EmbodiedBERT: Cognitively Informed Metaphor Detection Incorporating Sensorimotor Information

Yuxi Li[1,2] * and **Bo Peng**[1] † and **Yu-Yin Hsu**[1] ‡ and **Chu-Ren Huang**[1] §
[1]Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong, China
[2]School of Foreign Studies, Xi'an Jiaotong University, Xi'an, China

Abstract

The identification of metaphor is a crucial prerequisite for many downstream language tasks, such as sentiment analysis, opinion mining, and textual entailment. State-of-the-art systems of metaphor detection implement heuristic principles such as Metaphor Identification Procedure (MIP) (Pragglejaz Group, 2007) and Selection Preference Violation (SPV) (Wilks, 1975; Wilson, 2002). We propose an innovative approach that leverages the cognitive information of embodiment that can be derived from word embeddings, and explicitly models the process of *sensorimotor change* that has been demonstrated as essential for human metaphor processing. We showed that this cognitively motivated module is effective and can improve metaphor detection, compared with the heuristic MIP that has been applied previously.[1]

1 Introduction

Metaphor is a common type of figurative language that allows communicators to express novel construal (Shelley, 1890) and convey a myriad of implicit meanings (Gibbs, 2023). Effective metaphor processing is essential for natural language processing tasks (Rai and Chakraverty, 2020), such as sentiment analysis, machine translation, and textual entailment. (Bahdanau et al., 2014; Wu et al., 2018; Poria et al., 2016). As a result, NLP researchers have focused on the computational modeling of metaphor, which typically starts with the identification of metaphors.

The state-of-the-art systems of metaphor identification typically rely on two heuristic principles: the Metaphor Identification Procedure (MIP) (Pragglejaz Group, 2007), and Selection Preference Violation (SPV) (Wilks, 1975; Wilson, 2002). MIP identifies metaphors by recognizing that a word's metaphorical meaning differs from its basic, 'more concrete', 'related to bodily action', and 'historically older' meaning (Pragglejaz Group, 2007). SPV detects metaphors by identifying violations of words' semantic selection preferences in context. The modeling of MIP usually begins with the extraction of basic and contextual representations of target words and then learns their general differences (Li et al., 2023a; Choi et al., 2021), while SPV focuses on the relation between target words and their contexts (Song et al., 2021). Despite their effectiveness, they neglect the cognitive characteristics of metaphor.

Embodied cognition posits that all cognitive acts, including language processing, are rooted in perception and action (Meteyard et al., 2012). Psycholinguistic evidence supports that metaphor processing is also embodied (Gibbs et al., 2004; Khatin-Zadeh, 2023), but the contribution of embodiment is dynamic. Specifically, the embodiment levels of a metaphorical word often changes compared to the word's literal meaning during the online processing (Jamrozik et al., 2016). For example, in the metaphorical use of the verb 'drink' in (a), the embodied features of the action 'drink', such as 'consumed by mouth', and 'the object must be liquid' are abstracted away, unlike in its literal use in (b). This abstraction of sensorimotor information is essential for humans to derive a metaphorical sense of 'drink' (to consume a large amount quickly), especially in the early stage of a metaphor (Bowdle and Gentner, 2005)[2].

- (a) The students drink the knowledge.
- (b) The horse drinks the water.

*yu-xi.li@connect.polyu.hk
†peng-bo.peng@polyu.edu.hk
‡yu-yin.hsu@polyu.edu.hk
§churen.huang@polyu.edu.hk

[1]Our code can be found at https://github.com/EleoLI/EmbodiedBERT

[2]See in Appendix for a more detailed explanation of embodiment change by structural mapping theory and its relation to conventionality of metaphor

Findings of the Association for Computational Linguistics: EACL 2024, pages 16868–16876
November 12-16, 2024

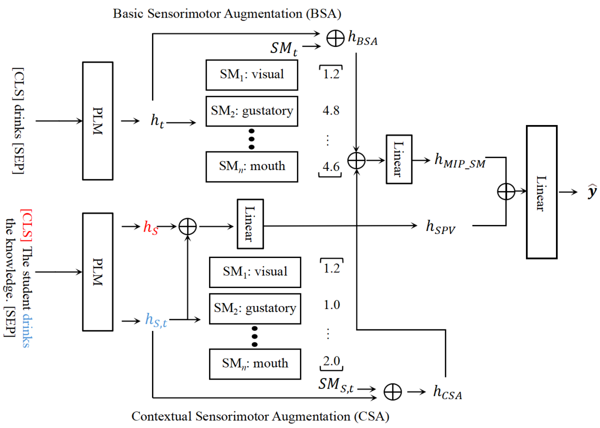

Figure 1: The architecture of EmbodiedBERT includes: two PLM encoders which generates the h_S (representation for `[CLS]`), $h_{S,t}$ (contextual representation of the target word), h_t (basic representation of the target word); ; a suite of sensorimotor regressors (n = 11) which generate $SM_{S,t}$ (sensorimotor representation of the contextual target) and SM_t (the sensorimotor representation of the basic target); a final binary classification module

Therefore, we hypothesize that the explicit modeling of embodiment change can enhance metaphor detection. To test this, we developed EmbodiedBERT, a metaphor identification system that explicitly models the process of *sensorimotor change*. Previous research has integrated sensorimotor information for metaphor identification, but most of them merely use it as word-level feature enrichment without considering its change in context (Bulat et al., 2017; Wan et al., 2023). Compared to general semantic change of word in context (MIP), sensorimotor change offers a more cognitively motivated and precise method for predicting metaphoricity. We show by extensive experiments that our cognitive module is indeed more effective for predicting metaphoricity than MIP.

2 EmbodiedBERT

2.1 Model architecture

EmbodiedBERT has four main components: two basic encoders for representing the target word's contextual and basic meaning; a suite of sensorimotor regressors that *maps distributional embeddings onto sensorimotor-related dimensions*; linear layers learning the function of MIP_SM (sensorimotor change) and SPV, and a final metaphoricity classifier.

Meaning Representation We use two `roberta-base` models (Liu et al., 2019) from Hugging Face [3] as the backbone encoder. Given a sentence $S = \{w_1, \ldots, w_n\}$, the first encoder outputs a set of contextualized embeddings $\{h_S, h_{S,1}, \ldots, h_{S,t}, \ldots, h_{S,n}\}$, where h_S stands for the global meaning of S and $h_{S,t}$ stands for the target's contextual meaning. To extract the target's basic meaning, we input the target word with special tokens into another encoder, resulting in the basic meaning embedding h_t.

The meaning representations are input into two linear functions: SPV and MIP_SM. Firstly, **SPV** aims to learn to contrast a word's contextual meaning with the meaning of its global context. It takes the concatenation of h_S and $h_{S,t}$ and learns their difference through the linear function.

MIP_SM transforms the encoder outputs before the concatenation operation to reflect the specific change in embodiment-related dimensions. It takes an additional step to map distributional word embeddings onto these embodiment-related dimensions. Specifically, we perform such a mapping for both h_t and $h_{S,t}$, to generate SM_t (basic sensorimotor embedding) and $SM_{S,t}$ (contextual sensorimotor embedding). Next, we concatenate the derived SM_t with h_t, and $SM_{S,t}$ with $h_{S,t}$ and input them to another linear function MIP_SM. (See the next section for further details).

Binary classification Finally, the output hidden vectors from SPV and MIP_SM are concatenated together and fed into a linear layer followed by a `sigmoid` function to predict the likelihood of a target being metaphorical (Eq.1). We minimize the binary cross entropy (Eq.2) and update model parameters via back propagation.

$$\hat{y} = \sigma\left(W^T\left(h_{SPV} \bigoplus h_{MIP_{SM}}\right) + b\right) \tag{1}$$

$$\mathcal{L} = \sum_{i=1}^{N} y_i log\hat{y_i} + (1 - y_i) log(1 - \hat{y_i}) \tag{2}$$

2.2 Sensorimotor regressors

We obtain embodiment-related information as inputs for MIP_SM by mapping distributional embeddings onto sensorimotor-related embeddings. There are 11 sensorimotor dimensions related to humans' embodied experience of the physical world, including: Auditory, Gustatory, Olfactory, Visual, Tactile, Interoceptive, Hand_Arm, Foot_Leg, Head, Mouth, and Torso. A word is assigned a value for each dimension which reflects how strongly the

[3] https://huggingface.co/FacebookAI/roberta-base

lexicalized concept is experienced by the respective sensor or affector (Lynott et al., 2020). We trained 11 mapping regressors that can automatically deduct these values for each word from a word's BERT embedding layer (layer 0). Each of the 11 regressors is a neural network mapping a 768-dimension embedding to a single dimension value (two fully connected hidden layers of the size of 384 and 192 respectively, both activated by ReLU). The training and evaluation details are in the Appendix.

3 Experiments

3.1 Dataset

We used the VUA-family datasets (VUA-18, VUA-20) provided by Choi et al. (2021) for training and testing. Moreover, to examine our model's generalizability to non-VUA datasets, we also tested our model on MOH (Mohammad et al., 2016) and Trofi (Birke and Sarkar, 2006) in a zero-shot transfer setting [4]. For all the datasets, we adopted the existing split of train, dev, test.

3.2 Baseline models

For a thorough comparison, we selected six baseline models:

MelBERT (Choi et al., 2021) incorporates SPV and MIP for metaphoricity prediction. EmbodiedBERT differs from it by substituting MIP with MIP_SM.

SGNN (Wan et al., 2023) simply incorporates sensorimotor information as word-level feature enrichment. It concatenates words' GloVe embeddings and sensorimotor values from Lancaster Sensorimotor norm as input for a recurrent neural network for metaphoricity prediction.

MrBERT (Song et al., 2021) explores the relations between metaphorical verbs and their various contexts, and predicts whether the relations are likely to be metaphorical.

MisNet (Zhang and Liu, 2022) implements MIP and SPV with different encoding and feature concatenation strategies.

BasicBERT (Li et al., 2023b) also proposes a new variant MIP, which can better model the meaning discrepancy between target word in context and its basic meaning. Compared with their model, EmbodiedBERT offers a cognitively motivated measure of contextual meaning change.

FrameBERT (Li et al., 2023a) also attempts to leverage external knowledge base FrameNet. It augments word embedding with self-trained FrameNet embedding for modelling MIP and SPV.

For all the baseline models except MelBERT, we directly obtain the performance of these baselines from the previous publications. We used our reproduced results of MelBERT, and ran two-tailed t-tests to compare it with our model.

	Model	**Prec**	**Rec**	**F1**
VUA-18	MrBERT	**82.7**	72.5	77.2
	MelBERT	81.2	74.7	77.8
	MisNet	80.4	<u>78.4</u>	**79.4**
	FrameBERT	**82.7**	75.3	78.8
	BasicBERT	79.5	**78.5**	<u>79.0</u>
	SGNN	76.7	75.5	76.1
	EmbodiedBERT	79.9 ± 1.1	77.9 ± 1.1	$78.9^{*}\pm0.2$
VUA-20	MrBERT	-	-	-
	MelBERT	72.9	69.5	71.0
	MisNet	-	-	-
	FrameBERT	**79.1**	67.7	<u>73.0</u>
	BasicBERT	73.3	**73.2**	**73.3**
	SGNN	-	-	-
	EmbodiedBERT	<u>73.6</u> ±1.9	<u>72.1</u> ±1.7	$72.8^{*}\pm0.2$

Table 1: Evaluation of metaphor identification systems on VUA datasets. **Bold** indicates the best, <u>underline</u> indicates the second best. * denotes our model is significantly better than MELBERT with $p < 0.05$ in two-tailed t-test

3.3 Implementation

We finetuned the hyperparameters with grid search. We increased our learning rate from 0 to 4e-5 during the first two epochs and gradually decreased it. We used the dropout rate of 0.2. The final model was trained with a batch size of 50 by three epochs, using Adam optimizer. We adopted precision, recall and f1-score as matrix for automatic evaluation. The final model's performance was obtained by averaging the results of five runs with random seeds. The experiments have been run on two NVIDA GeForce RTX 3090 GPUs, with a total of 48GB memory.

3.4 Results and discussion

Table 1 shows the automatic evaluation of our system compared with the baseline systems for metaphor detection in terms of precision, recall and f1 score.

VUA datasets For VUA-18, EmbodiedBERT achieves the third best f1 score, outperforming FrameBERT, MelBERT, MrBERT and SGNN. For

[4] Both MOH and Trofi contain exclusively verb metaphors, with the minor difference that the sentences in Trofi are generally longer than those in MOH.

VUA-20, our system still significantly outperforms MelBERT, but lags behind FrameBERT and BasicBERT. The consistently significant improvements over MelBERT in VUA datasets show that modelling sensorimotor change (MIP_SM) is indeed effective for detecting metaphors, for the major difference between EmbodiedBERT and MelBERT is the substitution of MIP by MIP_SM (see table 1). The results also validate our initial hypothesis that sensorimotor change of a word in context is closely associated with its metaphoricity. Given this, it would be interesting to examine whether the hypothesis holds across words of different part-of-speeches (POS) and genres. Therefore, we then break down the test results of VUA-18 by POS and genre, using two strongest models on VUA18: MisNet and MelBERT, for comparison.

Break-down analysis by POS When breaking down the results by Part-of-speech (POS) (see table 3 in Appendix), we find that our system achieves significant improvements over MelBERT in all categories except adverb. In particular, we find that EmbodiedBERT performs the best in verb metaphors (f1 = 76.1, 1.3% gain over MelBERT), though still lagging behind MisNet. Importantly, our system achieves the best result (f1 = 68.7) in one of the most challenging POS categories: adjective.

Break-down analysis by genre When dividing the results by genre (see table 4 in Appendix), our system outperforms all other systems in academic writings (f1 = 84.3, 0.5% gain over MelBERT) and achieves the second best in news (f1 = 78.9, 2.2% gain over MelBERT). However, our system does not beat MelBERT significantly in the genres of conversation and fiction.

The break-down analyses show that the incorporation of sensorimotor change is particularly useful for certain lexical categories and genres, which makes our system even outperform the strongest model on VUA-18 sometimes. The results thus warrant a more in-depth investigation into the complicated interactions between metaphoricity and other linguistic variables in influencing embodiment change in the future.

Transfer to non-VUA datasets We also tested our system's transferability to non-VUA datasets (see table 5 in Appendix), like TroFi and MOH-X, and the overall results are shown in the following table. Our system outperforms MelBERT in both datasets but not significantly. In general, the transfer ability of our system is not particularly strong [5].

Case analysis Finally, we intend to qualitatively reveal how the integration of sensorimotor change can help the model reduce both false positives and false negatives, so we compared our model's predictions for VUA20-test with the predictions of MelBERT (see more in table 6 in Appendix). For the reduction of false positives, EmbodiedBERT does not identify literal phrases with a minimal sensorimotor change as metaphor. For example, in the phrase 'MODERN trams, as most continental Europeans know, neither shake nor rattle, nor do they roll.', 'shake' and 'rattle' are supposed to be literal description of the tram's movement, but MelBERT predicts them to be metaphor. For the reduction of false negatives, our system is more skilled at identifying embodiment-based metaphors. For example, it can successfully identify visual metaphors like 'hazy' in 'a poet's sense of other people's very hazy', which represents cognitive incapacity by visual haziness. As 'hazy' is used metaphorically to denote cognitive property, the perceptual strength on the visual dimension should be particularly lower, along with the general shrinking of sensorimotor strength, which is accurately captured by the output from our regressors for this example (as shown in the figures 2 and 3).

4 Conclusion

In this study, we contribute a cognitively motivated system for metaphor detection EmbodiedBERT, which implements the idea that metaphorical words tend to show unique patterns of sensorimotor change in context. We have demonstrated quantitatively and qualitatively that incorporating the cognitive module MIP_SM can lead to performance improvements over systems simply using MIP. Based on our results, we envision that the incorporation of embodiment information cannot only benefit metaphor detection, but also many other language understanding tasks that require embodied experience.

Limitations

There are some limitations to be addressed in the future research. First, the modelling of sensorimotor change highly depends on the representations of basic meaning and contextual meaning of the

[5] As a reviewer points out, different annotation styles between VUA-family datasets and non-VUA datasets may be an obstacle for performance transfer

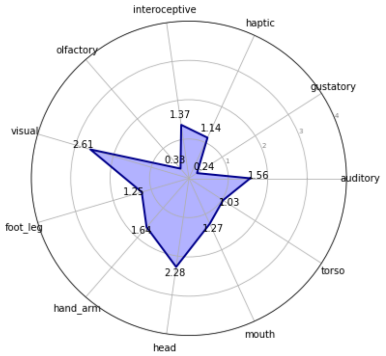

Figure 2: basic sensorimotor values of 'hazy'

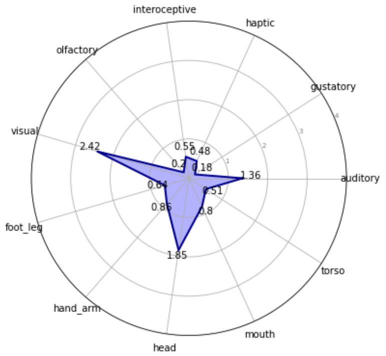

Figure 3: contextual sensorimotor values of 'hazy'

target word. We currently used the output by feeding single words into the encoder to represent their basic meaning, but a more precise basic meaning representation will be beneficial, which has been investigated by some researchers (e.g. Li et al. (2023a), Zhang and Liu (2022)).

Second, we currently used a relatively simple method to derive contextual and basic sensorimotor representation. We envision that a more sophisticated way of integrating sensorimotor change will not only improve the performance on existing datasets, but could also be beneficial for increasing the system's transfer ability to detect novel metaphors in new datasets.

Third, factors like conventionality of a metaphor (Bowdle and Gentner, 2005) and abstractness of its original meaning are also likely to influence its sensorimotor change in context. However, these annotations are currently not available in the metaphor datasets. We will focus on the annotation of existing datasets with more linguistic dimensions and examine the performance variation accordingly.

Finally, compared with BERT, recent large language models presumably contain more embodiment knowledge due to more sufficient training and more diverse inputs, which could be a more ideal source for deriving embodiment representation in the future.

Acknowledgments

We would like to thank for all the anonymous reviewers who contributed insightful suggestions on the revision of this paper.

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473.

Julia Birke and Anoop Sarkar. 2006. A clustering approach for nearly unsupervised recognition of nonliteral language. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (EACL)*.

Brian F. Bowdle and Dedre Gentner. 2005. The career of metaphor. *Psychological Review*, 112(1):193.

Ludmila Bulat, Stephen Clark, and Ekaterina Shutova. 2017. Modelling metaphor with attribute-based semantics. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 523–528, Valencia, Spain. Association for Computational Linguistics.

M. Choi, S. Lee, E. Choi, H. Park, J. Lee, D. Lee, and J. Lee. 2021. Melbert: Metaphor detection via contextualized late interaction using metaphorical identification theories. *arXiv preprint arXiv:2104.13615*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Dedre Gentner. 1983. Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2):155–170.

R. W. Jr. Gibbs, P. L. Costa Lima, and E. Francozo. 2004. Metaphor is grounded in embodied experience. *Journal of Pragmatics*, 36(7):1189–1210.

Raymond W. Jr. Gibbs. 2023. Pragmatic complexity in metaphor interpretation. *Cognition*, 237:Article 105455.

Anja Jamrozik, Marguerite McQuire, Eileen R. Cardillo, and Anjan Chatterjee. 2016. Metaphor: Bridging embodiment to abstraction. *Psychonomic Bulletin & Review*, 23(4):1080–1089.

Omid Khatin-Zadeh. 2023. Embodied metaphor processing: A study of the priming impact of congruent and opposite gestural representations of metaphor schema on metaphor comprehension. *Metaphor and Symbol*, 38(1):70–80.

Y. Li, S. Wang, C. Lin, and F. Guerin. 2023a. Metaphor detection via explicit basic meanings modelling. In *The 61st Annual Meeting of the Association for Computational Linguistics*.

Y. Li, S. Wang, C. Lin, F. Guerin, and L. Barrault. 2023b. Framebert: Conceptual metaphor detection with frame embedding learning. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1558–1563, Dubrovnik, Croatia. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Dermot Lynott, Louise Connell, Marc Brysbaert, James Brand, and James Carney. 2020. The lancaster sensorimotor norms: multidimensional measures of perceptual and action strength for 40,000 english words. *Behavior Research Methods*, 52:1271–1291.

Lotte Meteyard, Sara Rodriguez Cuadrado, Bahador Bahrami, and Gabriella Vigliocco. 2012. Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, 48(7):788–804. Language and the Motor System.

Saif Mohammad, Ekaterina Shutova, and Peter D. Turney. 2016. Metaphor as a medium for emotion: An empirical study. In *Proceedings of the Joint Conference on Lexical and Computational Semantics (*SEM@ACL)*.

Soujanya Poria, Erik Cambria, and Alexander Gelbukh. 2016. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, 108:42–49.

Pragglejaz Group. 2007. Mip: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22(1):1–39.

Sunny Rai and Shampa Chakraverty. 2020. A survey on computational metaphor processing. *ACM Computing Surveys*, 53(2).

Percy Shelley. 1890. *A Defense of Poetry*. Ginn, Boston.

W. Song, S. Zhou, R. Fu, T. Liu, and L. Liu. 2021. Verb metaphor detection via contextual relation learning. 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 4240–4251.

Chuandong Su, Fumiyo Fukumoto, Xiaoxi Huang, Jiyi Li, Rongbo Wang, and Zhiqun Chen. 2020. Deepmet: A reading comprehension paradigm for token-level metaphor detection. In *The Second Workshop on Figurative Language Processing (FigLang@ACL)*, pages 30–39.

Jacob Turton, Robert Elliott Smith, and David Vinson. 2021. Deriving contextualised semantic features from BERT (and other transformer model) embeddings. In *Proceedings of the 6th Workshop on Representation Learning for NLP (RepL4NLP-2021)*, pages 248–262, Online. Association for Computational Linguistics.

Mingyu Wan, Qi Su, Kathleen Ahrens, and Chu-Ren Huang. 2023. Perceptional and actional enrichment for metaphor detection with sensorimotor norms. *Natural Language Engineering*, page 1–29.

Yorick Wilks. 1975. A preferential, pattern-seeking, semantics for natural language inference. *Artificial Intelligence*, 6(1):53–74.

Margaret Wilson. 2002. Six views of embodied cognition. *Psychonomic Bulletin Review*, 9:625–636.

C. Wu, F. Wu, Y. Chen, S. Wu, Z. Yuan, and Y. Huang. 2018. Neural metaphor detecting with cnn-lstm model. In *Proceedings of the Workshop on Figurative Language Processing*, pages 110–114.

Shenglong Zhang and Ying Liu. 2022. Metaphor detection via linguistics enhanced Siamese network. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4149–4159, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

A Appendix

A.1 Structure Mapping Theory and Career of Metaphor

Structure mapping theory (Gentner, 1983) aims to offer a general way of accounting for conceptual analogy, of which metaphor is a specific category. It proposes that any kind of analogy involves two processing stages: structural alignment and projection. To process analogies, human begin to take two entities in an analogy into comparison and structurally align their corresponding properties. The alignment process observes three principles: one-to-one mapping, parallel connectivity, and systematicity. Features of the source concept which fail to connect to the aligned system due to the violation of the principles will be shed away from source representation, and thus cannot be projected to the target representation. In the case of metaphor processing, sensorimotor features that do not connect to the aligned system will be inhibited, resulting in an overall change of sensorimotor levels.

Career of Metaphor framework (Bowdle and Gentner, 2005) predicts that metaphor processing mode will switch as a metaphor becomes conventionalized. Namely, understanding a conventionalized metaphor does not require the structural alignment stage, which makes embodiment change happening in this stage less likely to occur.

A.2 Text representation

We use the byte-pair encoding (BPE) to tokenize S. Following Choi et al. (2021), we use the position embedding to distinguish target word and its local context. Also, following Su et al. (2020), after adding special tokens `[CLS]` and `[SEP]` to the beginning and the end of S, we utilize the part of speech (POS) information of the target word by appending its POS after `[SEP]`. Finally, we feed the element-wise addition of BPE token embedding, position embedding and segment embedding of S as input into the first encoder.

A.3 Training and evaluation of sensorimotor regressors

To train the regressors, we used Lancaster Sensorimotor Norm (Lynott et al., 2020), which contains contains 11-dimension sensorimotor information for 39,707 English words. We use word embedding from BERT embedding layer as input (Devlin et al., 2019). The size of overlapping vocabulary of Lancaster Sensorimotor Norm and BERT vocabulary is 11,402, and we split it into training and testing with the ratio of 8:2. We use mean squared error as criterion for calculating loss and adopt Adam optimizer for parameter updating. Our initial learning rate is 0.001 and gradually decreased by the factor of 0.1 with the patience of 10. We perform 5-fold cross-validation and use early stopping to save the best model based on the loss on validation set. For evaluation, we use Pearson correlations of models' predicted values with human rating. Overall, the relatively high correlations suggest that our regressors can reliably deduct sensorimotor information from word embeddings (see table 2). However, whether the regressors can predict sensorimotor values for a word in context remains an issue to be examined, though a similar approach by Turton et al. (2021) have shown its potentials.

Dimension	**BERT**
auditory	0.76
gustatory	0.78
haptic	0.79
interoceptive	0.81
olfactory	0.75
visual	0.72
foot_leg	0.74
hand_arm	0.73
head	0.61
mouth	0.73
torso	0.69
by-word	0.88

Table 2: Correlations of sensorimotor prediction with human judgement

POS	**Model**	**F1**	**Prec**	**Rec**
ADJ	MelBERT	65.6	<u>71.5</u>	60.6
	MisNet	<u>67.0</u>	68.8	<u>65.2</u>
	EB	**68.7*** $\pm$1.0	**70.0**$\pm$2.3	**67.6**$\pm$2.4
ADV	**MelBERT**	72.6	**79.1**	67.0
	MisNet	**73.3**	76.4	**70.5**
	EB	<u>72.7</u>$\pm$2.0	<u>77.2</u>$\pm$2.4	<u>68.7</u>$\pm$1.5
NOUN	MelBERT	<u>67.5</u>	**77.1**	60.0
	MisNet	**70.6**	74.4	**67.2**
	EB	**70.6*** $\pm$0.7	<u>76.3</u>$\pm$0.6	<u>65.7</u>$\pm$1.4
VERB	MelBERT	75.2	<u>76.7</u>	73.8
	MisNet	**77.6**	**77.5**	**77.6**
	EB	<u>76.1</u>* $\pm$0.5	74.8$\pm$1.4	<u>77.5</u>$\pm$1.0

Table 3: POS-specific evaluation of VUA-18 testing results. **Bold** indicates the best, <u>underline</u> indicates the second best. EB refers to EmbodiedBERT. * denotes our model is significantly better than MELBERT with $p < 0.05$ in two-tailed t-test

Genre	Model	F1	Prec	Rec
Acad	MelBERT	83.5	<u>87.5</u>	9.8
	MisNet	<u>83.8</u>	85.1	<u>82.5</u>
	EB	**84.3*** $\pm$ 0.3	**85.8** $\pm$ 0.1	**82.9** $\pm$ 0.9
Conv	MelBERT	69.6	<u>70.5</u>	8.7
	MisNet	**71.9**	**71.8**	**72.0**
	EB	<u>70.0</u> $\pm$ 0.5	69.7 $\pm$ 1.1	<u>70.4</u> $\pm$ 1.3
Fict	MelBERT	<u>74.5</u>	<u>74.4</u>	74.7
	MisNet	**76.0**	**74.5**	**77.5**
	EB	<u>75.2</u> $\pm$ 0.8	73.0 $\pm$ 2.0	<u>77.7</u> $\pm$ 1.0
News	MelBERT	77.1	83.6	71.5
	MisNet	**79.7**	**82.6**	**77.0**
	EB	<u>78.9</u>* $\pm$ 0.7	**82.6** $\pm$ 1.0	<u>75.6</u> $\pm$ 1.7

Table 4: Genre-specific evaluation of VUA-18 testing results. Acad: academic; Conv: conversation; Fict: fiction. **Bold** indicates the best, <u>underline</u> indicates the second best. EB refers to EmbodiedBERT. * denotes our model is significantly better than MELBERT with p < 0.05 in two-tailed t-test

Dataset	Model	F1	Prec	Rec
TroFi	MelBERT	61.9	53.6	73.4
	MrBERT	<u>72.9</u>	**73.9**	72.1
	MisNet	-	-	-
	FrameBERT	**74.2**	<u>70.7</u>	**78.2**
	EB	62.4 $\pm$ 0.6	52.9 $\pm$ 0.9	76.2 $\pm$ 1.0
MOH-X	MelBERT	78.2	78.7	78.4
	MrBERT	**84.2**	<u>84.1</u>	**85.6**
	MisNet	83.4	**84.2**	84.0
	FrameBERT	<u>83.8</u>	83.2	<u>84.4</u>
	EB	78.6 $\pm$ 1.2	75.6 $\pm$ 2.5	82.4 $\pm$ 2.0

Table 5: Zero-shot transfer to non-VUA datasets. **Bold** indicates the best, <u>underline</u> indicates the second best. EB refers to EmbodiedBERT

Sentence	**True**	**EB**	**MB**
This violent event, described at **length** in hysterically colourful terms, is the only piece of history to be woven convincingly into the plot.	0	0	1
Hardly a page goes by without the hapless Francis noticing something which **reminds** him, improbably, of something else.	0	0	1
There are strict time **limits**: generally, six years from when damage first occurred...	0	0	1
A solicitor **fails** to draw up a will within a reasonable time for a client who subsequently dies.	0	0	1
Children still would not have full political **status**.	0	0	1
That, says Mr Tyson, has been their only **blessing**.	1	1	0
But '**posturing** and pretending' went far beyond the unions.	1	1	0
But the chief result of all this farming was to produce huge food **mountains** which we could then refuse to give to the Third World	1	1	0
Nowadays, we all **swoon** with pleasure at the sight of a cow.	1	1	0
Though individuals are nailed, the greatest **villain** of all is the system.	1	1	0
Berry's songs are plausible **emblems** of rock'n'roll rebellion or, at any rate, youthful hedonism.	1	1	0

Table 6: Case analysis based on VUA-20 testing: reduction of false positives and negatives by EmbodiedBERT (EB) compared with MelBERT (MB). **Bold** indicates the target word.