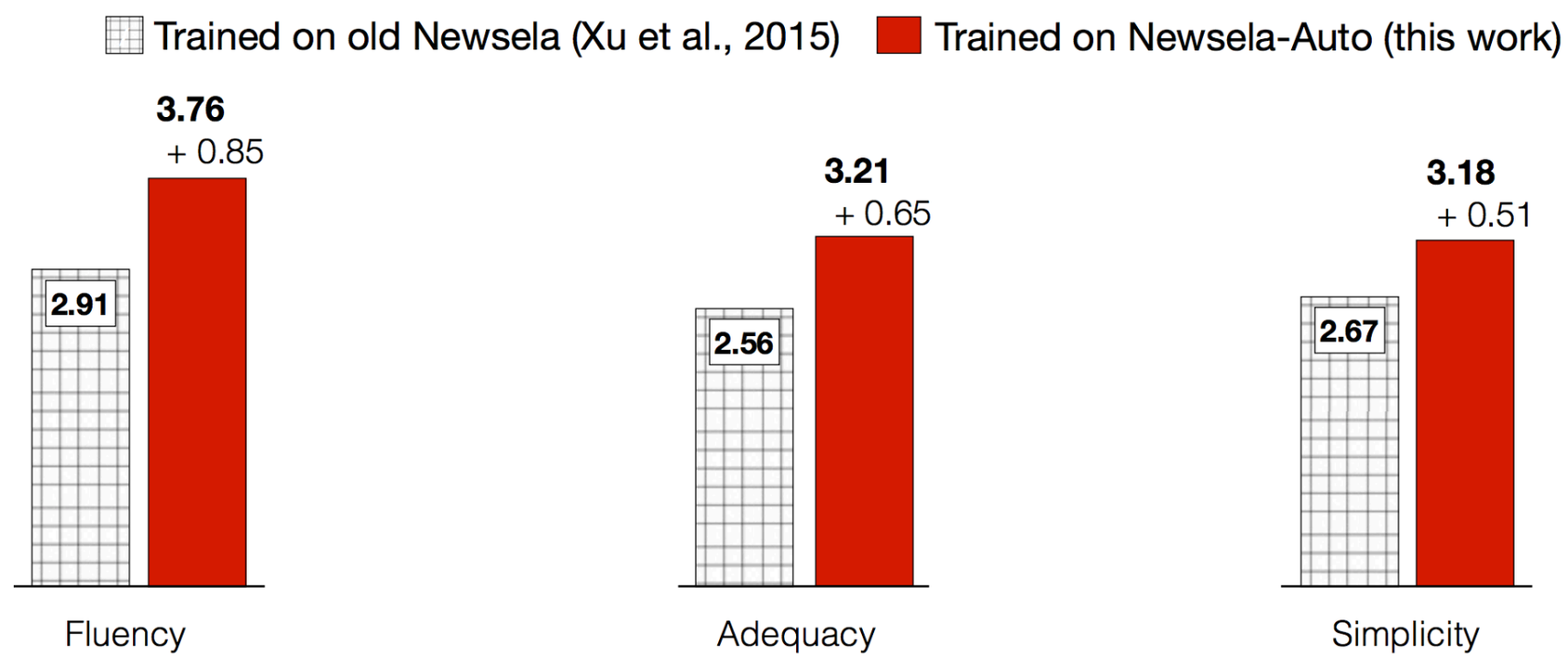
Our neural CRF aligner has the SOTA performance !

[1] Chao Jiang, Mounica Maddela, **Wuwei Lan**, Yang Zhong and Wei Xu, Neural CRF Model for Sentence Alignment in Text Simplification. (ACL 2020)

Towards large-scale sentence pair dataset

- Newsela-Auto (666k complex-simple sentence pairs)
- Wiki-Auto (468k complex-simple sentence pairs)

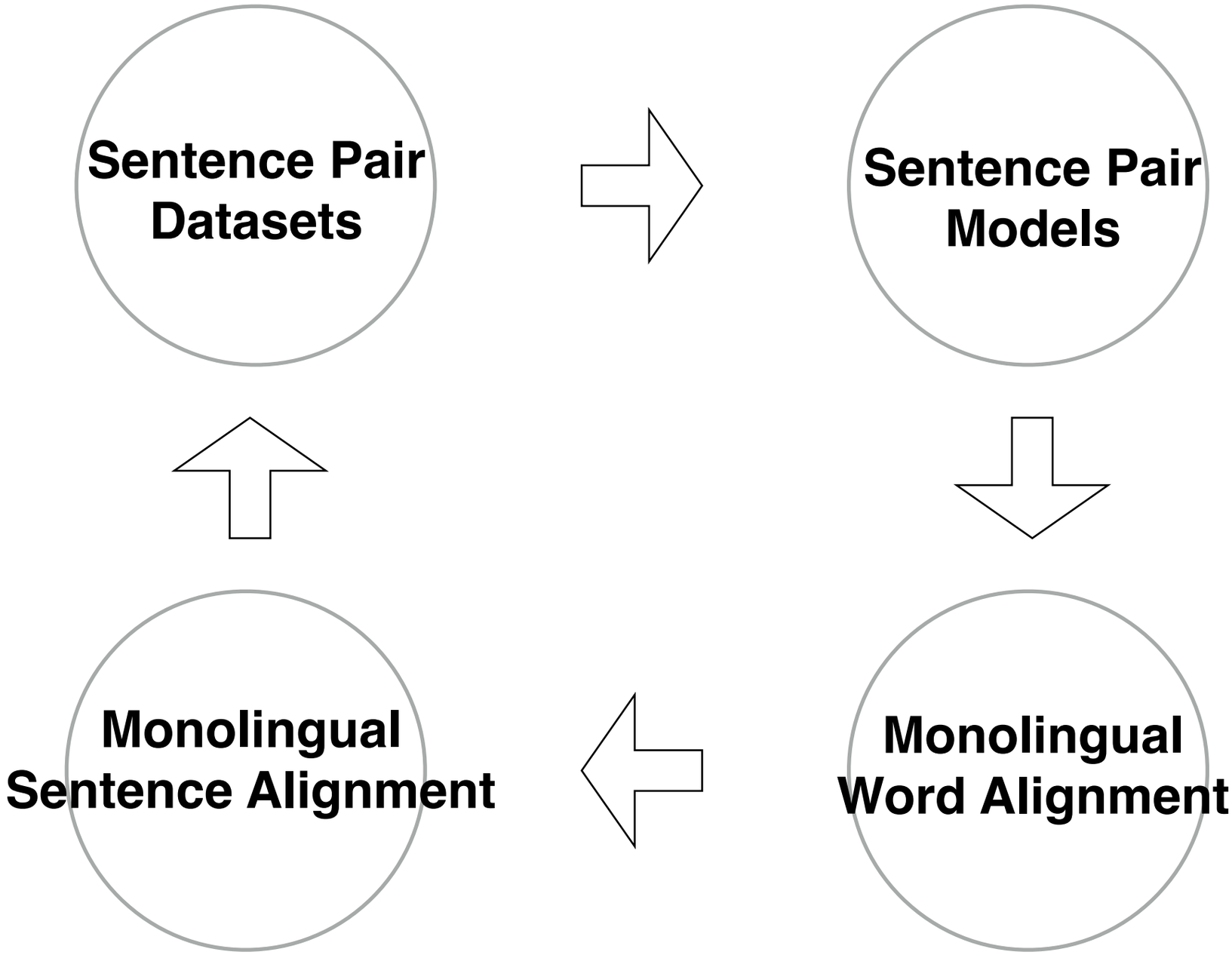
Human evaluation for text simplification system:



Transformer$_{BERT}$ model

(In 5-point Likert scale)

Summary


Sentence Pair Datasets
Sentence Pair Models
Monolingual Word Alignment
Monolingual Sentence Alignment


Q
&
A