Sebastian Reimann, Tatjana Scheffler

The Struggles of Large Language Models with Zero- and Few-Shot (Extended) Metaphor Detection

Abstract

Extended metaphor is the use of multiple metaphoric words that express the same domain mapping. Although it would provide valuable insight for computational metaphor processing, detecting extended metaphor has been rather neglected. We fill this gap by providing a series of zero- and few-shot experiments on the detection of all linguistic metaphors and specifically on extended metaphors with LLaMa and GPT models. We find that no model was able to achieve satisfactory performance on either task, and that LLaMa in particular showed problematic overgeneralization tendencies. Moreover, our error analysis showed that LLaMa is not sufficiently able to construct the domain mappings relevant for metaphor understanding.

1 Introduction

Mappings between an often concrete source domain (e.g. MONEY) and a more abstract target domain (e.g. TIME), so-called conceptual metaphors, structure the way how humans think, according to the Conceptual Metaphor Theory (CMT) of Lakoff and Johnson (1980). These conceptual mappings manifest in language as linguistic metaphors, such as "spending time at home". Extended metaphor, according to Reijnierse, Burgers, Krennmayr, and Steen (2020), represents a special case of linguistic metaphor where multiple metaphor-related words (MRWs) express the same mapping from a source domain to a target domain. Reijnierse et al. (2020) use example (1), taken from a newspaper report on Welsh rugby, to illustrate this. Three MRWs map from the source domain of FIRE to the target domain of RUGBY: a risky arrangement of club fixtures is compared to playing with fire and the negative result is described as being consumed in a conflagration.

(1) They were playing with fire when they decided to arrange a couple of club fixtures and they have been duly consumed in conflagration of their own making.

The automatic detection of single MRWs has received considerable attention within the NLP community. However, extended metaphor remains a phenomenon that has only received little attention in computational work on metaphor. Ge, Mao, and Cambria (2023) explicitly state that there is a lack of work on the detection of extended metaphor, which also extends to the availability of suitable datasets. In such scenarios, zero- (without any labeled examples) and few-shot (with few labeled examples) prompting

the current generation of generative, decoder-only large language models (LLMs) like (Chat)GPT and LLaMa has become a useful alternative, as demonstrated in NLP tasks such as sentiment analysis and named entity recognition (Qin et al., 2024).

Ge et al. (2023) moreover note that detecting extended metaphor would require an understanding of domain mappings according to CMT. Consequently, evaluating the performance of LLMs on extended metaphor detection would not just provide insight into the metaphor detection qualities of LLMs but also demonstrate whether the metaphor processing by LLMs follows the assumption made by Lakoff and Johnson (1980) on metaphor and human cognition. We thus make the following contributions:

- We present a series of experiments on metaphor and extended metaphor detection using models from the two most common LLM families and various prompts, where we find that referencing CMT helped overall and that especially LLaMa heavily overgeneralized the positive class. The code for these experiments is publicly available.[1]
- We conduct an extensive error analysis in order to interpret the behavior of LLMs when prompted for extended metaphor detection, which raises serious doubts that the current LLaMa models are able to actually construct domain mappings according to CMT.

2 Previous Work

2.1 Metaphor and LLMs

Finetuning of encoder-only, pre-trained transformer LMs like BERT has been extensively employed for the task of automatic metaphor detection. Often, multiple encoders were combined to model the linguistic theories of Metaphor Identification Procedure (MIP, Pragglejaz Group, 2007), focusing on a semantic clash between the contextual and a more basic meaning and Selectional Preference Violations (SPV, Wilks, 1975), focusing on the clash between a metaphoric word and its context (Babieno, Takeshita, Radisavljevic, Rzepka, & Araki, 2022; Choi et al., 2021; Li, Wang, Lin, & Guerin, 2023; Zhang & Liu, 2023).

The metaphor identification and interpretation abilities of generative LLMs were so far mostly tested on smaller data. Wachowiak and Gromann (2023) found that GPT-3 was able to predict the source domain of metaphors, mostly from the Master Metaphor List by George Lakoff[2], with an accuracy of 60.22%. Schuster and Markert (2023) included ChatGPT in their zero- and few-shot experiments on metaphor detection in adjective-noun pairs, and found that its zero-shot performance was however outperformed by smaller models that were fine-tuned on labeled data. Goren and Strapparava (2024) tested the ability of GPT-3.5 to identify metaphors in English and Italian proverbs,

[1] https://github.com/SFB-1475/C04-LLMFails-Metaphor
[2] https://www.lang.osaka-u.ac.jp/~sugimoto/MasterMetaphorList/metaphors/index.html

where prompts that asked the model to identify the meaning before identifying the metaphorical parts and the inclusion of larger contexts led to the best results.

The most elaborate approach to metaphor detection via LLMs, called TSI (Theory-Guided Scaffolding Instruction), was put forward by Tian, Xu, and Mao (2024). In order to fill slots in a knowledge graph, TSI prompts GPT-3.5 with a series of questions (either grounded in MIP, SPV or CMT), on the source and target domain of a word and whether these are different. After comparing the structure of the knowledge graphs, TSI provides a label (metaphoric or not). On the TroFi (Birke & Sarkar, 2006) and MOH-X (Mohammad, Shutova, & Turney, 2016) datasets, TSI outperformed several prompting-based methods and BERT models. However, Tian et al. (2024) state that for a large-scale evaluation of their method, they so far lacked the resources.

2.2 Computational Approaches to Extended Metaphor

Although the specific automatic identification of extended metaphor in particular has not yet been tackled, some works on metaphor detection have touched upon the concept of extended metaphor. Jang, Maki, Hovy, and Rosé (2017) present a method of automatically finding metaphors that particularly emphasizes extended metaphor. They use an unlabeled corpus and seed words that represent a source domain and its facets (e.g. the domain JOURNEY and *long*) to extract further potential seed words and repeat the procedure several times. Ultimately, features based on these clusters were added to input vectors for an SVM and helped to improve metaphor detection for the JOURNEY domain in posts from a forum of cancer patients.

Reimann and Scheffler (2024) provide a dataset of posts from Christian subreddits annotated via MIPVU and DMIP (Deliberate Metaphor Identification Procedure) (Reijnierse, Burgers, Krennmayr, & Steen, 2018). The latter requires a reason why a metaphorical expression is considered potentially deliberate (i.e. used and intended to be understood "as metaphor"), and a word's status as an extended metaphor was among the possible choices. It comprises annotations for 16,540 tokens, where 3,523 are MRWs, and is further subdivided into a test set of 14,981 tokens and a small training section (used as additional training data in transfer scenarios) of 1,559 tokens. They use the dataset to evaluate the cross-genre transfer capabilities of metaphor detection systems. Additionally, they look at the share of detected potentially deliberate MRWs and find that extended metaphors pose great problems for BERT-based state-of-the-art metaphor detection systems.

3 Data and Setup

For evaluation purposes, the test set of Reimann and Scheffler (2024) was a logical choice, given its annotations on extended metaphor. They also provide metaphor annotation on entire Reddit posts and thus larger discourse contexts, which is useful since extended metaphors may stretch over multiple sentences (Reimann & Scheffler, 2024). In total, the dataset contains 281 posts, out of which 72 contain extended metaphor.

To put the results on extended metaphor into the context of general metaphor understanding, we will additionally carry out a token-based identification of all MRWs. This has not yet been tried for the Reddit dataset that we use and, to the best of our knowledge, the detection of metaphoric tokens in text has not been attempted on a larger dataset since both datasets used by Tian et al. (2024) were smaller and focused on metaphoric word pairs.

Including theoretical ideas from CMT into the prompts had a beneficial effect in Tian et al. (2024). For our definition of *extended metaphor*, the notions of source and target domain play a crucial role. Consequently, we design our prompt involving CMT very similarly to the prompt of Tian et al. (2024) without knowledge graphs and scaffolding, which explains the terms metaphor, source domain and target domain according to CMT with the help of an example. In the second part of our prompt, we then define extended metaphor according to Reijnierse et al. (2020). Finally, we also ask the model if an extended metaphor is present and of which MRWs it is made up. We provide all prompts in the appendix.

In our experiments, we use models from the two most frequently used LLM families: LLaMa (Touvron et al., 2023) and GPT (Brown et al., 2020). For LLaMa, we specifically use the instruction-tuned Llama-3.1-8B-Instruct and its larger counterpart Llama-3.1-70B-Instruct, obtained via the HuggingFace transformers library (Wolf et al., 2020). For GPT, we use GPT-4o-mini, which is, at the time of writing, claimed by OpenAI to be more powerful than GPT-3 and GPT-3.5 (used in the approaches mentioned in section 2.1) and advertised as the most cost-efficient version of GPT, which we access via the OpenAI API. The API has a limit of 10,000 requests per day for GPT-4o-mini. Thus, for the word-based, general metaphor detection with GPT, we will thus use only 2,500 tokens from the Reddit dataset. We aim for predictable and reproducible behavior and thus used low values of 0.1 for the top_p and temperature hyperparameters, which control the creativity of the model. For all other hyperparameters, we choose the default values.

4 Results

The left side of Table 1 shows the results for the token-based automatic metaphor detection. All models appear to show a large tendency to overgeneralize the metaphor label, with recall much higher than precision. Given the limitations of the OpenAI API and the smaller test set used for the GPT model, a direct comparison of the performance of the two models is not entirely possible. However, it seems that overgeneralization of what may be considered a metaphoric token is much more prominent for LLaMa, compared to GPT. Introducing ideas from CMT had a slight positive effect on precision for all models.

In the extended metaphor detection experiments (right side of Table 1), we observe that the two LLM families exhibit strikingly different behavior. The LLaMa models prompted in a zero-shot fashion exhibit a tendency to overgeneralize by reaching satisfactory levels of recall but largely doing so at the cost of precision. For GPT-4o-

	Detection of all MRWs						Extended Metaphor Detection								
	No CMT			CMT			Zero			Zero+CMT			Few+CMT		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
LLaMa 3.1 8b	23	98	37	25	97	40	38	67	48	36	79	49	46	39	42
LLaMa 3.1 70b	26	98	42	33	93	49	46	62	53	41	72	53	46	50	48
GPT-4o-mini	32	80	46	42	60	49	53	33	41	56	32	41	59	22	32

Table 1: Precision, Recall and F1-scores for the metaphor class in experiments for the detection of all metaphors (left) and extended metaphor (right).

mini, this was the other way around. It was notably more careful in its predictions but failed to recognize a wide range of extended metaphors.

Choosing the 70b version over the 8b version only led to small improvements, suggesting that model size only has a minor impact on the metaphor understanding capabilities of LLaMa 3.1. A notable improvement in precision was achieved by showing the models two examples of extended metaphor. This, however, led vice versa to massive drops in recall, which hints that the models then were not able to properly generalize from the provided examples.

Finally, when evaluating the extraction of the exact MRWs constituting the extended metaphor, the models failed entirely. Only 8% (8b) and 13% (70b) of the MRWs extracted by the LLaMa models were actual MRWs and only 9% of the MRWs extracted by the GPT model were actually labelled as such. Conversely, the LLaMa models also only found 38% (8b) and 31% (70b) of the MRWs in the entire dataset respectively and GPT-4o-mini only detected 18% of the MRWs in the examples containing extended metaphors.

5 Error Analysis and Discussion

(2) Before you turn your backyard into a garden or homeless shelter, you need to check city and possibly neighborhood ordinances.

(3) I 've been burned by the hook-up culture many times before . I still have trouble completely renouncing it honestly. What should I do?

(4) Jesus took back the keys of hell at the cross.

In order to better understand the behavior of LLaMa demonstrated in section 4, we had a closer look at the false positives (examples that were falsely classified as extended metaphor) and identified three main categories of errors: (i) overinterpretation of entirely literal text, such as in the non-metaphoric example (2), where LLaMa 8b considered *backyard*, *garden*, *homeless* and *shelter* as MRWs; (ii) MRWs from different

	Over-interpretation	Different Domains	Wrong Boundaries	Other	Total
8b	44	39	17	2	102
70b	23	33	14	4	74

Table 2: Distribution of the different error types.

domains, as in (3), where *burned* and *hook-up* were correctly identified as MRWs but express different domain mappings and thus not an extended metaphor; and (iii) wrong boundaries like example (4) that only contains a single MRW (*keys*) but where models recognized an extended metaphor and classified further terms (here: *hell*).

Additionally, for cases that did not completely fit any of the aforementioned categories, we used the category "other issues". Table 2 shows how the different error types are distributed. For the smaller version of LLaMa, the main problem is still its strong tendency to consider a wide range of literal terms to be metaphoric. This changes slightly for the 70B version as the amount of non-metaphoric misclassified examples drops by almost half of the amount of the 8B version. The amount of false positives related to a domain confusion on the other hand remained stable.

This raises doubts whether the models really are able to understand the notion of a domain mapping according to CMT. At first glance, introducing CMT appeared to have improved performance in our experiments. However, the prominence of errors like example (3), that could not even be fixed by choosing a larger model, suggests otherwise. This is partially in line with the findings of Wachowiak and Gromann (2023) for GPT, where the main source of error in predicting the underlying conceptual metaphor was selecting a wrong source domain. In several cases, this happened because GPT-3 was triggered by non-metaphoric words related to the source domain, suggesting again that it lacked the capability to differentiate between domains.

However, it is still hard to discuss the overgeneralizing behavior of LLaMa within the context of previous work since, on the one hand, in previous work on the metaphor understanding capabilities of generative LLMs, mostly models from the GPT family were used. On the other hand, previous studies also employed smaller and more balanced datasets, which may to some extent overshadow such an overuse of the metaphor label as experienced in our case. The Reddit dataset of Reimann and Scheffler (2024) is much larger in size but notably less balanced with only around 20% of the words being metaphors. It can be argued that the large number of non-metaphoric examples is more representative of metaphor use in everyday language (Steen et al., 2010), and thus larger, authentic datasets are more useful for evaluating the metaphor detection and understanding capabilities of LLMs.

Finally, since metaphor annotation is also a challenging task for human annotators, we look at cases where, during the DMIP annotation of Reimann and Scheffler (2024), the two annotators disagreed on extended metaphor. This happened in 40 posts. In 33

of these cases, the annotators decided on the positive label. Seven of them, however, were eventually not considered to express extended metaphor. Out of these seven examples, the two LLaMa models labeled four (8B) and five (70B) as containing an extended metaphor in the best prompting scenario. These overgeneralization cases may thus also be explained by these examples being also ambiguous to human annotators.

6 Conclusion and Future Work

We evaluated the capabilities of two state-of-the-art LLM families to find metaphorically used words and extended metaphors. We then carried out a systematic error analysis of the output of the best performing model-prompt-combination. We found that the LLMs failed in two different ways: We observed a general strong bias towards the metaphor and extended metaphor labels, especially with the LLaMa models. Moreover, a closer look at these overgeneralization errors in extended metaphor detection suggests that the models failed to construct the domain mapping required to understand extended metaphor.

Thus, for future work, we suggest to further investigate and find the source of the overgeneralization bias that has plagued all experiments involving LLaMa. Moreover, a more complex prompting approach, similar to for example what Tian et al. (2024) were aiming for, might be worth trying out in order to address the difficulties of the models to understand and connect the source domains of MRWs.

Funding Statement

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1475 – Project ID 441126958.

References

Babieno, M., Takeshita, M., Radisavljevic, D., Rzepka, R., & Araki, K. (2022). Miss roberta wilde: Metaphor identification using masked language model with wiktionary lexical definitions. *Applied Sciences*, *12*(4). Retrieved from `https://www.mdpi.com/2076-3417/12/4/2081` doi: 10.3390/app12042081

Birke, J., & Sarkar, A. (2006, April). A clustering approach for nearly unsupervised recognition of nonliteral language. In D. McCarthy & S. Wintner (Eds.), *11th conference of the European chapter of the association for computational linguistics* (pp. 329–336). Trento, Italy: Association for Computational Linguistics. Retrieved from `https://aclanthology.org/E06-1042`

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... Amodei, D. (2020). *Language models are few-shot learners.* Retrieved from `https://arxiv.org/abs/2005.14165`

Choi, M., Lee, S., Choi, E., Park, H., Lee, J., Lee, D., & Lee, J. (2021, June). MelBERT: Metaphor detection via contextualized late interaction using metaphorical identifi-

cation theories. In K. Toutanova et al. (Eds.), *Proceedings of the 2021 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 1763–1773). Online: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2021.naacl-main.141 doi: 10.18653/v1/2021.naacl-main.141

Ge, M., Mao, R., & Cambria, E. (2023). A survey on computational metaphor processing techniques: from identification, interpretation, generation to application. *Artificial Intelligence Review*, 1–67.

Goren, G., & Strapparava, C. (2024, May). Context matters: Enhancing metaphor recognition in proverbs. In N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, & N. Xue (Eds.), *Proceedings of the 2024 joint international conference on computational linguistics, language resources and evaluation (lrec-coling 2024)* (pp. 3825–3830). Torino, Italia: ELRA and ICCL. Retrieved from https://aclanthology.org/2024.lrec-main.338

Jang, H., Maki, K., Hovy, E., & Rosé, C. (2017, August). Finding structure in figurative language: Metaphor detection with topic-based frames. In K. Jokinen, M. Stede, D. DeVault, & A. Louis (Eds.), *Proceedings of the 18th annual SIGdial meeting on discourse and dialogue* (pp. 320–330). Saarbrücken, Germany: Association for Computational Linguistics. Retrieved from https://aclanthology.org/W17-5538 doi: 10.18653/v1/W17-5538

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago [u.a.]: Univ. of Chicago Press.

Li, Y., Wang, S., Lin, C., & Guerin, F. (2023, July). Metaphor detection via explicit basic meanings modelling. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 91–100). Toronto, Canada: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2023.acl-short.9 doi: 10.18653/v1/2023.acl-short.9

Mohammad, S., Shutova, E., & Turney, P. (2016, August). Metaphor as a medium for emotion: An empirical study. In C. Gardent, R. Bernardi, & I. Titov (Eds.), *Proceedings of the fifth joint conference on lexical and computational semantics* (pp. 23–33). Berlin, Germany: Association for Computational Linguistics. Retrieved from https://aclanthology.org/S16-2003 doi: 10.18653/v1/S16-2003

Pragglejaz Group. (2007, January). MIP: A Method for Identifying Metaphorically Used Words in Discourse. *Metaphor and Symbol*, *22*(1), 1–39. doi: 10.1080/10926480709336752

Qin, L., Chen, Q., Feng, X., Wu, Y., Zhang, Y., Li, Y., ... Yu, P. S. (2024). *Large language models meet nlp: A survey.* Retrieved from https://arxiv.org/abs/2405.12819

Reijnierse, W. G., Burgers, C., Krennmayr, T., & Steen, G. J. (2018). DMIP: A method for identifying potentially deliberate metaphor in language use. *Corpus Pragmatics*, *2*(2), 129–147.

Reijnierse, W. G., Burgers, C., Krennmayr, T., & Steen, G. J. (2020). The role of

co text in the analysis of potentially deliberate metaphor. *Drawing attention to metaphor: Case studies across time periods, cultures and modalities*, 15–38.

Reimann, S., & Scheffler, T. (2024, May). Metaphors in online religious communication: A detailed dataset and cross-genre metaphor detection. In N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, & N. Xue (Eds.), *Proceedings of the 2024 joint international conference on computational linguistics, language resources and evaluation (lrec-coling 2024)* (pp. 11236–11246). Torino, Italia: ELRA and ICCL. Retrieved from https://aclanthology.org/2024.lrec-main.982

Schuster, J., & Markert, K. (2023, September). Nut-cracking sledgehammers: Prioritizing target language data over bigger language models for cross-lingual metaphor detection. In E. Breitholtz, S. Lappin, S. Loaiciga, N. Ilinykh, & S. Dobnik (Eds.), *Proceedings of the 2023 clasp conference on learning with small data (lsd)* (pp. 98–106). Gothenburg, Sweden: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2023.clasp-1.12

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., & Pasma, T. (2010). *A Method for Linguistic Metaphor Identification: From MIP to MIPVU* (Vol. 14). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Tian, Y., Xu, N., & Mao, W. (2024, June). A theory guided scaffolding instruction framework for LLM-enabled metaphor reasoning. In K. Duh, H. Gomez, & S. Bethard (Eds.), *Proceedings of the 2024 conference of the north american chapter of the association for computational linguistics: Human language technologies (volume 1: Long papers)* (pp. 7738–7755). Mexico City, Mexico: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2024.naacl-long.428 doi: 10.18653/v1/2024.naacl-long.428

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., ... Lample, G. (2023). *Llama: Open and efficient foundation language models.* Retrieved from https://arxiv.org/abs/2302.13971

Wachowiak, L., & Gromann, D. (2023, July). Does GPT-3 grasp metaphors? identifying metaphor mappings with generative language models. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 1018–1032). Toronto, Canada: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2023.acl-long.58 doi: 10.18653/v1/2023.acl-long.58

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

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... Rush, A. (2020, October). Transformers: State-of-the-art natural language processing. In Q. Liu & D. Schlangen (Eds.), *Proceedings of the 2020 conference on empirical methods in natural language processing: System demonstrations* (pp. 38–45). Online: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2020.emnlp-demos.6 doi: 10.18653/v1/2020.emnlp-demos.6

Zhang, S., & Liu, Y. (2023, July). Adversarial multi-task learning for end-to-end

metaphor detection. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Findings of the association for computational linguistics: Acl 2023* (pp. 1483–1497). Toronto, Canada: Association for Computational Linguistics. Retrieved from `https://aclanthology.org/2023.findings-acl.96` doi: 10.18653/v1/2023.findings-acl.96

7 Appendix: LLM Prompts

Strategy	Prompt
CMT	"According to conceptual metaphor theory, metaphor facilitates a mapping of attributes or characteristics from source domain to target domain. For example, the word 'invested' in the sentence 'I have invested a lot of time in her' is a metaphorical expression. The source domain implied by this metaphor is the domain of money and the target domain implied by this metaphor is the domain of time. In the sentence [sentence] decide if the word [word] is a metaphorical expression. If yes, output only label 1, otherwise output only 0."
No CMT	"In the sentence {sent} decide if the word {word} is a metaphorical expression. If yes, output only label 1, otherwise output only 0."

Table 3: Prompt templates for the detection of all metaphoric tokens

Strategy	Prompt
Zero	"Extended metaphor represents a particuar case of metaphor where several metaphors express the same mapping from a source to a target domain. Based on the above information, decide whether an extended metaphor is expressed in the following text: [post]"
Zero+CMT	"According to conceptual metaphor theory, metaphor facilitates a mapping of attributes or characteristics from source domain to target domain. For example, the word 'invested' in the sentence 'I have invested a lot of time in her' is a metaphorical expression. The source domain implied by this metaphor is the domain of money and the target domain implied by this metaphor is the domain of time. Extended metaphor represents a particuar case of metaphor where several metaphors express the same mapping from a source to a target domain. Based on the above information, decide whether an extended metaphor is expressed in the following text: [post]"
Few-Shot	"According to conceptual metaphor theory, metaphor facilitates a mapping of attributes or characteristics from source domain to target domain. For example, the word 'invested' in the sentence 'I have invested a lot of time in her' is a metaphorical expression. The source domain implied by this metaphor is the domain of money and the target domain implied by this metaphor is the domain of time. Extended metaphor represents a particuar case of metaphor where several metaphors express the same mapping from a source to a domain. This is illustrated in the following two examples: 'Another time I heard someone describe Jesus as God ´s character in an MMO. He´s still God , but he´s playing on our server , and Jesus is how we see him in the game.' 'Like the closeness between him and God are such that one is the Father and the other is His Son. In this sense it gives a greater meaning to Jesus (peace be upon him) and his relationship to God.' In the first example, the words 'character', 'MMO', 'playing', 'server' and 'game' all express a mapping from the source domain of gaming to the domain of religion and transcendence. In the second example, the relationship between God and Jesus is mapped onto the family terms 'Father' and 'Son'. Based on this information, decide whether an extended metaphor is expressed in the following text: [post]"

Table 4: Prompt templates for the detection of extended metaphor

Correspondence

Sebastian Reimann

Ruhr-Universität Bochum
Germanistisches Institut
Bochum, Germany
sebastian.reimann@rub.de

Tatjana Scheffler

Ruhr-Universität Bochum
Germanistisches Institut
Bochum, Germany
tatjana.scheffler@rub.de