

# CMDAG: A Chinese Metaphor Dataset with Annotated Grounds as CoT for Boosting Metaphor Generation

Yujie Shao<sup>2\*</sup>   Xinrong Yao<sup>3\*</sup>   Xingwei Qu<sup>1,6\*</sup>  
Chenghua Lin<sup>6</sup>   Shi Wang<sup>8</sup>   Stephen W. Huang<sup>9</sup>   Ge Zhang<sup>1,4,5,7</sup>  
Jie Fu<sup>1,5</sup>

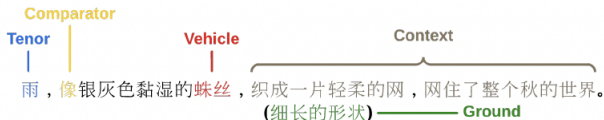
<sup>1</sup>HKUST   <sup>2</sup>University of California, San Diego   <sup>3</sup>Massachusetts Institute of Technology  
<sup>4</sup>University of Waterloo  
<sup>5</sup>Multimodal Art Projection Research Community   <sup>6</sup>University of Manchester  
<sup>7</sup>Stardust.ai  
<sup>8</sup>Institute of Computing Technology, Chinese Academy of Sciences   <sup>9</sup>harmony.ai

May 3, 2024

# Metaphors

## Definition

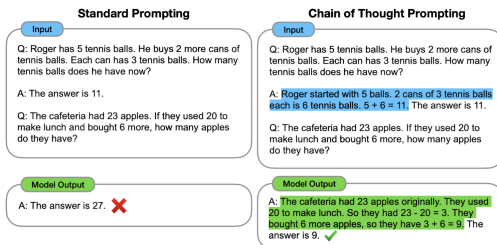
A **Metaphor** is a linguistic device in which a word or phrase literally denoting one kind of object or idea is used in place of another to suggest a likeness or analogy between them.



The **rain** is **like** silver-gray sticky **spider silk**, weaving into a soft net that captures the entire realm of autumn.  
(elongated shape)

# Overview

- We present CMDAG: a unique Chinese metaphor corpus, comprising of 28K annotated sentences, wherein a key feature is the inclusion of GROUNDS.
- We introduce a metaphor annotation pipeline by leveraging academically specialized annotators' expertise.
- We propose the first work introducing Chain-of-Thought (CoT) into metaphor generations.



# Related Work

- Chinese metaphor corpora: Liu et al. (2019), Zhang et al. (2021), Li et al. (2022b), Yang et al. (2023).
- Boosting NLG via CoT: Liu et al. (2022, 2023), Yue et al. (2023), Jiang et al. (2023), Chang et al. (2023).

# Corpora Comparison

Statistic characteristics and annotation information of main existing Chinese metaphor/simile datasets of metaphor and simile and CMDAG dataset. W and F separately denote the tenor/vehicle words and the corresponding feature words.

Dataset	# Nums	Tenor	Vehicle	Ground	Context	Open-source
		W/F	W/F		Above/Below	
Poetry (Liu et al., 2019)	43,051	—/—	—/—	—	✓/—	✓
Lyrics (Liu et al., 2019)	246,669	—/—	—/—	—	✓/—	✓
CS (Zhang et al., 2021)	5,490,721	—/—	—/—	—	✓/✓	✓
CMC (Li et al., 2022b)	2,787	✓/—	✓/—	—	—/—	✓
GraCe (Yang et al., 2023)	61,360	✓/✓	✓/✓	✓	✓/✓	—
CMDAG	27,989	✓/✓	✓/✓	✓	✓/✓	✓

# Data Collection

Collected a raw set of  $\sim 153\text{K}$  probable metaphoric sentences from various Chinese literary sources online.

Heuristic rules to detect likely-metaphors:

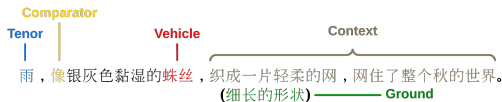
- Sentence contains Chinese simile comparators (“像”, “好似”, etc.).
- Sentence’s subject and object are not highly related (cosine similarity score  $\leq 0.575$ ) and do not have a hyponym/hypernym relationship (WordNet). (Su et al.(2017))

Source Type	# Literature Works	# Likely-Metaphors	# Annotated Metaphors
Prose/Poem	3,459	28,553	5,294
Song Lyrics	102,197	109,827	21,276
Contemporary Poem	4,494	7,268	939
HipHop/Rap Lyrics	3,004	7,603	480
<b>Total</b>	<b>113,154</b>	<b>153,251</b>	<b>27,989</b>

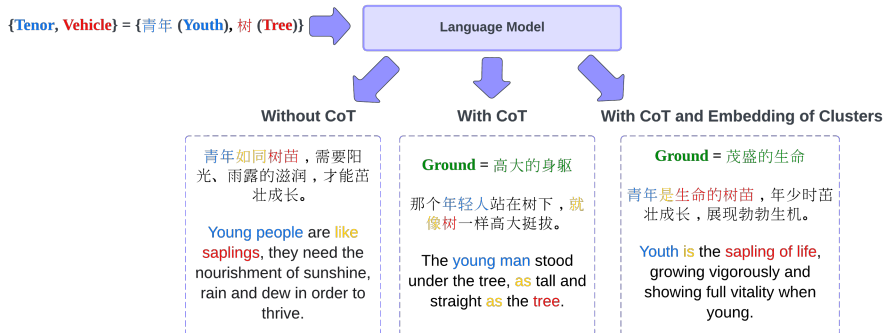
# Experiment

- Task 1 Setting 0: prompt the model with TENOR-VEHICLE pairs and for each pair we ask it to generate a corresponding metaphor.
- Task 1 Setting 1:
  - Prompt the model with TENOR-VEHICLE pairs, as well as annotated examples selected based on our first clustering method, and for each pair we ask it to generate a corresponding GROUND.
  - Prompt the model again with the same TENOR-VEHICLE pairs and annotated examples, as well as the inferred GROUND, and for each pair we ask it to generate a corresponding metaphor.
- Task 1 Setting 2: similar process as in Setting 1, except we select the annotated examples based on our second clustering method.
- Task 2: similar procedure as in Task 1, but instead we prompt the model and provide annotated examples with TENOR-GROUND pairs, and ask it to infer the corresponding VEHICLE for each pair in Settings 1 and 2.

# Flowchart Example: Task 1



The **rain** is **like** silver-gray sticky **spider silk**, weaving into a soft net that captures the entire realm of autumn.  
(**elongated shape**)





# Evaluation Metrics

Annotators evaluate generated metaphors based on:

- **Clarity:** the degree to which a statement is expressed without ambiguity, ensuring its comprehensibility.
- **Creativity:** the originality of the given statement, differentiating between novel concepts and clichéd ideas.
- **Authentic Expression:** the degree to which a statement aligns with expressions that are considered authentic or native-like by the evaluators.

# Results

Model Name	Setting	Clarity	Creativity	Authentic Expression	Final Score
Baichuan	⊙	2.94	2.06	2.36	2.4
Baichuan	◇	2.98	2.09	2.29	2.49
Baichuan	★	2.98	2.07	2.20	2.32
Belle	⊙	2.61	1.71	2.18	2.07
Belle	◇	2.83	1.9	2.37	2.33
Belle	★	2.97	1.69	2.23	2.17
GPT-4	⊙	2.92	1.64	2.16	2.25
GPT-4	◇	2.96	1.6	2.11	2.21
GPT-4	★	2.98	1.66	2.24	2.36
GPT-3.5	⊙	2.99	1.78	2.23	2.21
GPT-3.5	◇	2.99	1.75	2.16	2.25
GPT-3.5	★	2.98	1.45	1.94	2.03
Chinese-alpaca-33B	⊙	2.99	1.83	2.14	2.28
Chinese-alpaca-33B	◇	2.97	1.68	2.14	2.11
Chinese-alpaca-33B	★	2.99	1.86	2.29	2.20
ERNIE	⊙	2.87	1.86	2.30	2.27
ERNIE	◇	2.97	1.56	2.16	2.27
ERNIE	★	2.90	1.73	2.02	2.17

⊙ is the symbol of Setting 0, ◇ is the symbol of Setting 1 and ★ represents the Setting 2.

Model Name	Setting	Task1	Task2
Belle	⊙	0.112	0.236
Belle	◇	0.12	0.268
Belle	★	0.14	0.216
GPT-4	⊙	0.38	0.484
GPT-4	◇	0.448	0.548
GPT-4	★	0.448	0.548
GPT-3.5	⊙	0.372	0.384
GPT-3.5	◇	0.392	0.416
GPT-3.5	★	0.32	0.368

**Table:** Percentage of model-generated sentences that are reasonable Chinese metaphors.

# Conclusion

- An annotated Chinese Metaphor Dataset, encompassing approximately 28K sentences sourced from a wide array of Chinese literary forms, including poems, prose, and song lyrics.
- A thorough set of guidelines, which are instrumental in aiding annotators in the identification of tenors, vehicles, and grounds.
- An evaluation method for metaphor sentence generation that leverages a Chain of Thoughts (CoT) framework.
- Open-source LLMs are employed to test the capability of the corpus's capability to facilitate the generation of creative and linguistically metaphors.