

Evaluating Text Generation from Discourse Representation Structures

Chunliu Wang Rik van Noord Arianna Bisazza Johan Bos



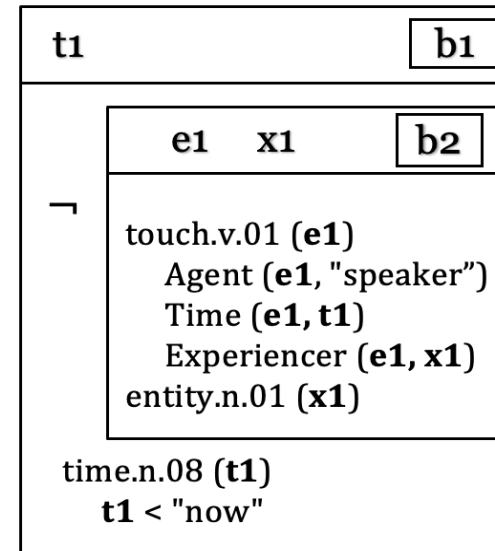
Contributions of this talk

- A new method for meaning-to-text generation
- Five semantic challenge sets for NLG
- A new evaluation metric: ROSE

Discourse Representation Structures (DRS)

- Proposed in Discourse Representation Theory
- Formatting DRSs with boxes and clauses
- Using various letters for variables
- Concepts are represented by WordNet synsets
- Semantic relations are represented by Verbnet roles

Box Format of DRS

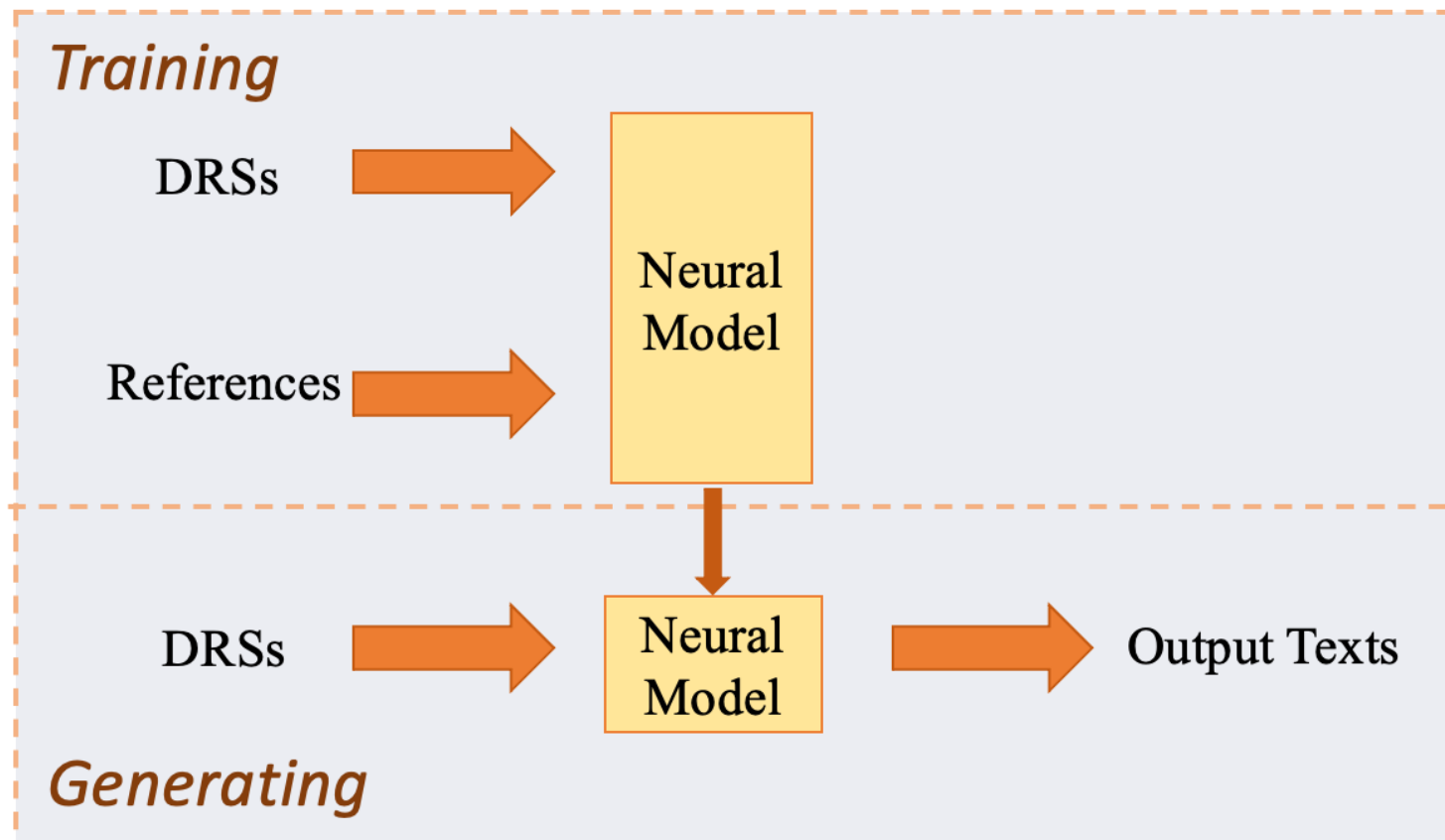


Flatten

Clausal Format of DRS

b1 NEGATION **b2**
b1 REF **t1**
b1 TPR **t1** "now"
b1 time "n.08" **t1**
b2 REF **e1**
b2 Agent **e1** "speaker"
b2 Experiencer **e1** **x1**
b2 Time **e1** **t1**
b2 touch "v.01" **e1**
b2 REF **x1**
b2 entity "n.01" **x1**

DRS-to-Text Generation



Source Representation: DRSs

➤ Represent meaningful units as atomic entities

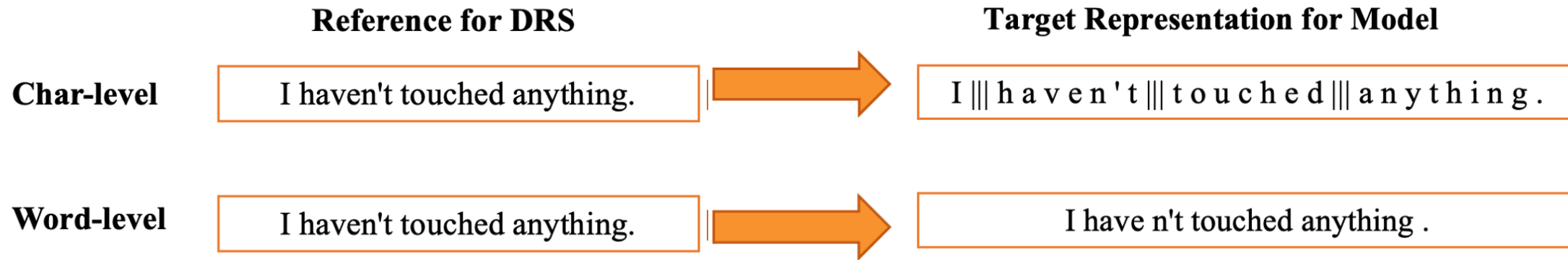
- variable indices
- DRS operators
- semantic relations
- deictic constants
- concepts

➤ Represent remaining parts as character-level

- time/date expressions (“ 1 9 6 8 ”)
- value expressions (“ 2 – 0 ”)
- quantities (“ 2 6 0 0 ”)
- proper names (“ b r a d ~ p i t t ”)

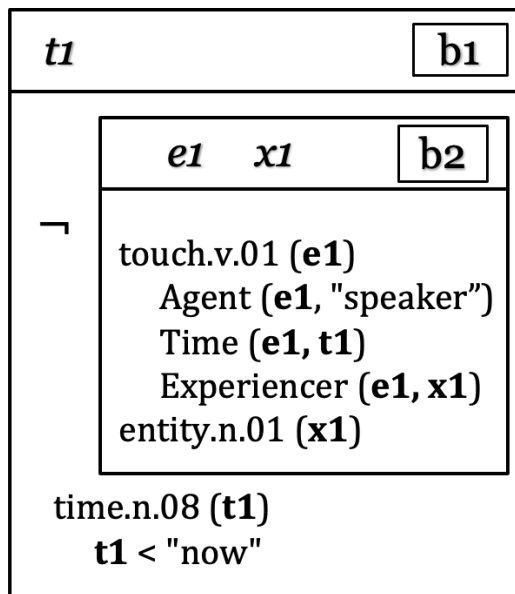
Clausal Format of DRS		Input Representation for Model
b1 NEGATION b2		\$NEW NEGATION \$NEW
b1 REF t1		\$-1 REF
b1 TPR t1 "now"		\$-1 TPR @0 "now"
b1 time "n.08" t1		\$-1 time.n.08 @0
b2 REF e1	Preprocess	\$0 REF
b2 Agent e1 "speaker"	➔	\$0 Agent @0 "speaker"
b2 Experiencer e1 x1		\$0 Experiencer @0 @1
b2 Time e1 t1		\$0 Time @0 @-1
b2 touch "v.01" e1		\$0 touch.v.01 @0
b2 REF x1		\$0 REF
b2 entity "n.01" x1		\$0 entity.n.01 @0

Target Representation: Text



Methodology

Box Format of DRS



Flatten



Clausal Format of DRS

b1 NEGATION **b2**
b1 REF **t1**
b1 TPR **t1** "now"
b1 time "n.08" **t1**
b2 REF **e1**
b2 Agent **e1** "speaker"
b2 Experiencer **e1** **x1**
b2 Time **e1** **t1**
b2 touch "v.01" **e1**
b2 REF **x1**
b2 entity "n.01" **x1**

Preprocess



Input Representation for Model

\$NEW NEGATION \$NEW
\$-1 REF
\$-1 TPR @0 "now"
\$-1 time.n.08 @0
\$0 REF
\$0 Agent @0 "speaker"
\$0 Experiencer @0 @1
\$0 Time @0 @-1
\$0 touch.v.01 @0
\$0 REF
\$0 entity.n.01 @0

Reference for DRS

I haven't touched anything.

Preprocess



Target Representation for Model

I ||| h a v e n ' t ||| t o u c h e d ||| a n y t h i n g .

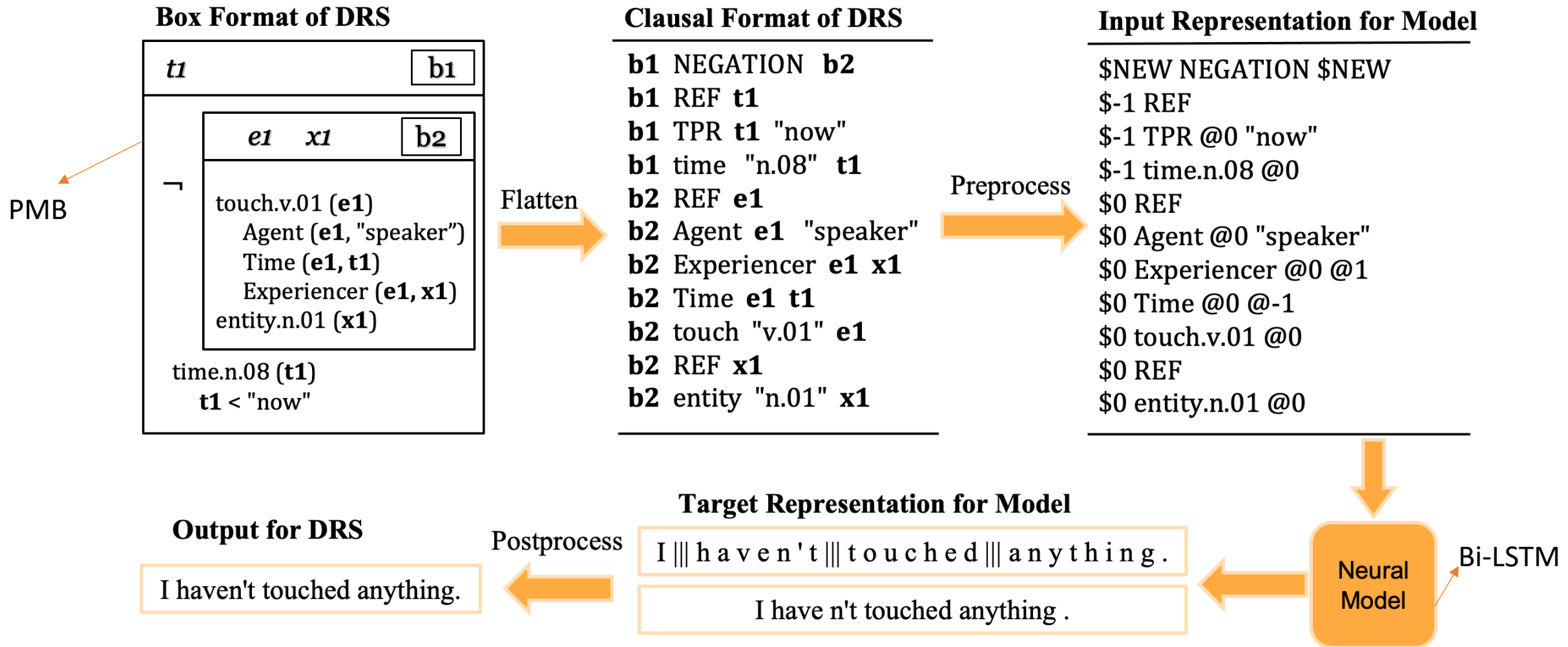
I have n't touched anything .




Neural
Model



Methodology

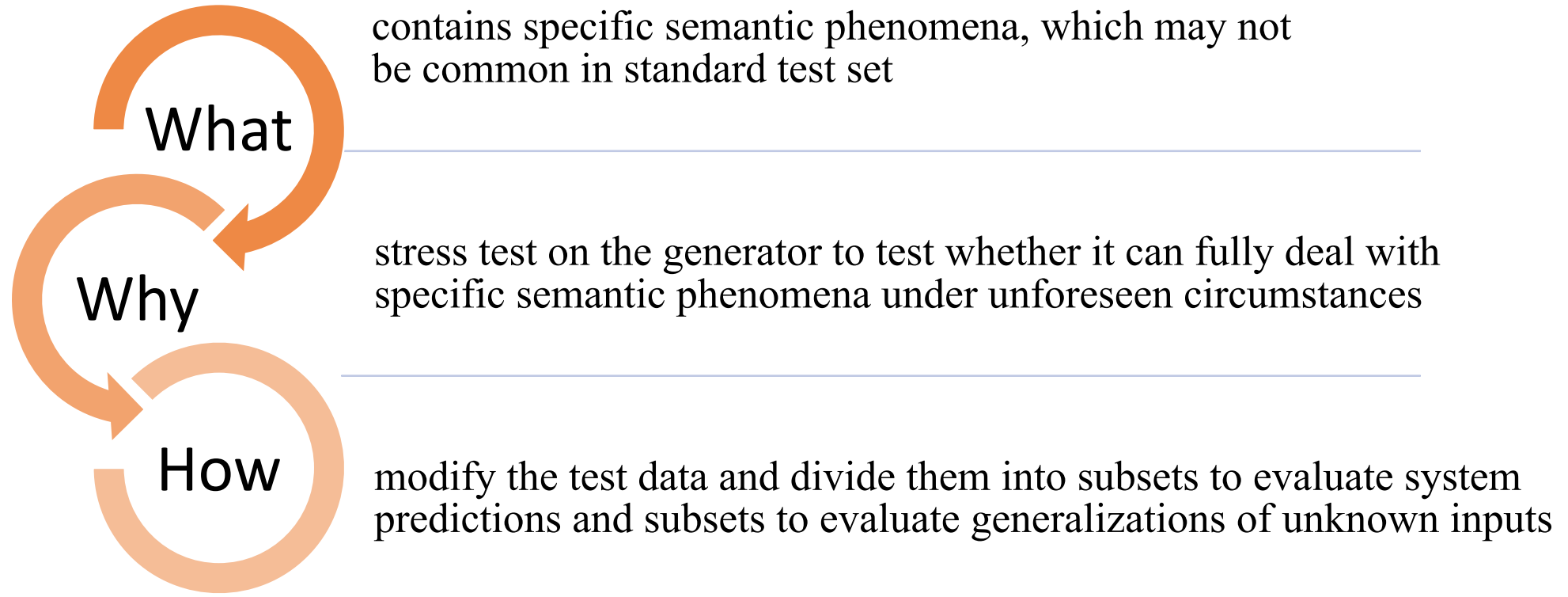


Standard Automatic Metrics Results



	BLEU	METEOR	ROUGE
Char-level (raw)	69.3	51.8	84.9
Word-level (tok)	64.7	47.8	81.8

Semantic Challenge set



Semantic Challenge set

- **Tense Change**
 - past ($t < \text{now}$)
 - present ($t = \text{now}$)
 - future ($t > \text{now}$)
- **Polarity Change**
 - operator: NEGATION
- **Grammatical Number Change**
 - a number greater than one \rightarrow one
 - one \rightarrow a number greater than one
- **Names Change**
 - name entities \rightarrow not occur in the training data
- **Quantities Change**
 - numbers \rightarrow unknown quantity expressions

Original	Tom has three thousand books.
-----------------	-------------------------------

Tense	Tom had three thousand books.
--------------	-------------------------------

Polarity	Tom does not have three thousand books.
-----------------	---

Number	Tom has one book.
---------------	-------------------

Names	Kirk has three thousand books.
--------------	--------------------------------

Quantity	Tom has 3,200 books.
-----------------	----------------------

Semantic Challenge set : Tense Change

DRS: Original

b1 REF x1
b1 Name x1 "tom"
b1 PRESUPPOSITION b2
b1 male "n.02" x1
b2 REF e1
b2 REF t1
b2 EQU t1 "now"
b2 Pivot e1 x1
b2 Theme e1 x2
b2 Time e1 t1
b2 have "v.04" e1
b2 time "n.08" t1
b2 REF x2
b2 Quantity x2 "3000"
b2 book "n.02" x2

Reference:

Tom has three thousand books.



DRS: Tense change

b1 REF x1
b1 Name x1 "tom"
b1 PRESUPPOSITION b2
b1 male "n.02" x1
b2 REF e1
b2 REF t1
b2 **TPR** t1 "now"
b2 Pivot e1 x1
b2 Theme e1 x2
b2 Time e1 t1
b2 have "v.04" e1
b2 time "n.08" t1
b2 REF x2
b2 Quantity x2 "3000"
b2 book "n.02" x2

Reference:

Tom had three thousand books.

Semantic Challenge set : Polarity Change

DRS: Original

b1 REF x1
b1 Name x1 "tom"
b1 PRESUPPOSITION b2
b1 male "n.02" x1
b2 REF e1
b2 REF t1
b2 EQU t1 "now"
b2 Pivot e1 x1
b2 Theme e1 x2
b2 Time e1 t1
b2 have "v.04" e1
b2 time "n.08" t1
b2 REF x2
b2 Quantity x2 "3000"
b2 book "n.02" x2

Reference:

Tom has three thousand books.



DRS: Polarity change

b1 REF x1
b1 Name x1 "tom"
b1 PRESUPPOSITION b2
b1 male "n.02" x1
b2 REF e1
b2 REF t1
b2 EQU t1 "now"
b2 NEGATION b3
b3 REF e1
b3 Pivot e1 x1
b3 Theme e1 x2
b3 Time e1 t1
b3 have "v.04" e1
b3 REF x2
b3 Quantity x2 "3000"
b3 book "n.02" x2

Reference:

Tom does not have three thousand books..

Semantic Challenge set : Number Change

DRS: Original

b1 REF x1
b1 Name x1 "tom"
b1 PRESUPPOSITION b2
b1 male "n.02" x1
b2 REF e1
b2 REF t1
b2 EQU t1 "now"
b2 Pivot e1 x1
b2 Theme e1 x2
b2 Time e1 t1
b2 have "v.04" e1
b2 time "n.08" t1
b2 REF x2
b2 Quantity x2 "3000"
b2 book "n.02" x2

Reference:

Tom has three thousand books.



DRS: Number change

b1 REF x1
b1 Name x1 "tom"
b1 PRESUPPOSITION b2
b1 male "n.02" x1
b2 REF e1
b2 REF t1
b2 EQU t1 "now"
b2 Pivot e1 x1
b2 Theme e1 x2
b2 Time e1 t1
b2 have "v.04" e1
b2 time "n.08" t1
b2 REF x2
b2 Quantity x2 "1"
b2 book "n.02" x2

Reference:

Tom has one book.

Semantic Challenge set : Names Change

DRS: Original

b1 REF x1
b1 Name x1 "tom"
b1 PRESUPPOSITION b2
b1 male "n.02" x1
b2 REF e1
b2 REF t1
b2 EQU t1 "now"
b2 Pivot e1 x1
b2 Theme e1 x2
b2 Time e1 t1
b2 have "v.04" e1
b2 time "n.08" t1
b2 REF x2
b2 Quantity x2 "3000"
b2 book "n.02" x2

Reference:

Tom has three thousand books.



DRS: Name change

b1 REF x1
b1 Name x1 "kirk"
b1 PRESUPPOSITION b2
b1 male "n.02" x1
b2 REF e1
b2 REF t1
b2 EQU t1 "now"
b2 Pivot e1 x1
b2 Theme e1 x2
b2 Time e1 t1
b2 have "v.04" e1
b2 time "n.08" t1
b2 REF x2
b2 Quantity x2 "3000"
b2 book "n.02" x2

Reference:

Kirk has three thousand books.

Semantic Challenge set : Quantities Change

DRS: Original

b1 REF x1
b1 Name x1 "tom"
b1 PRESUPPOSITION b2
b1 male "n.02" x1
b2 REF e1
b2 REF t1
b2 EQU t1 "now"
b2 Pivot e1 x1
b2 Theme e1 x2
b2 Time e1 t1
b2 have "v.04" e1
b2 time "n.08" t1
b2 REF x2
b2 Quantity x2 "3000"
b2 book "n.02" x2

Reference:

Tom has three thousand books.



DRS: Quantity change

b1 REF x1
b1 Name x1 "tom"
b1 PRESUPPOSITION b2
b1 male "n.02" x1
b2 REF e1
b2 REF t1
b2 EQU t1 "now"
b2 Pivot e1 x1
b2 Theme e1 x2
b2 Time e1 t1
b2 have "v.04" e1
b2 time "n.08" t1
b2 REF x2
b2 Quantity x2 "3200"
b2 book "n.02" x2

Reference:

Tom has 3,200 books.

ROSE: A New Expert Assessment

Semantics

- score 1 if the meaning of the output reflects that of the underlying meaning representation

Grammaticality

- score 1 if the sentence is grammatical and free of spelling mistakes

Phenomenon

- score 1 if the phenomenon of control is generated at all

ROSE exemplified

Reference text	Generated text	Sem.	Gram.	Phen.	ROSE
(a) She liked short skirts.	She liked short tomical.	0	0	1	0
(b) Tom does not have three thousand books.	Tom never has three thousand books.	0	1	1	0
(c) The small skirt will be pink.	The small skirt was pink.	0	1	0	0
(d) He left 157 minutes ago.	He left fifteen minutes ago.	0	1	0	0
(e) I checked it nine times.	I checked it nine.	0	0	1	0
(f) We are painting the house green.	I paint the house green.	1	1	1	1
(g) That hat cost around fifty dollars.	This hat cost about 50 dollars.	1	1	1	1
(h) When I painted this picture, I was 23 years old.	I painted the picture when I was twenty-three years old.	1	1	1	1

Computing the ROSE Scores

		Sem.		Gram.		Phen.		ROSE	
	#	Orig	Chal	Orig	Chal	Orig	Chal	Orig	Chal
Tense	200	80.0	71.0	92.0	87.5	99.5	86.5	78.0	64.0
Polarity	100	80.0	52.0	96.0	81.0	100.0	99.0	78.0	49.0
Number	100	80.0	79.0	95.0	84.0	100.0	95.0	77.0	69.0
Names	50	82.0	76.0	94.0	84.0	100.0	98.0	82.0	74.0
Quantity	50	88.0	72.0	98.0	90.0	92.0	84.0	86.0	70.0

Comparing with Automatic Metrics Scores

		BLEU		METEOR		ROUGE		ROSE	
	#	Orig	Chal	Orig	Chal	Orig	Chal	Orig	Chal
Tense	200	68.4	55.8	50.9	44.8	85.0	76.1	78.0	64.0
Polarity	100	68.1	37.4	50.8	37.9	85.0	66.1	78.0	49.0
Number	100	72.5	69.2	53.7	53.4	85.7	86.4	77.0	69.0
Names	50	69.1	71.9	53.0	53.5	87.2	87.8	82.0	74.0
Quantity	50	69.7	68.0	56.4	50.6	86.0	83.4	86.0	70.0

Evaluating Text Generation from Discourse Representation Structures

Chunliu Wang Rik van Noord Arianna Bisazza Johan Bos

Conclusion

- Character-level achieves higher standard automatic metrics scores than word-level
- Negation and tense are the most challenging phenomena
- Changes in grammatical number and generalizations to unseen quantities or names are well handled by the model.

