

Controllable Text Simplification with Explicit Paraphrasing

Mounica Maddela, Fernando Alva Manchego and Wei Xu



Text-to-text Generation

Input sentence:

According to Ledford, Northrop executives said they would build substantial parts of the bomber in Palmdale, creating about 1,500 jobs.



Simplification:

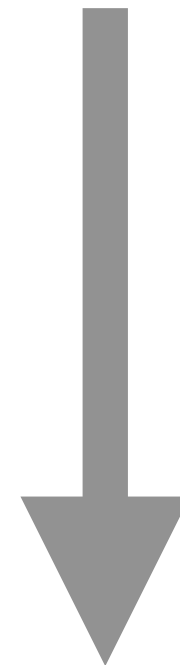
According to Ledford, Northrop said they would build most of the bomber parts in Palmdale .
It would create 1,500 jobs .

Text-to-text Generation

Such as text simplification, often requires complex rewriting with 3 operations.

Input sentence:

According to Ledford, Northrop **executives** said they would build **substantial parts of the bomber** in Palmdale, **creating about** 1,500 jobs.



Simplification:

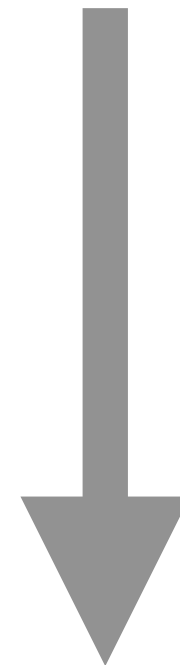
According to Ledford, Northrop said they would build **most of the bomber parts** in Palmdale .
It would **create** 1,500 jobs .

Text-to-text Generation

Such as text simplification, often requires complex rewriting with 3 operations.

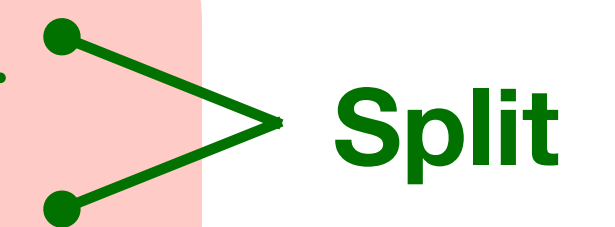
Input sentence:

According to Ledford, Northrop **executives** said they would build **substantial parts of the bomber** in Palmdale, **creating about** 1,500 jobs.



Simplification:

According to Ledford, Northrop said they would build **most of the bomber parts** in Palmdale .
It would **create** 1,500 jobs .



Text-to-text Generation

Such as text simplification, often requires complex rewriting with 3 operations.

Input sentence:

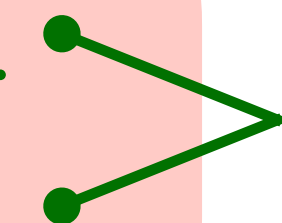
Delete

According to Ledford, Northrop ~~executives~~ said they would build **substantial parts of the bomber** in Palmdale, ~~creating about~~ 1,500 jobs.



Simplification:

According to Ledford, Northrop said they would build **most of the bomber parts** in Palmdale .
It would **create** 1,500 jobs .

 Split

Text-to-text Generation

Such as text simplification, often requires complex rewriting with 3 operations.

Input sentence:

Delete

According to Ledford, Northrop ~~executives~~ said they would build **substantial parts of the bomber** in Palmdale, **creating** ~~about~~ 1,500 jobs.

Simplification:

According to Ledford, Northrop said they would build **most of the bomber parts** in Palmdale .

It would **create** 1,500 jobs .

Paraphrase

Split

Automatic Text Simplification

But, SOTA neural generation models perform mostly deletion.

Input sentence:

According to Ledford, Northrop executives said they would build substantial parts of the bomber in Palmdale, creating about 1,500 jobs.

Generated output:

Programmer-interpreter
(Dong et al., 2019)

ledford **is a big group** of bomber in palmdale.

Rerank
(Kriz et al., 2019)

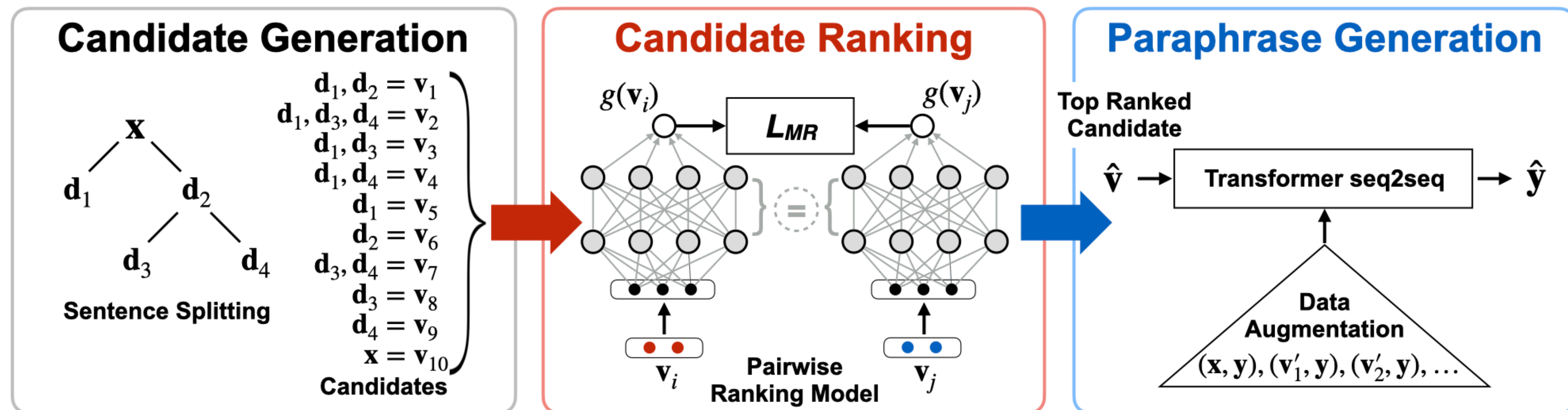
ledford **is** northrop.

Reinforcement Learning
(Zhang & Lapata, 2017)

, said they would build **palmdale** parts of **the substantial in creating**.

Our Work - Controllable Text Generation

- **Control over 3 edit operations** - deletion, splitting and paraphrasing.
- Incorporate linguistic rules with neural generation models.
- New setup to evaluate generation models's capability over these edit operations.

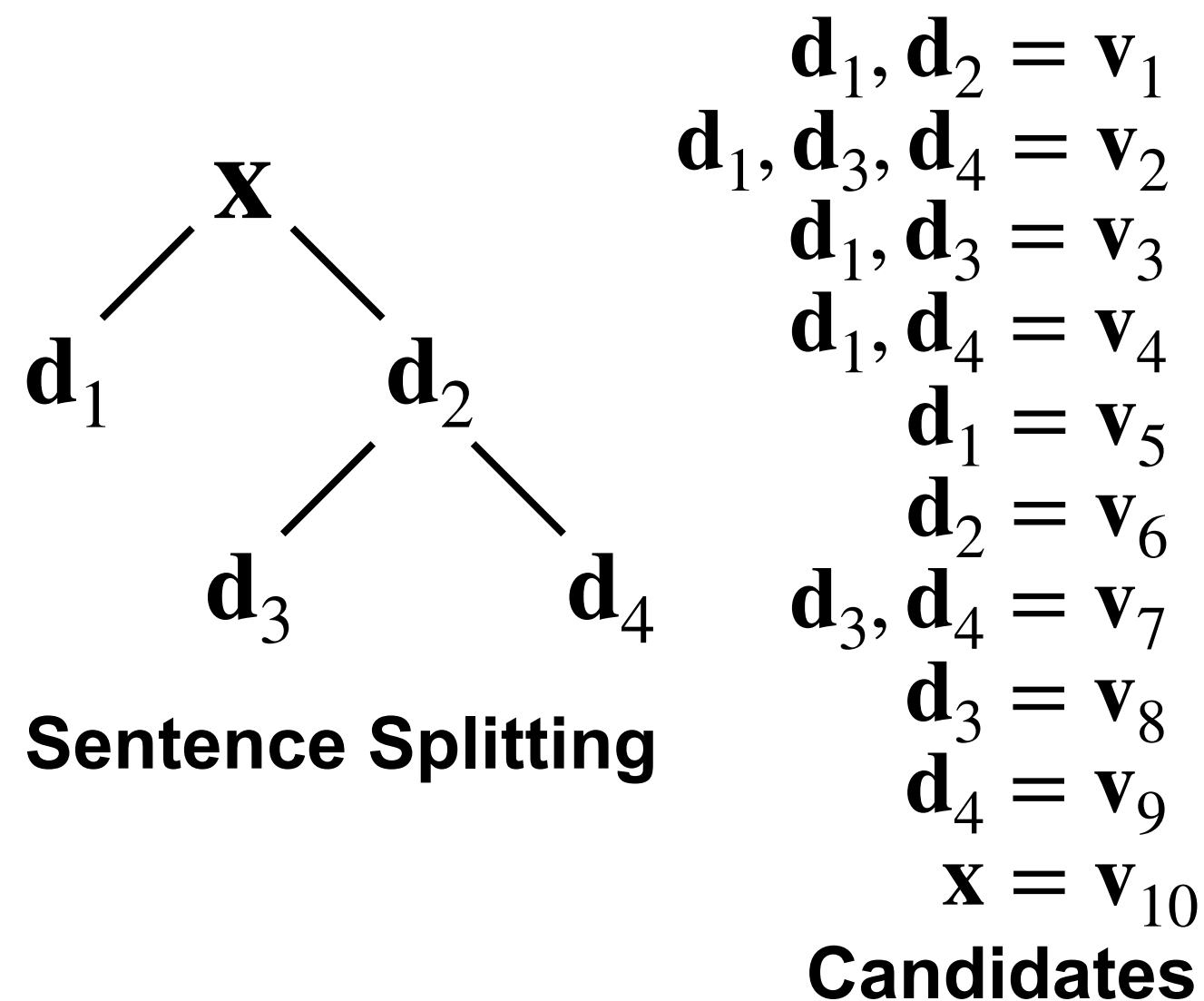


Step 1 —

We use a rule-based method (Niklaus et al., 2019) + a seq2seq model for splitting and deletion.

- 35 hand-crafted grammar rules for English based on Stanford's parser (Socher et al., 2013).
- successfully split 92% of sentences with ≥ 20 words and make only 6.8% errors*.

Candidate Generation

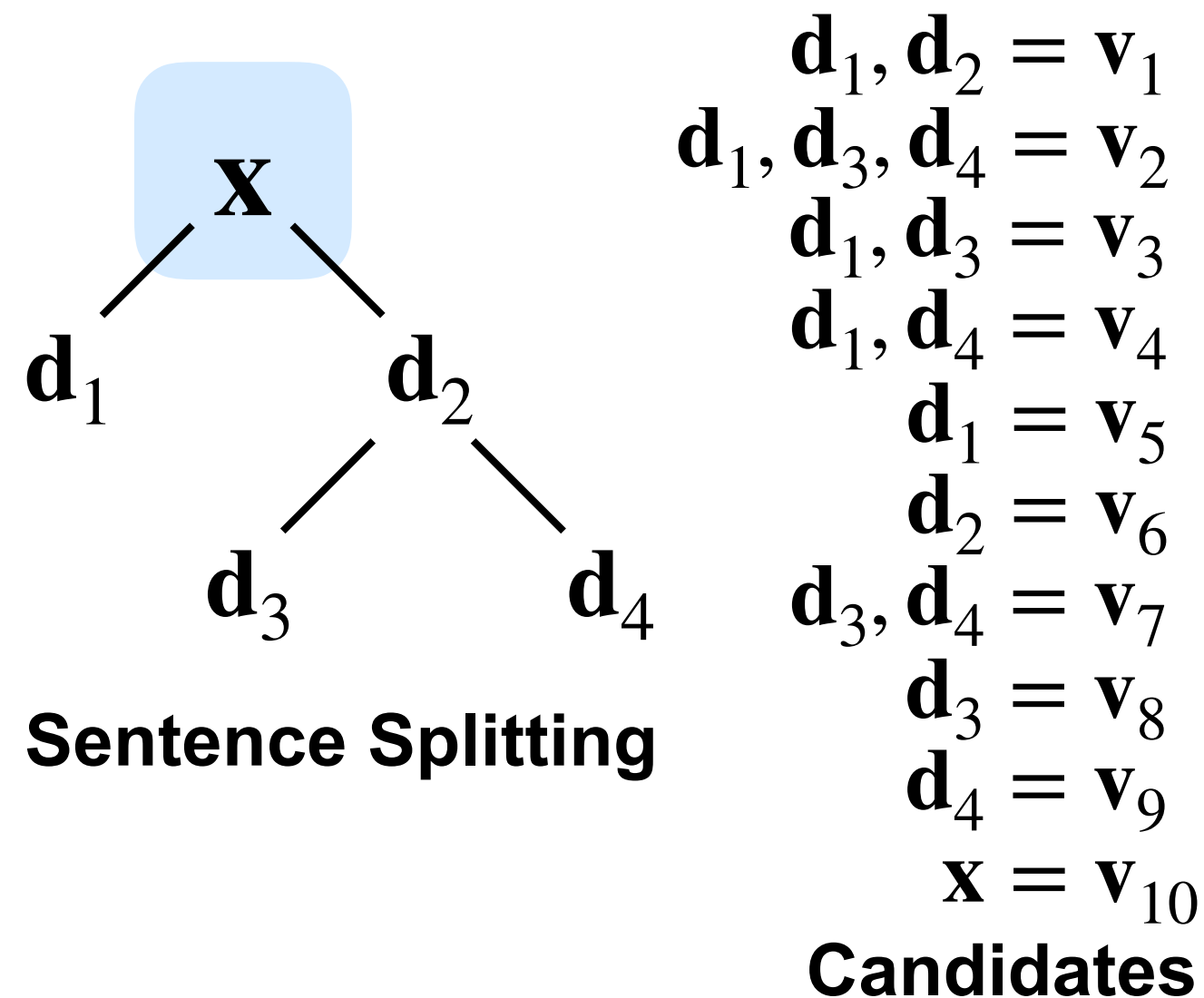


* Based on manual inspection on 100 random sentences.

Step 1 —

We use a rule-based method (Niklaus et al., 2019) + a seq2seq model for splitting and deletion.

Candidate Generation



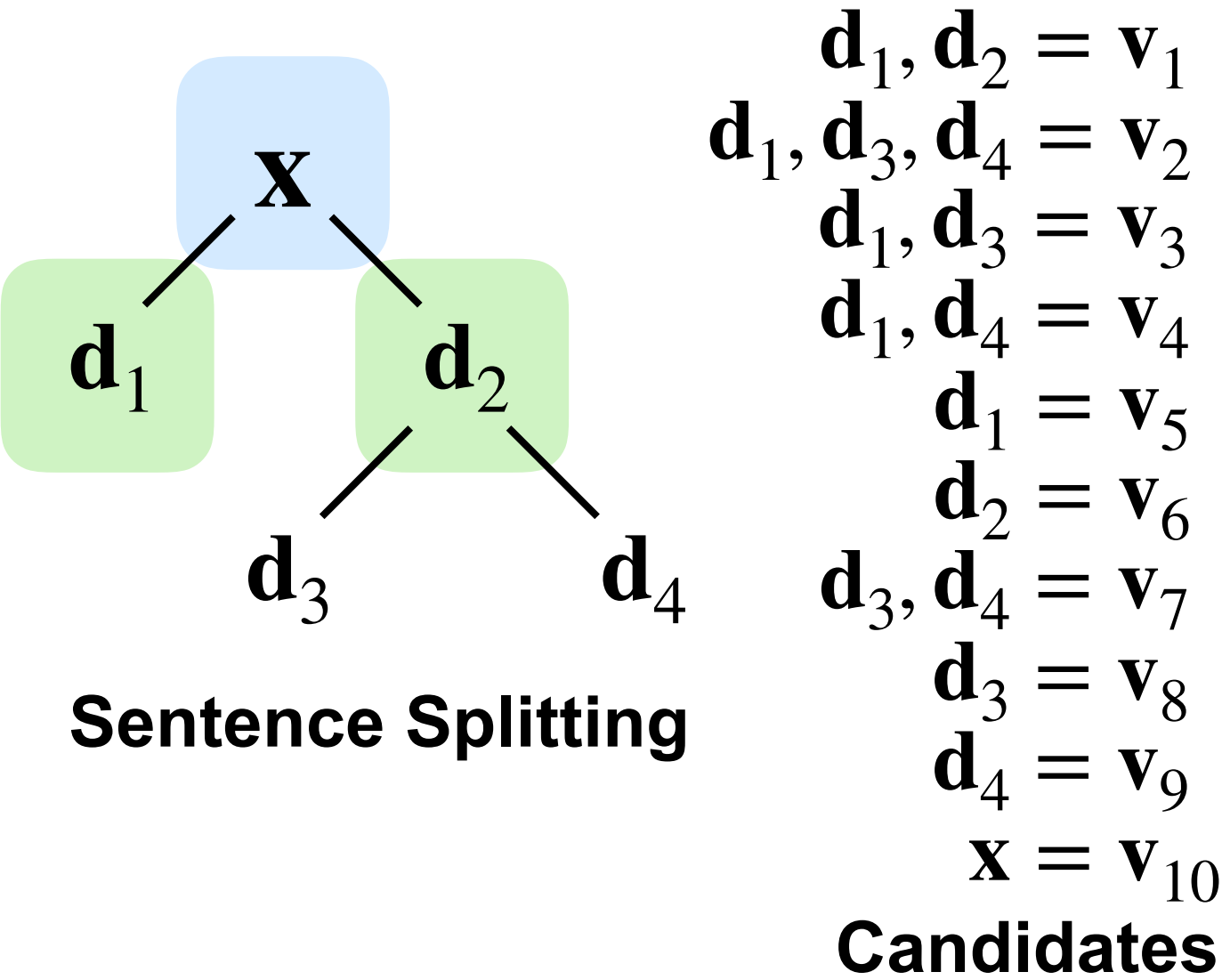
Input sentence:

The exhibition, which opened Oct. 8 and runs through Jan. 3, features 27 self-portraits.

Step 1 —

We use a rule-based method (Niklaus et al., 2019) + a seq2seq model for splitting and deletion.

Candidate Generation



Input sentence:

The exhibition, which opened Oct. 8 and runs through Jan. 3, features 27 self-portraits.

Split sentences:

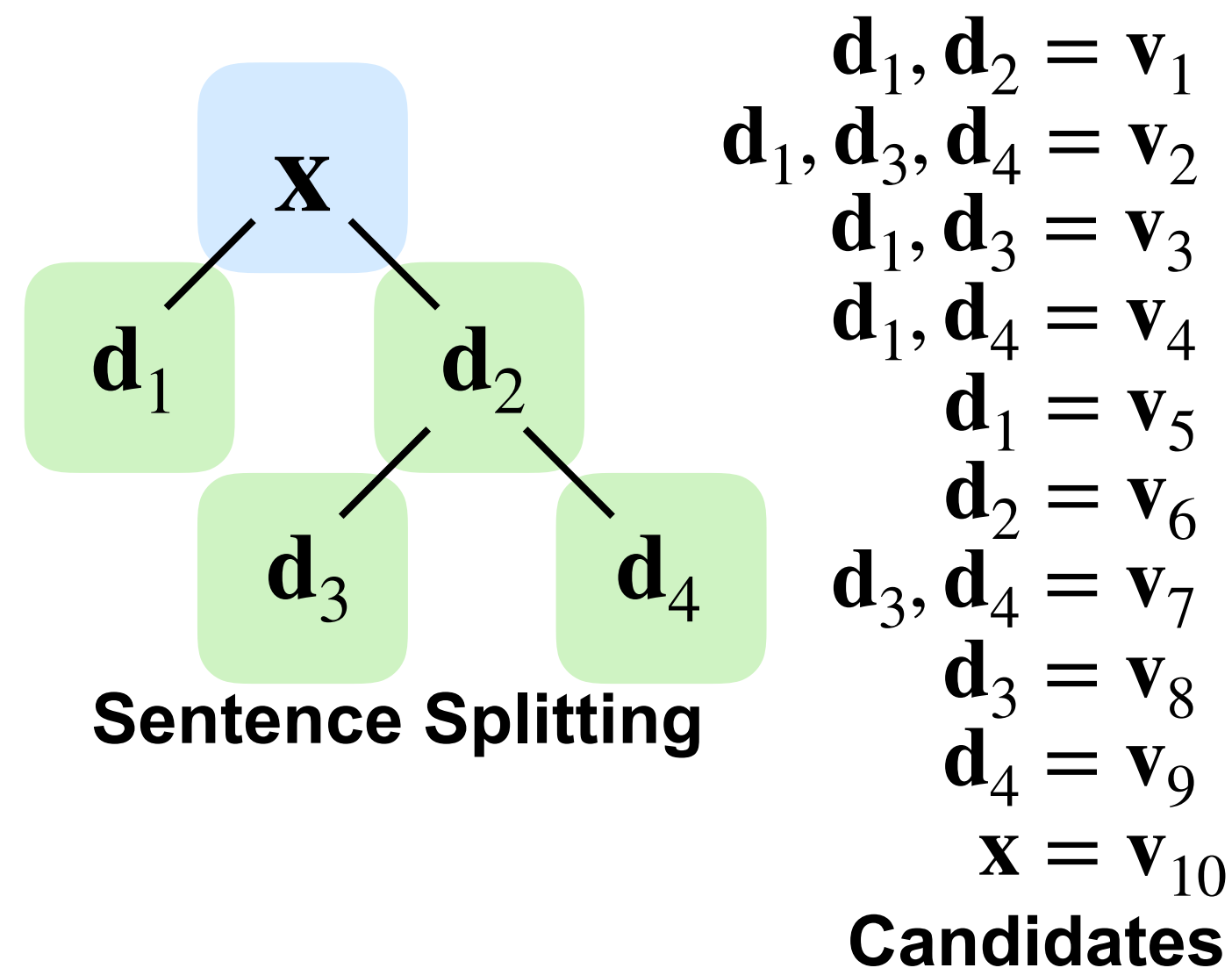
The exhibition features 27 portraits.

The exhibition opened Oct. 8 and runs through Jan. 3.

Step 1 —

We use a rule-based method (Niklaus et al., 2019) + a seq2seq model for splitting and deletion.

Candidate Generation



Input sentence:

The exhibition, which opened Oct. 8 and runs through Jan. 3, features 27 self-portraits.

Split sentences:

The exhibition features 27 portraits.

The exhibition opened Oct. 8 and runs through Jan. 3.

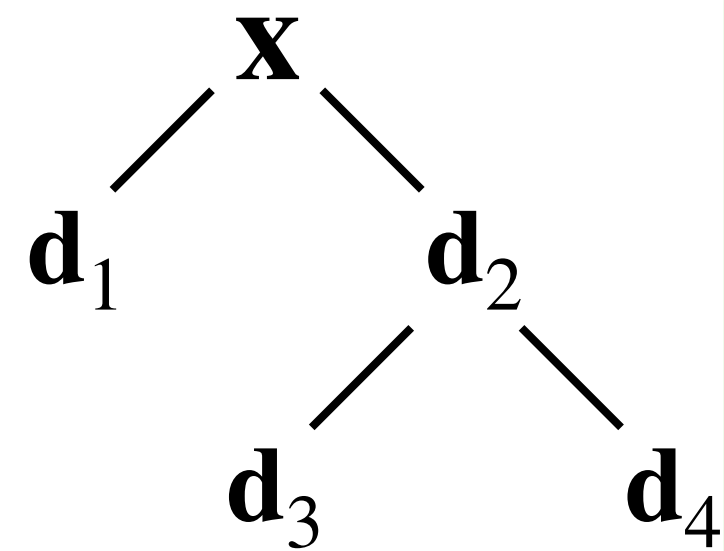
The exhibition opened Oct. 8.

The exhibition runs through Jan. 3.

Step 1 —

We use a rule-based method (Niklaus et al., 2019) + a seq2seq model for splitting and deletion.

Candidate Generation



Sentence Splitting

$d_1, d_2 = v_1$
 $d_1, d_3, d_4 = v_2$
 $d_1, d_3 = v_3$
 $d_1, d_4 = v_4$
 $d_1 = v_5$
 $d_2 = v_6$
 $d_3, d_4 = v_7$
 $d_3 = v_8$
 $d_4 = v_9$
 $x = v_{10}$

Candidates

Candidates:

The exhibition features 27 portraits. The exhibition opened Oct. 8 and runs through Jan. 3.

The exhibition opened Oct. 8 and runs through Jan. 3.

The exhibition features 27 portraits.

The exhibition opened Oct. 8. The exhibition runs through Jan. 3.

The exhibition features 27 portraits. The exhibition opened Oct. 8.

... (and more)

Step 2 —

Then, we rank all the intermediate outputs (after splitting & deletion).

Candidates:

The exhibition features 27 portraits. The exhibition opened Oct. 8 and runs through Jan. 3.

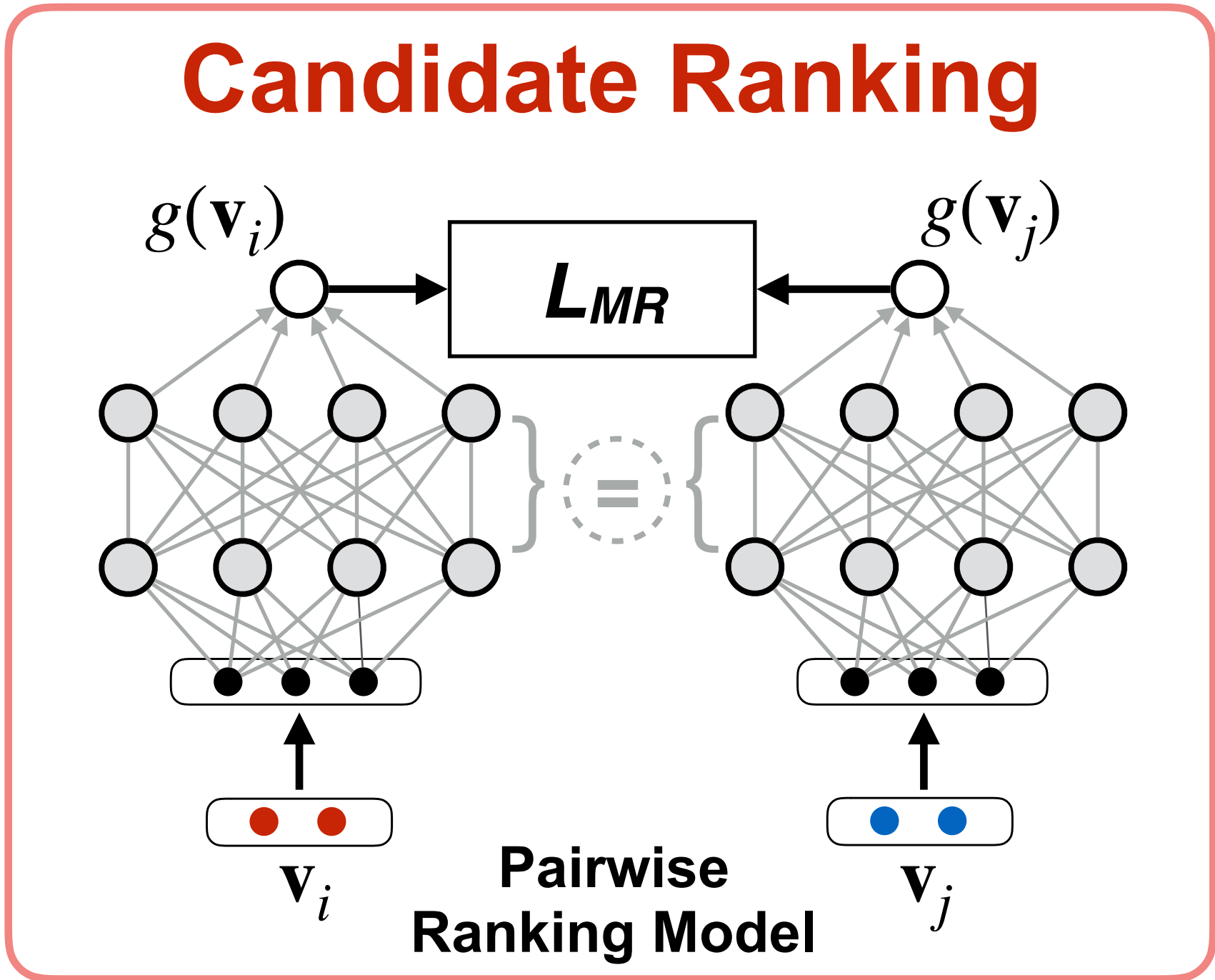
The exhibition opened Oct. 8 and runs through Jan. 3.

The exhibition features 27 portraits.

The exhibition opened Oct. 8. The exhibition runs through Jan. 3.

The exhibition features 27 portraits. The exhibition opened Oct. 8.

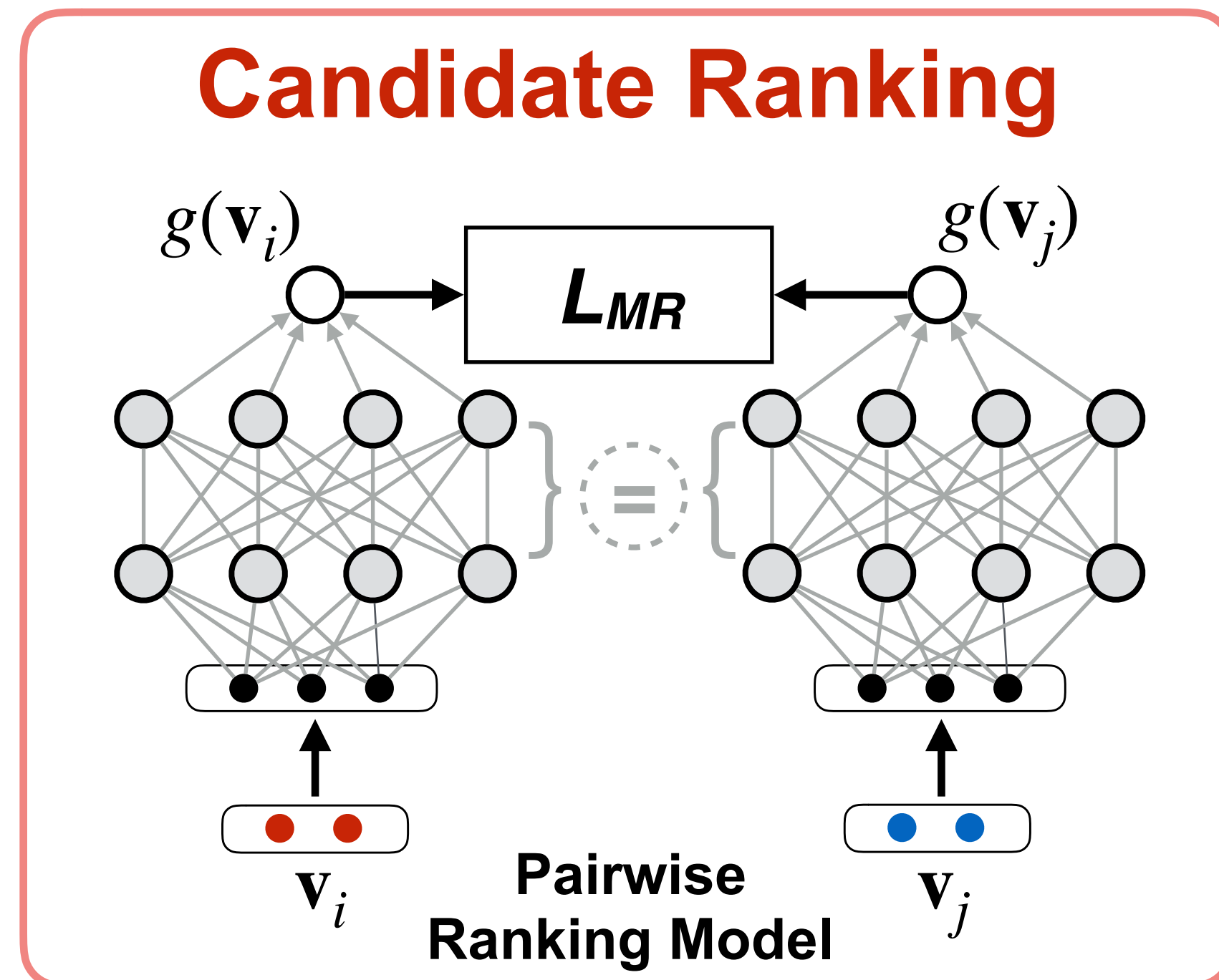
... (and more)



Features: number of words in \mathbf{v}_i and \mathbf{x} , compression ratio of \mathbf{v}_i with respect to \mathbf{x} , Jaccard similarity between \mathbf{v}_i and \mathbf{x} , the rules applied on \mathbf{x} to obtain \mathbf{v}_i , and the number of rule applications.

Step 2 —

Then, we rank all the intermediate outputs (after splitting & deletion).



Candidates:

The exhibition opened Oct. 8. The exhibition runs through Jan. 3.

The exhibition opened Oct. 8 and runs through Jan. 3.

The exhibition features 27 portraits. The exhibition opened Oct. 8.

The exhibition features 27 portraits. The exhibition opened Oct. 8 and runs through Jan. 3.

The exhibition features 27 portraits.

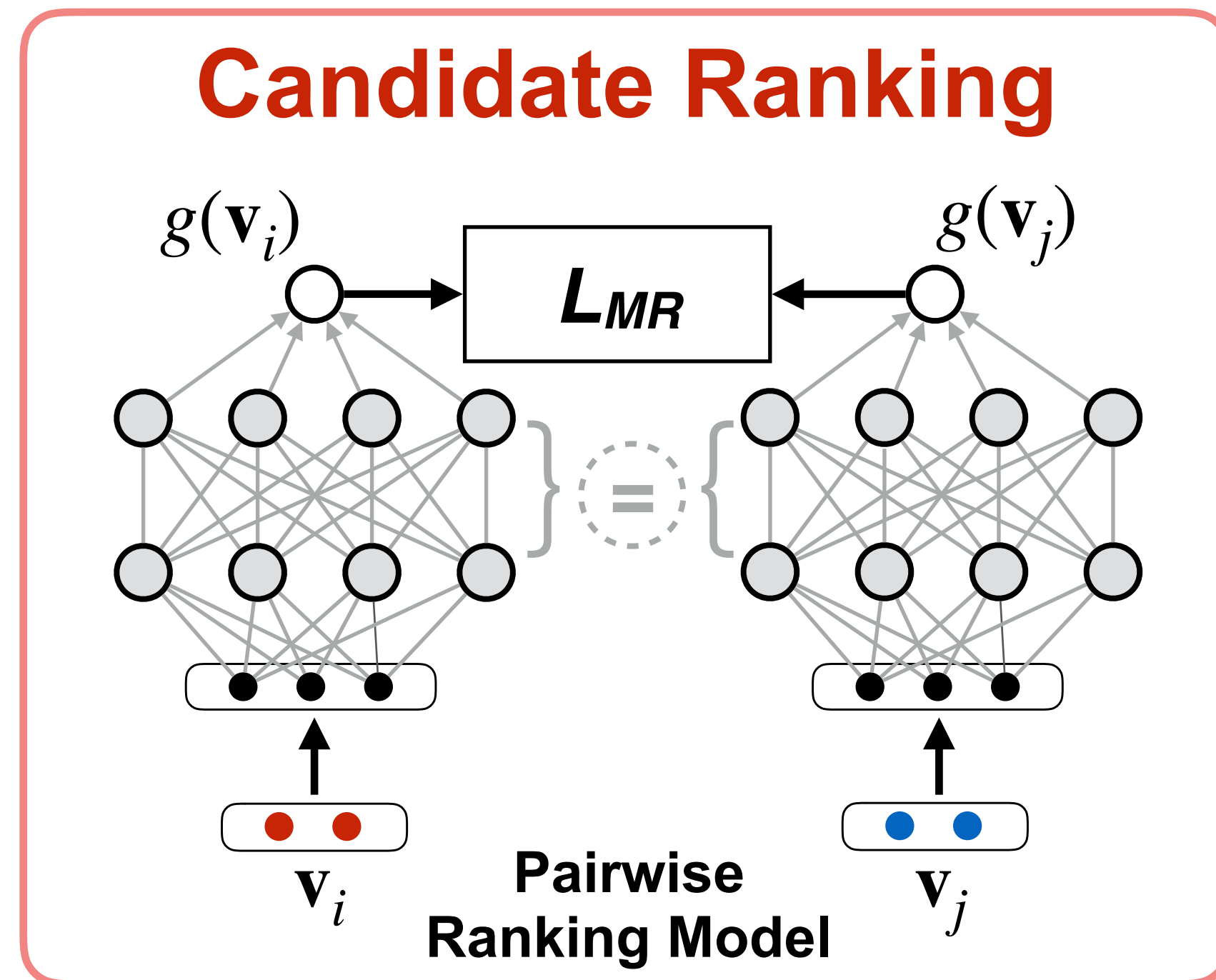
... (and more)

Human reference:

The show started Oct. 8. It ends. Jan 3.

Step 2 —

During training, we access each candidate using BERTScore (Zhang et al. 2019) with length penalty.



Gold Scoring function:

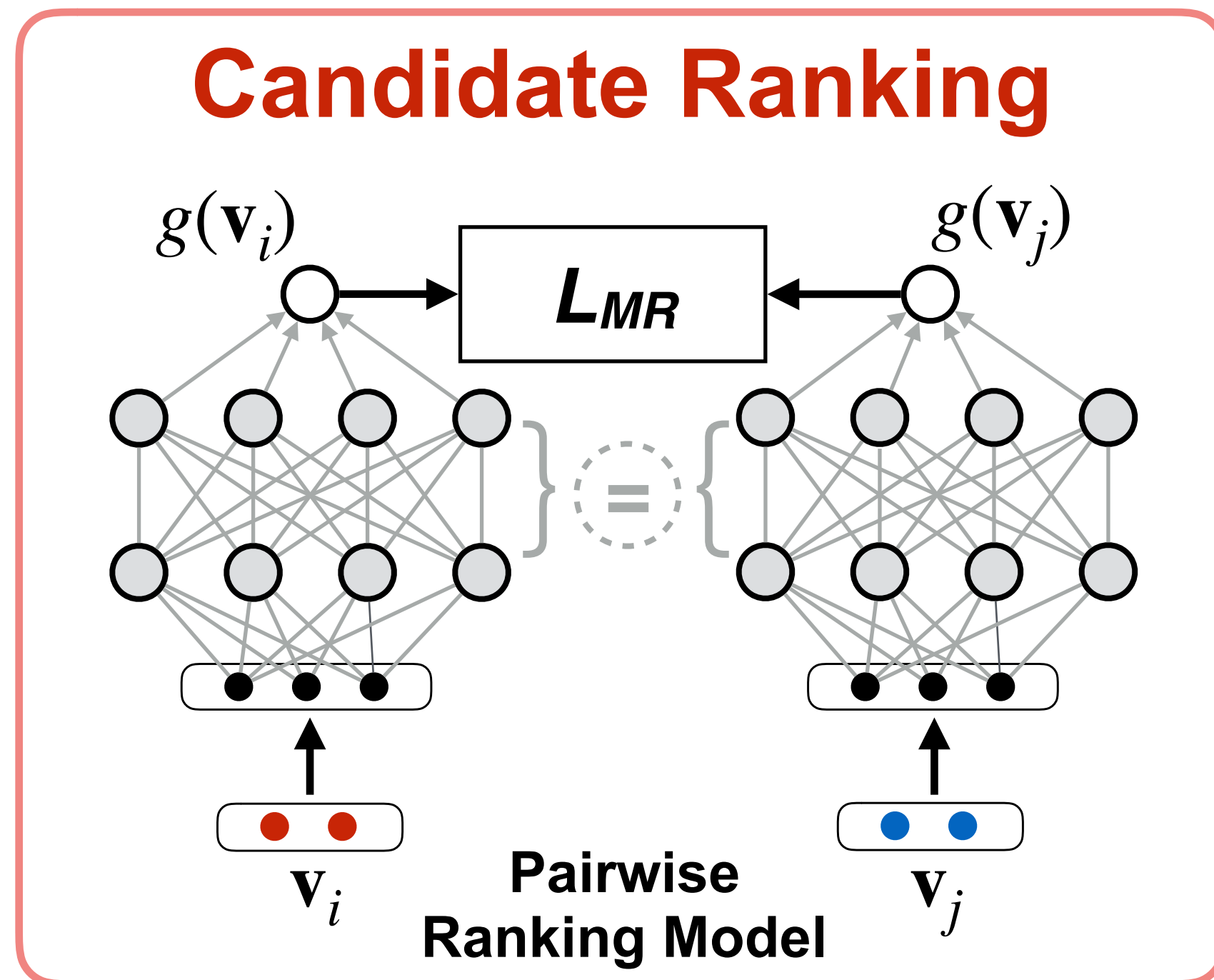
target compression ratio

$$g^*(\mathbf{v}_i, \mathbf{y}) = e^{-\lambda \|\phi_{\mathbf{v}_i} - \phi_{\mathbf{y}}\|} \times$$
$$BERTScore(\mathbf{v}_i, \mathbf{y})$$

candidate reference

Step 2 —

During training, we access each candidate using BERTScore (Zhang et al. 2019) with length penalty.



Loss function:

$$L_{MR} = \frac{1}{m} \sum_{k=1}^m \frac{1}{n_k^2} \sum_{i=1}^{n_k} \sum_{j=1, i \neq j}^{n_k} \max(0, 1 - l_{ij}^k d_{ij}^k)$$

$$d_{ij}^k = g(\mathbf{v}_i^k) - g(\mathbf{v}_j^k)$$

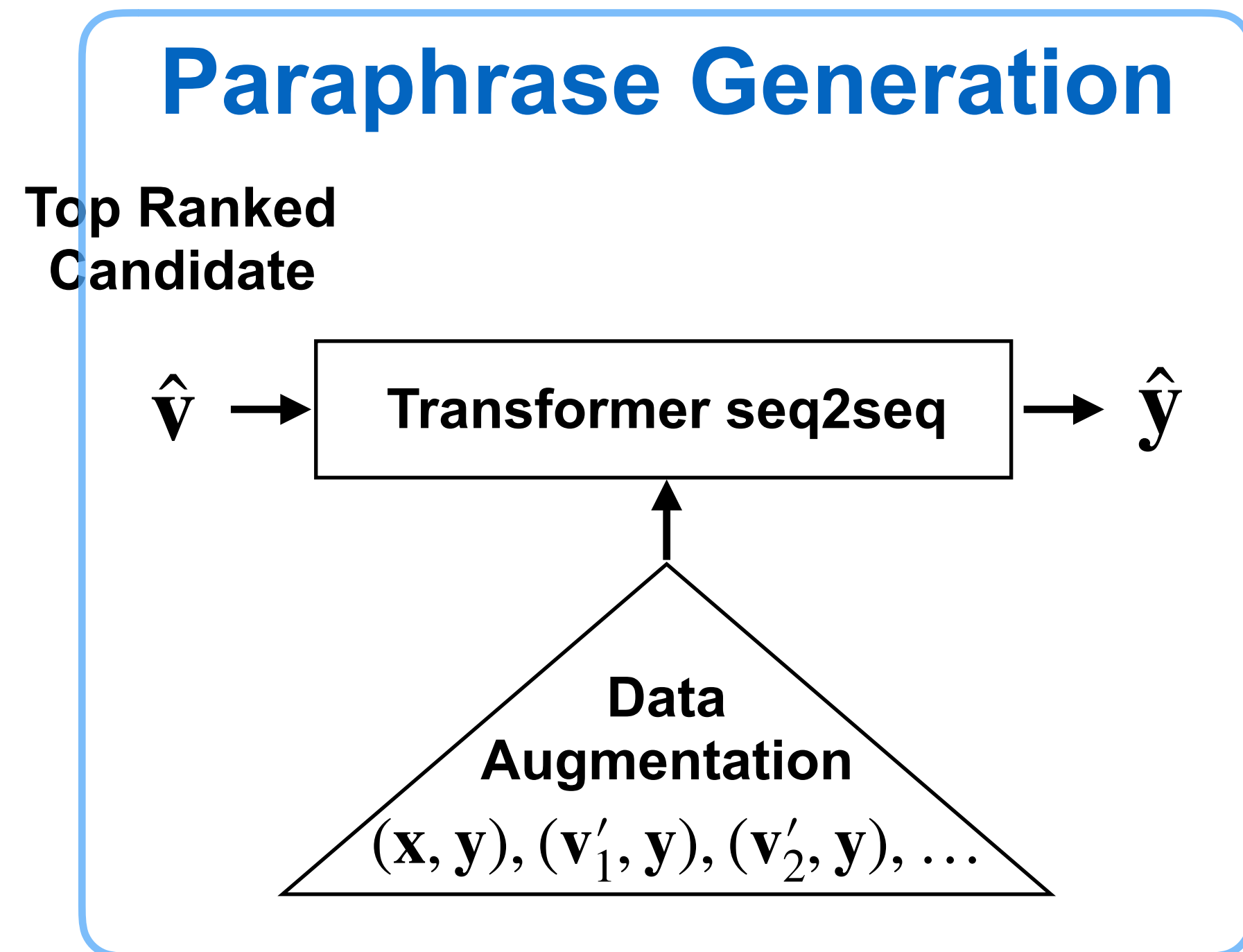
$$l_{ij}^k = \text{sign} \left(\underbrace{g^*(\mathbf{v}_i^k, \mathbf{y}^k)}_{\text{Ranker score}} - \underbrace{g^*(\mathbf{v}_j^k, \mathbf{y}^k)}_{\text{Length-penalized BERTScore}} \right)$$

Ranker score

Length-penalized BERTScore

Step 3 —

Finally, we have a paraphrase generation model trained with augmented training data.

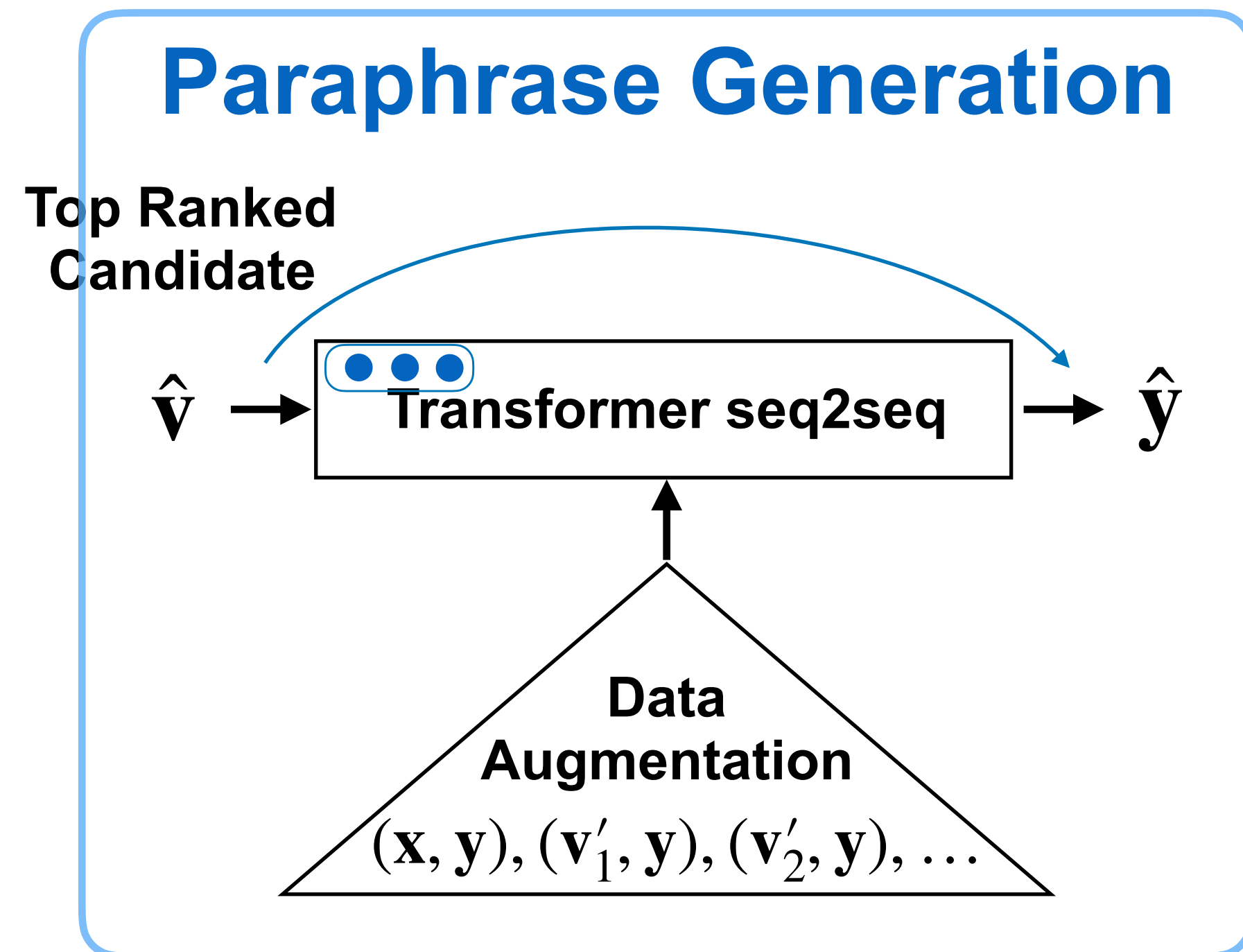


Data Augmentation

- Few selected candidates, in addition to the original input, are paired with the human reference.
- Encourages diverse paraphrases.

Step 3 —

Finally, we have a paraphrase generation model trained with augmented training data.



Additional control over the degree of paraphrasing:

- A copy-control token as soft constraint.
- An auxiliary task (whether a word should be copied) using a monolingual word aligner to derive noisy training labels.

Controllable Text Generation

Input: Since 2010, project researchers have uncovered documents in Portugal that have revealed who owned the ship

Reference: Scientists have found documents in Portugal. They have also found out who owned the ship.

Our Model (<i>split</i> , <i>cp</i> = 0.6)	scientists have found a secret deal . they have discovered who owned the ship.
Our Model (<i>split</i> , <i>cp</i> = 0.7)	scientists have found documents in portugal. they have also found out who owned the ship.
Our Model (<i>split</i> , <i>cp</i> = 0.8)	scientists have found a documents in portugal. they have discovered who owned the ship.

Controllable Text Generation

Input: Since 2010, project researchers have uncovered documents in Portugal that have revealed who owned the ship

Reference: Scientists have found documents in Portugal. They have also found out who owned the ship.

Our Model (<i>split</i> , <i>cp</i> = 0.6)	scientists have found a secret deal . they have discovered who owned the ship.
Our Model (<i>split</i> , <i>cp</i> = 0.7)	scientists have found documents in portugal. they have also found out who owned the ship.
Our Model (<i>split</i> , <i>cp</i> = 0.8)	scientists have found a documents in portugal. they have discovered who owned the ship.
Hybrid-NG	since 2010, project researchers have uncovered documents in portugal that have about who owns the ship.

Controllable Text Generation

Input: Since 2010, project researchers have uncovered documents in Portugal that have revealed who owned the ship

Reference: Scientists have found documents in Portugal. They have also found out who owned the ship.

Our Model (<i>split, cp = 0.6</i>)	scientists have found a secret deal . they have discovered who owned the ship.
Our Model (<i>split, cp = 0.7</i>)	scientists have found documents in portugal. they have also found out who owned the ship.
Our Model (<i>split, cp = 0.8</i>)	scientists have found a documents in portugal. they have discovered who owned the ship.
Hybrid-NG	since 2010, project researchers have uncovered documents in portugal that have about who owns the ship.
LSTM	since 2010, scientists have uncovered documents in portugal that have revealed who owned the ship.

Controllable Text Generation

Input: Since 2010, project researchers have uncovered documents in Portugal that have revealed who owned the ship

Reference: Scientists have found documents in Portugal. They have also found out who owned the ship.

Our Model (<i>split</i> , <i>cp</i> = 0.6)	scientists have found a secret deal . they have discovered who owned the ship.
Our Model (<i>split</i> , <i>cp</i> = 0.7)	scientists have found documents in portugal. they have also found out who owned the ship.
Our Model (<i>split</i> , <i>cp</i> = 0.8)	scientists have found a documents in portugal. they have discovered who owned the ship.
Hybrid-NG	since 2010, project researchers have uncovered documents in portugal that have about who owns the ship.
LSTM	since 2010, scientists have uncovered documents in portugal that have revealed who owned the ship.
Transformer	they discovered that the ship had been important .

Controllable Text Generation

Input: Since 2010, project researchers have uncovered documents in Portugal that have revealed who owned the ship

Reference: Scientists have found documents in Portugal. They have also found out who owned the ship.

Our Model (<i>split</i> , <i>cp</i> = 0.6)	scientists have found a secret deal . they have discovered who owned the ship.
Our Model (<i>split</i> , <i>cp</i> = 0.7)	scientists have found documents in portugal. they have also found out who owned the ship.
Our Model (<i>split</i> , <i>cp</i> = 0.8)	scientists have found a documents in portugal. they have discovered who owned the ship.
Hybrid-NG	since 2010, project researchers have uncovered documents in portugal that have about who owns the ship.
LSTM	since 2010, scientists have uncovered documents in portugal that have revealed who owned the ship.
Transformer	they discovered that the ship had been important .
EditNTS	since 2010, project researchers have uncovered documents in portugal. have revealed who owned the ship

Controllable Text Generation

Input: Experts say China’s air pollution exacts a tremendous toll on human health.

Reference: China’s air pollution is very unhealthy.

Our Model (cp = 0.6)	experts say china’s air pollution is a big problem for human health.
Our Model (cp = 0.7)	experts say china’s air pollution can cause a lot of damage on human health.
Our Model (cp = 0.8)	experts say china’s air pollution is a huge toll on human health.
Hybrid-NG	experts say government’s air pollution exacts a tremendous toll on human health.
LSTM	experts say china’s air pollution exacts a tremendous toll on human health.
Transformer	experts say china’s air pollution exacts a tremendous effect on human health.
EditNTS	experts say china’s air pollution can cause human health.

Experiments on Text Simplification

- Evaluation setup
 - Standard Evaluation on **Newsela-Auto** and **Wikipedia-Auto** (Jiang et al. 2020).



259, 778 train
32, 689 valid
33, 391 test
sentence pairs

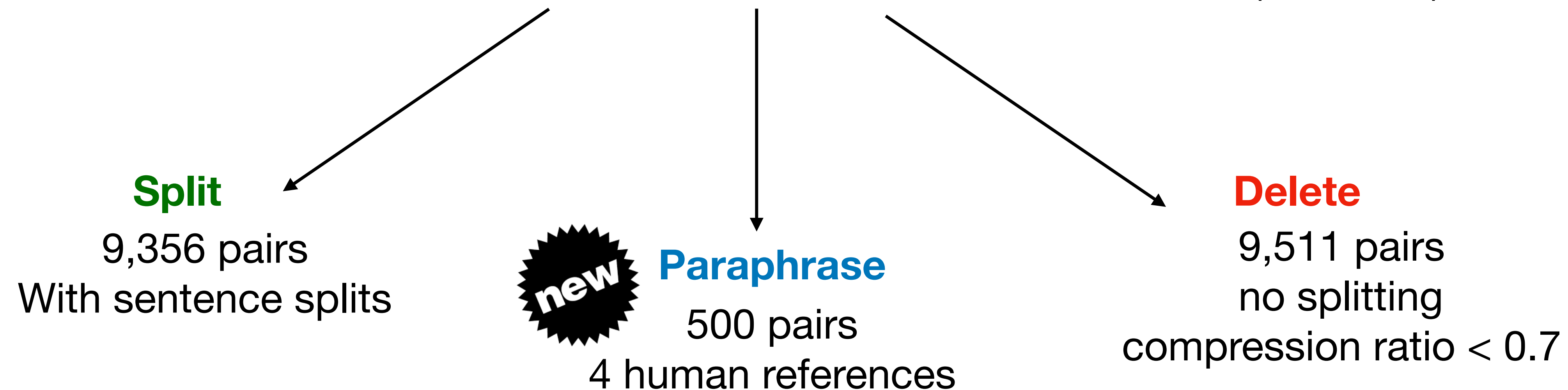


See our paper!

Experiments on Text Simplification

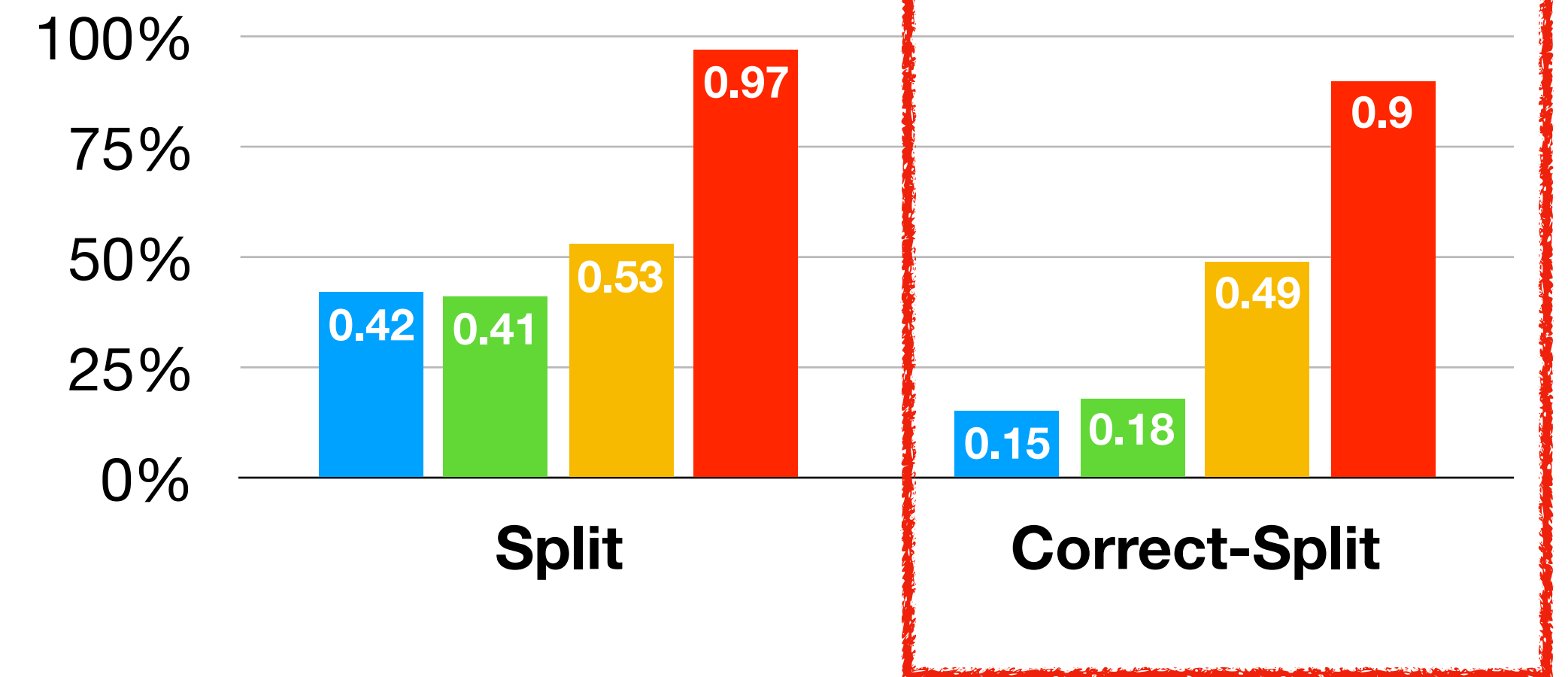
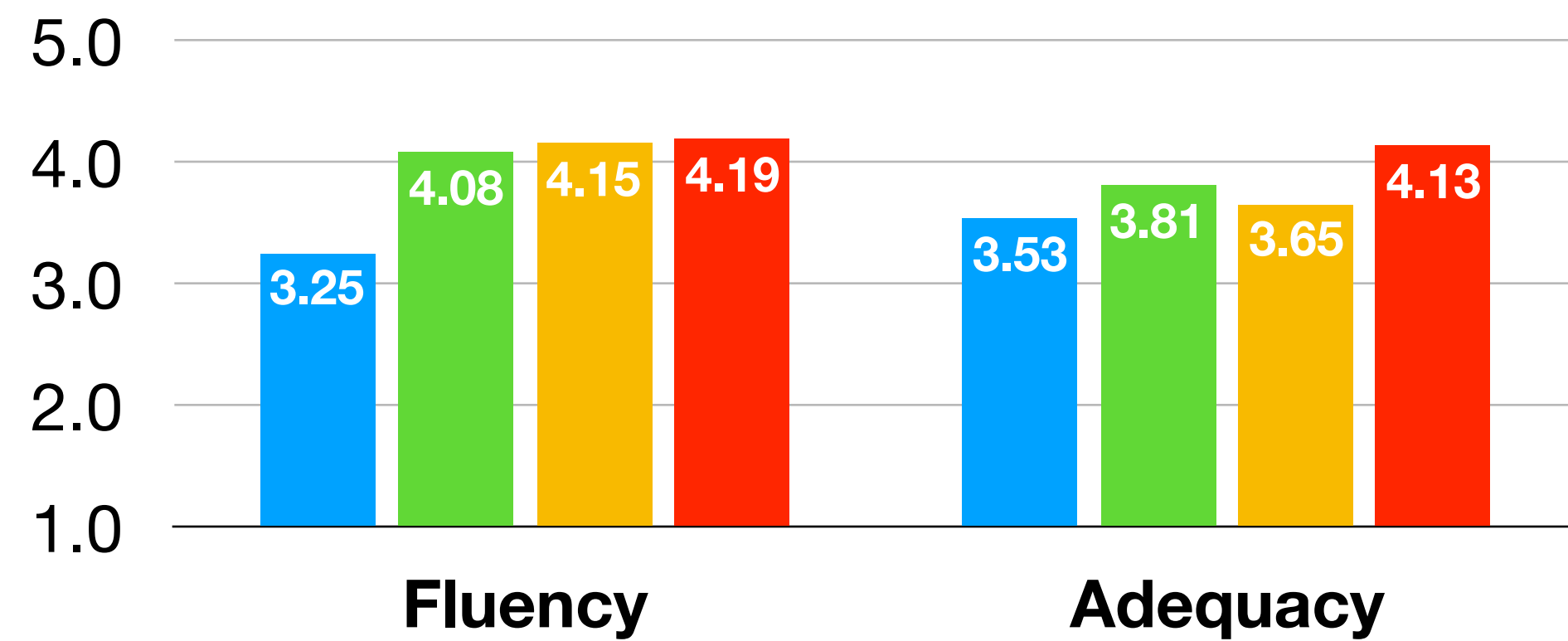
- Evaluation setup

- Standard Evaluation on **Newsela-Auto** and **Wikipedia-Auto** (Jiang et al. 2020).
- Edit-focused Evaluation on different sections of test set (Our work).



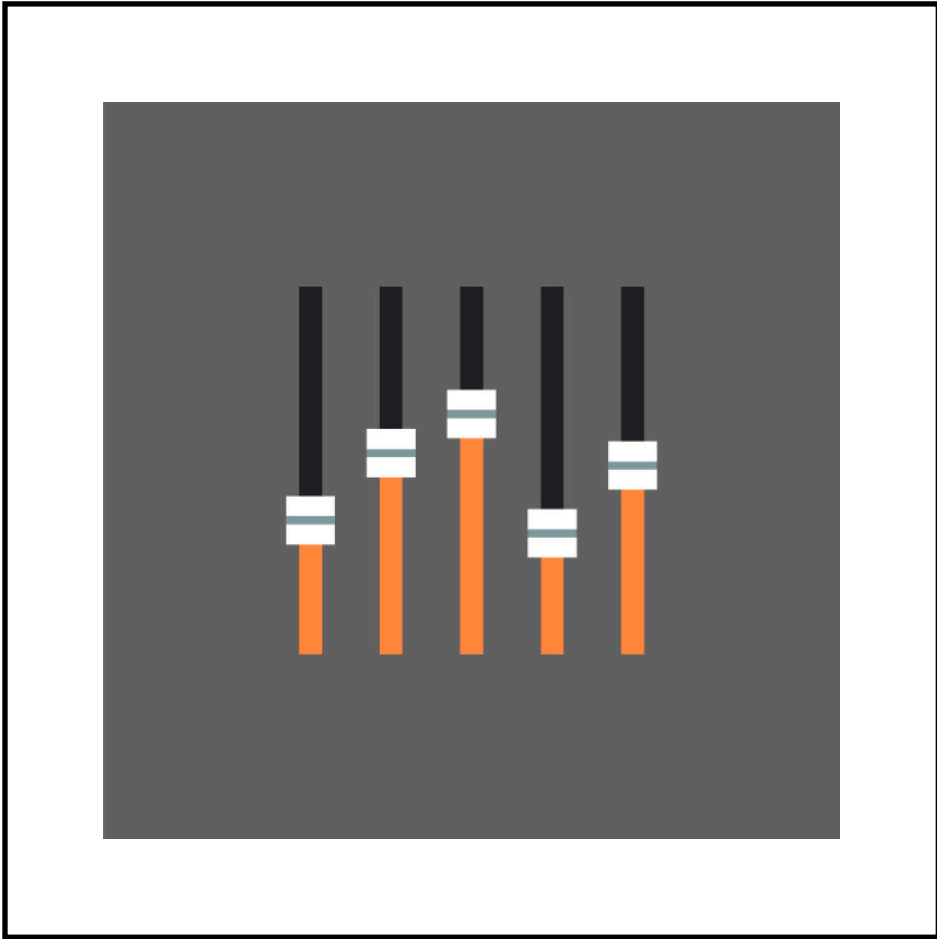
More Syntactic Transformations

Human evaluation (1-5 Likert scale) on sentences where simplification involves splitting.



- Hybrid (Narayan & Gardent, 2014)
- Programmer-Interpreter (Dong et al., 2019)
- Transformer (Jiang et al., 2020 — also our work)
- ControllableTS (this work)

Controllable Generation & Evaluation



Models	SARI	add	keep	del	FK	SLen	OLen	CR	%split	s-BL	%new	%eq
Complex (input)	22.3	0.0	67.0	0.0	12.8	23.3	23.5	1.0	0.0	100.0	0.0	100.0
Simple (reference)	62.3	44.8	68.3	73.9	11.1	23.8	23.5	1.01	0.0	48.5	24.1	0.0
Hybrid-NG	38.2	2.8	57.0	54.8	10.7	21.6	23.1	0.98	7.0	57.2	9.1	1.4
Transformer _{bert}	36.0	3.3	54.9	49.8	8.9	16.1	20.2	0.87	23.0	58.7	13.3	7.6
EditNTS	36.4	1.1	59.1	48.9	9.9	17.5	20.6	0.88	17.0	70.6	5.2	3.2
Our Model	38.1	3.9	55.1	55.5	8.8	16.6	20.2	0.86	19.6	50.4	15.7	0.0
Our Model (no split; $cp = 0.6$)	39.0	3.8	57.7	55.6	11.2	22.1	22.9	0.98	0.2	55.9	18.0	1.0
Our Model (no split; $cp = 0.7$)	41.0	3.4	63.1	56.6	11.5	22.2	22.9	0.98	0.0	60.4	10.4	4.2
Our Model (no split; $cp = 0.8$)	40.6	2.9	65.0	54.0	11.8	22.4	23.0	0.99	0.0			

paraphrasing

Table 2: Automatic evaluation results on [NEWSELA-TURK that focuses on paraphrasing](#) (500 complex sentences with 4 human written paraphrases). We control the extent of paraphrasing of our models by specifying the percentage of words to be copied (cp) from the input as a soft constraint.

Models	SARI	add	keep	del	FK	SLen	OLen	CR	%split	s-BL	%new	%eq
Complex (input)	17.0	0.0	51.1	0.0	14.6	30.0	30.2	1.0	0.0	100.0	0.0	100.0
Simple (reference)	93.0	89.9	91.6	97.5	7.0	13.4	28.6	0.98	100.0	36.8	29.7	0.0
Hybrid-NG	37.1	2.2	44.9	64.1	11.6	25.5	30.1	1.0	17.3	57.7	8.7	1.6
Transformer _{bert}	39.5	4.2	47.3	67.0	8.8	17.1	25.3	0.85	39.7	57.7	11.9	5.2
EditNTS	38.5	1.1	48.3	66.1	9.6	18.3	24.7	0.83	32.8	67.7	3.7	1.5
Our Model	39.4	4.0	46.6	67.6	8.7	17.5	25.5	0.85	40.6	48.3	15.6	0.1
Our Model (w/ split)	42.1	5.6	50.6	70.1	8.1	15.3	30.3	1.02	93.5	60.7	12.4	

splitting

Table 3: Automatic evaluation results on a subset of [NEWSELA-AUTO test set that focuses on splitting](#) (9,356 complex-simple sentence pairs with splitting). Our model chooses only candidate simplifications that have undergone splitting during the ranking step of the pipeline.

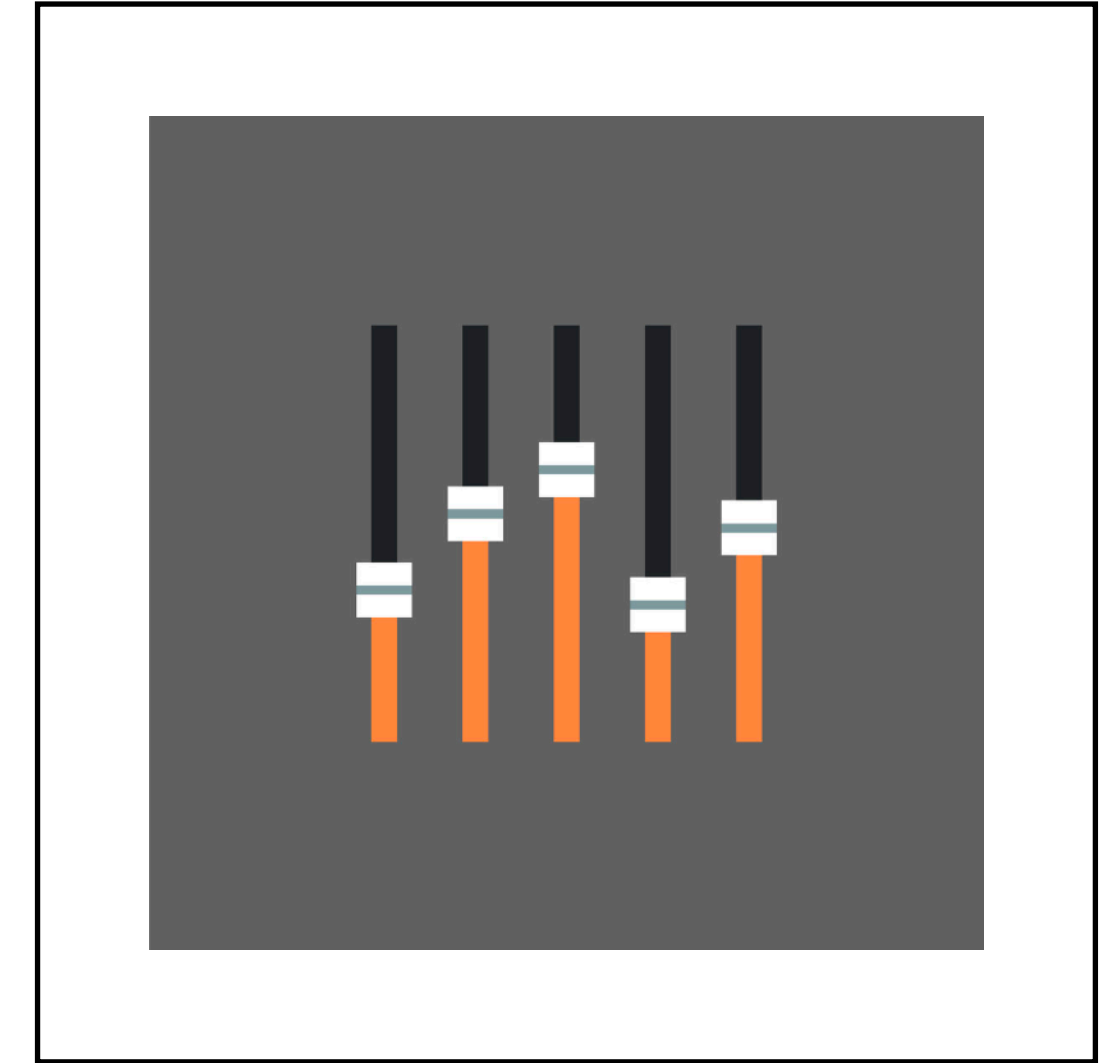
Models	SARI	add	keep	del	FK	SLen	OLen	CR	%split	s-BL	%new	%eq
Complex (input)	9.6	0.0	28.8	0.0	12.9	25.8	26.0	1.0	0.0	100.0	0.0	100.0
Simple (reference)	85.7	82.7	76.0	98.6	6.7	12.6	12.6	0.5	0.0	19.6	32.6	0.0
Hybrid-NG	35.8	1.4	27.0	79.1	10.6	22.7	25.9	1.0	13.3	58.9	8.7	3.6
Transformer _{bert}	36.8	2.2	29.6	78.7	8.4	16.2	21.7	0.85	27.7	57.9	12.3	8.2
EditNTS	37.4	0.9	29.8	81.5	9.2	17.5	22.0	0.86	24.1	68.9	4.6	2.5
Our Model	39.2	2.4	29.8	85.3	8.2	16.4	21.9	0.85	29.1	48.8	15.6	0.4
Our Model (no split; $CR < 0.7$)	38.2	2.0	28.5	84.1	8.6	16.8	17.5	0.68	0.1	42.0	12.5	

deletion

Table 4: Automatic evaluation results on a subset of [NEWSELA-AUTO test set that focuses on deletion](#) (9,511 complex-simple sentence pairs with compression ratio < 0.7 and no sentence splits). Our model selects only candidates with similar compression ratio and no splits during ranking.

Takeaways

- Novel approach to control edit operations.
- Evaluation setup for edit operations.
- New dataset to evaluate lexical paraphrasing.
- Cleaner and larger training data for simplification.



Previous work: Neural CRF Model for Sentence Alignment in Text Simplification

- Check the code/data at https://github.com/mounicam/controllable_simplification
- Contact: Mounica Maddela (mmaddela3@gatech.edu)