MULTI-FIGURATIVE LANGUAGE GENERATION

Huiyuan Lai
h.lai@rug.nl

Malvina Nissim
m.nissim@rug.nl

university of groningen

NLG in the Lowlands
Utrecht, June 22nd 2023

FIGURATIVE LANGUAGE

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

Metaphor: He made a road of my broken works.

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*.

WHY FIGURATIVE LANGUAGE?

Figurative language 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: *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: *My niece will babysit for you for a little bit of money.*
Idiom: *My niece will babysit for you for pin money.*

Literal: *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

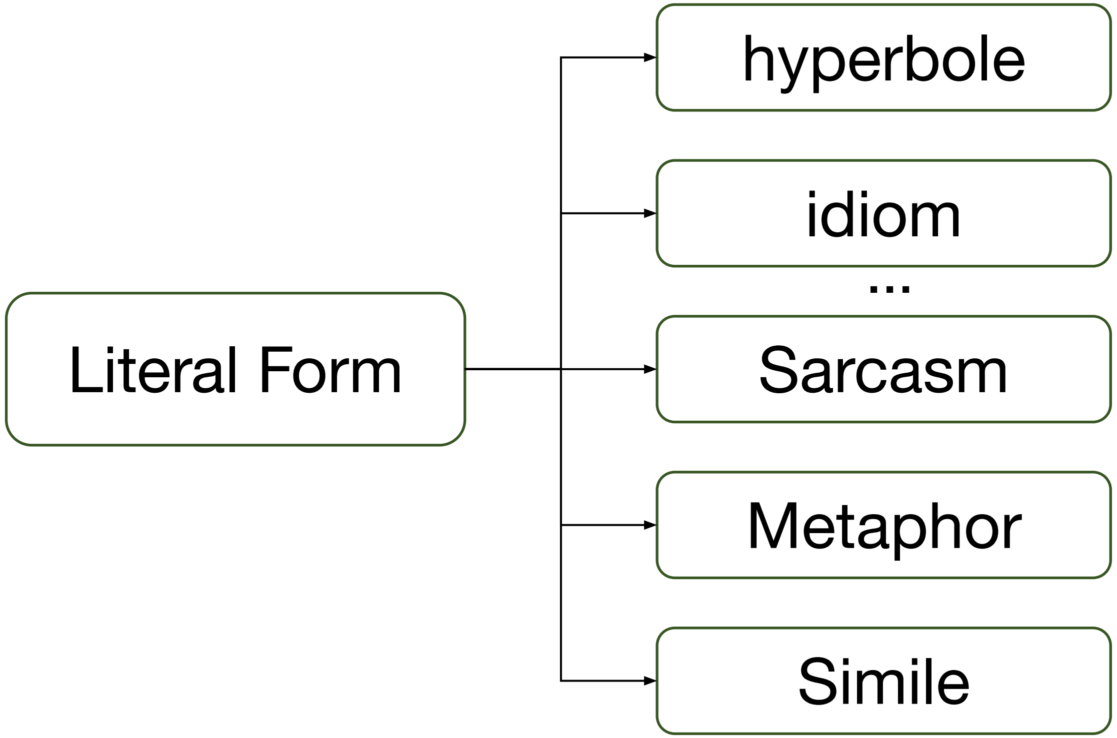
Literal: *Old Mr. Smith has been teaching here for a very long time.*
Hyperbole: *Old Mr. Smith has been teaching here since the Stone Age.*

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

Literal: *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.*

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

CURRENT STATUS



[Tian et al., 2021; Zhang and Wan, 2022]

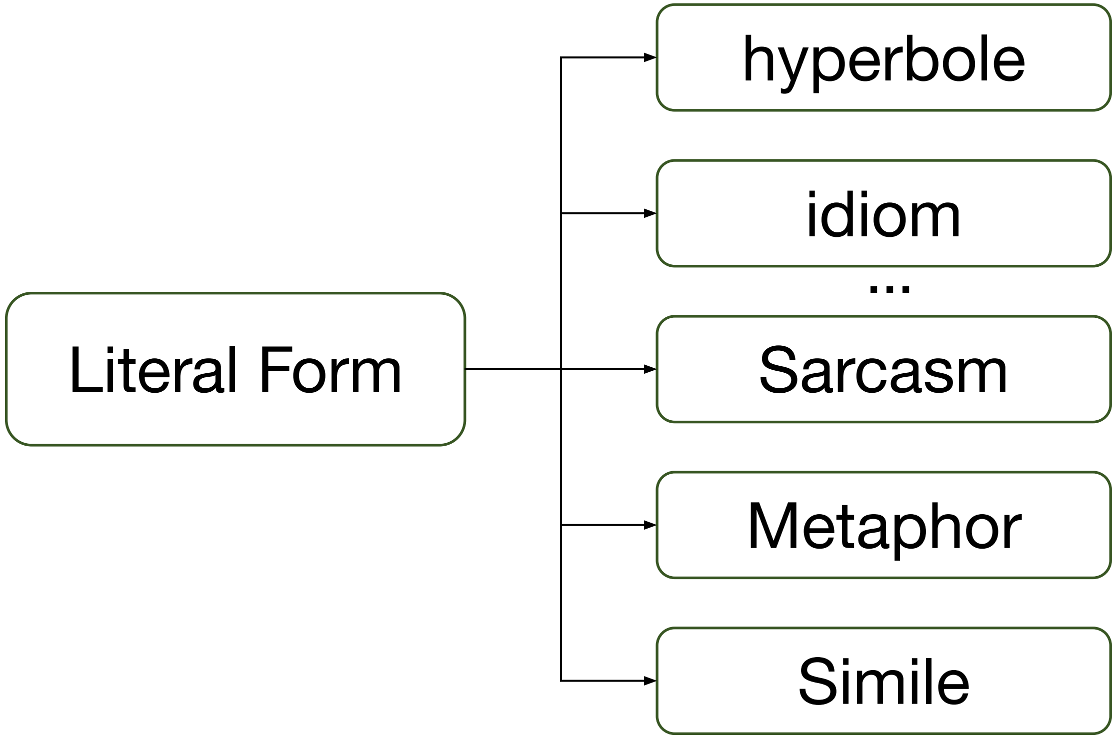
[Zhou et al., 2021]

[Zhu et al., 2021; Chakrabarty et al., 2020]

[Abe et al., 2006; Stowe et al., 2021]

[Chakrabarty et al., 2020]

CURRENT STATUS



[Tian et al., 2021; Zhang and Wan, 2022]

[Zhou et al., 2021]

[Zhu et al., 2021; Chakrabarty et al., 2020]

[Abe et al., 2006; Stowe et al., 2021]

[Chakrabarty et al., 2020]

- ❖ Separate models
- ❖ No cross-figurative knowledge transfer

MULTI-FIGURATIVE LANGUAGE GENERATION

- Can we model multiple figures of speech **jointly**?

MULTI-FIGURATIVE LANGUAGE GENERATION

- Can we model multiple figures of speech **jointly**?

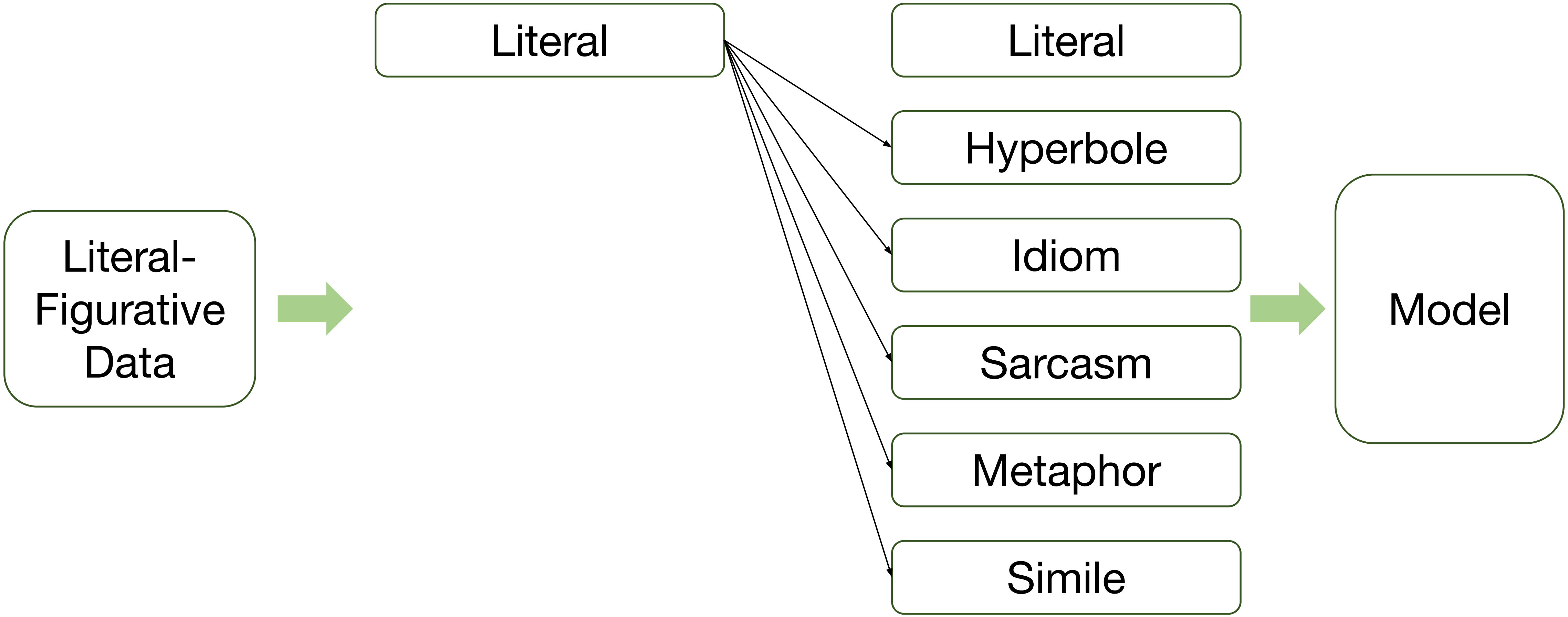
COLING 2022

Multi-Figurative Language Generation

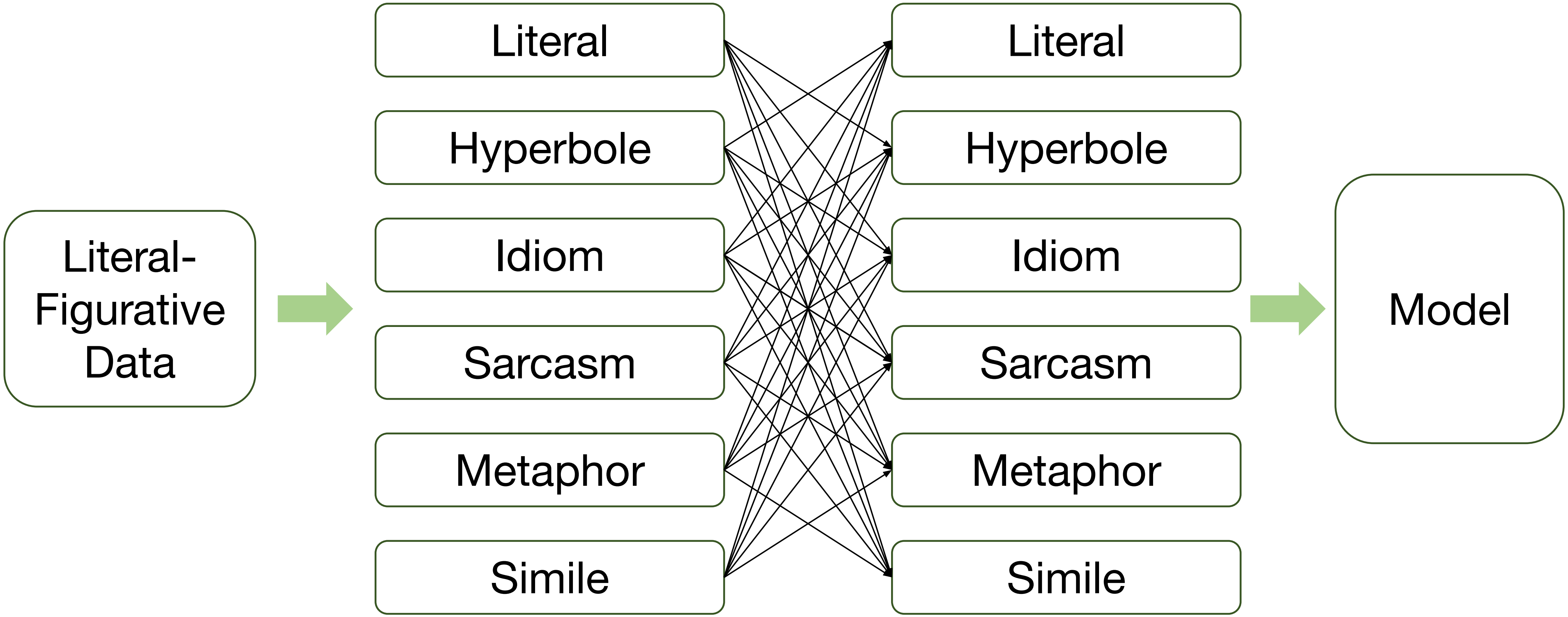
Huiyuan Lai and **Malvina Nissim**
Center for Language and Cognition (CLCG)
University of Groningen / The Netherlands
{h.lai, m.nissim}@rug.nl

Proceedings of the 29th International Conference on Computational Linguistics, pages 5939–5954
October 12–17, 2022.

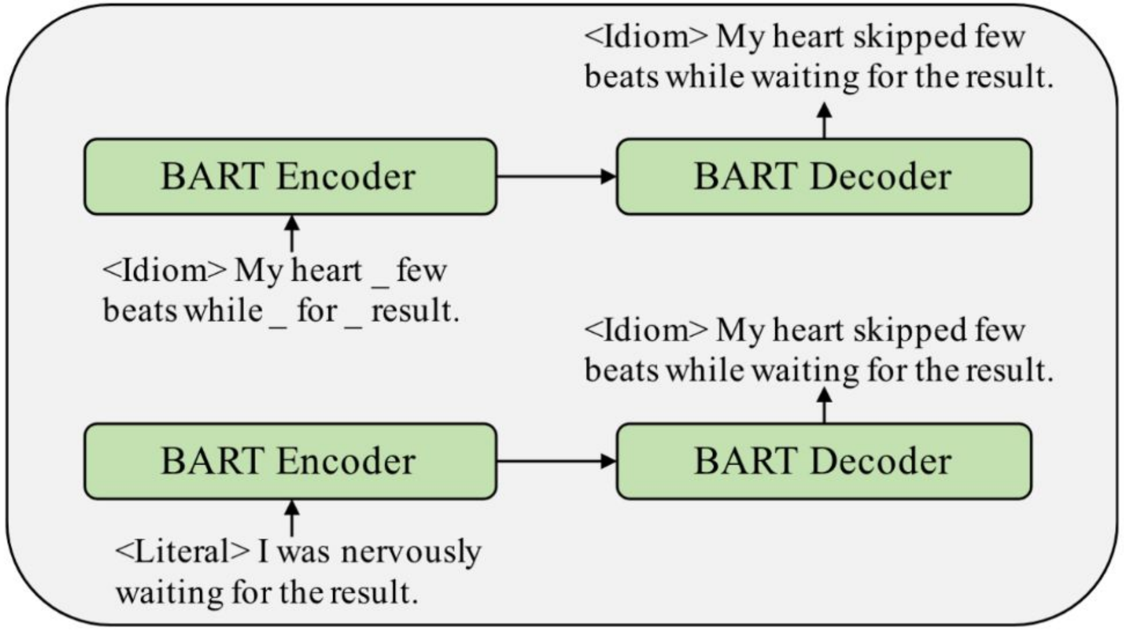
MULTI-FIGURATIVE LANGUAGE GENERATION



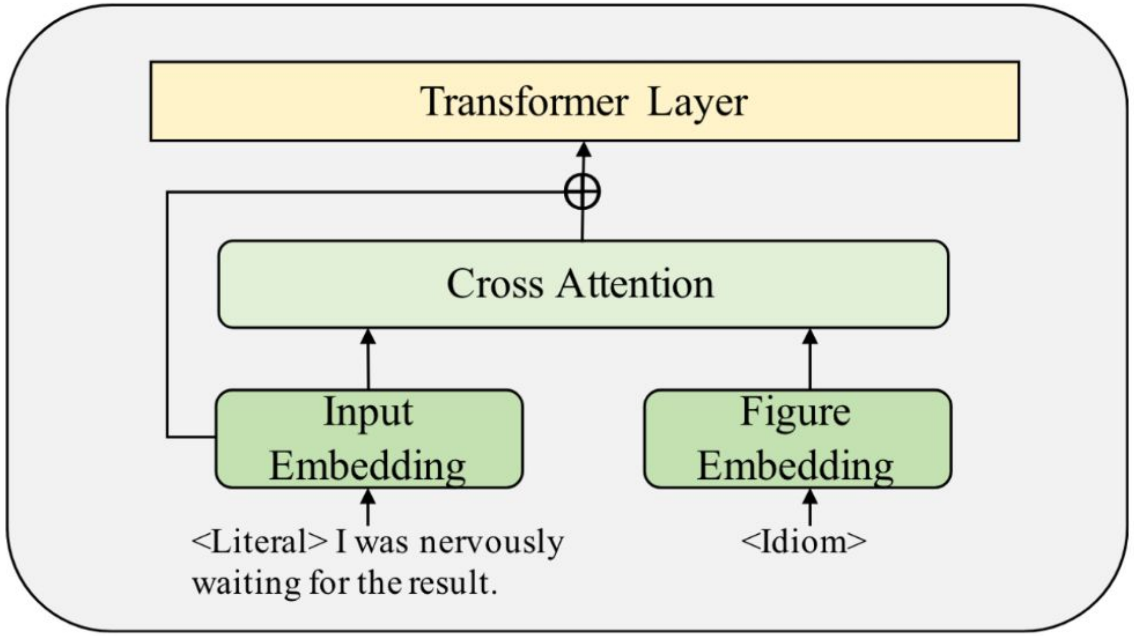
MULTI-FIGURATIVE LANGUAGE GENERATION



MULTI-FIGURATIVE LANGUAGE GENERATION



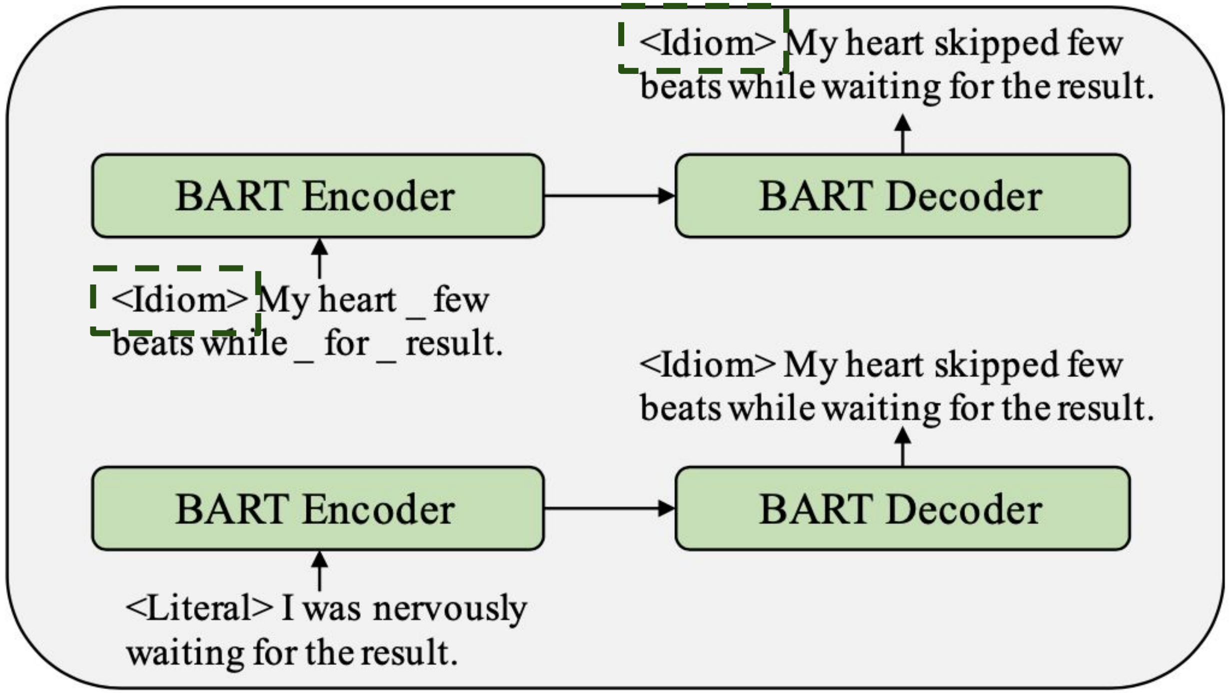
(a) Multi-figurative language denoising pre-training and fine-tuning.



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

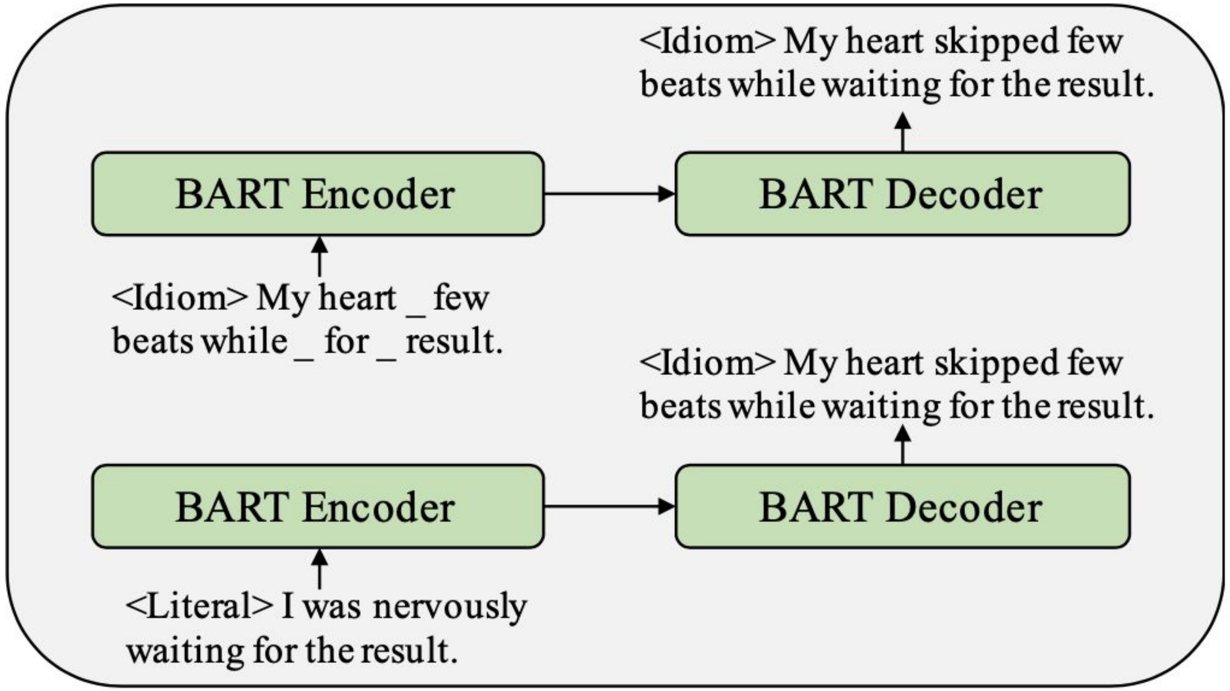
Overview of **m**ulti-**f**igurative **la**nguage **g**eneration (**mFLAG**).

MULTI-FIGURATIVE LANGUAGE GENERATION



Denoising Training:
Selected Paraphrase Data
[PARABANK 2]

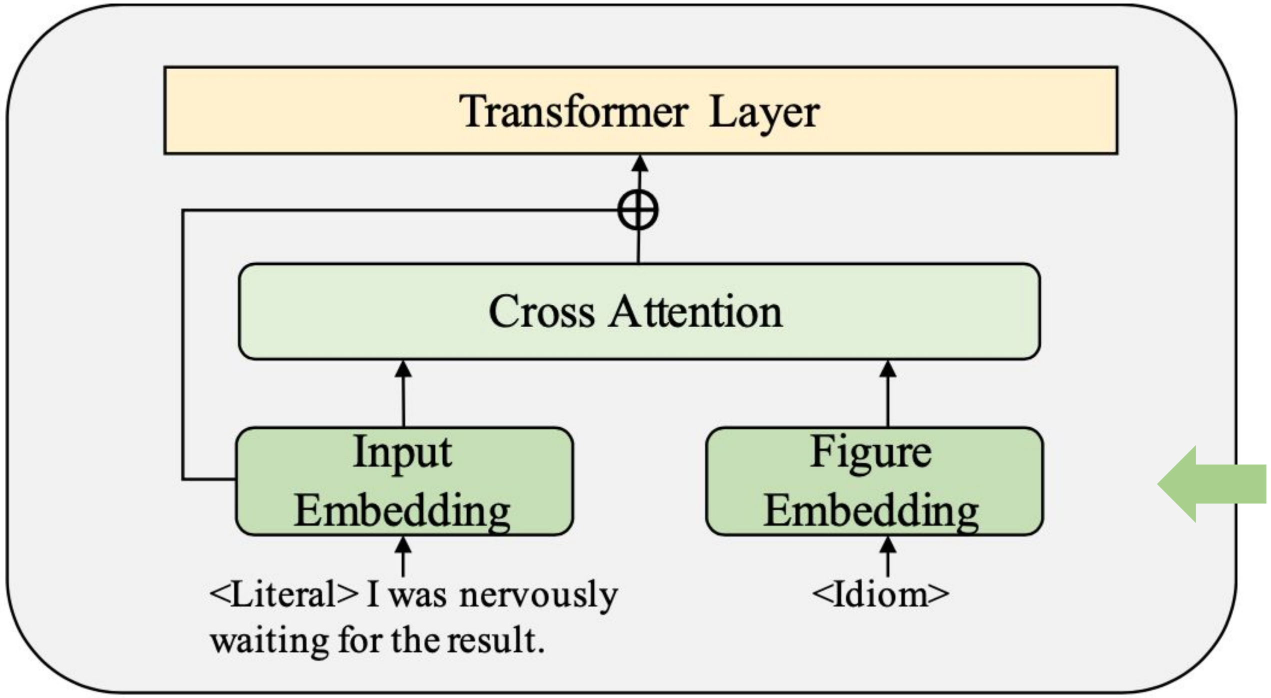
MULTI-FIGURATIVE LANGUAGE GENERATION



Denoising Training: Selected Paraphrase Data [PARABANK 2]

Fine-Tuning

MULTI-FIGURATIVE LANGUAGE GENERATION



A mechanism can *leak* the information of the desired figure of speech to the encoder with a figurative embedding as additional input.

OUR MODEL AND BASELINES

- mFLAG: Our model
- PT-to-FT: mFLAG without figurative attention mechanism

OUR MODEL AND BASELINES

- mFLAG: Our model
- PT-to-FT: mFLAG without figurative attention mechanism
- BART-Multi: We concatenate the five parallel training sets and fine-tune BART for multi-figurative language modelling
- BART-Single: For each figure of speech, we fine- tune BART on the corresponding parallel literal-figurative data (figurative-literal-figurative generation)

AUTOMATIC EVALUATION

- Form Strength: BERT based classifier

AUTOMATIC EVALUATION

- Form Strength: BERT based classifier

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

AUTOMATIC EVALUATION

- 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)

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	**0.791**	**0.855**	**0.595**	**0.808**	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	**0.582**	**0.733**	**0.379**	**0.490**	**0.686**	**0.769**	0.765	0.841	0.536	0.738	**0.767**
mFLAG	**0.844**	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	**0.491**	**0.611**	**0.052**	**0.188**	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	**0.765**	0.485	0.609	0.040	0.162	**0.594**	0.867	**0.643**	**0.812**	**0.493**	0.842	**0.738**
mFLAG	0.762	0.487	0.609	0.043	0.169	**0.594**	**0.880**	0.628	0.809	0.490	**0.844**	0.733
	Literal Form→Simile						Figurative→Literal Form					
BART-Single	0.647	0.724	0.720	**0.017**	**0.321**	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	**0.801**	0.634	0.766	**0.542**	0.544	**0.708**
mFLAG	**0.953**	**0.745**	**0.727**	-0.021	0.220	**0.836**	0.796	**0.637**	**0.769**	0.375	**0.681**	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.

LITERAL-TO-FIGURATIVE GENERATION

- BART-Multi outperforms BART-Single on most generation directions

LITERAL-TO-FIGURATIVE GENERATION

- BART-Multi outperforms BART-Single on most generation directions
- Compared to BART-Single and BART-Multi, both of our proposed models PT-to-FT and mFLAG have consistently stronger results

LITERAL-TO-FIGURATIVE GENERATION

- BART-Multi outperforms BART-Single on most generation directions
- Compared to BART-Single and BARTMulti, both of our proposed models PT-to-FT and mFLAG have consistently stronger results
- The performance of the two models mFLAG and PT-to-FT is very close

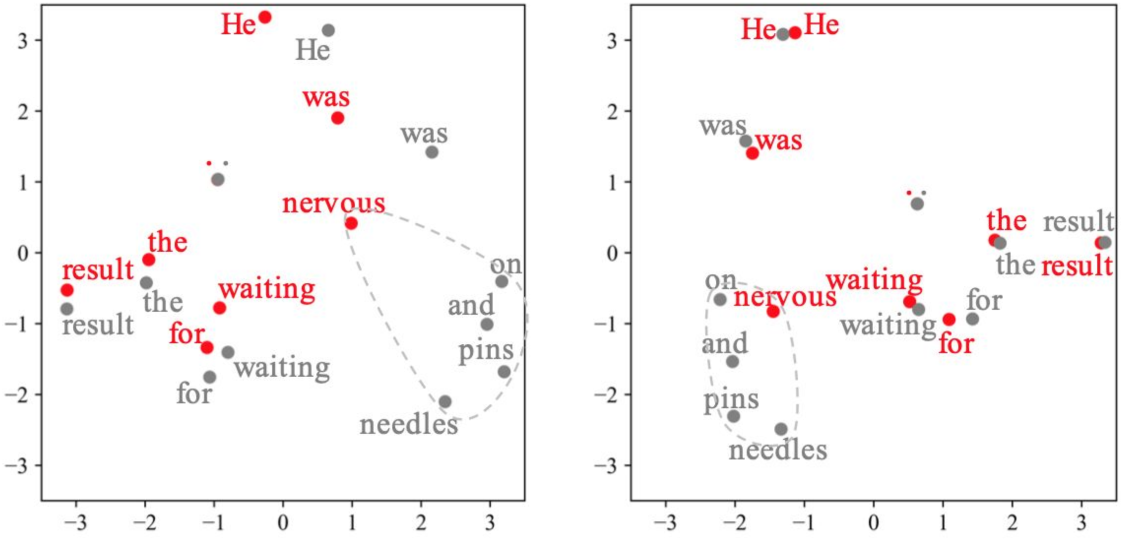
FIGURATIVE-TO-FIGURATIVE GENERATION

- BART-Single performs better than BART-Multi and PT-to-FT

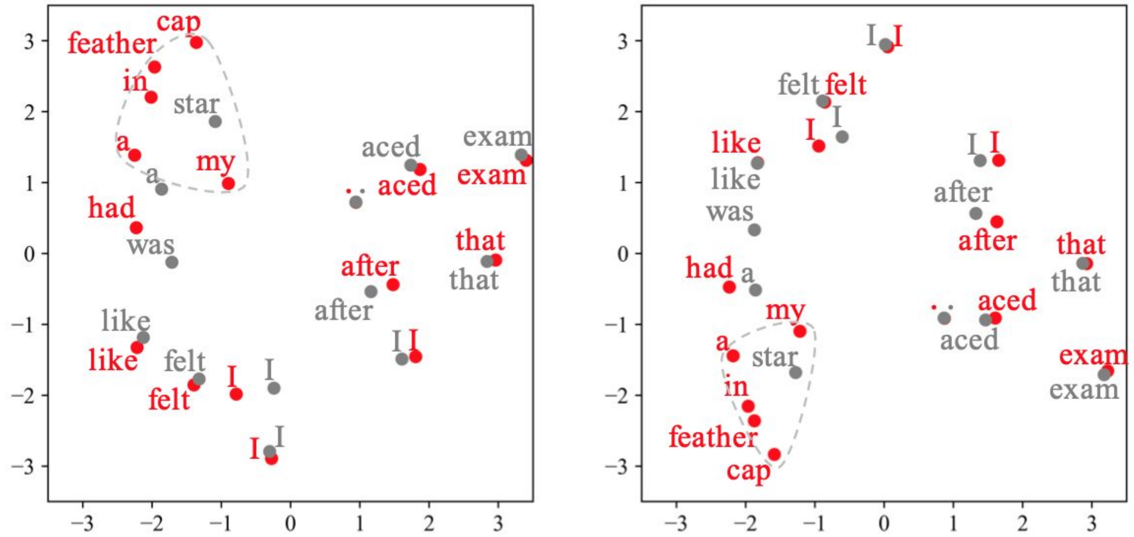
FIGURATIVE-TO-FIGURATIVE GENERATION

- BART-Single performs better than BART-Multi and PT-to-FT
- mFLAG achieves the best performance on most directions

FIGURATIVE INFORMATION IN ENCODER



(a) Result for PT-to-FT. (b) Result for mFLAG.



(c) Result for PT-to-FT. (d) Result for mFLAG.

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

Idiom: *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?

HOW SIMILAR ARE DIFFERENT FORMS?

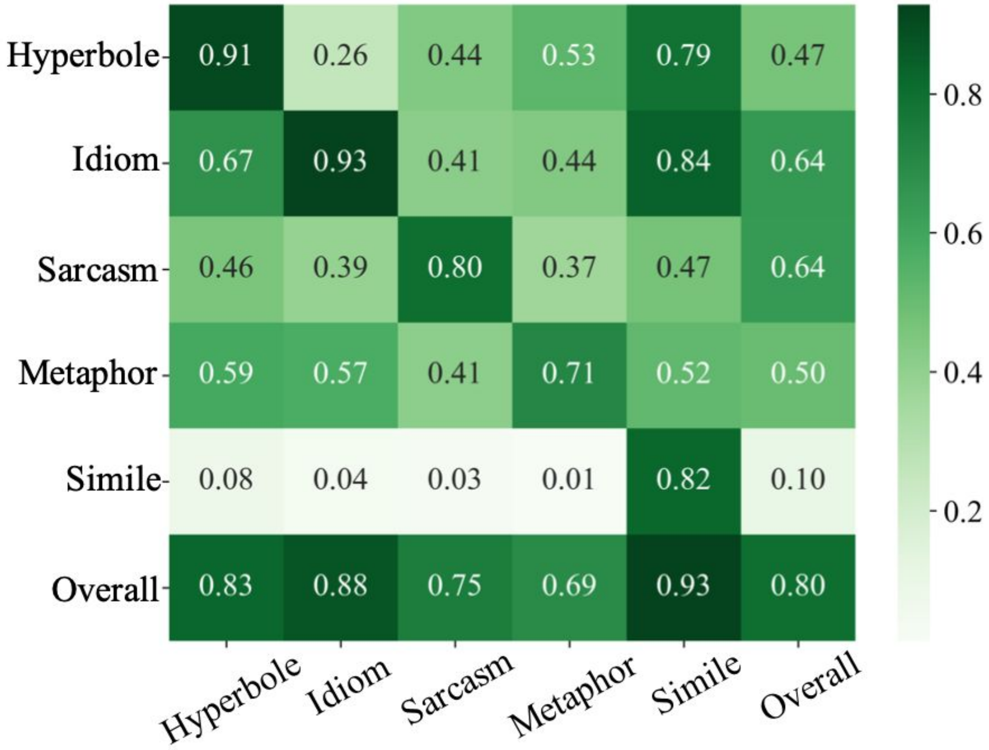


Figure 3: Performances (F1 score) of classifiers on different figurative forms. Each row represents results of a classifier tested on each/all figurative form(s).

HOW SIMILAR ARE DIFFERENT FORMS?

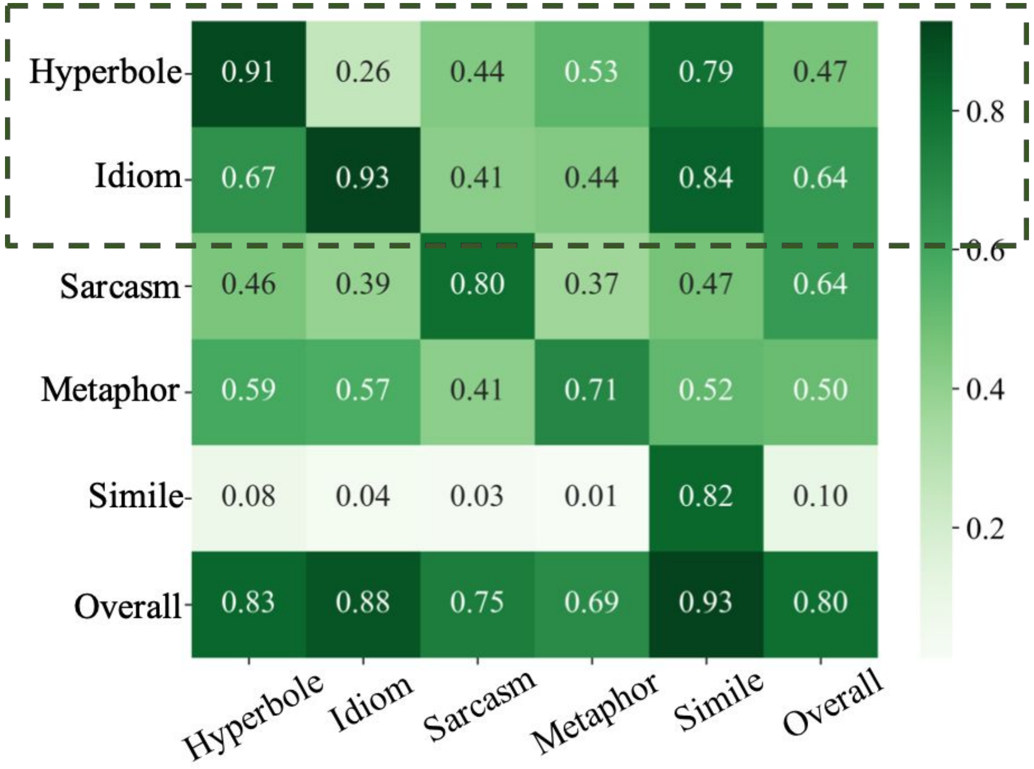


Figure 3: Performances (F1 score) of classifiers on different figurative forms. Each row represents results of a classifier tested on each/all figurative form(s).

HOW SIMILAR ARE DIFFERENT FORMS?

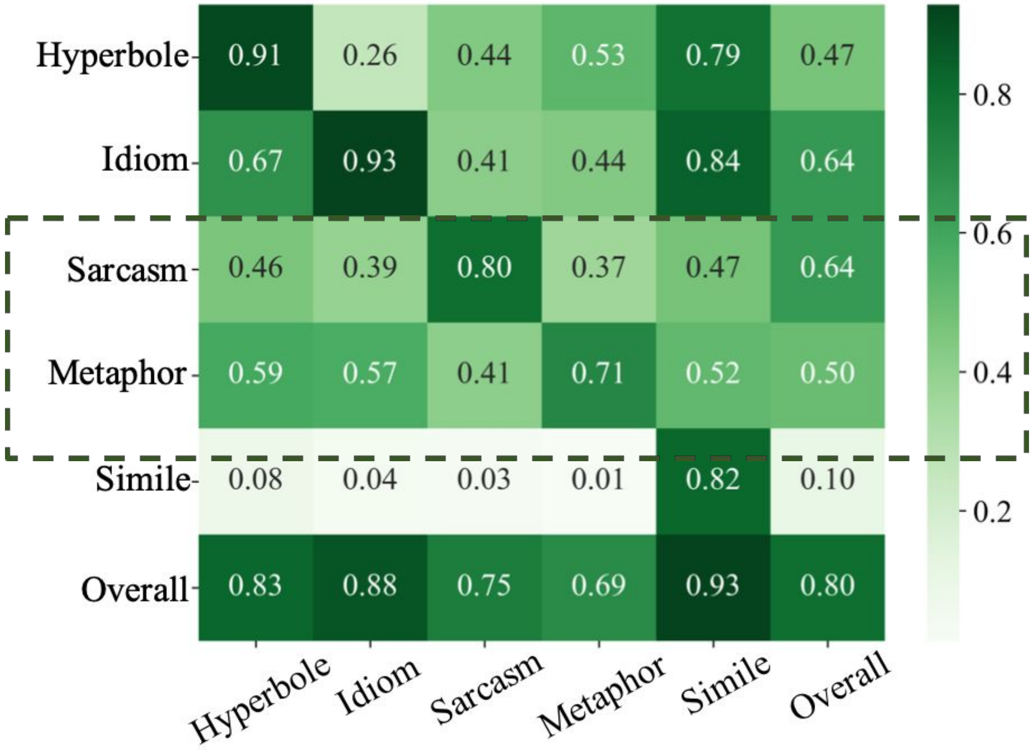
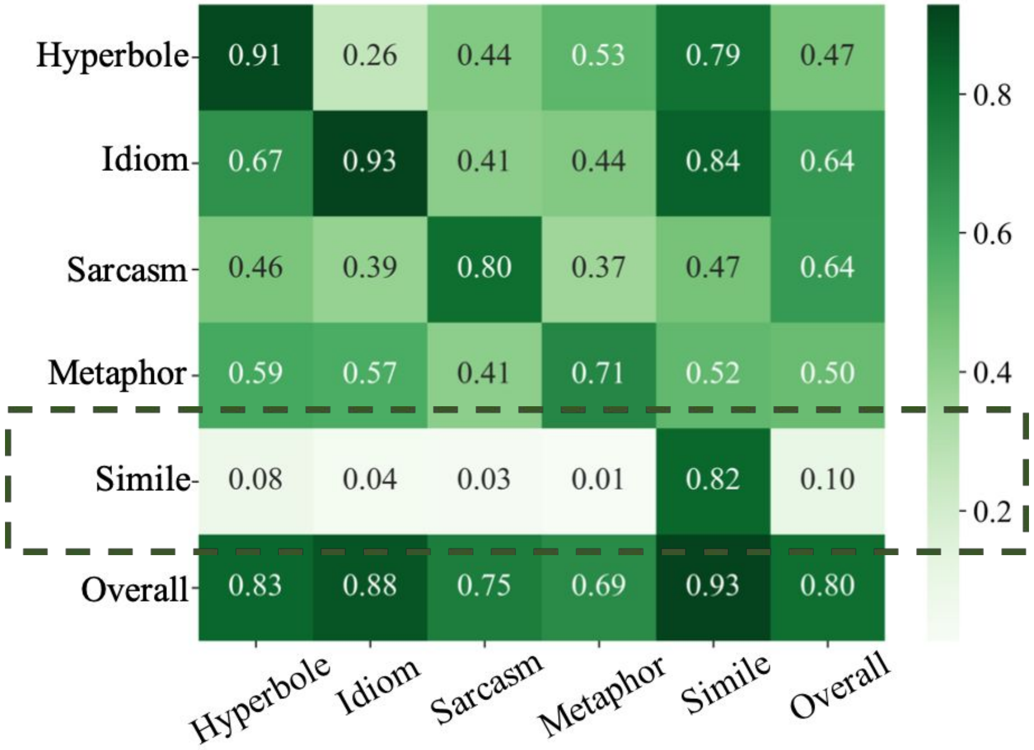


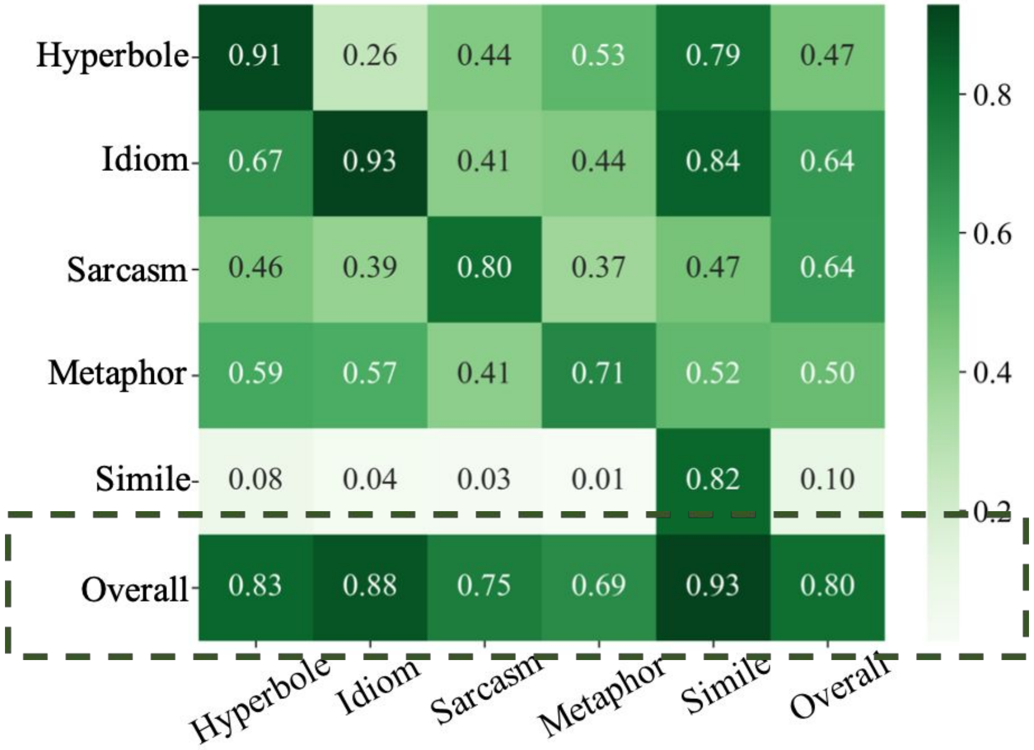
Figure 3: Performances (F1 score) of classifiers on different figurative forms. Each row represents results of a classifier tested on each/all figurative form(s).

HOW SIMILAR ARE DIFFERENT FORMS?



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

HOW SIMILAR ARE DIFFERENT FORMS?



Literal VS Figurative

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?

- Can we jointly model multi-figurative language detection across multiple languages?

HOW SIMILAR ARE DIFFERENT FORMS?

- Can we jointly model multi-figurative language detection across multiple languages?

ACL 2023 (Findings)

Multilingual Multi-Figurative Language Detection

Huiyuan Lai, Antonio Toral, Malvina Nissim
CLCG, University of Groningen / The Netherlands
{h.lai, a.toral.ruiz, m.nissim}@rug.nl

CONCLUSION AND OUTLOOK

This Work:

- A novel task of multi-figurative language generation, and a 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

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	**0.791**	**0.855**	**0.595**	**0.808**	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	**0.582**	**0.733**	**0.379**	**0.490**	**0.686**	**0.769**	0.765	0.841	0.536	0.738	**0.767**
mFLAG	**0.844**	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	**0.491**	**0.611**	**0.052**	**0.188**	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	**0.765**	0.485	0.609	0.040	0.162	**0.594**	0.867	**0.643**	**0.812**	**0.493**	0.842	**0.738**
mFLAG	0.762	0.487	0.609	0.043	0.169	**0.594**	**0.880**	0.628	0.809	0.490	**0.844**	0.733
	Literal Form→Simile						Figurative→Literal Form					
BART-Single	0.647	0.724	0.720	**0.017**	**0.321**	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	**0.801**	0.634	0.766	**0.542**	0.544	**0.708**
mFLAG	**0.953**	**0.745**	**0.727**	-0.021	0.220	**0.836**	0.796	**0.637**	**0.769**	0.375	**0.681**	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.

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	**0.507**	**0.732**	**0.438**	**0.407**	0.342
mFLAG-DR	**0.922**	0.608	**0.815**	**0.893**	**0.753**	**0.836**	**0.696**	0.411	0.633	0.036	-0.105	0.490
mFLAG-BT	0.482	**0.644**	0.539	0.702	0.253	0.246	0.586	0.421	0.662	0.169	0.093	**0.509**
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	**0.760**	**0.860**	**0.646**	**0.715**	0.325
mFLAG-DR	**0.910**	0.400	**0.901**	**0.940**	**0.822**	**0.869**	**0.554**	0.694	0.799	0.328	0.375	0.507
mFLAG-BT	0.328	**0.409**	0.724	0.831	0.491	0.566	0.523	0.703	0.816	0.490	0.569	**0.517**

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.

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	**0.593**	-0.049	-0.014	0.324
PT-to-FT	0.464	0.252	0.863	0.891	0.613	0.774	0.390	**0.468**	0.592	**-0.031**	**0.000**	0.328
mFLAG-DR	**0.840**	0.438	**0.907**	**0.928**	**0.813**	**0.872**	0.591	0.442	0.563	-0.198	-0.143	0.440
mFLAG-BT	0.583	**0.481**	0.808	0.831	0.460	0.604	**0.605**	0.430	0.554	-0.164	-0.133	0.**454**
						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	**0.763**	0.369	**0.648**	**0.824**	**0.507**	**0.665**	0.365
mFLAG-DR	**0.795**	0.518	**0.697**	**0.846**	**0.614**	0.706	**0.594**	0.516	0.758	0.320	0.410	0.517
mFLAG-BT	0.387	**0.557**	0.502	0.734	0.329	0.434	0.528	0.496	0.743	0.317	0.417	**0.525**
						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	**0.765**	**0.818**	**0.262**	**0.415**	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	**0.440**	0.685	**0.849**	**0.884**	**0.637**	**0.690**	**0.758**	0.589	0.698	-0.016	-0.057	0.633
mFLAG-BT	0.132	**0.687**	0.606	0.670	-0.069	-0.064	0.644	0.672	0.766	0.163	0.250	**0.679**

EXAMPLE OUTPUTS

Forms	Models	Sentences	Suc.
Hyperbole	Input	I am not happy that he urged me to finish all the hardest tasks in the world.	-
Literal	BART-Single	I am not happy that he urged me to finish all the hard tasks in the world.	✗
	BART-Multi	I am 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	I am not happy that he urged me to finish all the difficult tasks.	✓
Idiom	BART-Single	I am not happy that he urged me to finish all the hard tasks in the world.	✗
	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.	✗
	mFLAG-DR	I am not happy that he urged me to tick off all the hardest tasks in the world.	✓
	mFLAG-BT	I am not happy that he urged me to get cracking 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.

EXAMPLE OUTPUTS

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