

# Multi-Figurative Language Generation

Huiyuan Lai

[h.lai@rug.nl](mailto:h.lai@rug.nl)

Malvina Nissim

[m.nissim@rug.nl](mailto:m.nissim@rug.nl)

University of Groningen / The Netherlands



university of  
 groningen

COLING 2022

# Figurative Language

**Hyperbole:** Old Mr. Smith has been teaching here since the Stone Age.

**Sarcasm:** I love when they run the same commercial twice in a row.

**Simile:** You can publish the whole thing like a diary.

**Idiom:** My niece will babysit for you for pin money.

**Metaphor:** He made a road of my broken works.

# Figurative Language

Figurative language, such as metaphors, can make an expression stand out by making it **more interesting and captivating**, and evoke **strong emotions**.

Each figure of speech is used to accomplish a constellation of **communicative goals** (e.g. speakers can be humorous by using hyperbole).

**Richard M. Roberts and Roger J. Kreuz.**

Why do people use figurative language? *Psychological Science*. 1994

# Figurative Language Generation

Literal Input: *Old Mr. Smith has been teaching here for a very long time.*

Hyperbole: *Old Mr. Smith has been teaching here **since the Stone Age**.*

---

Literal Input: *My niece will babysit for you for a little bit of money.*

Idiom: *My niece will babysit for you for **pin money**.*

---

Literal Input: *I hate it when they run the same commercial twice in a row.*

Sarcasm: *I **love** when they run the same commercial twice in a row.*

# Figurative Language Generation

Reformulating a given text in the **desired figure of speech** while still being faithful to the **original context**.

---

Literal Input: *Old Mr. Smith has been teaching here for a very long time.*

Hyperbole: *Old Mr. Smith has been teaching here **since the Stone Age**.*

---

Literal Input: *My niece will babysit for you for a little bit of money.*

Idiom: *My niece will babysit for you for **pin money**.*

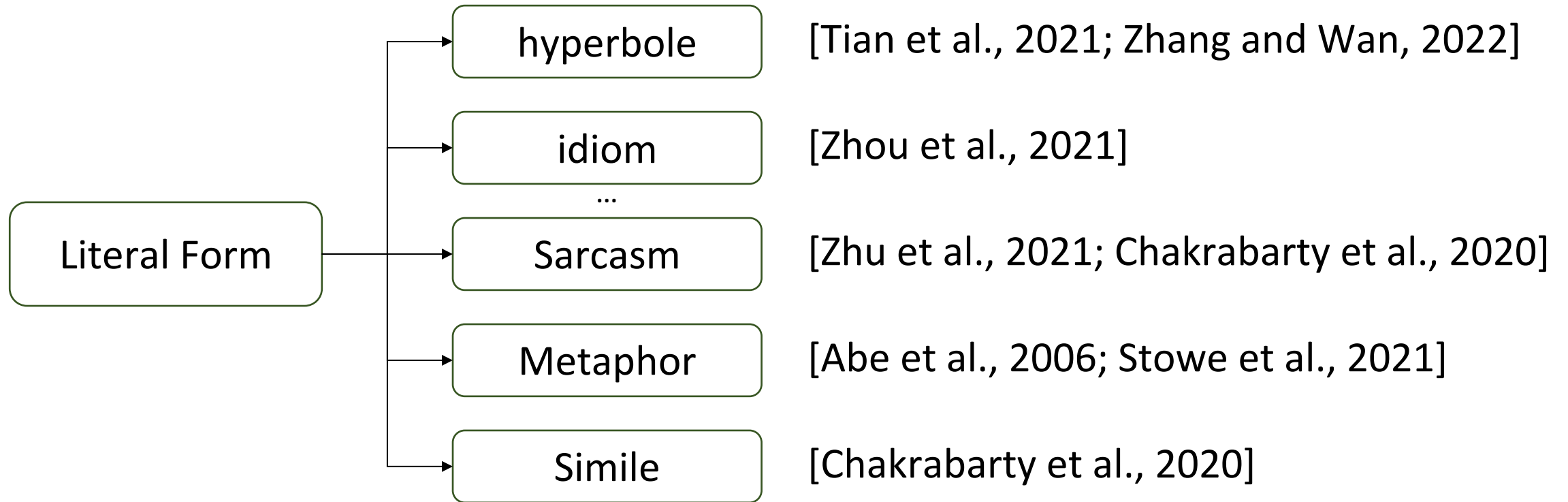
---

Literal Input: *I hate it when they run the same commercial twice in a row.*

Sarcasm: *I **love** when they run the same commercial twice in a row.*

---

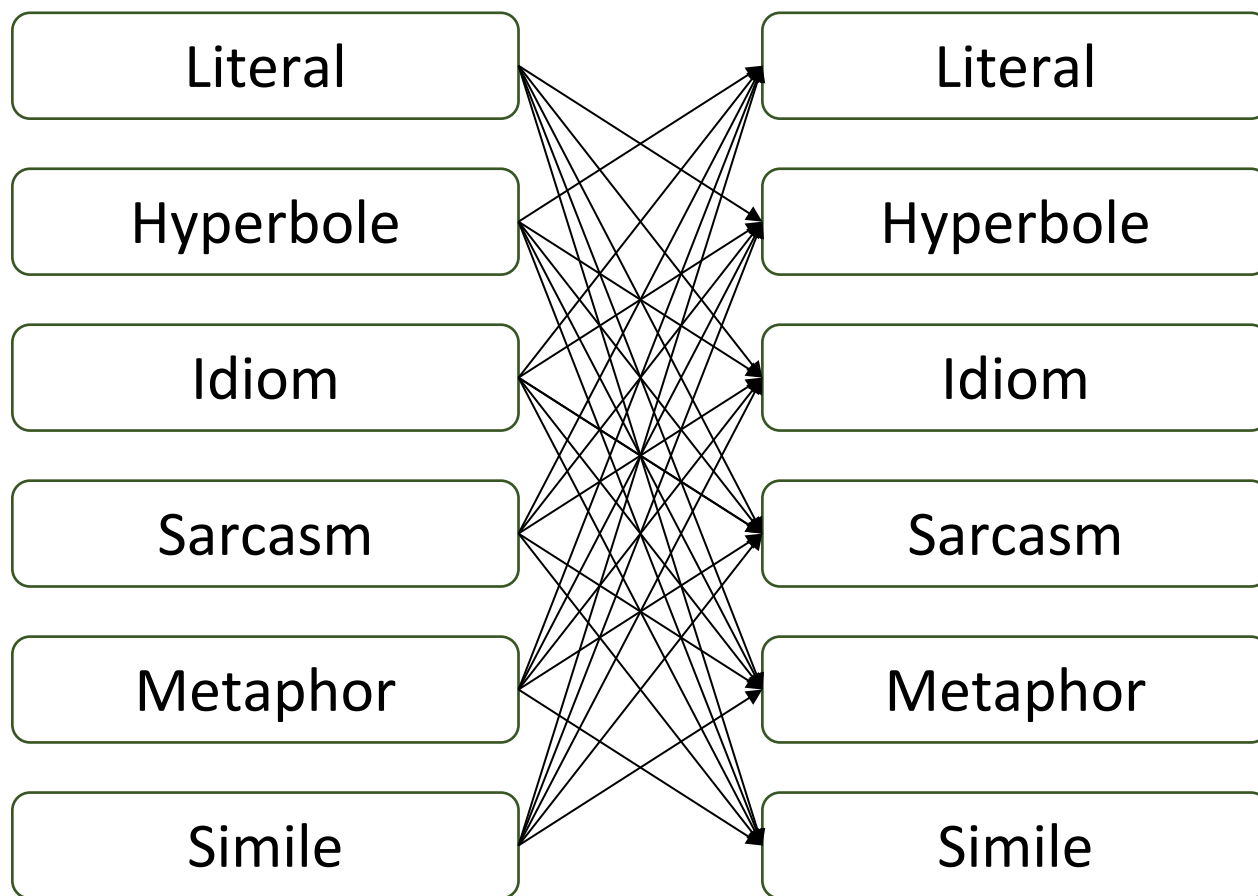
# Figurative Language Generation: **Current Status**



# Figurative Language Generation: **Current Status**

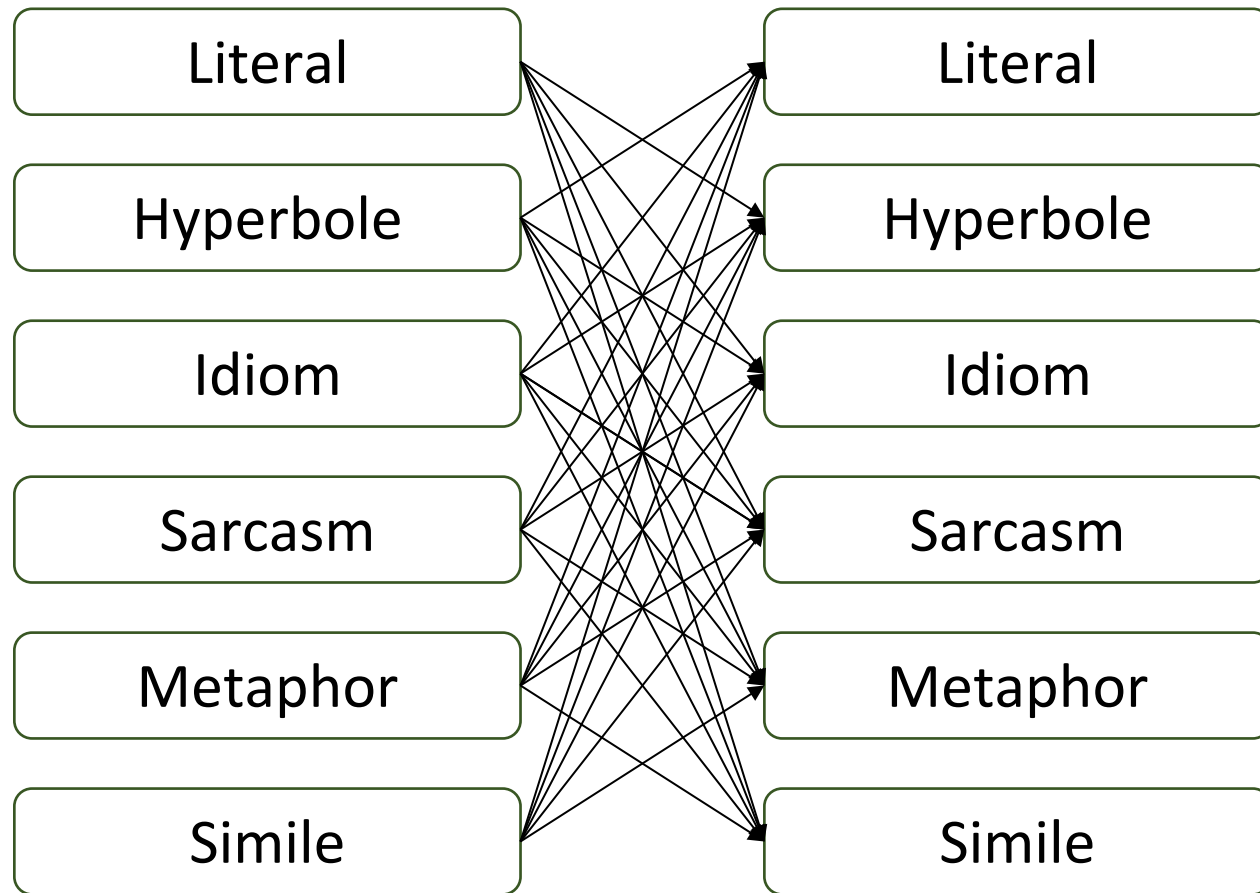
- Can we train a model to generate text with the desired figurative form from another figurative form (e.g. generating an idiomatic text from the hyperbolic one)
- Can we model multiple figures of speech jointly?

# Figurative Language Generation: **This Work**

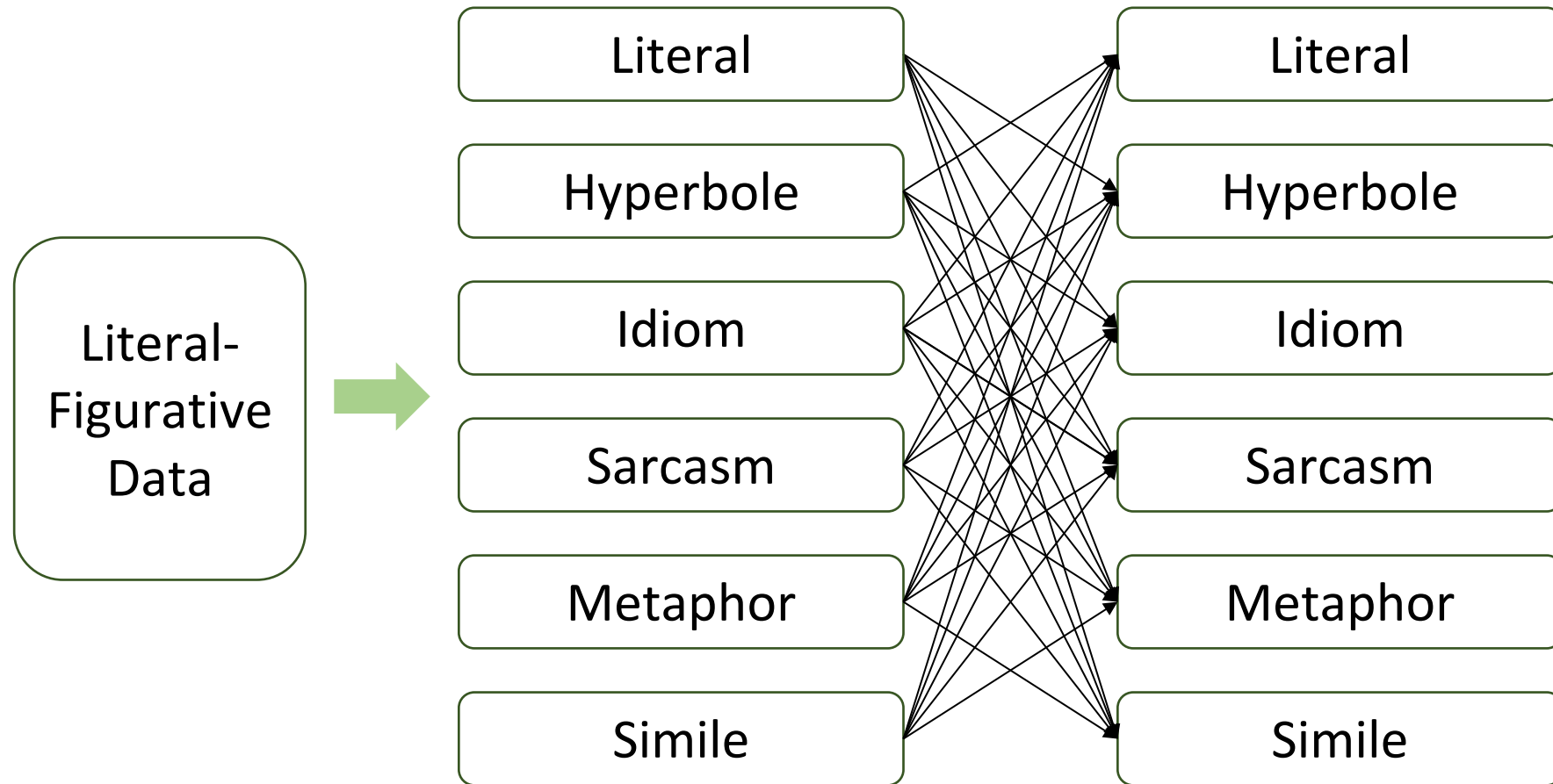




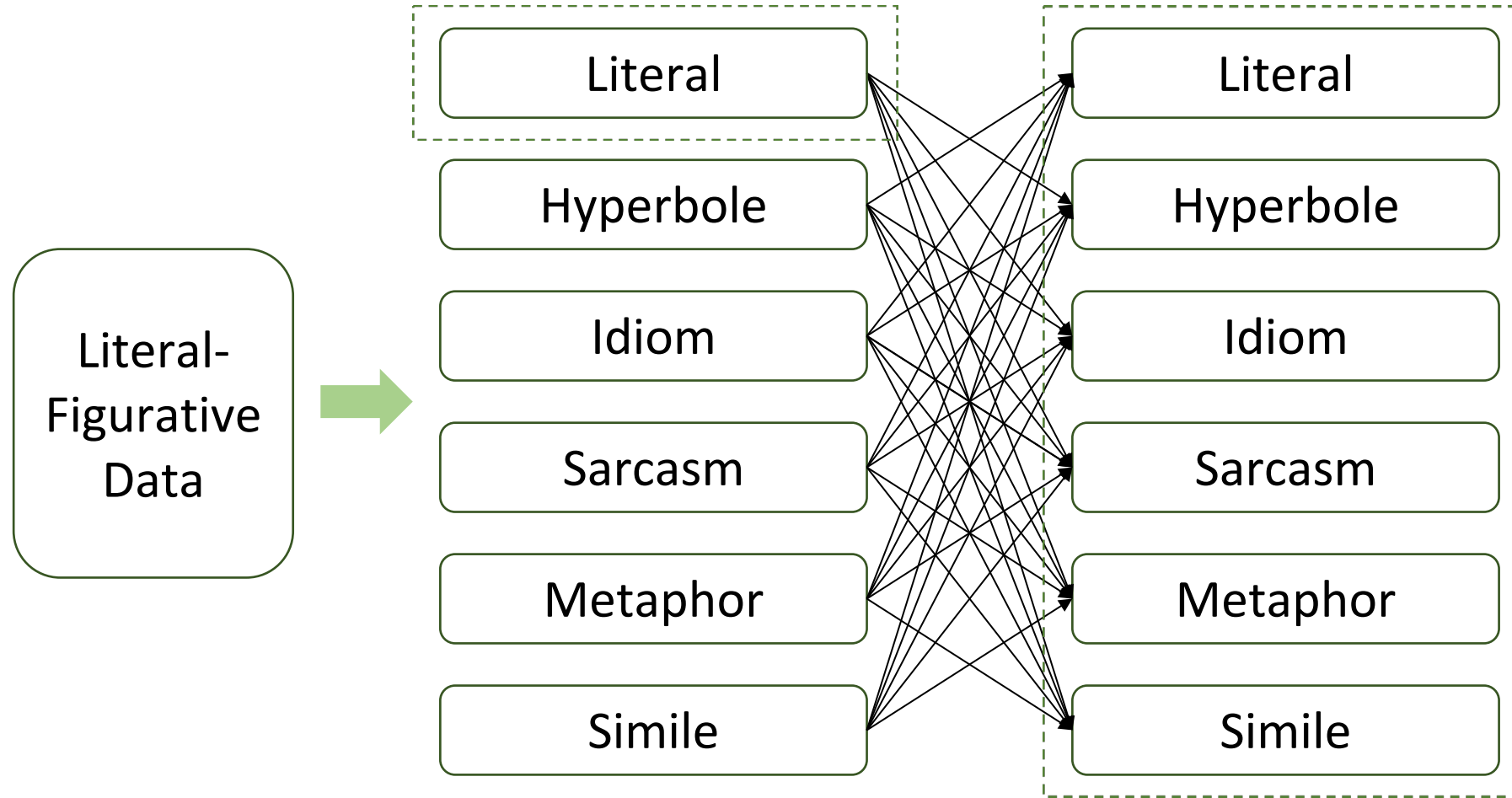
# Multi-Figurative Language Generation



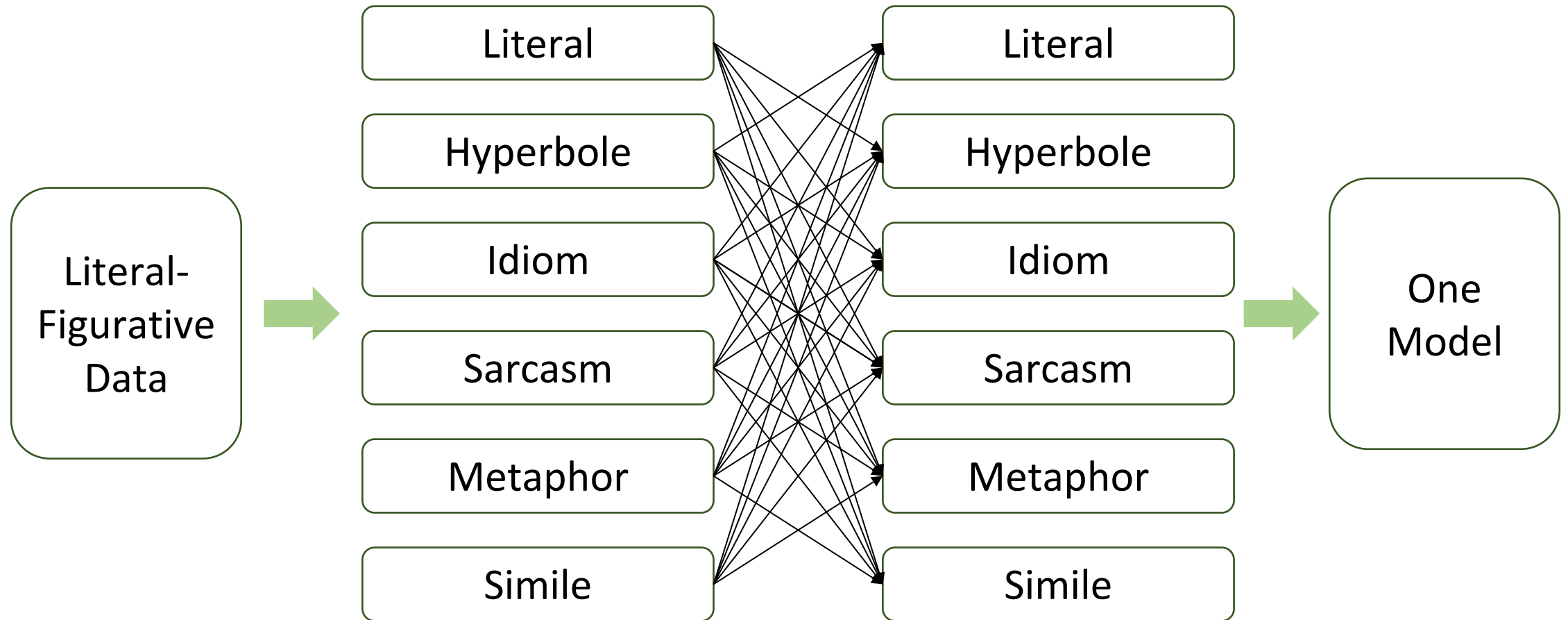
# Multi-Figurative Language Generation



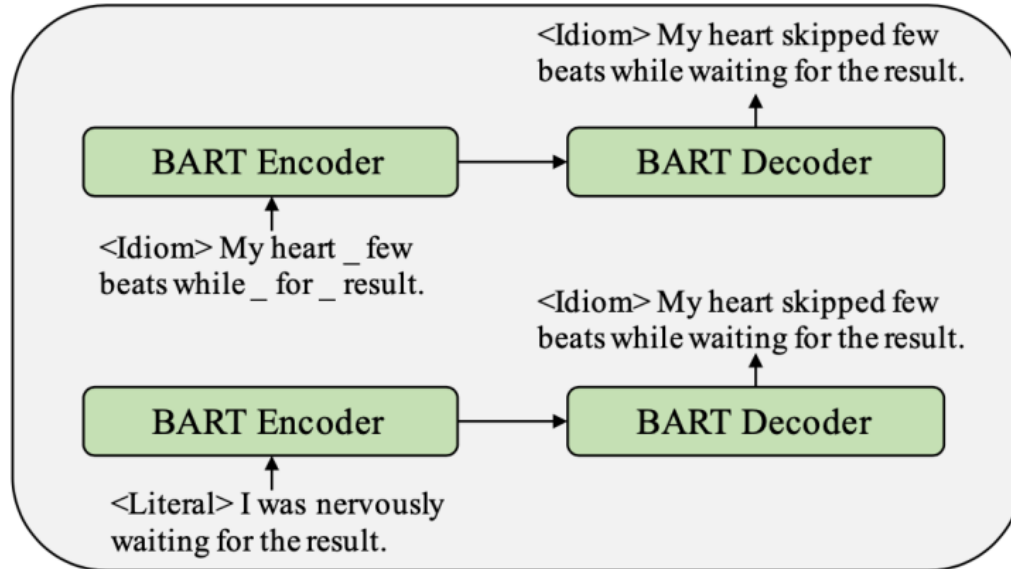
# Multi-Figurative Language Generation



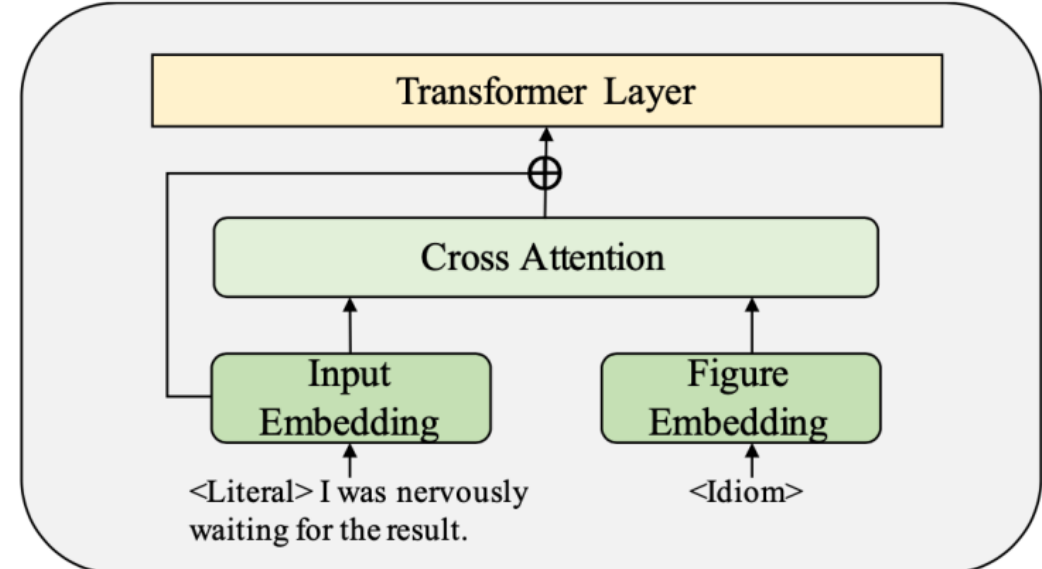
# Multi-Figurative Language Generation



# Modelling: Overview



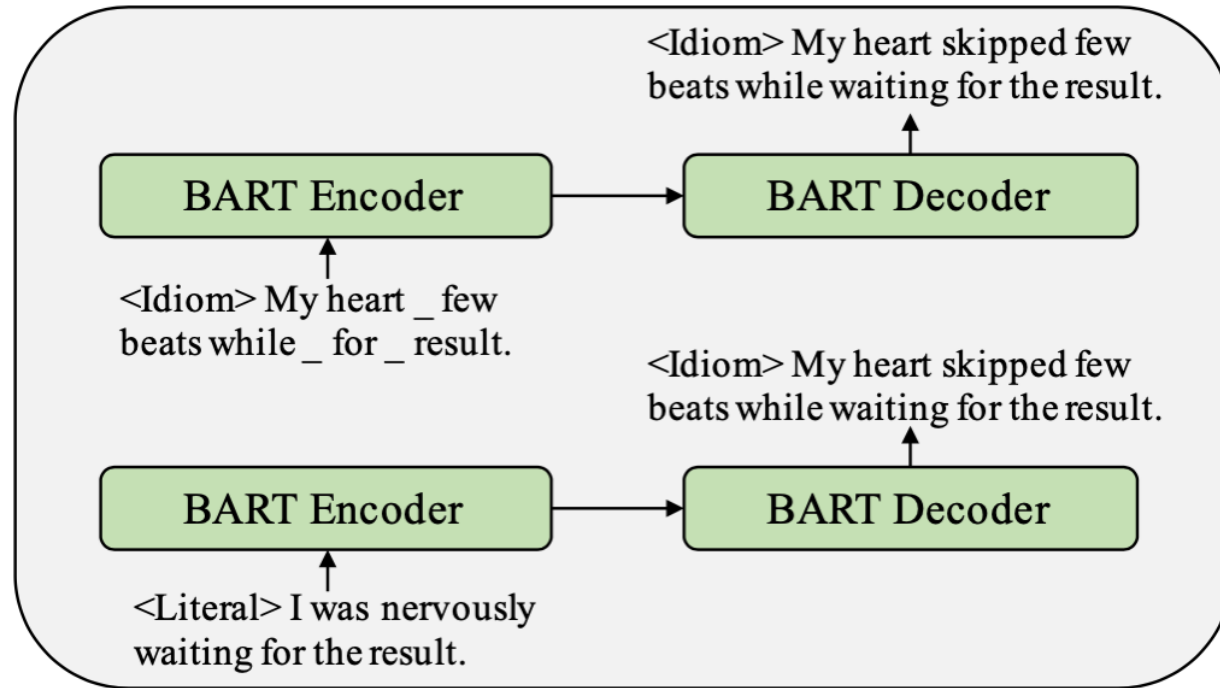
(a) Multi-figurative language denoising pre-training and fine-tuning.



(b) An overview of the mechanism for injecting the figurative information into the Encoder.

Overview of **multi-figurative language generation (mFLAG)**.

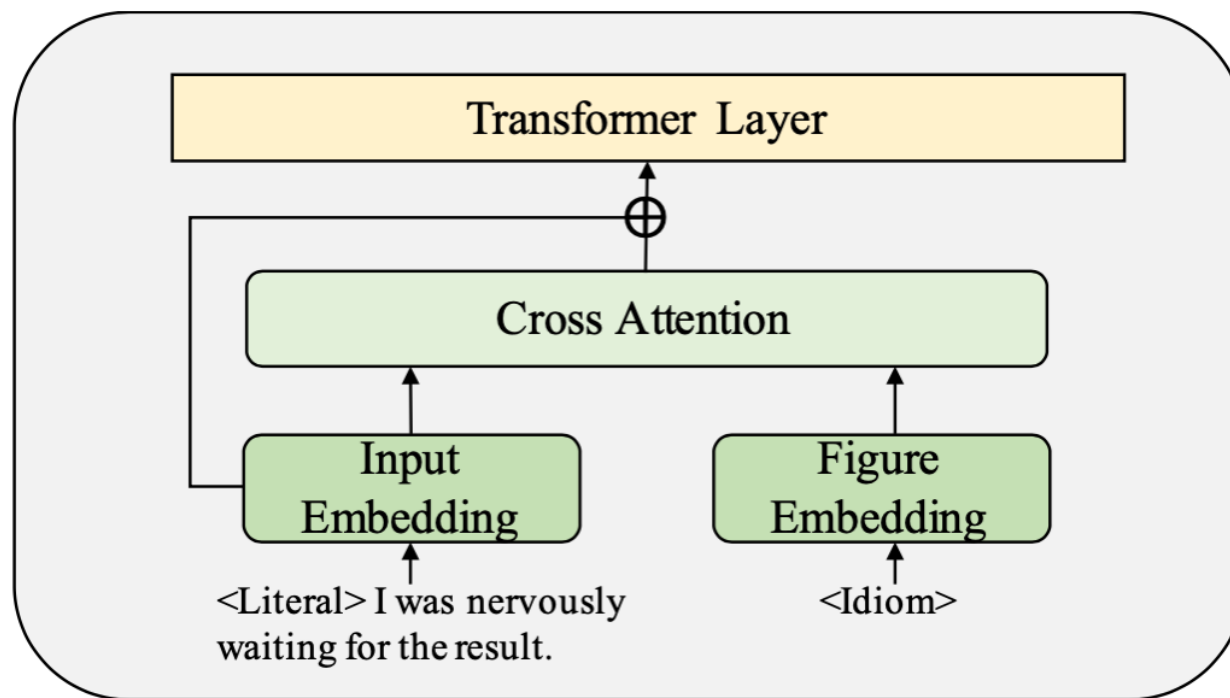
# Modelling: Pre-Training and Fine-Tuning



Denoising Training:  
Selected Paraphrase Data  
[PARABANK 2]

Fine-Tuning

# Modelling: Figurative Attention Mechanism



$$\text{CrossAttn}(\mathbf{W}, \mathbf{F}) = \text{softmax}\left(\frac{\mathbf{W}\mathbf{F}^T}{\sqrt{d}}\right)\mathbf{F}$$

$$\mathbf{C} = \text{CrossAttn}(\mathbf{W}, \mathbf{F}) + \mathbf{W}$$

where  $\mathbf{W} \in \mathbb{R}^{m \times d}$  represents the embedding of the source sentence.  $\mathbf{F} \in \mathbb{R}^{1 \times d}$  is the embedding of the target form code  $T$ .

# Modelling: **mFLAG** and Baselines

## **mFLAG:**

- mFLAG-DR: figurative-figurative generation
- mFLAG-BT: figurative-literal-figurative generation

**PT-to-FT:** mFLAG without figurative attention mechanism



# Modelling: mFLAG and Baselines

**mFLAG:** Our model

**PT-to-FT:** mFLAG without figurative attention mechanism

**BART-Single:** For each figure of speech, we fine-tune BART on the corresponding parallel literal-figurative data (figurative-literal-figurative generation)

**BART-Multi:** We concatenate the five parallel training sets and fine-tune BART for multi-figurative language modelling

# Evaluation: Automatic Metrics

**Form Strength:** BERT based classifier

**Context Preservation:**

- BLEU; BERT Score (following previous work)
- BLEURT; COMET (learnable metrics)

**Overall:** HM (the harmonic mean of figurative accuracy and BLEU score)

# Results: Literal-to-Figurative Generation

	TGT	BLEU	BERT	BLEURT	COMET	HM	TGT	BLEU	BERT	BLEURT	COMET	HM
	Literal Form→Hyperbole						Literal Form→Idiom					
BART-Single	0.627	0.513	0.693	0.280	0.461	0.564	0.711	<b>0.791</b>	<b>0.855</b>	<b>0.595</b>	<b>0.808</b>	0.749
BART-Multi	0.707	0.541	0.698	0.260	0.352	0.613	0.637	0.747	0.829	0.498	0.706	0.688
PT-to-FT	0.833	<b>0.582</b>	<b>0.733</b>	<b>0.379</b>	<b>0.490</b>	<b>0.686</b>	<b>0.769</b>	0.765	0.841	0.536	0.738	<b>0.767</b>
mFLAG	<b>0.844</b>	0.556	0.726	0.349	0.463	0.670	0.764	0.761	0.839	0.539	0.735	0.762
	Literal Form→Sarcasm						Literal Form→Metaphor					
BART-Single	0.679	<b>0.491</b>	<b>0.611</b>	<b>0.052</b>	<b>0.188</b>	0.570	0.720	0.595	0.771	0.364	0.720	0.652
BART-Multi	0.743	0.483	0.598	0.011	0.137	0.585	0.767	0.577	0.780	0.434	0.785	0.659
PT-to-FT	<b>0.765</b>	0.485	0.609	0.040	0.162	<b>0.594</b>	0.867	<b>0.643</b>	<b>0.812</b>	<b>0.493</b>	0.842	<b>0.738</b>
mFLAG	0.762	0.487	0.609	0.043	0.169	<b>0.594</b>	<b>0.880</b>	0.628	0.809	0.490	<b>0.844</b>	0.733
	Literal Form→Simile						Figurative→Literal Form					
BART-Single	0.647	0.724	0.720	<b>0.017</b>	<b>0.321</b>	0.683	0.733	0.606	0.742	0.284	0.455	0.663
BART-Multi	0.420	0.658	0.681	-0.025	0.178	0.513	0.725	0.622	0.762	0.364	0.522	0.670
PT-to-FT	0.907	0.729	0.722	-0.021	0.219	0.808	<b>0.801</b>	0.634	0.766	<b>0.542</b>	0.544	<b>0.708</b>
mFLAG	<b>0.953</b>	<b>0.745</b>	<b>0.727</b>	-0.021	0.220	<b>0.836</b>	0.796	<b>0.637</b>	<b>0.769</b>	0.375	<b>0.681</b>	0.707

Table 4: Results of literal↔figurative form generation. TGT represents the accuracy of output labeled as the target form by the classifier; the results of figurative→literal form generation are averaged across all figures of speech.

# Results: Literal-to-Figurative Generation

	TGT	BLEU	BERT	BLEURT	COMET	HM	TGT	BLEU	BERT	BLEURT	COMET	HM
	Literal Form→Hyperbole						Literal Form→Idiom					
BART-Single	0.627	0.513	0.693	0.280	0.461	0.564	0.711	<b>0.791</b>	<b>0.855</b>	<b>0.595</b>	<b>0.808</b>	0.749
BART-Multi	0.707	0.541	0.698	0.260	0.352	0.613	0.637	0.747	0.829	0.498	0.706	0.688
PT-to-FT	0.833	<b>0.582</b>	<b>0.733</b>	<b>0.379</b>	<b>0.490</b>	<b>0.686</b>	<b>0.769</b>	0.765	0.841	0.536	0.738	<b>0.767</b>
mFLAG	<b>0.844</b>	0.556	0.726	0.349	0.463	0.670	0.764	0.761	0.839	0.539	0.735	0.762
	Literal Form→Sarcasm						Literal Form→Metaphor					
BART-Single	0.679	<b>0.491</b>	<b>0.611</b>	<b>0.052</b>	<b>0.188</b>	0.570	0.720	0.595	0.771	0.364	0.720	0.652
BART-Multi	0.743	0.483	0.598	0.011	0.137	0.585	0.767	0.577	0.780	0.434	0.785	0.659
PT-to-FT	<b>0.765</b>	0.485	0.609	0.040	0.162	<b>0.594</b>	0.867	<b>0.643</b>	<b>0.812</b>	<b>0.493</b>	0.842	<b>0.738</b>
mFLAG	0.762	0.487	0.609	0.043	0.169	<b>0.594</b>	<b>0.880</b>	0.628	0.809	0.490	<b>0.844</b>	0.733
	Literal Form→Simile						Figurative→Literal Form					
BART-Single	0.647	0.724	0.720	<b>0.017</b>	<b>0.321</b>	0.683	0.733	0.606	0.742	0.284	0.455	0.663
BART-Multi	0.420	0.658	0.681	-0.025	0.178	0.513	0.725	0.622	0.762	0.364	0.522	0.670
PT-to-FT	0.907	0.729	0.722	-0.021	0.219	0.808	<b>0.801</b>	0.634	0.766	<b>0.542</b>	0.544	<b>0.708</b>
mFLAG	<b>0.953</b>	<b>0.745</b>	<b>0.727</b>	-0.021	0.220	<b>0.836</b>	0.796	<b>0.637</b>	<b>0.769</b>	0.375	<b>0.681</b>	0.707

Table 4: Results of literal↔figurative form generation. TGT represents the accuracy of output labeled as the target form by the classifier; the results of figurative→literal form generation are averaged across all figures of speech.

# Results: Literal-to-Figurative Generation

	TGT	BLEU	BERT	BLEURT	COMET	HM	TGT	BLEU	BERT	BLEURT	COMET	HM
	Literal Form→Hyperbole						Literal Form→Idiom					
BART-Single	0.627	0.513	0.693	0.280	0.461	0.564	0.711	<b>0.791</b>	<b>0.855</b>	<b>0.595</b>	<b>0.808</b>	0.749
BART-Multi	0.707	0.541	0.698	0.260	0.352	0.613	0.637	0.747	0.829	0.498	0.706	0.688
PT-to-FT	0.833	<b>0.582</b>	<b>0.733</b>	<b>0.379</b>	<b>0.490</b>	<b>0.686</b>	<b>0.769</b>	0.765	0.841	0.536	0.738	<b>0.767</b>
mFLAG	<b>0.844</b>	0.556	0.726	0.349	0.463	0.670	0.764	0.761	0.839	0.539	0.735	0.762
	Literal Form→Sarcasm						Literal Form→Metaphor					
BART-Single	0.679	<b>0.491</b>	<b>0.611</b>	<b>0.052</b>	<b>0.188</b>	0.570	0.720	0.595	0.771	0.364	0.720	0.652
BART-Multi	0.743	0.483	0.598	0.011	0.137	0.585	0.767	0.577	0.780	0.434	0.785	0.659
PT-to-FT	<b>0.765</b>	0.485	0.609	0.040	0.162	<b>0.594</b>	0.867	<b>0.643</b>	<b>0.812</b>	<b>0.493</b>	0.842	<b>0.738</b>
mFLAG	0.762	0.487	0.609	0.043	0.169	<b>0.594</b>	<b>0.880</b>	0.628	0.809	0.490	<b>0.844</b>	0.733
	Literal Form→Simile						Figurative→Literal Form					
BART-Single	0.647	0.724	0.720	<b>0.017</b>	<b>0.321</b>	0.683	0.733	0.606	0.742	0.284	0.455	0.663
BART-Multi	0.420	0.658	0.681	-0.025	-0.178	0.513	0.725	0.622	0.762	0.364	0.522	0.670
PT-to-FT	0.907	0.729	0.722	-0.021	0.219	0.808	<b>0.801</b>	0.634	0.766	<b>0.542</b>	0.544	<b>0.708</b>
mFLAG	<b>0.953</b>	<b>0.745</b>	<b>0.727</b>	-0.021	0.220	<b>0.836</b>	0.796	<b>0.637</b>	<b>0.769</b>	0.375	<b>0.681</b>	0.707

Table 4: Results of literal↔figurative form generation. TGT represents the accuracy of output labeled as the target form by the classifier; the results of figurative→literal form generation are averaged across all figures of speech.

# Results: Figurative-to-Figurative Generation

	Form Strength		Source Text					Literal Text				
	SRC	TGT	BLEU	BERT	BLEURT	COMET	HM	BLEU	BERT	BLEURT	COMET	HM
Hyperbole→Others												
BART-Single	0.470	0.425	0.665	0.782	0.459	0.472	0.519	0.488	0.700	0.294	0.248	0.454
BART-Multi	0.328	0.242	0.602	0.761	0.455	0.443	0.345	0.505	0.731	0.427	0.385	0.327
PT-to-FT	0.252	0.258	0.590	0.749	0.437	0.420	0.359	<b>0.507</b>	<b>0.732</b>	<b>0.438</b>	<b>0.407</b>	0.342
mFLAG-DR	<b>0.922</b>	0.608	<b>0.815</b>	<b>0.893</b>	<b>0.753</b>	<b>0.836</b>	<b>0.696</b>	0.411	0.633	0.036	-0.105	0.490
mFLAG-BT	0.482	<b>0.644</b>	0.539	0.702	0.253	0.246	0.586	0.421	0.662	0.169	0.093	<b>0.509</b>
Idiom→Others												
BART-Single	0.290	0.309	0.783	0.864	0.575	0.646	0.443	0.749	0.844	0.578	0.659	0.438
BART-Multi	0.273	0.204	0.785	0.873	0.602	0.674	0.324	0.758	0.859	0.630	0.701	0.408
PT-to-FT	0.204	0.207	0.771	0.867	0.594	0.662	0.326	<b>0.760</b>	<b>0.860</b>	<b>0.646</b>	<b>0.715</b>	0.325
mFLAG-DR	<b>0.910</b>	0.400	<b>0.901</b>	<b>0.940</b>	<b>0.822</b>	<b>0.869</b>	<b>0.554</b>	0.694	0.799	0.328	0.375	0.507
mFLAG-BT	0.328	<b>0.409</b>	0.724	0.831	0.491	0.566	0.523	0.703	0.816	0.490	0.569	<b>0.517</b>

Table 5: Results of figurative↔figurative form generation. Notes: (i) SRC (TGT) represents the accuracy of output labeled as the source (target) form by the classifier of the source (target) form; (ii) results for each block are averaged for all generations from one figurative language to others.



# Results: Figurative-to-Figurative Generation

	Form Strength		Source Text					Literal Text				
	SRC	TGT	BLEU	BERT	BLEURT	COMET	HM	BLEU	BERT	BLEURT	COMET	HM
Sarcasm→Others												
BART-Single	0.577	0.370	0.877	0.899	0.650	0.792	0.520	0.454	0.579	-0.088	-0.051	0.408
BART-Multi	0.569	0.247	0.903	0.923	0.701	0.838	0.388	0.471	<b>0.593</b>	-0.049	-0.014	0.324
PT-to-FT	0.464	0.252	0.863	0.891	0.613	0.774	0.390	<b>0.468</b>	0.592	<b>-0.031</b>	<b>0.000</b>	0.328
mFLAG-DR	<b>0.840</b>	0.438	<b>0.907</b>	<b>0.928</b>	<b>0.813</b>	<b>0.872</b>	0.591	0.442	0.563	-0.198	-0.143	0.440
mFLAG-BT	0.583	<b>0.481</b>	0.808	0.831	0.460	0.604	<b>0.605</b>	0.430	0.554	-0.164	-0.133	<b>0.454</b>
Metaphor→Others												
BART-Single	0.163	0.314	0.603	0.776	0.412	0.555	0.413	0.575	0.773	0.381	0.486	0.406
BART-Multi	0.255	0.249	0.647	0.825	0.554	0.723	0.360	0.632	0.820	0.550	0.689	0.357
PT-to-FT	0.147	0.254	0.671	0.832	0.599	<b>0.763</b>	0.369	<b>0.648</b>	<b>0.824</b>	<b>0.507</b>	<b>0.665</b>	0.365
mFLAG-DR	<b>0.795</b>	0.518	<b>0.697</b>	<b>0.846</b>	<b>0.614</b>	0.706	<b>0.594</b>	0.516	0.758	0.320	0.410	0.517
mFLAG-BT	0.387	<b>0.557</b>	0.502	0.734	0.329	0.434	0.528	0.496	0.743	0.317	0.417	<b>0.525</b>
Simile→Others												
BART-Single	0.057	0.607	0.469	0.559	-0.406	-0.429	0.529	0.588	0.667	0.160	-0.102	0.597
BART-Multi	0.007	0.272	0.629	0.686	-0.043	-0.051	0.380	<b>0.765</b>	<b>0.818</b>	<b>0.262</b>	<b>0.415</b>	0.401
PT-to-FT	0.000	0.314	0.622	0.671	-0.031	-0.067	0.417	0.754	0.804	0.244	0.394	0.443
mFLAG-DR	<b>0.440</b>	0.685	<b>0.849</b>	<b>0.884</b>	<b>0.637</b>	<b>0.690</b>	<b>0.758</b>	0.589	0.698	-0.016	-0.057	0.633
mFLAG-BT	0.132	<b>0.687</b>	0.606	0.670	-0.069	-0.064	0.644	0.672	0.766	0.163	0.250	<b>0.679</b>

# Results: Figurative-to-Figurative Generation

	Form Strength		Source Text					Literal Text				
	SRC	TGT	BLEU	BERT	BLEURT	COMET	HM	BLEU	BERT	BLEURT	COMET	HM
Hyperbole→Others												
BART-Single	0.470	0.425	0.665	0.782	0.459	0.472	0.519	0.488	0.700	0.294	0.248	0.454
BART-Multi	0.328	0.242	0.602	0.761	0.455	0.443	0.345	0.505	0.731	0.427	0.385	0.327
PT-to-FT	<u>0.252</u>	<u>0.258</u>	<u>0.590</u>	<u>0.749</u>	<u>0.437</u>	<u>0.420</u>	<u>0.359</u>	<b>0.507</b>	<b>0.732</b>	<b>0.438</b>	<b>0.407</b>	<u>0.342</u>
mFLAG-DR	<b>0.922</b>	0.608	<b>0.815</b>	<b>0.893</b>	<b>0.753</b>	<b>0.836</b>	<b>0.696</b>	0.411	0.633	0.036	-0.105	0.490
mFLAG-BT	0.482	<b>0.644</b>	0.539	0.702	0.253	0.246	0.586	0.421	0.662	0.169	0.093	<b>0.509</b>
Idiom→Others												
BART-Single	0.290	0.309	0.783	0.864	0.575	0.646	0.443	0.749	0.844	0.578	0.659	0.438
BART-Multi	0.273	0.204	0.785	0.873	0.602	0.674	0.324	0.758	0.859	0.630	0.701	0.408
PT-to-FT	<u>0.204</u>	<u>0.207</u>	<u>0.771</u>	<u>0.867</u>	<u>0.594</u>	<u>0.662</u>	<u>0.326</u>	<b>0.760</b>	<b>0.860</b>	<b>0.646</b>	<b>0.715</b>	<u>0.325</u>
mFLAG-DR	<b>0.910</b>	0.400	<b>0.901</b>	<b>0.940</b>	<b>0.822</b>	<b>0.869</b>	<b>0.554</b>	0.694	0.799	0.328	0.375	0.507
mFLAG-BT	0.328	<b>0.409</b>	0.724	0.831	0.491	0.566	0.523	0.703	0.816	0.490	0.569	<b>0.517</b>

Table 5: Results of figurative↔figurative form generation. Notes: (i) SRC (TGT) represents the accuracy of output labeled as the source (target) form by the classifier of the source (target) form; (ii) results for each block are averaged for all generations from one figurative language to others.



# Results: Figurative-to-Figurative Generation

	Form Strength		Source Text					Literal Text				
	SRC	TGT	BLEU	BERT	BLEURT	COMET	HM	BLEU	BERT	BLEURT	COMET	HM
Sarcasm→Others												
BART-Single	0.577	0.370	0.877	0.899	0.650	0.792	0.520	0.454	0.579	-0.088	-0.051	0.408
BART-Multi	0.569	0.247	0.903	0.923	0.701	0.838	0.388	0.471	<b>0.593</b>	-0.049	-0.014	0.324
PT-to-FT	0.464	0.252	0.863	0.891	0.613	0.774	0.390	<b>0.468</b>	<b>0.592</b>	<b>-0.031</b>	<b>0.000</b>	0.328
mFLAG-DR	<b>0.840</b>	0.438	<b>0.907</b>	<b>0.928</b>	<b>0.813</b>	<b>0.872</b>	0.591	0.442	0.563	-0.198	-0.143	0.440
mFLAG-BT	0.583	<b>0.481</b>	0.808	0.831	0.460	0.604	<b>0.605</b>	0.430	0.554	-0.164	-0.133	<b>0.454</b>
Metaphor→Others												
BART-Single	0.163	0.314	0.603	0.776	0.412	0.555	0.413	0.575	0.773	0.381	0.486	0.406
BART-Multi	0.255	0.249	0.647	0.825	0.554	0.723	0.360	0.632	0.820	0.550	0.689	0.357
PT-to-FT	0.147	0.254	0.671	0.832	0.599	<b>0.763</b>	0.369	<b>0.648</b>	<b>0.824</b>	<b>0.507</b>	<b>0.665</b>	0.365
mFLAG-DR	<b>0.795</b>	0.518	<b>0.697</b>	<b>0.846</b>	<b>0.614</b>	0.706	<b>0.594</b>	0.516	0.758	0.320	0.410	0.517
mFLAG-BT	0.387	<b>0.557</b>	0.502	0.734	0.329	0.434	0.528	0.496	0.743	0.317	0.417	<b>0.525</b>
Simile→Others												
BART-Single	0.057	0.607	0.469	0.559	-0.406	-0.429	0.529	0.588	0.667	0.160	-0.102	0.597
BART-Multi	0.007	0.272	0.629	0.686	-0.043	-0.051	0.380	<b>0.765</b>	<b>0.818</b>	<b>0.262</b>	<b>0.415</b>	0.401
PT-to-FT	0.000	0.314	0.622	0.671	-0.031	-0.067	0.417	0.754	0.804	0.244	0.394	0.443
mFLAG-DR	<b>0.440</b>	0.685	<b>0.849</b>	<b>0.884</b>	<b>0.637</b>	<b>0.690</b>	<b>0.758</b>	0.589	0.698	-0.016	-0.057	0.633
mFLAG-BT	0.132	<b>0.687</b>	0.606	0.670	-0.069	-0.064	0.644	0.672	0.766	0.163	0.250	<b>0.679</b>

# Results: Example Outputs

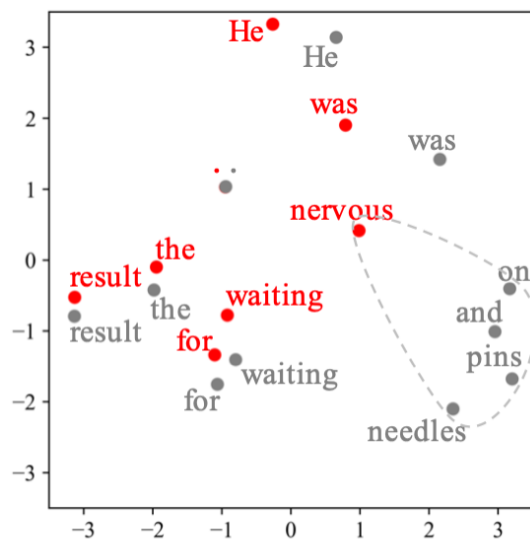
Forms	Models	Sentences	Suc.
Hyperbole	Input	I am not happy that he urged me to finish <b>all the hardest tasks in the world.</b>	-
Literal	BART-Single	I am not happy that he urged me to finish all the hard tasks in the world.	X
	BART-Multi	I am glad that he urged me to finish all the hardest tasks in the world.	X
	PT-to-FT	I am not happy that he urged me to complete all the difficult tasks in the world.	X
	mFLAG	I am not happy that he urged me to finish <b>all the difficult tasks.</b>	✓
Idiom	BART-Single	I am not happy that he urged me to finish all the hard tasks in the world.	X
	BART-Multi	I am so glad that he urged me to finish all the hardest tasks in the world.	X
	PT-to-FT	I am not happy that he urged me to complete all the difficult tasks.	X
	mFLAG-DR	I am not happy that he urged me to <b>tick off</b> all the hardest tasks in the world.	✓
	mFLAG-BT	I am not happy that he urged me to <b>get cracking</b> on all the difficult tasks.	✓

Table 6: Examples outputs generated by various models from hyperbolic text, where **red** denotes appropriate words/phrases for desired forms. Suc.==Successful.

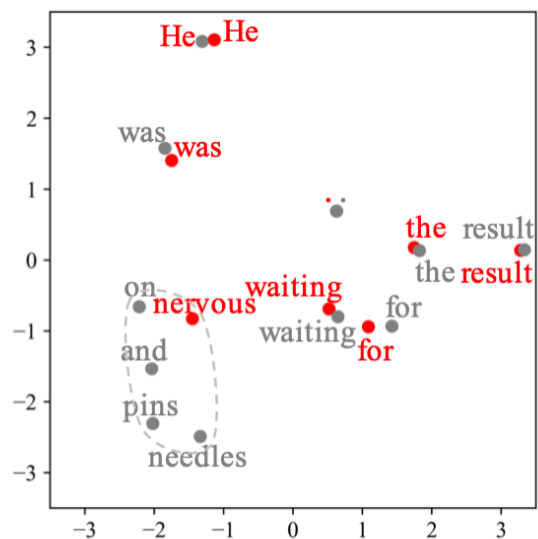
# Results: Example Outputs

Forms	Models	Sentences	Suc.
Hyperbole	Input	I am not happy that he urged me to finish <b>all the hardest tasks in the world.</b>	-
Sarcasm	BART-Single	I'm <b>so glad</b> my dad encouraged me to finish all the hard tasks in the world.	✓
	BART-Multi	<b>Thank you for encouraging me</b> to finish all the hardest tasks in the world.	✓
	PT-to-FT	I am <b>happy</b> that he urged me to finish all the hardest tasks in the world.	✓
	mFLAG-DR	<b>Glad</b> he urged me to finish all the hardest tasks in the world.	✓
	mFLAG-BT	<b>Glad</b> he urged me to finish all the difficult tasks.	✓
Metaphor	BART-Single	I am not happy that he urged me to <b>bear</b> all the difficult tasks.	✓
	BART-Multi	I am so glad that he urged me to finish all the hardest tasks in the world.	✗
	PT-to-FT	I am not happy that he urged me to complete all the difficult tasks in the world.	✗
	mFLAG-DR	I am not happy that he urged me to <b>bear</b> all the hardest tasks in the world.	✓
	mFLAG-BT	I am not happy that he pressed me to finish all the difficult tasks.	✗
Simile	BART-Single	I am not happy that he urged me to finish all the difficult tasks.	✗
	BART-Multi	I am so glad that he urged me to finish all the hardest tasks in the world.	✗
	PT-to-FT	I am not happy that he urged me to complete all the difficult tasks in the world.	✗
	mFLAG-DR	I am not happy that he urged me to finish all the <b>like a million things.</b>	✓
	mFLAG-BT	I am not happy that he urged me to finish all the difficult tasks.	✗

# Probing Figurative Information for Encoder (PCA)



(a) Result for PT-to-FT.

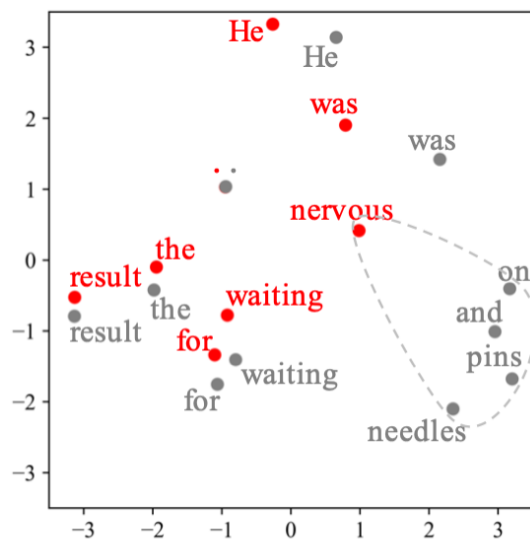


(b) Result for mFLAG.

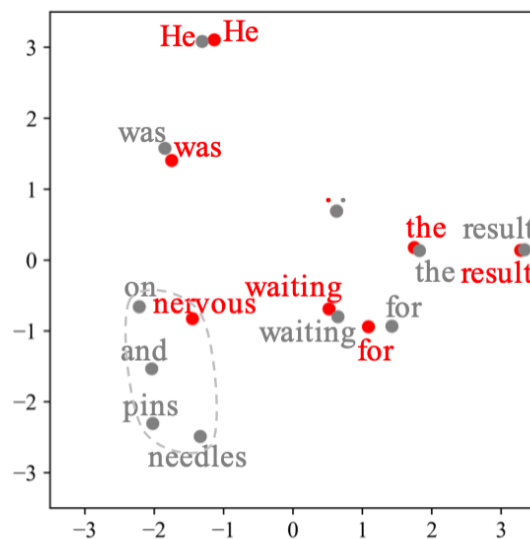
**Literal:** *He was nervous waiting for the result.*

**Hyperbole:** *He was on pins and needles waiting for the result.*

# Probing Figurative Information for Encoder (PCA)



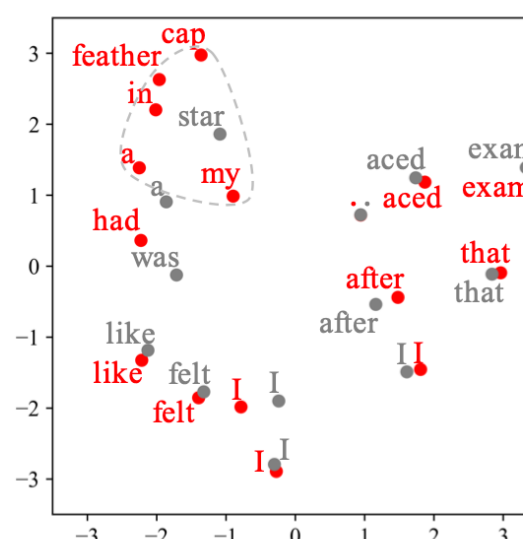
(a) Result for PT-to-FT.



(b) Result for mFLAG.

**Literal:** *He was nervous waiting for the result.*

**Hyperbole:** *He was on pins and needles waiting for the result.*



(c) Result for PT-to-FT.



(d) Result for mFLAG.

**Idiomatic:** *I felt like I had a feather in my cap after I aced that exam.*

**Hyperbole:** *I felt like I was a star after I aced that exam.*

# How Similar are Different Forms?

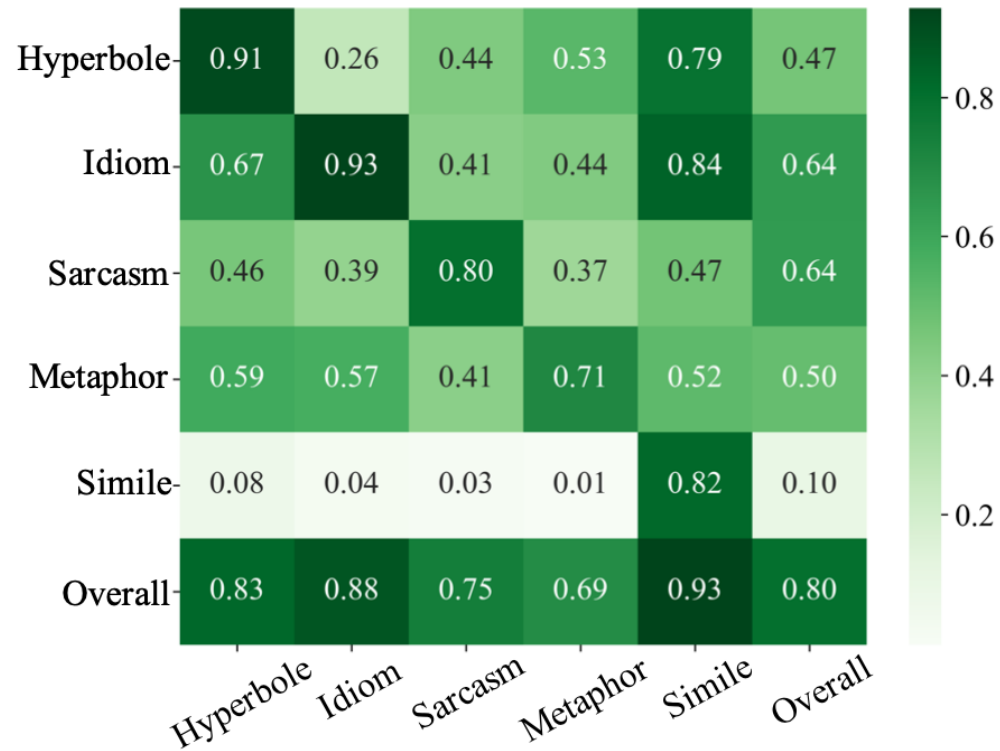


Figure 3: Performances (F1 score) of classifiers on different figurative forms. Each row represents results of a classifier tested on each/all figurative form(s).

# How Similar are Different Forms?

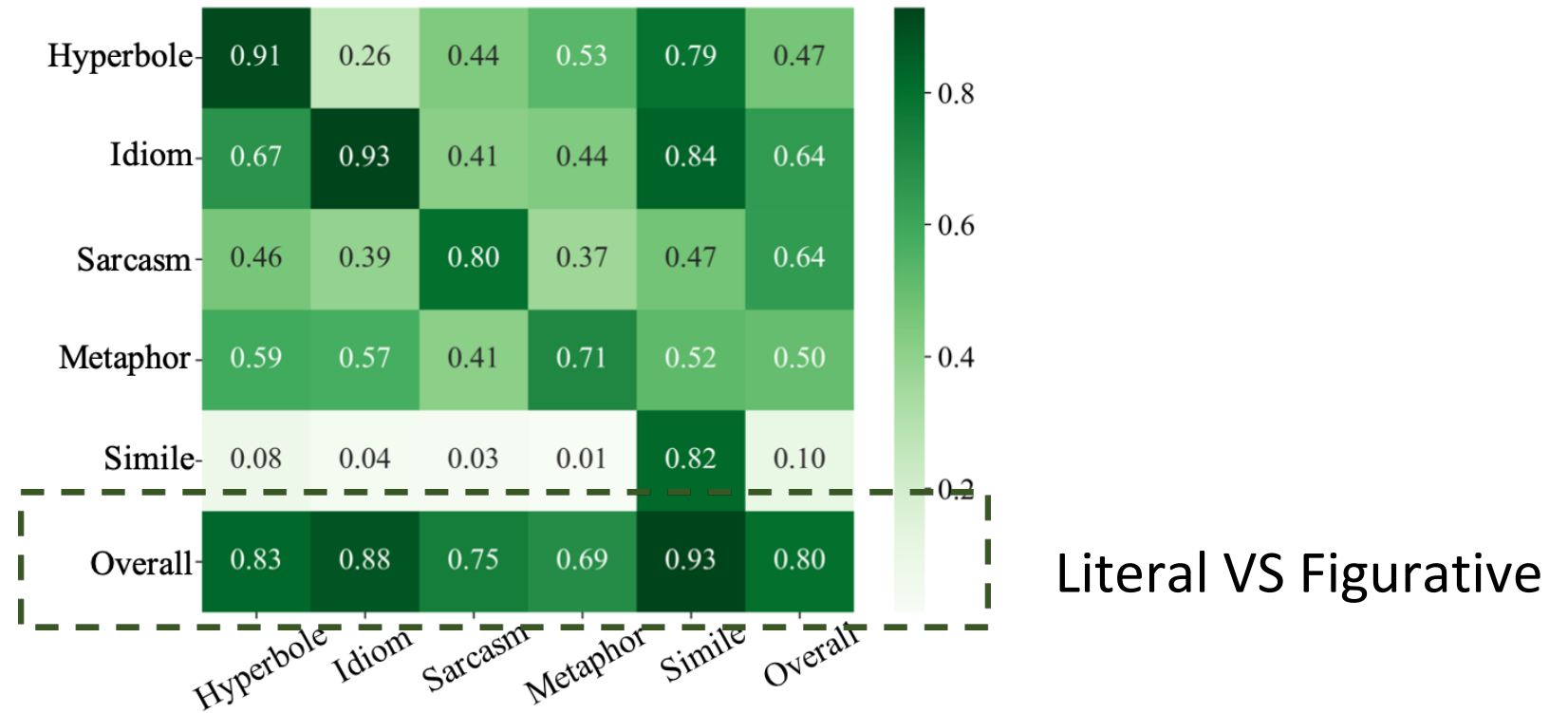


Figure 3: Performances (F1 score) of classifiers on different figurative forms. Each row represents results of a classifier tested on each/all figurative form(s).



# How Similar are Different Forms?

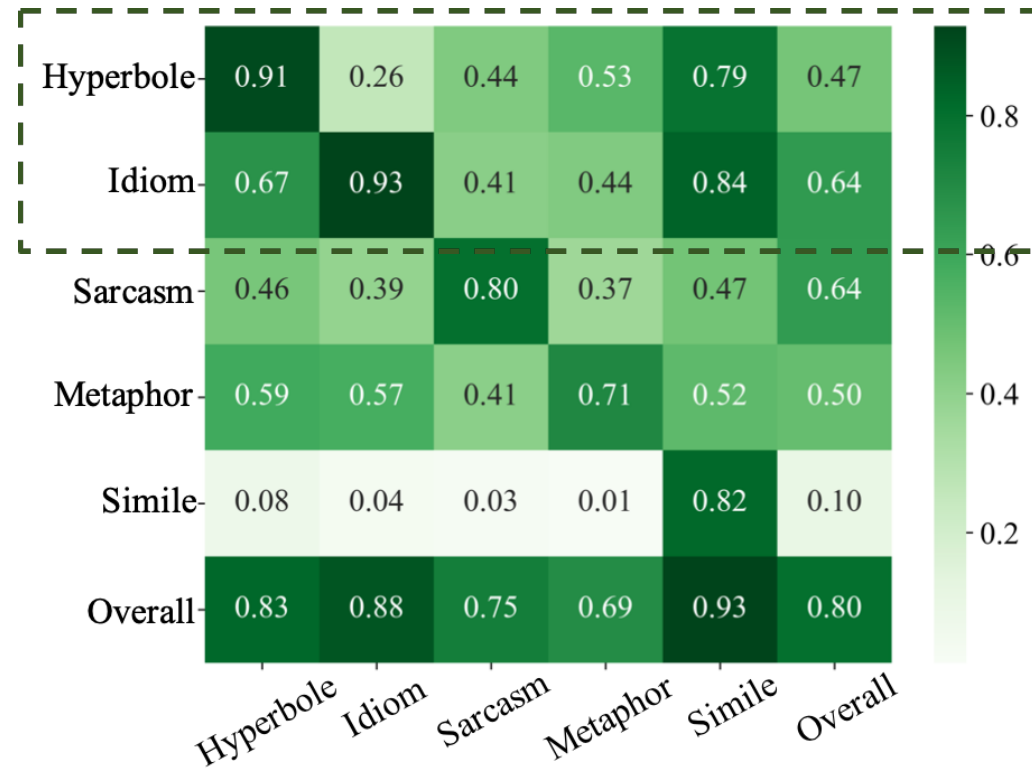


Figure 3: Performances (F1 score) of classifiers on different figurative forms. Each row represents results of a classifier tested on each/all figurative form(s).



# How Similar are Different Forms?

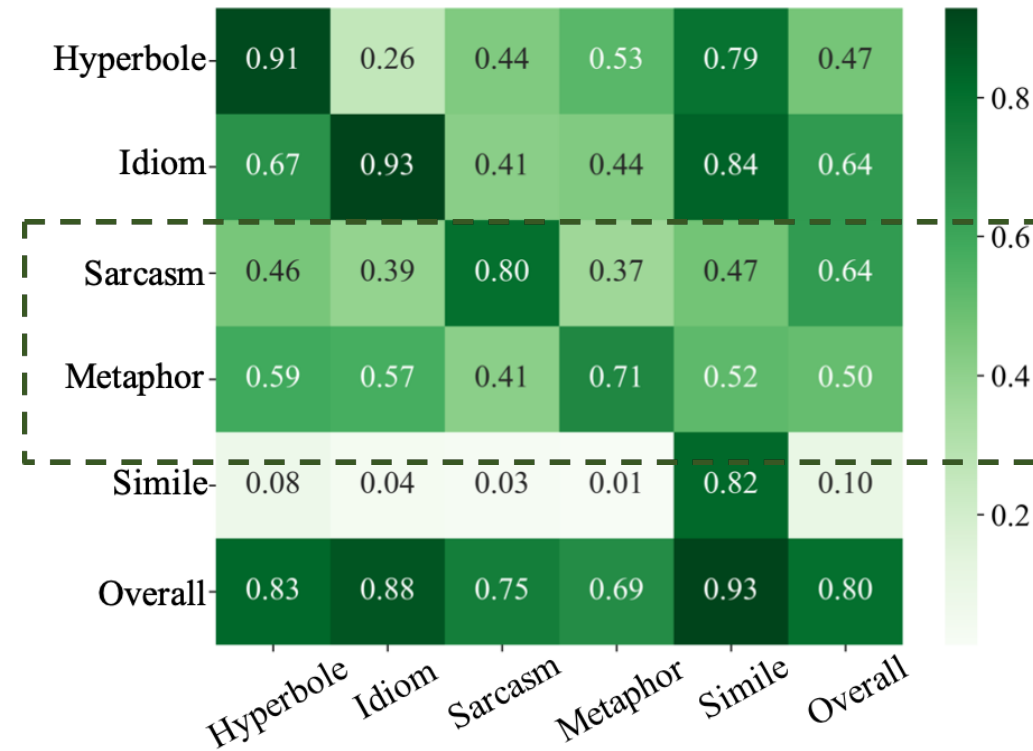
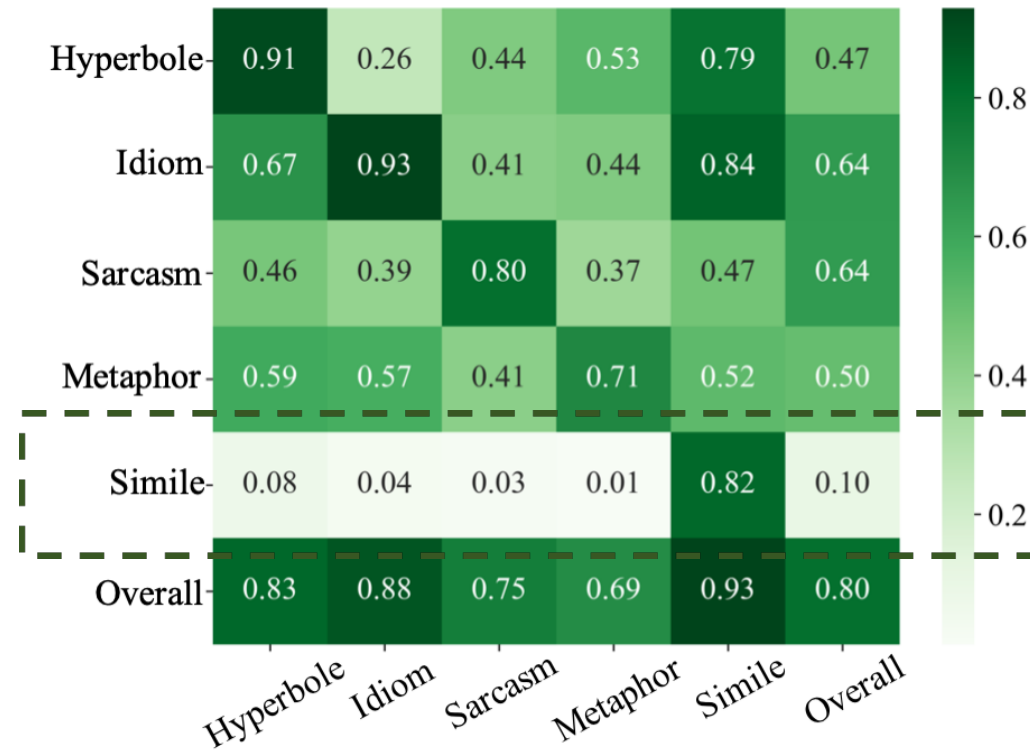


Figure 3: Performances (F1 score) of classifiers on different figurative forms. Each row represents results of a classifier tested on each/all figurative form(s).

# How Similar are Different Forms?



Simile: You can publish the whole thing *like a* diary.

# Conclusion and Outlook

## This Work:

- A novel task of multi-figurative language generation, and a corresponding benchmark
- An approach for this task with no need for parallel figurative-figurative data
- Data, code, model: <https://github.com/laihuiyuan/mFLAG>

## Outlook:

- More figures of speech
- Multilingual modelling
- Evaluation method



@HuiyuanLai @MalvinaNissim @GroNlp



huiyuanlai.l@gmail.com