Multimodal CL and NLP: Combining Language and Vision for (Computational) Semantics

Carina Silberer

IMS, University of Stuttgart

CL Fall School 2024, Passau
16 - 20 September 2024

Outline

Outline

Metaphors

What does the image mean?

source: `www.mytraveldiaryusa.de/redensarten-in-den-usa`

Metaphors

What does the image mean?

source: `www.mytraveldiaryusa.de/redensarten-in-den-usa`

"a piece of cake"

Metaphors: Terminology [Saeed, 2022]

Figurative language understanding

Interpretation of non-literal textual expressions

Metaphors

Figurative use of language / visuals; Expression where one concept is understood "through the lens" of another

Metaphors: Terminology [Saeed, 2022]

Figurative language understanding

Interpretation of non-literal textual expressions

Metaphors

Figurative use of language / visuals; Expression where one concept is understood "through the lens" of another

- Analogical mapping between two domains (Lakoff and Johnson, 1980; Conceptual Metaphor Theory, CMT):

Metaphors: Terminology [Saeed, 2022]

Figurative language understanding

Interpretation of non-literal textual expressions

Metaphors

Figurative use of language / visuals; Expression where one concept is understood "through the lens" of another

- Analogical mapping between two domains (Lakoff and Johnson, 1980; Conceptual Metaphor Theory, CMT):
  - **Source:** Comparison concept, analogy (aka *vehicle*)
    e.g., ease or effort of accomplishment
  - **Target:** Described concept, starting point (aka *tenor*)
    e.g., food, piece of cake

Metaphors: Terminology [Saeed, 2022]

Figurative language understanding

Interpretation of non-literal textual expressions

Metaphors

Figurative use of language / visuals; Expression where one concept is understood "through the lens" of another

- Analogical mapping between two domains (Lakoff and Johnson, 1980; Conceptual Metaphor Theory, CMT):
  - **Source:** Comparison concept, analogy (aka *vehicle*) e.g., ease or effort of accomplishment
  - **Target:** Described concept, starting point (aka *tenor*) e.g., food, piece of cake

No agreement on the definition and mechanics of processing of metaphors

Metaphors: Terminology

Metaphors

- **Visual metaphors:** images/videos are used to convey metaphorical meaning, e.g., `http://www.vismet.org/VisMet/display.php`
- **Multimodal metaphors:** often in memes, e.g., `https://www.kaggle.com/datasets/liaolianfoka/met-meme?select=E_text.csv`

Metaphors

Metaphors: Why relevant? [Ge et al., 2023]

- Metaphors are parts of our daily discourse
- Mechanism that draws upon familiar and concrete concepts to facilitate the communication of abstract or complex ideas
- They shape our understanding of abstract concepts
- Systems, (e.g., machine translation, sentiment analysis) are prone to misunderstandings if they do not understand metaphors

Metaphors

Metaphors: Why relevant? [Ge et al., 2023]

- Metaphors are parts of our daily discourse
- Mechanism that draws upon familiar and concrete concepts to facilitate the communication of abstract or complex ideas
- They shape our understanding of abstract concepts
- Systems, (e.g., machine translation, sentiment analysis) are prone to misunderstandings if they do not understand metaphors

Metaphor Comprehension

- **General Task:** Given a sentence/phrase/image, does it contain a metaphor?
- **Challenges:** Deep interpretation of both textual and visual elements; requires commonsense knowledge

Metaphors

What does the image mean?

source: www.mytraveldiaryusa.de/redensarten-in-den-usa

"a piece of cake" "a walk in the park"

Metaphors

What does the image mean?

source: `www.mytraveldiaryusa.de/redensarten-in-den-usa`

"a piece of cake" "a walk in the park" "Das ist ein Kinderspiel"

Metaphors in Multimodal NLP/CL

Black Holes and White Rabbits: Metaphor Identification with Visual Features [Shutova et al., 2016]

Overview

- The first visual-linguistic approach, very simple
- **Goal:** Metaphor identification
- **Task:** Given a phrase, classify it as *literal* or *metaphorical*
  e.g., "foggy brain", "foggy night", "cold beer", "honest meal"
  "wine breathes", "person breathes", "breath life"

Metaphors in Multimodal NLP/CL

Black Holes and White Rabbits: Metaphor Identification with Visual Features [Shutova et al., 2016]

Overview

- The first visual-linguistic approach, very simple
- **Goal:** Metaphor identification
- **Task:** Given a phrase, classify it as *literal* or *metaphorical*
  e.g., "foggy brain", "foggy night", "cold beer", "honest meal"
  "wine breathes", "person breathes", "breath life"
- **Question: What aspect do these metaphors underlie (how can they be explained)?**

Metaphors in Multimodal NLP/CL

Black Holes and White Rabbits: Metaphor Identification with Visual Features [Shutova et al., 2016]

Overview

- The first visual-linguistic approach, very simple
- **Goal:** Metaphor identification
- **Task:** Given a phrase, classify it as *literal* or *metaphorical*
  e.g., "foggy brain", "foggy night", "cold beer", "honest meal"
  "wine breathes", "person breathes", "breath life"
- **Method:**
  Use of multimodal phrase embeddings constructed with multimodal word embeddings
- **Results:**
  - Joint model more effective than unimodal (textual or visual) models, and comparable to knowledge-based methods
  - Models more effective in Adj + Noun than on Verb + Noun
    Possible reasons: adjectives, nouns more concrete than verbs, and perceptual features more relevant for adjectives, nouns

Metaphors in Multimodal NLP/CL

I Spy a Metaphor: Large Language Models and Diffusion Models Co-Create Visual Metaphors [Chakrabarty et al., 2023]

Overview

- Cross-modal metaphor generation
- **Goal:** Visual metaphor generation from linguistic metaphors
- **Task:** Given a linguistic metaphor (LinM), generate its visual correspondence (VisM)
  $\Rightarrow$ Convey meaning through symbolism

Metaphors in Multimodal NLP/CL

I Spy a Metaphor: Large Language Models and Diffusion Models Co-Create Visual Metaphors [Chakrabarty et al., 2023]

Overview

- Cross-modal metaphor generation
- **Goal:** Visual metaphor generation from linguistic metaphors
- **Task:** Given a linguistic metaphor (LinM), generate its visual correspondence (VisM)
  $\Rightarrow$ Convey meaning through symbolism

Metaphors in Multimodal NLP/CL

I Spy a Metaphor: Large Language Models and Diffusion Models Co-Create Visual Metaphors [Chakrabarty et al., 2023]

Overview

- Cross-modal metaphor generation
- **Goal:** Visual metaphor generation from linguistic metaphors
- **Task:** Given a linguistic metaphor (LinM), generate its visual correspondence (VisM)
  $\Rightarrow$ Convey meaning through symbolism
- **Method:** Generate visual elaboration of LinM through Chain-of-Thought prompting of LLM (GPT-3), then generate image with Diffusion Model

Metaphors in Multimodal NLP/CL

[Chakrabarty et al., 2023]

Experiments: Results

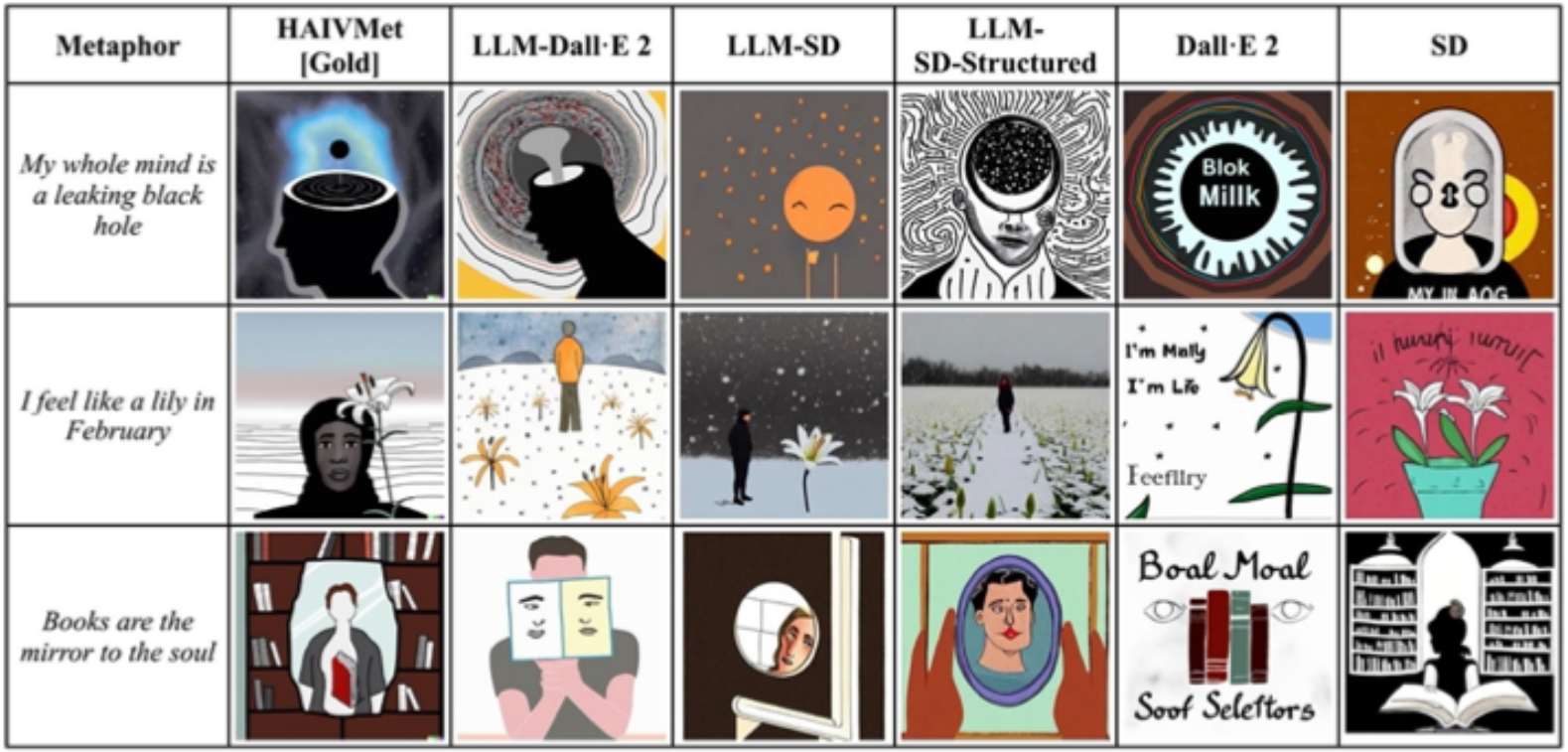


Figure 4: Examples of output from each model described in Section 4.1 for three randomly chosen metaphors. **HAIVMet** is our gold standard. More examples are available in Figure 8 in the Appendix.

Metaphors/Idioms/Similes: IRFL Task [Yosef et al., 2023]

Task: Idioms

- Given four images and an idiom
- Choose the image that conveys a/the definition(s) of the idiom
- The figurative part is the caption (**language**)

Example

on a full stomach

1. Directly after eating
2. Straight after a meal

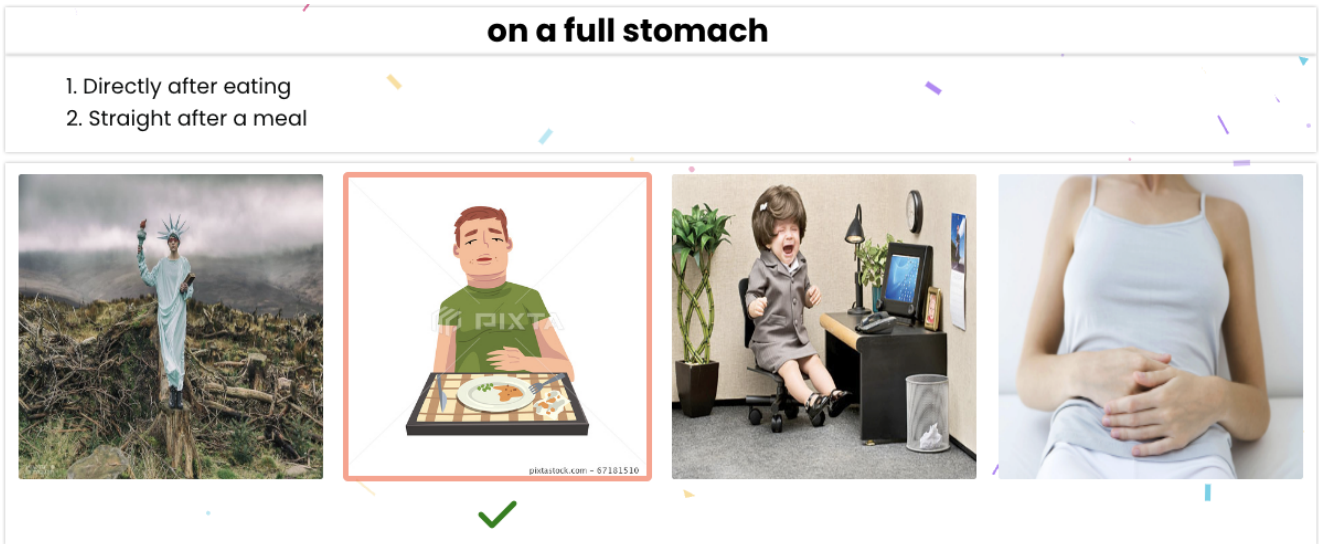



Metaphors/Idioms/Similes: IRFL Task [Yosef et al., 2023]

Task: Idioms

- Given four images and an idiom
- Choose the image that conveys a/the definition(s) of the idiom
- The figurative part is the caption (**language**)

Example

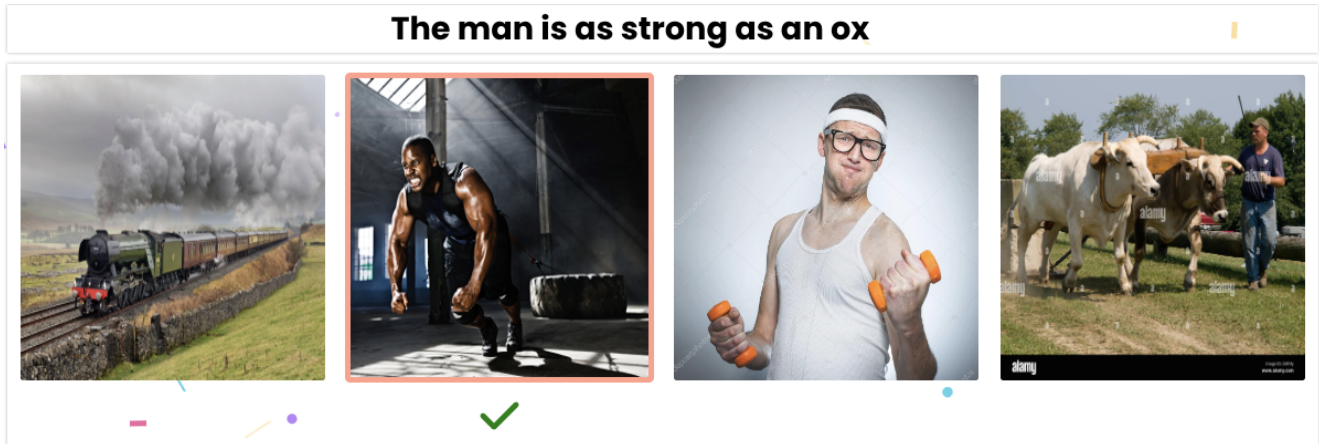


Metaphors/Idioms/Similes: IRFL Task [Yosef et al., 2023]

Task: Similes

- Given four images and a simile
- Choose the image that matches the simile
- The figurative part is the caption (**language**)

Example



Metaphors/Idioms/Similes: IRFL Task [Yosef et al., 2023]

Task: Metaphors

- Given four images and a metaphor
- Choose the image that conveys the metaphorical message
- The figurative part is the caption (**language**)

Examples

Metaphors/Idioms/Similes: IRFL Task [Yosef et al., 2023]

Task: Metaphors

- Given four images and a metaphor
- Choose the image that conveys the metaphorical message
- The figurative part is the caption (**language**)

Examples

Metaphors/Idioms/Similes: IRFL Task [Yosef et al., 2023]

Task: Metaphors

- Given four images and a metaphor
- Choose the image that conveys the metaphorical message
- The figurative part is the caption (**language**)

Examples

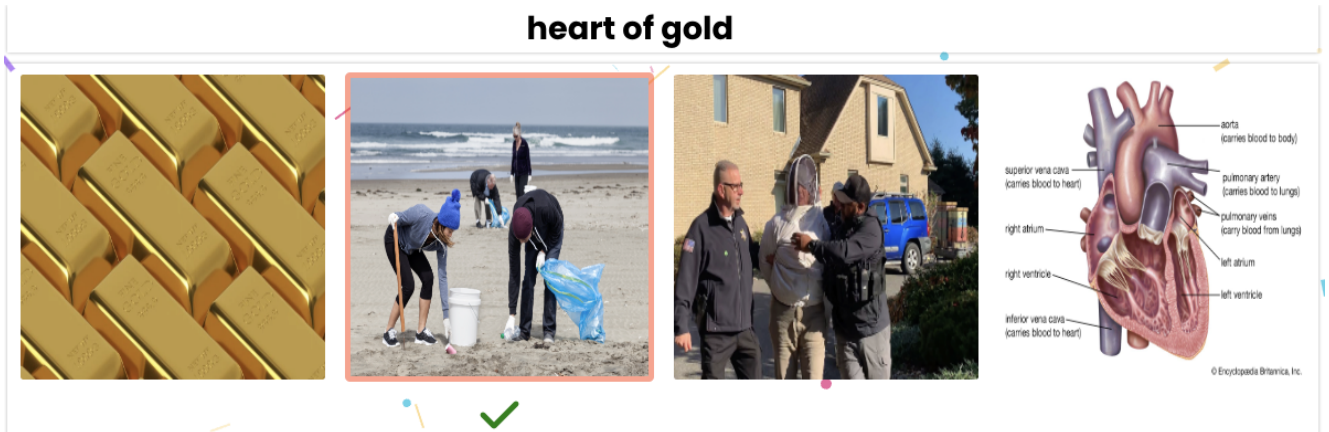
heart of gold

Metaphors/Idioms/Similes: IRFL Task [Yosef et al., 2023]

Task: Metaphors

- Given four images and a metaphor
- Choose the image that conveys the metaphorical message
- The figurative part is the caption (**language**)

Examples



Metaphors/Idioms/Similes: IRFL Task [Yosef et al., 2023]

Approaches: Zero-shot baselines

- Encode image i and text t with a model
- Compute matching score $\mathcal{M}(i, t)$
- Choose the image $\hat{i}$ with the highest score:
  $\hat{i} = \arg\max_{i \in \mathcal{I}} \mathcal{M}(i_k, t)$
- Models: CLIP, ViLT, BLIP2 (i.a.)
  Chatbot models LLaVA, InstructBLIP not applicable,
  OpenFlamingo not effectiveness
- Results: CLIP best on metaphors, but all models prefer partially literal images over figurative ones

Metaphors/Idioms/Similes: IRFL Task [Yosef et al., 2023]

Approaches: Supervised Models

- Figurative classification of idioms (binary task)
- Based on CLIP: concatenate the textual (idiom) and image embeddings of each text–image pair
- Train a binary classifier on top of the VL embeddings

FigLang 2024 Shared Task on Multimodal Figurative Language

[Kulkarni et al., 2024]

Task: Explainable Figurative Visual Entailment

- Given an image (premise) and text (hypothesis)
  - Does the image support or contradict the hypothesis?
  - Provide an explanation.
- V-FLUTE dataset [Saakyan et al., 2024] of annotated image–text pairs, created from 5 existing datasets on figurative language

FigLang 2024 Shared Task on Multimodal Figurative Language

[Kulkarni et al., 2024]

Example: IRFL Dataset

Hypothesis:
Heart of gold
Premise:

Label: **entailment**

Explanation: *The image depicts a group of people wearing VOLUNTEER t-shirts, packing food items such as canned goods and fresh fruits into paper bags, likely for a charitable cause or community service effort. The metaphor "heart of gold" is entailed by the image as it symbolizes kindness and generosity, which is demonstrated by the actions of the volunteers as they help prepare food donations for those in need. Their acts of service align with having a compassionate and giving nature, traits associated with having a "heart of gold".*

FigLang 2024 Shared Task on Multimodal Figurative Language

[Kulkarni et al., 2024]

Example: VisMet (HAIVMet) Dataset [Chakrabarty et al., 2023]

Image (Premise)	Claim (Hypothesis)	Label and Explanation
	The faculty meeting was peaceful.	**Label:** Contradiction *Explanation:* The image shows a faculty meeting transformed into a dramatic battlefield scene, with members dressed as knights discussing academic content on boards behind them as if they were battle tactics. This visual metaphor suggests the faculty meeting was like a war, and not peaceful.

FigLang 2024 Shared Task on Multimodal Figurative Language

[Kulkarni et al., 2024]

Approaches

- FigCLIP:
  (1) obtain multimodal embeddings for classification and explanation generation: image and text are both embedded with CLIP, converted through cross-modal attention and then concatenated
  (2) classification with a binary linear layer
  (3) GPT-2 generates explanation
- Mapper:
  uses LLaVA-7B, LoRa and CLIP

MET-Meme: A Multimodal Meme Dataset Rich in Metaphors [Xu et al., 2022]

- MET-Meme dataset: 10K text–image pairs annotated with metaphor, sentiment, intention, offensiveness (along with corresponding tasks)[1]
- Metaphor features used for meme studies

V dominant
source: cauliflower
target: sheep

T dominant
source: mistakes
target: experience

V–T complementary
source: green life
target: offence

[1] 4K English, 6K Chinese memes

Metaphors: Works in VL I

- MetaClue [Akula et al., 2023]: `https://metaclue.github.io`
  - Set of vision tasks on visual metaphor (visual metaphor classification, localisation, understanding (retrieval, question answering, captioning) and generation (text-to-image synthesis))
  - Four types of metaphors: contextual, juxtaposition, hybrid, multimodal, multiple tasks
  - Find that existing VL models fail at visual metaphor understanding

Metaphors: Works in VL II

- A Textual Modal Supplement Framework for Understanding Multi-Modal Figurative Language [Chen et al., 2024]
- MultiMET [Zhang et al., 2021]: A Multimodal Dataset for Metaphor Understanding [Zhang et al., 2021]
  `https://aclanthology.org/2021.acl-long.249.pdf`
  Multimodal metaphors (image–text pairs)
  Tasks: metaphor detection, sentiment analysis, intent detection
- MemeCap [Hwang and Shwartz, 2023] and MET-Meme [Xu et al., 2022] (the figurative part is the image)
- Sky + Fire = Sunset Exploring Parallels Between Visually Grounded Metaphors and Image Classifiers (Bizzoni & Dobnik, 2020)
  `https://aclanthology.org/2020.figlang-1.19.pdf`
- Visual metaphor generation (generated from sentences using DALL·E 2 or Stable Diffusion) [Chakrabarty et al., 2023]

Metaphors: Works in VL III

- Multimodal Captioning and Figurative Language Understanding: A Survey [Kalarani and Bhattacharyya, 2024]
- IRFL [Yosef et al., 2023]: Metaphors, similes, idioms + images
  `https://irfl-dataset.github.io/`
- FigLang 2024: The Fourth Workshop on Figurative Language Processing: Shared task on figurative language understanding through visual entailment.
  `https://sites.google.com/view/figlang2024`
  `https://www.codabench.org/competitions/1970/`

Metaphors: Works in NLP and Videos

- A Survey on Computational Metaphor Processing Techniques: From Identification, Interpretation, Generation to Application [Ge et al., 2023]
- Adjective–noun phrases [Tsvetkov et al., 2014]:
  e.g., "bitter cold"
  `https://github.com/ytsvetko/metaphor`
- Verb phrases [Birke and Sarkar, 2006]:
  e.g., "He just might dance into the 1990s", TroFi
- Video metaphor captioning [Kalarani et al., 2024]

Metaphors in NLP/CL/Multimodal NLP I

- Machines and Metaphors: Challenges for the Detection, Interpretation and Production of Metaphors by Computer Programs [Hesse, 2023]
- Survey on Computational Metaphor Processing [Rai and Chakraverty, 2020]
- **Black Holes and White Rabbits: Metaphor Identification with Visual Features** [Shutova et al., 2016]
- Sky + Fire = Sunset — Exploring Parallels Between Visually Grounded Metaphors and Image Classifiers [Bizzoni and Dobnik, 2020]
- Improving Neural Metaphor Detection with Visual Datasets [Kehat and Pustejovsky, 2020]
- Multimodal Metaphor Detection Based on Distinguishing Concreteness [Su et al., 2021]
- CoMeta: A Corpus for Metaphor Detection in Spanish `ixa-ehu.github.io/cometa` [?]

References I

Akula, A. R., Driscoll, B., Narayana, P., Changpinyo, S., Jia, Z., Damle, S., Pruthi, G., Basu, S., Guibas, L., Freeman, W. T., et al. (2023).
Metaclue: Towards comprehensive visual metaphors research.
In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23201–23211.

Birke, J. and Sarkar, A. (2006).
A clustering approach for nearly unsupervised recognition of nonliteral language.
In *11th Conference of the European chapter of the association for computational linguistics*, pages 329–336.

Bizzoni, Y. and Dobnik, S. (2020).
Sky+ Fire= Sunset. Exploring Parallels between Visually Grounded Metaphors and Image Classifiers.
In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 126–135.

Chakrabarty, T., Saakyan, A., Winn, O., Panagopoulou, A., Yang, Y., Apidianaki, M., and Muresan, S. (2023).
I spy a metaphor: Large language models and diffusion models co-create visual metaphors.
In Rogers, A., Boyd-Graber, J., and Okazaki, N., editors, *Findings of the Association for Computational Linguistics: ACL 2023*, pages 7370–7388, Toronto, Canada. Association for Computational Linguistics.

Chen, J., Yang, Q., Dong, X., Mao, X., and Hao, T. (2024).
A textual modal supplement framework for understanding multi-modal figurative language.
In Ghosh, D., Muresan, S., Feldman, A., Chakrabarty, T., and Liu, E., editors, *Proceedings of the 4th Workshop on Figurative Language Processing (FigLang 2024)*, pages 85–91, Mexico City, Mexico (Hybrid). Association for Computational Linguistics.

References II

Ge, M., Mao, R., and Cambria, E. (2023).

A survey on computational metaphor processing techniques: From identification, interpretation, generation to application.
Artificial Intelligence Review, 56(Suppl 2):1829–1895.

Hesse, J. (2023).

Machines and metaphors: Challenges for the detection, interpretation and production of metaphors by computer programs.
Theoria, 89(5):607–624.

Hwang, E. and Shwartz, V. (2023).

MemeCap: A dataset for captioning and interpreting memes.
In Bouamor, H., Pino, J., and Bali, K., editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 1433–1445, Singapore. Association for Computational Linguistics.

Kalarani, A. R. and Bhattacharyya, P. (2024).

Multimodal captioning and figurative language understanding: A survey.

Kalarani, A. R., Bhattacharyya, P., and Shekhar, S. (2024).

Seeing the unseen: Visual metaphor captioning for videos.
arXiv preprint arXiv:2406.04886.

Kehat, G. and Pustejovsky, J. (2020).

Improving neural metaphor detection with visual datasets.
In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 5928–5933.

References III

Kulkarni, S., Saakyan, A., Chakrabarty, T., and Muresan, S. (2024).

A report on the FigLang 2024 shared task on multimodal figurative language.
In Ghosh, D., Muresan, S., Feldman, A., Chakrabarty, T., and Liu, E., editors, *Proceedings of the 4th Workshop on Figurative Language Processing (FigLang 2024)*, pages 115–119, Mexico City, Mexico (Hybrid). Association for Computational Linguistics.

Rai, S. and Chakraverty, S. (2020).

A Survey on Computational Metaphor Processing.
ACM Computing Surveys (CSUR), 53(2):1–37.

Saakyan, A., Kulkarni, S., Chakrabarty, T., and Muresan, S. (2024).

V-flute: Visual figurative language understanding with textual explanations.
arXiv e-prints, pages arXiv–2405.

Saeed, J. (2022).

Semantics.
Introducing Linguistics. Wiley.

Shutova, E., Kiela, D., and Maillard, J. (2016).

Black holes and white rabbits: Metaphor identification with visual features.
In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 160–170.

Su, C., Chen, W., Fu, Z., and Chen, Y. (2021).

Multimodal metaphor detection based on distinguishing concreteness.
Neurocomputing, 429:166–173.

References IV

Tsvetkov, Y., Boytsov, L., Gershman, A., Nyberg, E., and Dyer, C. (2014).

Metaphor detection with cross-lingual model transfer.

In Toutanova, K. and Wu, H., editors, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 248–258, Baltimore, Maryland. Association for Computational Linguistics.

Xu, B., Li, T., Zheng, J., Naseriparsa, M., Zhao, Z., Lin, H., and Xia, F. (2022).

MET-Meme: A multimodal meme dataset rich in metaphors.

In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, pages 2887–2899.

Yosef, R., Bitton, Y., and Shahaf, D. (2023).

IRFL: Image recognition of figurative language.

In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 1044–1058, Singapore. Association for Computational Linguistics.

Zhang, D., Zhang, M., Zhang, H., Yang, L., and Lin, H. (2021).

Multimet: A multimodal dataset for metaphor understanding.

In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3214–3225.