Pak. j. life soc. Sci. (2024), 22(1): 5629-5639 E-ISSN: 2221-7630;P-ISSN: 1727-4915

Pakistan Journal of Life and Social Sciences

www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2024-22.1.00415

RESEARCH ARTICLE

# A Study of Corpus Construction for English Metaphorical Translation Using Natural Language Processing

Honglei Zhu[1*], Omer Hassan Ali Mahfoodh[2], Debbita Tan Ai Lin[3]

[1,2,3]School of Languages, Literacies and Translation, Universiti Sains Malaysia, Malaysia

***ARTICLE INFO***

Received: May 21, 2024

Accepted: Jun 24, 2024

***Keywords***

English metaphorical translation

Machine learning

Natural language processing

CNN

Semantic role labeling

BERT

Accuracy

***Corresponding Author:**

zhuhonglei123@student.usm.my

**ABSTRACT**

Expressed sophisticated thoughts enrich metaphors; nonetheless, their intricacies provide a challenge for Natural Language Processing (NLP). This project develops a specialized corpus for English Metaphorical Translation (EMT) using recent NLP approaches such as Dependency Parsing (DP), Bidirectional Encoder Representations from Transformers (BERT), Semantic Role Labelling (SRL), and Convolutional Neural Networks (CNNs). The corpus picked from various sources, which includes literary works, academic publications, media stories, and web information, provides improved metaphor recognition and translation accuracy. Our research findings are deemed to be significant since they demonstrate the efficiency of our approach and provide valuable insights for both academic research and industrial Machine Translation (ML) applications.

## INTRODUCTION

People claim metaphorical language is essential to human interaction since it helps people communicate with other people better and effectively and provides us with a chance to express intricate concepts through symbols in language [1,19]. However, there are significant challenges in Natural Language Processing (NLP) caused by the intricate nature of metaphors, particularly when it comes to precisely identifying and expressing these expressions [2,16]. The prediction can be difficult to comprehend and interact well across languages and contexts when employing metaphors because of the significant historical and psychological implications that reach above their concrete meanings [3,18]. The current study aims to address those issues through the development of an original corpus for English Metaphorical Translation (MEL), which employs advanced NLP techniques to improve translation and comprehension of metaphorical language.

The investigation is essential as it could result in improved MT systems and language analysis through the introduction of a deeper understanding of metaphorical language. Communication, logical thought, and innovation are greatly enhanced by metaphors, which are prevalent in standard and professional language [4-5]. The scope and integrity of interactions between cultures rely on accurate translation. In the present research, researchers attempt to enhance past approaches to metaphor proof of identity, knowledge, and translation using *state-of-the-art* Machine Learning (ML) frameworks and language techniques.

In order to address the problems of metaphor translation, this investigation uses several types of modern NLP techniques that have been effective at processing metaphorical language. Recognized for their deep analysis of context capabilities are Bidirectional Encoder Representations from Transformers (BERT) and Convolutional Neural Networks (CNNs), which are adept at recognizing structural patterns in text. An additional all-encompassing approach to metaphor recognition and translation is attained by assessing syntax and lexical phrases using linguistic operations like DP and Semantic Role Labelling (SRL) [6-8].

Several sources of information, such as academic papers, literature, news articles, and online data, have been organized into a unique and coherent corpus of data by this recommended method. The wide range of metaphorical language that had been carefully collected into the above collection makes it suitable across multiple fields and types of communication [9-10]. In order to collect valid texts, the data collecting method requires using particular extraction methods like web scraping for websites, Optical Character Recognition (OCR) for manuscripts, and an application programming interface for online social networking data.

This work discusses the approaches that will be used to construct our corpus, involving data collection, preliminary processing, and annotation. In this research, researchers present the findings of an analysis of multiple NLP models that have been developed to identify and translate distinct metaphors. The findings demonstrate that the recommended strategy has made significant improvements. The results we obtain are designed to lead to the creation of more sophisticated metaphor analysis tools in NLP. The method improves researchers in academia and practical use in ML and other domains [11, 12,17].

The paper is organized in the following order: section 2 presents the methods used in the research, section 3 presents the methodology, section 4 presents the analysis, and section 5 concludes the study.

## NLP Techniques Employed

In this study, the recognition and translation of metaphorical statements are optimized by focusing on a limited number of advanced NLP techniques [13-15]. The practices are selected based on the verified usefulness and pertinency in dealing with the details of metaphorical language in big text corpora.

### Machine Learning Models

We use two basic ML models, which are especially useful for the comprehension and classification of metaphors:

### Convolutional Neural Networks (CNNs)

Since CNNs are extremely good at detecting hierarchical patterns in data, which is very useful for text analysis tasks like metaphor recognition, they are used widely. In this case, for analyzing a text, CNNs use multiple layers of convolutional operations to extract features from word embeddings—dense vector representations of words that contain their semantic nuances.

*The process includes phases and procedures that are given as follows:*

1. **Convolutional Layers**: These different layers, which utilize filters on the input data, are responsible for emphasizing the relationships that identify particular features in the text, such as syntax hints or semantics hints that identify metaphorical phrases.

2. **Pooling Layers**: Pooling layers, which are used after convolution, decrease the dimensionality of the data. This results in the aggregation of identified features into a format that is simpler to manage while still maintaining the vital data.

3. **Fully Connected Layers**: In order to assess if phrases are metaphorical based on learned patterns, fully connected layers at the final point of the network combine the features that were collected by previous layers.

**Bidirectional Encoder Representations from Transformers**

A novel technique for managing the collection of textual data is symbolized by the practical contextual analysis features of BERT. It is probable that this model analyzes words in contexts instead of independently, which helps its high level of performance in recognizing metaphors.

The self-attention techniques that BERT contains allow it to actively consider the links that exist among all of the words that compose a sentence. About the line "Time flew by," for example, BERT can understand the word "flew" in a metaphorical sense, reading it as signifying the progression of time instead of the fundamental process of travel by looking at the terms "time" and "by" combined. Each of the many layers of Transformer blocks composing BERT is intended for laying and performing complex textual expressions. BERT comprises of Transformer blocks. The metaphorical expressions like "drowning in sorrow," where traditional models may overlook the non-literal interpretation, are identified and encoded with the help of the transformer blocks. To keep a sense of word order that is significant for comprehending sentences where the metaphorical meaning is dependent on sentence structure, like "breaking the ice, BERT is enabled by the Positional Encoding component. Positional encodings assist BERT in determining the value of each word's place in transmitting metaphorical meanings. We build on the pre-trained model by adding layers that are exclusively designed for detecting metaphors. This fine-tuning phase modifies the weights of BERT using a dataset annotated with metaphorical terms, helping the model to predict metaphors in a better way based on previously acquired patterns.

**Linguistic Algorithms**

In addition to these ML models, for dissecting and analyzing the text for syntactic and semantic patterns, linguistic algorithms are used:

**Dependency Parsing**

DP is an important NLP technique for analyzing sentences' grammatical structure by creating links between "head" words and words that change those heads since the parsing method helps clarify the roles and relationships of words tagged in sentences, which frequently contain metaphorical meanings.

DP converts phrases into a tree structure, with nodes representing words and edges defining the relationships between them. For example, "seize" is the root, and "day" is dependent on the metaphorical statement "seize the day," which is usually classified as a direct object relationship. The parser recognizes numerous grammatical relations, including subject, object, and modifiers.

Understanding these relationships is very important when identifying non-literal word use. For instance, in a context, if "fire" (a noun) is used, it does not technically entail igniting something; it may imply a metaphorical meaning, such as "firing enthusiasm."

Dependency parsers frequently use a score function, $s(T)$, for each possible tree '$T$' deriving from a sentence, EQU (1):

$$s(T) = \sum_{(h,m)\in T} w_{h,m} \tag{1}$$

where the head and dependent words are denoted by $'h'$ and $'m'$, respectively, and the weight assigned to the dependency relationship between them is $w_{h,m}$. Determining the most probable tree structure is to maximize this score. ML techniques may be used by advanced dependency parsers to generate the weights $w_{h,m}$ based on features extracted from the words, their parts of speech, and other local context signals. This method enables the parser to learn the dependency configurations that are most likely to reflect correct syntactic and semantic links, which include metaphors, adaptively.

### Semantic Role Labeling (SRL)

SRL is an important NLP approach. SRL determines the role that words and phrases play in relation to the main verbs (predicates) in sentences. Since SRL gives a comprehensive insight into how the various aspects within a sentence contribute to its overall meaning, which is critical for identifying non-literal language, this method is very good at detecting metaphors.

SRL is the process of mapping sentence components to their semantic roles, like Agent, Instrument, Patient, and so on. These responsibilities serve to clarify the meaning of the structures of sentences. For instance, SRL identifies "the committee" as the Agent, "shot down" as the Action, and "the proposal" as the Patient in the metaphor "The committee shot down the proposal," emphasizing the metaphorical use of "shoot" in a decision-making setting. SRL often uses Conditional Random Fields (CRFs) –deep learning models, or LSTM (Long Short-Term Memory) networks – the neural network topologies, to capture the contextual dependencies required for adequate role labeling.

SRL is formulated in a computational model by assigning a probability to each conceivable label configuration given in a sentence, EQU (2):

$$P(\mathbf{r} \mid \mathbf{w}) = \frac{\exp\left(\sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_k f_k(r_i, w_i, w_{i-1}, \ldots, w_{i+1})\right)}{\sum_{\mathbf{r}'} \exp\left(\sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_k f_k(r_i', w_i, w_{i-1}, \ldots, w_{i+1})\right)} \tag{2}$$

In this example, the words in the sentence are represented by $'\mathbf{w}'$, the roles assigned to each word are represented by *'**r**'*, the feature functions that connect the roles and words are $'f_k'$, and the weights learned during training are '$\lambda_k'$. Words and syntactic categories are not only features of SRL but also dependency parse outputs, location information, and interaction features that record the interactions between sentence components.

## METHODOLOGY

### Data Collection

The carefully selected texts noted for their rich metaphorical content, spanning different areas and modes of communication, form the corpus for this study. It contains classic and modern literary materials (novels, poems, plays) from academic pieces from the humanities and social sciences, distinct cultural backgrounds, social media posts that reflect current linguistic trends, and articles and blogs from major newspapers and magazines. These sources were chosen deliberately to give a wide range of formal to informal language and structured to unstructured communication, which is required for researching metaphorical translations.

Data extraction techniques are adapted to the format and source of the content. Python scripts are used to scrape and filter relevant textual content from web-based sources using packages such as Beautiful Soup and Scrapy. High-resolution scanners and OCR software such as Adobe Acrobat or ABBYY FineReader are used to digitise printed text and convert it into editable formats. APIs such as Twitter API and Facebook Graph API are used to collect postings systematically based on specific keywords, hashtags, or user accounts, ensuring that the collected data are relevant to the study's objectives. Table 1 displays the data collection sources.

**Table 1: Data collection sources and methods**

| Source Category | Description | Data Extraction Method |
|---|---|---|
| Literary Texts | Classic and contemporary novels, poetry, and plays from diverse traditions were chosen for their metaphorical language. | Automated scripts for digital formats; OCR for physical books. |
| Academic Articles | Research papers and reviews from humanities and social sciences. | Digital libraries and academic databases with text mining tools. |
| Media and Blogs | Articles from major newspapers, magazines, and popular blogs discuss cultural and social issues. | Web scraping using Beautiful Soup or Scrapy for websites; manual selection for non-digital archives. |
| Social Media Posts | Tweets and Facebook posts reflect current linguistic usage. | APIs (Twitter API, Facebook Graph API) to collect posts based on specific keywords, hashtags, or accounts. |

## Data Pre-processing

During the data processing step of corpus development, we work hard to ensure that the text is clean and prepared for advanced linguistic analysis. Initially, custom Python scripts employed text normalization. These scripts eliminate all HTML tags and extraneous formatting, correct common typographical errors, and standardize textual forms, such as converting all characters to lowercase and formatting dates and numbers consistently. After normalization, the text is tokenized and segmented using Python's Natural Language Toolkit (NLTK). This process involves dividing the cleaned text into individual words and sentences. For instance, phrases like "He spilled the beans during the meeting" are divided into discrete tokens: 'He', 'Spilled', 'the', 'beans', 'during', 'the', 'meeting', and punctuation marks.

Linguistic pre-processing is the final step of data processing. In this step, various NLP approaches are applied to tokenized text to add layers of linguistic metadata. Using the spaCy library, each token is tagged with the appropriate part-of-speech tag, allowing verbs, nouns, adjectives, and other words to be recognized. The text is then evaluated to identify syntactic dependencies and parse structures that help readers understand grammatical relationships inside phrases. Recognition and classification of correct nouns and other significant things can also be achieved by applying recognition of names. This provides primary semantic data to the preexisting corpus. The procedures that are performed at each phase of the data processing process are presented in Table 2, which demonstrates how each stage processes a hypothetical phrase.

**Table 2: Data processing phase that deals with an example sentence**

| Processing Step | Example Sentence | Action Taken |
|---|---|---|
| Text Normalization | "Mr. Smith met Ms. Brown in New York on Jan. 1st!" | Remove extraneous elements, correct typos, convert to lowercase, and standardize dates and places: "Mr. Smith met ms brown in New York on January 1." |
| Tokenization and Segmentation | "Mr Smith met ms brown in New York on January 1." | Break down into words and punctuation: ['Mr', 'smith', 'met', 'ms', 'brown', 'in', 'new', 'york', 'on', 'January', '1', '.'] |
| Linguistic Pre-processing | ['Mr', 'smith', 'met', 'ms', 'brown', 'in', 'new', 'York', 'on', 'January', '1', '.'] | Apply POS tagging, parsing, entity recognition:<br> - POS: ['NOUN', 'NOUN', 'VERB', 'NOUN', 'NOUN', 'ADP', 'PROPN', 'PROPN', 'ADP', 'PROPN', 'NUM', 'PUNCT']<br> - Named Entities: ['Mr. Smith', 'Ms. Brown', 'New York', 'January 1']<br> |

| | | |
|---|---|---|
| | | - Dependency Parse: ['compound', 'nsubj', 'ROOT', 'compound', 'dobj', 'prep', 'compound', 'pobj', 'prep', 'pobj', 'nummod', 'punct'] |

**Annotation Process**

When referring to research on metaphorical translation, the procedure for annotation is essential in order to be sure that the corpus is of excellent quality and can be appropriately utilized. Annotations are given with rules that are meant to help them constantly label and recognize metaphorical phrases that can be found in the text. In addition to providing examples, these criteria provide comprehensive descriptions of what defines a metaphor in the context of language that is literal.

- **Example of Metaphor**: This is precisely envisioned by the expression, "He drowned in an ocean of sadness". The employ of metaphor communicates the concept that one is experiencing a visceral feeling of sadness as if someone were floating in the ocean.
- **Example of Literal Language**: "He splashed in the ocean". The word 'swam' is applied in the literal sense here. Further, the standards define contextual factors that may be symptomatic of metaphorical language. These hints include unusual verb-noun combinations or adjectives that do not commonly define the nouns they adapt.

In Table 3, the rules for labeling metaphorical versus literal expressions have been highlighted, in addition to examples and definitions of how to recognize hints.

**Table 3: Annotation process using a sample sentence**

| Sentence Example | Classification | Guideline |
|---|---|---|
| "He drowned in an ocean of sadness." | Metaphor | Identify the metaphorical use of verbs and nouns indicating emotional states, not physical actions. |
| "He swam in the ocean." | Literal | Recognize literal actions directly describing physical activities. |
| "She planted the seeds of kindness." | Metaphor | Look for verbs and nouns used figuratively to describe abstract concepts. |
| "He runs a tight ship." | Metaphor | Note expressions where common phrases indicate management style, not literal actions. |
| "The scientist explored the cave." | Literal | Apply guidelines literally when actions and objects correspond directly to reality. |

In order to investigate metaphorical language precisely and frequently, the text describes a course of study for annotators, including experienced linguists and community workers. Metaphor recognition seminars, practice annotations, and review sessions are all included in this procedure. To help with group projects and systematic tests, annotators label metaphorical terms via tools like BRAT or intended software. Regular adjudication sessions and verification for inter-annotator agreements provide permanent annotations and manage conflicts. Given this recurrent annotation approach, the data set is versatile for NLP and associated domains.

**Corpus Design and Development**

Significant preparation and creation went into creating the corpus of texts used for investigating metaphorical translations, providing that it includes an extensive number of languages and includes every required metadata. Metadata such as writing, origin proof of identity, and publication date (as applicable) have been integrated into the corpus alongside detailed linguistic data like grammatical parsing data, semantic assignments, and part-of-speech tags. Linguistic and NLP research focuses significantly on the ability to perform advanced searches and detailed analyses rendered feasible by this data pooling.

Different methods are used to ensure that the corpus' metaphorical expressions are accurate and unique. Books from several types, times, and styles have been included in the thoughtfully selected selection guidelines. Literary works are selected from existing sources and historical periods to represent the evolution of metaphorical usage over time. In the same way, data from newspapers and magazines span political and artistic languages, indicating a broad spectrum of metaphor usage in many situations. Examples of selected metaphorical terms include phrases such as "drowning in debt" from financial articles and "a flood of emotions" from romantic literature, which demonstrate a wide range of metaphorical conceptualizations. Table 4 shows example sentences that demonstrate the precise structure, diversity, and multilingual integration of the corpus design and development.

**Table 4: Corpus creation and development with a sample sentence**

| Aspect | Example Sentence | Description |
|---|---|---|
| Linguistic and Metadata Elements | "He was lost in a sea of thoughts." | Includes part-of-speech tags (e.g., PRP, VBD, IN, DT, NN, IN, NNS), syntactic parsing, and semantic roles. |
| Diversity of Metaphorical Expressions | "The project hit a roadblock." | They were selected from numerous sources, like technical reports, showing metaphorical use in different contexts. |
| | "She has a heart of stone." | They were chosen from literary texts, illustrating emotional metaphors. |
| | "The company weathered the storm." | From business articles demonstrating financial and economic metaphors. |
| Representativeness | "He was on cloud nine after the news." | Reflects common metaphorical expressions used in everyday language, ensuring the corpus covers widespread usage. |
| | "The athlete broke new ground." | It represents metaphors from sports journalism and captures domain-specific language. |
| Integration of Multilingual Data | "Romper el hielo" (Spanish for "breaking the ice") | Shows parallel metaphorical expressions in different languages, highlighting cross-linguistic similarities and differences. |
| | "Un océan de tristesse" (French for "a sea of sadness") | Includes multilingual data to examine how metaphors translate and adapt across cultures and languages. |

## Metaphor Identification and Translation Using NLP Tools and Techniques

To recognize and interpret metaphorical language in massive text corpora, NLP techniques for metaphor detection and translation use complex ML models and linguistic algorithms. CNNs and BERT are the two most common models employed. CNNs find hierarchical patterns using convolutional, pooling, and fully connected layers, enabling them to recognize metaphorical expressions effectively. BERT excels at contextual analysis because it evaluates relationships between all words in a phrase using self-attention mechanisms, which allows it to discover nuanced metaphorical meanings. For example, BERT comprehends "Time flew by" by examining the context of "flew" and "time." Its numerous positional encoding and Transformer blocks enhance its ability to interpret metaphorical meanings in organized sentences. Aside from these models, linguistic algorithms such as DP and SRL are utilized. DP converts the phrases into tree structures, which helps to clarify grammatical relationships and detect metaphors. Table 5 demonstrates how each NLP tool handles a single line. "He drowned in a sea of sadness," gradually. To help readers find metaphors and clarify grammatical interactions, DP transforms words into tree structures. Table 5 demonstrates how each NLP tool addresses an individual line. "He died in the ocean of sadness," he said them.

**Table 5: NLP process in metaphor identification and translation**

| Process Step | Example Sentence Handling |
|---|---|
| CNNs | Sentence: "He drowned in an ocean of sadness." |
| Convolutional Layers | Action: Identify "drowned" and " ocean of sadness" as key features. |

| | |
|---|---|
| Pooling Layers | Action: Consolidate features related to "drowned" and " ocean of sadness". |
| Fully Connected Layers | Action: Classify "drowned" and " ocean of sadness " as metaphors. |
| BERT | Action: Understand "drowned" metaphorically in the context of " sadness ". |
| Self-Attention Mechanisms | Action: Assess how "drowned" relates to " ocean" and " sadness ". |
| Transformer Blocks | Action: Encode "drowned" as a metaphor when associated with "sadness ". |
| Positional Encoding | Action: Preserve the phrase structure of "an ocean of sadness ". |
| Dependency Parsing | Action: Identify "drowned" as the root and " ocean of sadness " as dependents. |
| Grammatical Relations | Action: Recognize "he" as the subject and " ocean of sadness " as a modifier of "drowned". |
| SRL | Action: Label "he" as Agent, "drowned" as Action, and " ocean of sadness " as Experiencer. |
| Role Assignment | Action: Confirm "he" as Agent and " ocean of sadness " as metaphorical context for "drowned". |

## ANALYSIS

### Corpus Statistics

With its large number of numerous and properly classified text sources, the metaphorical translation corpus provides an in-depth analysis of the various forms and sources of metaphor. In order to understand the scope and depth of the data and its possible uses, it is essential to search Table 6 for a statistical analysis of the corpus of information.

**Table 6: Statistics for the corpus generated**

| Statistical Category | Description | Value |
|---|---|---|
| Total Metaphors | Total number of metaphor entries | 11,863 |
| Source Languages | Languages from which texts are drawn | English |
| Conventional | Commonly used metaphors | 5,943 |
| Extended | More elaborated metaphors | 3,517 |
| Novel | New and creative metaphors | 2,403 |
| Literary Sources | Metaphors from novels, poetry | 3,250 |
| Academic Articles | Metaphors from scholarly publications | 2,150 |
| Media Articles | Metaphors from newspapers, magazines | 4,213 |
| Online Content | Metaphors from blogs, forums | 2,250 |

The corpus is a significant dataset for research, with 11,863 metaphor values. The text is presented in English only, thus rendering it suitable for analyzing English metaphors and additional language and cultural features. Of the 5,943 metaphors in the corpus, 3,517 have additional metaphors, 2,403 are distinct, and 3,250 are written sources. A complete understanding of the use of metaphor is made feasible by this broad collection of writings in the English language. To provide an accurate visualization of metaphorical language, the corpus of work has been collected from several text genres of literature, such as literary works and poems. Academic articles include 2,150 entries that reflect scholarly applications of metaphors to describe complicated ideas. The majority of the entries, i.e., 4,213, come from media articles such as newspapers and magazines, demonstrating how metaphors are employed to express ideas to the public. Online content, which includes blogs and forums, has 2,250 entries that capture modern and informal applications of metaphors in digital communication.

### Corpus Analysis

In the research of metaphor translation by utilizing the generated corpus, three basic ML models are used, each adapted to a different form of metaphor due to their distinct capabilities and strengths. An overview of each mode is given here.

**1. Support Vector Machine (SVM): Model A:** SVMs are particularly successful for traditional metaphors with well-defined patterns and less ambiguity. SVMs function by determining the best hyperplane to segregate data points in a high-dimensional space, making them both strong and effective in binary classification tasks. In metaphor translation, SVMs easily handle classical metaphors' structured, rule-based nature, ensuring great accuracy and reliability.

**2. Gradient Boosting Machines (GBM): Model B:** GBM is a robust ensemble learning algorithm that is excellent for extended metaphors requiring complicated, layered context awareness. GBMs construct an additive model in stages, reducing arbitrary differentiable loss functions. In metaphor translation, GBMs outdo manipulating the complexities and subtleties of extended metaphors, enhancing prediction accuracy based on nuanced language features that simpler models may miss.

**3. CNNs: Model C:** CNNs are employed for novel metaphors that are particularly difficult due to their inventive and context-dependent character. CNNs are DL algorithms capable of mining hierarchical features from big datasets. They are skilled at analyzing sequential data and finding patterns with significant spatial hierarchies, transforming them into ideal ones for decoding novel metaphors' complex and nuanced hints. This capacity enables CNNs to capture and translate the intended metaphorical meaning, though it differs significantly from ordinary usage.

**Table 7: Scores for each translation task demonstrate how each model performed across all sorts of metaphors**

| Translation Task | Model Used | Accuracy Before Corpus | Accuracy After Corpus | Improvement |
|---|---|---|---|---|
| Conventional Metaphors | Model A | 72.3% | 85.7% | +13.4% |
| | Model B | 72.3% | 84.1% | +11.8% |
| | Model C | 72.3% | 83.5% | +11.2% |
| Extended Metaphors | Model A | 65.0% | 77.8% | +12.8% |
| | Model B | 65.0% | 79.2% | +14.2% |
| | Model C | 65.0% | 78.3% | +13.3% |
| Novel Metaphors | Model A | 58.1% | 69.7% | +11.6% |
| | Model B | 58.1% | 71.2% | +13.1% |
| | Model C | 58.1% | 74.5% | +16.4% |

Table 7 displays a study of individual model scores across several translation tasks, which suggests that using the constructed corpus improves accuracy. Model A (SVM) has accuracy for conventional metaphors by 13.4%, from 72.3% to 85.7%. Model B (GBM) also showed a significant increase in accuracy, from 72.3% to 84.1%, or an 11.8% improvement. Similarly, Model C (CNN) increased its accuracy from 72.3% to 83.5%, or an 11.2% increase. When extended metaphors were examined, Model A's accuracy elevated from 65.0% to 77.8%, which is a 12.8% enhancement. Model B did slightly better, increasing its accuracy from 65.0% to 79.2%, a 14.2% improvement. Model C also improved significantly, with accuracy rising from 65.0% to 78.3% or a 13.3% gain. Novel metaphors are inherently more complex because of their innovative nature. Model A's accuracy rose from 58.1% to 69.7%, representing an 11.6% gain. Model B's performance improved even further, with accuracy improving from 58.1% to 71.2%, a 13.1% gain. Especially with accuracy increasing from 58.1% to 74.5%, Model C exhibited the most remarkable improvement for novel metaphors, which is a significant 16.4% increase.

## CONCLUSION AND FUTURE WORK

In the present investigation, researchers demonstrate how to use state-of-the-art Natural Language Processing (NLP) methods to develop a corpus specifically for English Metaphorical Translation

(EMT). Researchers significantly improve the accuracy of metaphor identification and translation through the use of CNNs, BERT, Dependency Parsing, and SRL. The data's quality and use have been enhanced by the numerous and accurate corpus, which is collected from multiple sources. All of the NLP models demonstrated improvement, which indicates that this approach performs. This research promotes NLP tools, which provide the possibility of translations that are more detailed and more adaptable to linguistic standards.

Further into the future, the corpus will be developed, and NLP models will be improved in order to enhance its metaphor analysis and translation capabilities.

## REFERENCES

1. Sayidov, S., & Mirzaeva, D. (2023, December). Unveiling the power of metaphor. In *Fergana State University Conference* (pp. 134-134).
2. Li, Y., Guerin, F., & Lin, C. (2024). Finding Challenging Metaphors that Confuse Pretrained Language Models. *arXiv preprint arXiv:2401.16012*.
3. Hong, W., & Rossi, C. (2021). The cognitive turn in metaphor translation studies: A critical overview. *Journal of Translation Studies*, *5*(2), 83-115.
4. Lakoff, G., & Johnson, M. (2020). Conceptual metaphor in everyday language. In *Shaping entrepreneurship research* (pp. 475-504). Routledge.
5. Csábi, S. (2023). Metaphor and stylistics. In *The Routledge Handbook of stylistics* (pp. 217-234). Routledge.
6. Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. J. Mach. Learn. Res., 21(140):1–67.
7. 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. arXiv preprint arXiv:1907.11692.
8. YukunZhu, RyanKiros, Richard S. Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In Proceedings of the IEEE International Conference on Computer Vision. IEEE Computer Society, 19–27. DOI: https: //doi.org/10.1109/ICCV.2015.11
9. Xinyi Wang, Sebastian Ruder, and Graham Neubig. 2022. Expanding trained models to thousands more languages via lexicon-based adaptation. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 863–877. DOI: https://doi.org/10.18653/v1/2022.acl-long.61
10. Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé et al. 2022. BLOOM: A 176B-parameter open-access multilingual language model. arXiv Preprint arXiv:2211.05100 (2022).
11. Bahdanau, D., Cho, K. and Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations*
12. Brun, C., Ehrmann, M. and Jacquet, G. (2007). XRCE-M: A hybrid system for named entity metonymy resolution. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*. Association for Computational Linguistics, pp. 488–491.10.3115/1621474.1621583
13. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. and Kuksa, P.P. (2011). Natur al language processing (almost) from scratch. Journal of Machine Learning Research 12, 2493–2537.

14. Farkas, R., Simon, E., Szarvas, G. and Varga, D. (2007). Gyder: Maxent metonymy resolution. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pp. 161–164.10.3115/1621474.1621507
15. Fu, T.-J., Li, P.-H. and Ma, W.-Y. (2019). GraphRel: Modeling text as relational graphs for joint entity and relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 1409–1418.
16. Kanval, N., Ihsan, H., Irum, S., & Ambreen, I. (2024). Human Capital Formation, Foreign Direct Investment Inflows, and Economic Growth: A Way Forward to Achieve Sustainable Development. Journal of Management Practices, Humanities and Social Sciences, 8(3), 48-61.
17. Rashid, A., Jehan, Z., & Kanval, N. (2023). External Shocks, Stock Market Volatility, and Macroeconomic Performance: An Empirical Evidence from Pakistan. Journal of Economic Cooperation & Development, 44(2), 1-26.
18. Jam, F. A., Rauf, A. S., Husnain, I., Bilal, H. Z., Yasir, A., & Mashood, M. (2014). Identify factors affecting the management of political behavior among bank staff. *African Journal of Business Management*, *5*(23), 9896-9904.
19. Jam, F. A., Singh, S. K. G., Ng, B., & Aziz, N. (2018). The interactive effect of uncertainty avoidance cultural values and leadership styles on open service innovation: A look at malaysian healthcare sector. *International Journal of Business and Administrative Studies*, *4*(5), 208-223.