AUTOMATIC IDENTIFICATION OF FILIPINO METAPHORS USING SENTENCE TRANSFORMERS AND AN AUTOMATIC DEFINITION SELECTOR

LORENZO MIGUEL L. OCAMPO

SUBMITTED TO THE FACULTY OF THE UNDERGRADUATE SCHOOL
UNIVERSITY OF THE PHILIPPINES LOS BAÑOS
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE
DEGREE OF

BACHELOR OF SCIENCE
(COMPUTER SCIENCE)

JUNE 2023

BIOGRAPHICAL SKETCH

Lorenzo Miguel L. Ocampo is currently pursuing a Bachelor of Science degree in Computer Science at the Institute of Computer Science, University of the Philippines Los Baños. He has a keen interest in the fields of Machine Learning, Financial Technology, and Data Science, continuously seeking opportunities to enhance his knowledge and skills in these and other areas as well.

LORENZO MIGUEL L. OCAMPO

ACKNOWLEDGEMENT

First and foremost, I would like to thank my immediate family for always encouraging me and believing in my abilities. Their continuous support and understanding have been extremely important in every stage of my academic journey.

I would like to express my deepest thanks to my adviser, Dr. Maria Art Antonette D. Clariño, for their guidance and support throughout this research journey. Their expertise, helpful feedback, and insightful discussions have been extremely important in helping to create this work.

I extend my appreciation to my friends for their constant support, encouragement, and understanding during the challenging phases of this research. They made the challenging times of this research bearable because of their company. I would also like to thank the people I have worked with during my university years. They are the people who helped shape who I am today.

Lastly, I would like to sincerely thank all the participants who generously volunteered their time for this study. Their active participation played a crucial role in the success and credibility of this research. I am deeply grateful for their contributions.

I am sincerely thankful to all those who have played a part, no matter how small, in this research. Their contributions, support, and encouragement have been invaluable, and this work would not have been possible without them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The rise of social media has led to the widespread use of online posts as valuable data for machine learning studies. However, these studies often encounter the challenge of dealing with metaphors embedded in the text, which can introduce inconsistencies in data processing since metaphors are known to change meaning of text. This challenge is even more prominent in high context languages such as Filipino. In response to this issue, multiple novelties were introduced in this study. These novelties are the Filipino metaphor corpus, the Filipino metaphor identifier created that utilizes an automatic definition selector, and finally a modified metaphor counting method in the evaluation.

INTRODUCTION

Metaphors are more than just creative figures of speech; they are intrinsic to language and culture, providing vivid imagery that helps us grasp abstract concepts such as emotions, ideas, and interpersonal relationships (Lakoff & Johnson, 1980). With the rise of social media, researchers are now exploring the valuable information contained in social media post in fields like natural language processing and machine learning (Catayna & Clariño, 2022). However, the presence of metaphors in these digital texts introduces unexpected challenges when interpreting their intended messages (Titong, 2014). Metaphors add layers of meaning and subjective nuances that can complicate the deciphering process, often altering the overall message conveyed (del Pilar Salas-Zárate et al., 2020).

Understanding metaphors in online content necessitates acknowledging the broader implications of language and context on meaning. Challenges such as context sensitivity, parser issues, and the inclusion of metaphoric usages in dictionaries highlight the complexity of interpreting metaphors in digital communication (Krishnakumaran & Zhu, 2007). This nuanced understanding becomes particularly significant in languages like Filipino, characterized as high-context languages where interpretation heavily relies on the usage in discourse (Wise, 2013).

Filipino language and culture encompass various aspects that make use of context-heavy or figurative language in communication (Mesiona, Agrabio, Dianasas, Ambrad, & Diones, 2022). Filipino songs, for example, often incorporate lyrics rich in metaphorical expressions, evoking deep emotions and imagery (Mesiona et al., 2022). Additionally, the popular concept of ”hugot lines” permeates the Filipino language, where individuals express their feelings by relating phrases or sentences to their personal experiences (ARIAS & SENCIL, 2017). Furthermore, the Filipino language exhibits different types of metaphors, such as metaphorical idioms and poetic metaphors, with the latter being particularly challenging to identify in discourse (GAITAN-BACOLOD, 2010).

By recognizing the influence of context-heavy and figurative language in Filipino communication, it becomes evident that analyzing common Filipino texts presents considerably more challenges. The presence of figurative language has the potential to significantly alter the intended meanings of these texts, necessitating an in-depth examination of figurative expressions in Filipino social media discourse to accurately interpret their content.

Objectives

Overall, the study implemented an algorithm that can detect Filipino metaphors. The specific objectives of the study were the following:

1. Build a Corpus via Twitter API containing Filipino sentences;
2. Develop a Filipino metaphor identification algorithm that utilizes both metaphor identification procedure and selectional preference violation; and
3. Measure the performance of the detection algorithm using precision, recall and F1 scores.

Significance of the study

The detection of metaphors in the Filipino language holds significant importance not only for natural language processing (NLP) tasks but also for future research in related fields (Lapitan, Batista-Navarro, & Albacea, 2016) (Fernandez & Adlaon, 2022). This study particularly benefits fields that rely on interpreting textual data, as identifying metaphors within text determines whether additional processing is needed to understand the true meaning of words. One such field is machine translation, where the identification of text segments containing metaphorical expressions that cannot be directly translated due to their distinct meanings is crucial (Mao, Lin, & Guerin, 2018). Accurately detecting metaphors in a text also contributes to sentiment analysis, allowing analysts to interpret texts based on their literal or metaphorical meanings (J. Liu, O'Hara, Rubin, Draelos, & Rudin, 2020). These findings not only advance NLP applications but also open avenues for further research in metaphor detection and its impact on language comprehension.

Scope and limitations

This study specifically focuses on detecting metaphors written in the Filipino language. While sentences may contain English words, only the metaphors written in Filipino language will be detected and counted. To achieve this, a dataset comprised of numerous online sources has been utilized. This dataset includes various text sources such as Twitter-based news articles and ”hugot” lines, which are popular expressions in Filipino culture where individuals relate phrases or sentences to their personal experiences.

REVIEW OF LITERATURE

Metaphors

Metaphors serve as cognitive mechanisms that enhance culture and language, enabling effective communication and expression. They provide a means to convey intricate emotions, ideas, and experiences that may be challenging to articulate using literal language. Metaphors play a vital role in shaping our cognitive processes and influencing our perception of the world, as they establish connections, draw comparisons, and offer explanations for abstract concepts (Lakoff & Johnson, 1980).

In the Philippines, metaphors can be divided into two categories: metaphorical idioms and poetic metaphors. Metaphorical idioms are metaphors that possess a single, fixed meaning and interpretation, and are typically used in a non literal sense. For instance, the phrase ”Ang kapal ng mukha” conveys the idea of arrogance when used metaphorically, regardless of the context in which it is used. On the other hand, poetic metaphors do not have fixed forms and its interpretation can differ based on who is reading it and how it is used (GAITAN-BACOLOD, 2010).

Metaphorical idioms in the Filipino language offer a clear and unambiguous interpretation, providing straightforward meanings that can be easily identified by checking against a predefined database. In contrast, poetic metaphors present a greater challenge due to their nuanced and subjective nature. The identification of poetic metaphors requires in-depth analysis of the surrounding context within the text to grasp how words or phrases are being used. This study focuses on the detection and identification of poetic metaphors.

Challenges in Metaphor identification

Due to the nature of metaphors, the identification of metaphors remains a difficult task. Some challenges in metaphor identification include context sensitivity, parser issues, and metaphorical usages in dictionaries (Krishnakumaran & Zhu, 2007).

Context Sensitivity

Many Filipino words have meanings that can vary greatly depending on the context in which they are used, leading to significant shifts in interpretation. For instance, the following sentence can be understood both literally and metaphorically, resulting in considerable differences in meaning.

Handa na ang mga pasabog para sa darating na pasko

If interpreted literally, the sentence can mean that there are explosives ready to be used for the incoming Christmas. Meanwhile, the metaphorical meaning is that the gimmicks are ready for the incoming Christmas.

Parser Issues

It remains a challenge to identify metaphors because the accuracy of the identification algorithm is highly dependent on the accuracy of the parser, models, and other technologies used in the identification process. Although certain technologies have proven effective for metaphor identification, additional factors need to be considered when selecting an appropriate technology, particularly when addressing specific tasks that may differ from established approaches and previous studies.

Metaphoric Usages in dictionaries

Some metaphoric uses of certain words have been used so often in everyday language that they have been officially recognized and documented in dictionaries.

Siya ay isang ahas

The metaphoric usage of the word ”ahas” is already included in the dictionary. These metaphors are also known as dead metaphors because their usage has become so commonplace that they are already part of official definitions.

Metaphor Identification

The term ”semantic gap” refers to the difference between different data representations of an object(Hein, 2010). The computation of the semantic gap is essential for various natural language processing tasks. Determining interpretations and relationships between semantic information is crucial for tasks such as text understanding, sentiment analysis, machine translation, and information retrieval (Jurafsky & Martin, 2014). This is also true for the task of metaphor identification since it is used to determine the metaphoricity (Babieno, Takeshita, Radisavljevic, Rzepka, & Araki, 2022).

Researchers have employed various methods to tackle the task of metaphor identification using semantic gap on different linguistic features. They have proven to have made significant progress though the use of different innovative approaches. One study utilized visual features to determine the metaphoricity of phrases or sentences. Average image embeddings were extracted from obtained images, and the metaphoricity was measured by computing the cosine similarities of word pairs within the phrase, such as verb-noun or adjective-noun combinations. This method achieved a precision score of 50%, recall score of 95%, and F1 score of 66% (Shutova, Kiela, & Maillard, 2016).

Another approach utilized measuring the semantic gap to identify metaphoricity of certain words. One example of an approach in this manner is the selectional preference violation (SPV) method, which identifies preferences among patterns in logical reasoning to determine the most appropriate interpretation based on evidence and pattern preferences (Wilks, 1975). The study incorporated selectional preference violation as a feature in metaphor detection, calculating the probability of certain nouns occurring with specific verbs. The results showed slight improvement over the baseline model(Haagsma & Bjerva, 2016).

In this study, both the selectional preference violation and another semantic gap approach called metaphor identification procedure (MIP) were used to detect metaphors. MIP measures the contrast between the target word's definition and its contextual usage (Group, 2007). These methods were employed as features in the model to compute the metaphoricity of sentences, resulting in an average precision score of 76%, average recall score of 73.7%, and average F1 score of 75% (Zhang & Liu, 2022).

Previous implementations

Several studies proposed different models for the task of metaphor detection. This study used recurrent neural networks and word embeddings to predict noun-adjective phrases and whether they are literal or metaphorical in its usage in the Polish language. They compared the performance of both bidirectional GRU and LSTM architecture, which resulted in the LSTM having better results. The final result of this architecture was an F1 score of 0.73 for metaphorical words and 0.58 for the data in their cross validation schema (Mykowiecka, Wawer, & Marciniak, 2018).

However, neural network-based solutions have limitations in representing certain word contexts due to their word-for-word prediction nature (Choi et al., 2021). To address this, other studies adopted contextualization-based approaches, such as BERT (Bidirectional

Encoder Representations from Transformers) (Devlin, Chang, Lee, & Toutanova, 2018) and RoBERTa (A Robustly Optimized BERT Pretraining Approach). RoBERTa, known for its superior performance compared to BERT (Y. Liu et al., 2019), was used as the language model for metaphor detection. Similar to (Zhang & Liu, 2022), they also used MIP and SPV as features to be input into the transformer. The obtained results were a precision score of 53.4, a recall score of 74.1, and an F1 score of 62.0. These results fall short of the performance of RoBERTa-base (Choi et al., 2021).

Building on the previous research model called MelBert, this study introduced MIss RoBERTa WiLDe, which revised several layers to enhance metaphor detection. Changes included using sentence embeddings from SBERT instead of the CLS token to represent the sentence and employing the target word's definition in the MIP layer. MIss RoBERTa WiLDe outperformed MelBert in some datasets, while MelBert excelled in others (Babieno et al., 2022).

METHODOLOGY

The model architecture is inspired by the MelBERT (Choi et al., 2021) and MIss RoBERTa WIlde (Babieno et al., 2022) since the accuracy of their detection showed significant improvement compared to other models. While this work is inspired by these previous works, some changes were made to this research to fit the setting of the study more appropriately. Specifically, the target language is now in Filipino rather than English, a different method in determining the target word's definition, a Filipino metaphor dataset was created for training, the combination of two models will be merged, and a different method of metaphor counting was used. The language model that will be used is XLM-RoBERTA to accommodate the use of filipino language and the presence of minimal english words (Conneau et al., 2019). Additionally, a sentence embedder (Velasco et al., 2022) will be used to generate the embeddings for the features.

Data collection

The first novelty of this study is the creation of a new Filipino metaphor corpus. The data collection process involved gathering tweets from the social media platform Twitter using the Python library Snscrape. Specifically, the data focused on Filipino tweets sourced from two Twitter accounts, namely DZMMTeleRadyo and mgapinoyhugot. These twitter accounts were chosen based on the availability of data and presence of metaphors within their content. DZMMTeleRadyo was chosen as the main source of non-metaphorical data. The account mgapinoyhugot was chosen since Filipino ”hugot” lines are known to contain metaphorical language since it is ”the Filipino way of making metaphors”(ARIAS & SENCIL, 2017). Among many ”hugot” accounts in twitter, this account was chosen based on the large amount of data in the account. To ensure data integrity and avoid overlapping with publicly available information, tweets were collected within the time

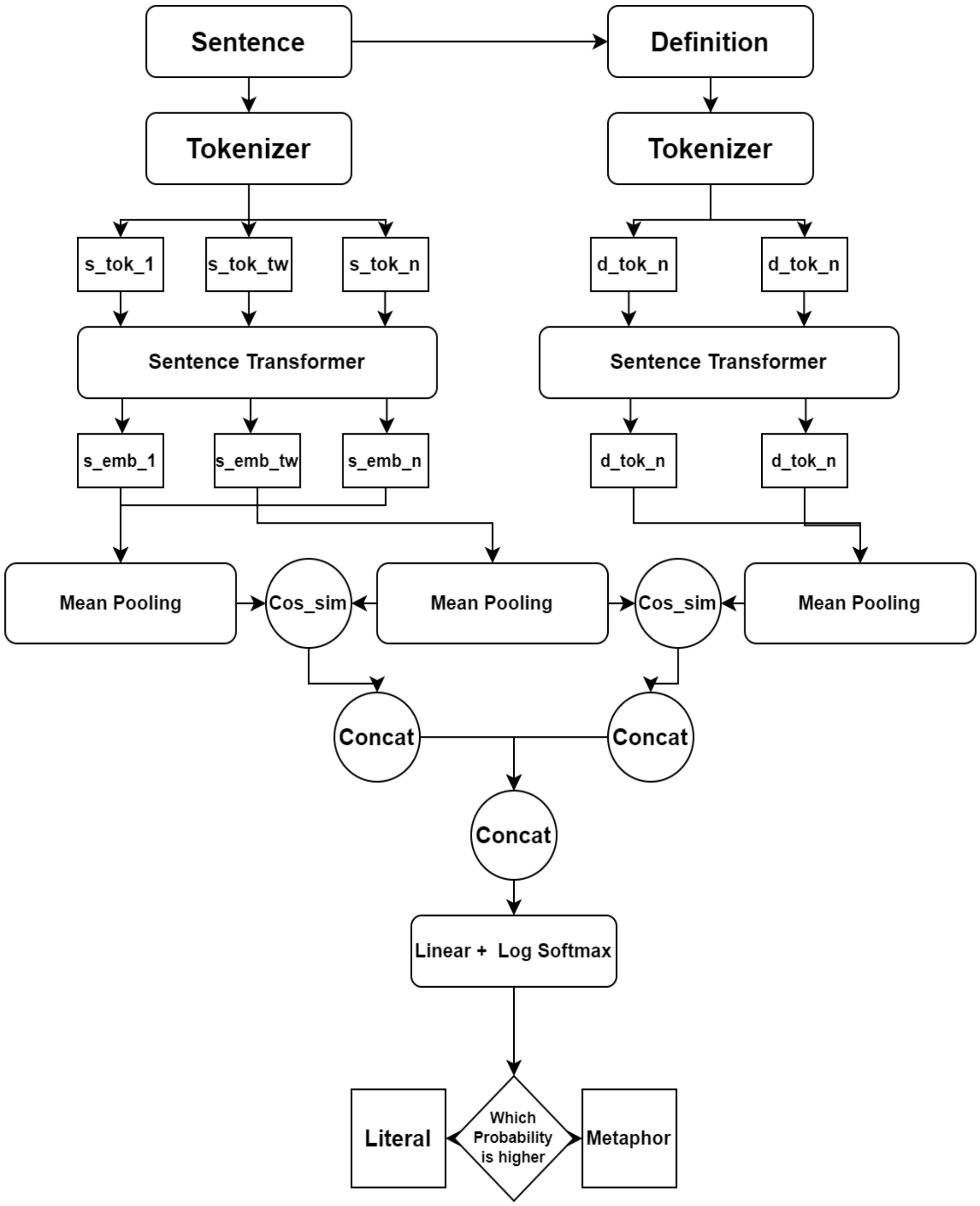

Figure 1. Model Architecture

frame of January 2021 to January 2023.

The collected tweets were stored and organized in a dataframe format using the pandas Python library. Subsequently, the data was exported as a tab separated values (tsv) file.

To ensure data quality and eliminate duplicates, a Python script was implemented to filter out redundant tweets. Approximately 5,000 tweets were identified and removed during this process, enhancing the uniqueness and consistency of the final dataset.

Data Pre-processing

During the data pre-processing phase, a series of steps were executed on each tweet using the Python programming language. These steps aimed to standardize the data and remove irrelevant information to ensure consistency and enhance the quality of subsequent analyses.

The initial step involved converting all tweets to lowercase, thereby achieving uniformity in the text data. This normalization process minimized potential discrepancies arising from variations in capitalization.

STEP	NEWS	HUGOT
Base Tweet	Tutok na sa #Sakto sa TeleRadyo kasama sina Amy Perez at @JohnsonManabat! FACEBOOK LIVE: https://t.co/g60oWZwhPP	Hinde Porket panget ka, Eh matatakot mo na ako! hahaha #JourneyToHappiness lol
Convert to lowercase	tutok na sa #sakto sa teleradyo kasama sina amy perez at @johnsonmanabat! facebook live: https://t.co/g60owzwhpp	hinde porket panget ka, eh matatakot mo na ako! hahaha #journeytohappiness lol
Remove Hashtags, Mentions, and Links	tutok na sa sa teleradyo kasama sina , amy perez at ! facebook live:	hinde porket panget ka, eh matatakot mo na ako! hahaha lol
Removing laughter text	tutok na sa sa teleradyo kasama sina , amy perez, at ! facebook live:	hinde porket panget ka, eh matatakot mo na ako!
Manual text normalization and spell checking	tutok na sa sa teleradyo kasama sina amy perez	hindi porke pangit ka, eh matatakot mo na ako!

Figure 2. Pre Processing Procedure

Following this, several components were eliminated from the tweets, including mentions, hashtags, and links, utilizing the regular expressions library in Python. Mentions referred to words or phrases starting with the ”@” symbol, which are used to tag or refer to other Twitter users. Hashtags, indicated by the ”#” symbol, are employed to categorize or identify specific topics. Links encompassed phrases starting with ”https:” or ”www” that directed users to external webpages. By removing these elements, the focus of the analysis remained on the core content of the tweets.

Additionally, expressions denoting laughter or similar emotions were also eliminated from the dataset. Examples of such expressions encompassed words like ”lol,” ”haha,” and their variations. The removal of these laughter-related terms aimed to minimize their potential influence on subsequent training and these were deemed impossible to be metaphors. The diagram illustrating the data pre-processing can be found in figure 2.

Data Annotation

The data annotation process was carried out manually by the researcher and four other native Filipino data annotators, three students majoring in Development Communication from UPLB and one student majoring in European Languages from UPD. During the annotation process, similar tweets were identified and removed to avoid duplication and maintain the uniqueness of the dataset.

To enhance the quality and consistency of the annotated data, manual text normalization and spell checking was performed. Text normalization aimed to standardize the language used in the dataset, while spell checking ensured the accuracy of word spellings. The online dictionary used served as a reference for this process, enabling the correction of any spelling errors or inconsistencies in the text.

The annotated dataset was annotated according to a modified Metaphor Identification Procedure VU (MIP VU). This methodology is a recognized approach for identifying metaphors (Steen et al., 2010). Additionally, it was modified to accommodate the Common

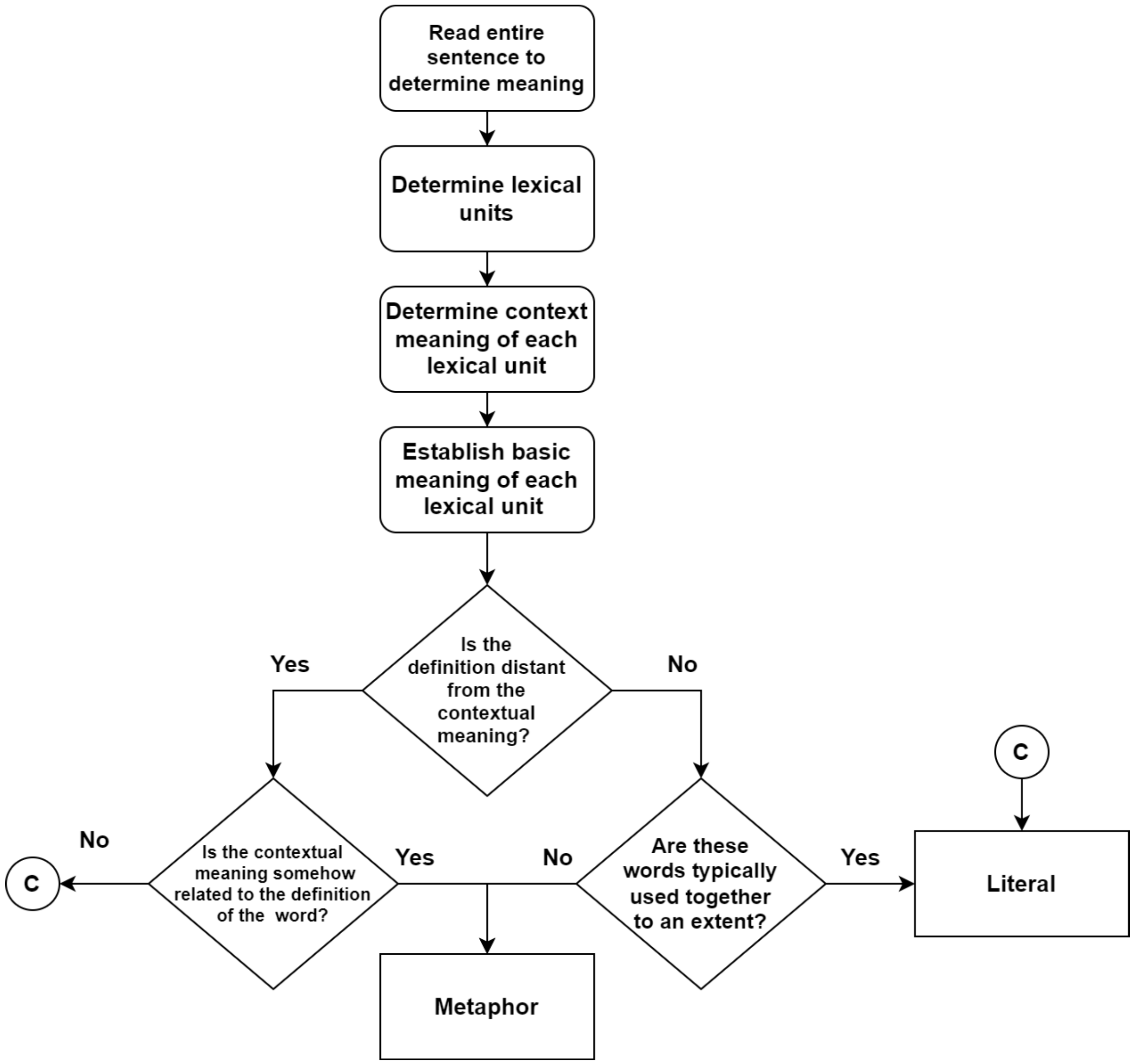

Figure 3. Modified MIPVU Procedure

Association Acquisition (CAA) procedures proposed by (Panicheva, Mamaev, & Litvinova, 2023). The flowchart illustrating the annotation process can be found in figure 3.

Following the annotation process, the dataset was randomized, while keeping the same sentences together, and divided into 5 folds for training and 5 folds for testing. This was done using the sklearn.model_selection module in Python to prepare the data for k-fold cross validation. K-fold cross validation is a model evaluation technique that helps avoid overfitting by training on multiple subsets of the data and testing on the remaining subsets.

Table 1. Sample Data

Index	Label	Sentence	w_index
14191	0	true love forever, nagbreak ayun nganga	0
14191	0	true love forever, nagbreak ayun nganga	1
14191	0	true love forever, nagbreak ayun nganga	2
14191	0	true love forever, nagbreak ayun nganga	3
14191	0	true love forever, nagbreak ayun nganga	4
14191	1	true love forever, nagbreak ayun nganga	5
4958	0	bagyong maaaring tumama sa metro manila	0
4958	0	bagyong maaaring tumama sa metro manila	1
4958	0	bagyong maaaring tumama sa metro manila	2
4958	0	bagyong maaaring tumama sa metro manila	3
4958	0	bagyong maaaring tumama sa metro manila	4
4958	0	bagyong maaaring tumama sa metro manila	5

In this case, each fold was used as a test set once, while the remaining 4 folds were used as training sets. This process was repeated 5 times, so that each fold was used as a test set once and as a training set 4 times. The average accuracy of the model across all 5 folds was used as a measure of its performance. (Marcot & Hanea, 2020). An example of the final data can be found in table 1.

Definition Selection

The second novelty of this study is the creation of an automatic definition selection algorithm. The definition of the target word is obtained to be used as an input in the Metaphor Identification Procedure (MIP) feature. This study implements this differently from 2 previous different studies that also employs MIP. One study utilized manual human annotators to choose from the closest dictionary definitions of each target word and use that as an input(Wan, Lin, Du, Shen, & Zhang, 2021). While another study automatically collected the definitions of each target word but only used the first definition since it they said that it would be the basic definition of the word (Babieno et al., 2022). In this study, a combination of the previous two approaches was applied by automatically searching and selecting from all definitions. The definition of the target word is obtained from the online dictionary called diksiyonaryoṗh. To determine the base form of the target word, a stemmer is applied similar to (Babieno et al., 2022). Unlike the previous methods used, this model considers multiple definitions of the target word found in the dictionary. To determine the appropriate definition to be used, the sentence embeddings of all the definitions was taken using a sentence transformer (Velasco et al., 2022) and the cosine similarities between each of the sentence emdeddings and the tweet was computed. The definition with the highest cosine similarity was selected using argmax and used an input. Figure 4 illustrates this entire process.

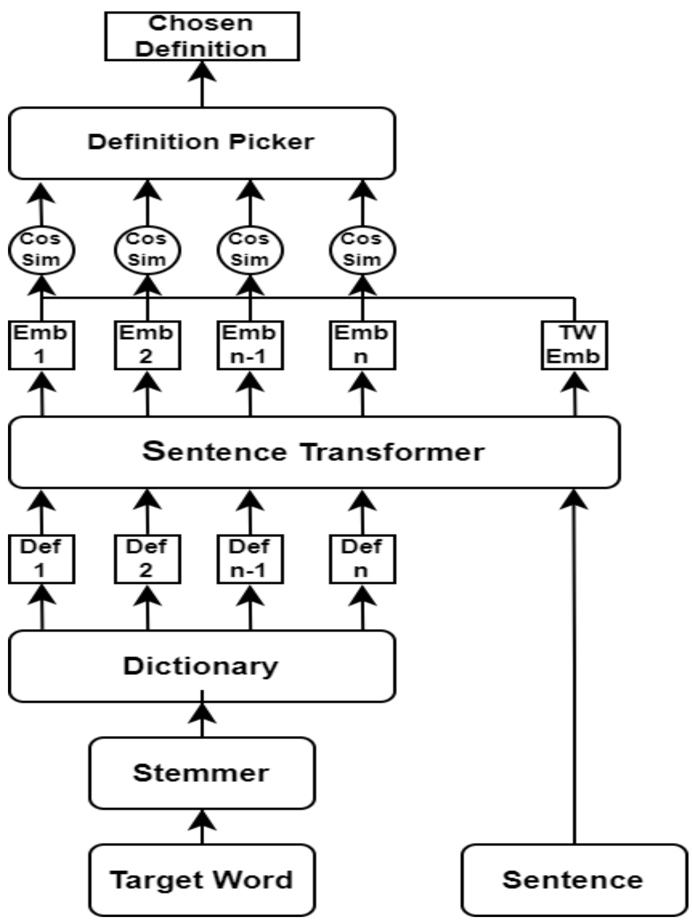

Figure 4. Definition Choosing Method

Metaphor Identification Model

This model will employ the combination of previous models that used sentence transformers and cosine similarity for metaphor detection.The Selectional Preference Violation (SPV) is computed by concatenating the mean of the encoded sentence using the language model, the pooled value of the sentence obtained from the sentence embedder, and the cosine similarity score between them. Similarly, the Metaphor Identification Procedure (MIP) is computed by concatenating the mean of the encoded sentence using the language model, the output of the definition after passing through the sentence embedder, and the cosine similarity between them. The resulting concatenation is linearly transformed to obtain a combined value. This combined value will once again be linearly transformed by concatenating it with the SPV values. Finally, the logarithmic softmax function is applied to the resulting linear value to obtain the final value for analysis. The final value is the probability for each of the categories, with the greater one being chosen as the predicted label. Figure 5 contains a sample output of the model.

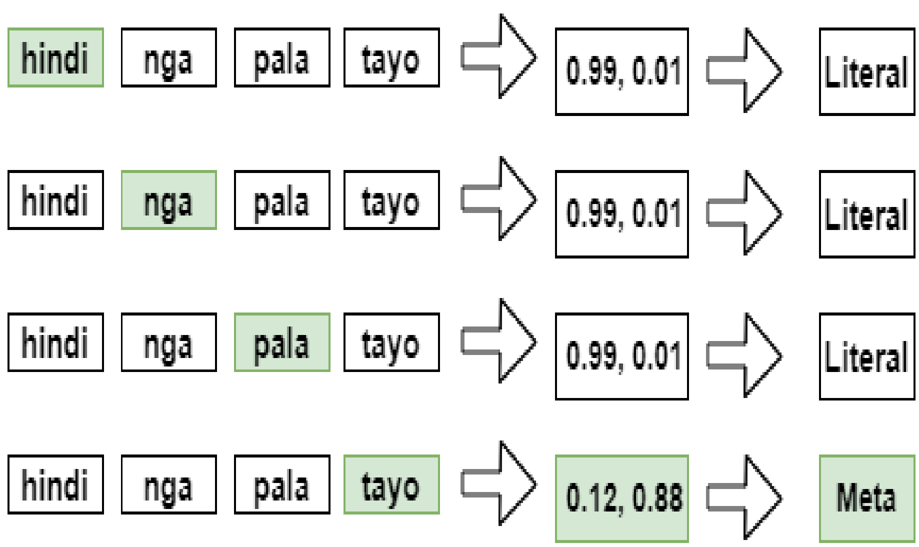

Figure 5. Sample Output

Training and testing

The hyperparameters used in this study are similar to those used in the study (Babieno et al., 2022). The model's maximum sequence length is set to 150, and it is trained for three epochs using the AdamW optimizer with an initial learning rate of 1x10-5. The third epoch is designated as the warm-up epoch. The training batch size is 32, while the testing batch size is 8. The training is performed on the V100 GPU provided by Google Colab. Using these parameters, this model was trained for a total of 30 hours.

Evaluation

The algorithm's performance was evaluated using precision, recall, and F1 metrics, which are commonly used measures for assessing classification models' effectiveness (Powers, 2020). In this evaluation, the counting of data was conducted differently compared to previous studies (Choi et al., 2021) (Babieno et al., 2022) which counted the metrics on a per-word basis.

The final novelty of this study is the modification of the typical correct metaphor counting method. The counting method for correctly guessed metaphors was modified from focusing on individual words to focusing on the identification of entire metaphorical sequences. This change means that rather than counting each part of the metaphor sequence, the new approach considers phrases or expressions as a whole.

For example, in the previous counting method, the phrase ”Kapal ng mukha” used metaphorically would count as three separate words employed metaphorically. However, in the updated approach, the entire sequence is counted as a single unit.

This revised method will only count a correct prediction only if all the words within the phrase are accurately predicted and forms a complete metaphorical phrase.

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1 = \frac{2xPrecisionxRecall}{Precision+Recall}$$

Figure 6. Evaluation Metrics

RESULTS AND DISCUSSION

Data Collection

The objective of this research was to develop an algorithm for identifying Filipino figures of speech in sentences. A dataset of 30,000 raw tweets was collected from the Twitter accounts DZMMTeleRadyo and @mgapinoyhugot between January 2022 and January 2023, as these accounts were known for regularly posting Filipino tweets. Snscrape, a Python library, was used to filter the collected tweets based on language and posting dates. During the data annotation process, approximately 15,000 repetitive tweets were removed to avoid redundancy in computation, resulting in a final dataset of approximately 14,500 tweets.

Table 2. Total Tweets

Dataset	Total	Non-Literal	Literal
News	9, 524	304	9, 145
Hugot	4, 976	582	4, 605
Total	14, 500	886	13, 750

Table 2 provides an overview of the dataset used in the research paper. There were 9, 524 distinct tweets taken from the twitter news source. 9,145 tweets that do not contain figurative language and 304 tweets that contain figurative language. There were also 5,000 distinct tweets taken from the twitter hugot source. 4,976 tweets that do not contain figurative language and 582 that contain figurative language. Table 3 and 4 provides an overview of the train and test dataset respectively once they have been divided into 5 folds. Similar to table 2, the columns of these tables contain the total number of tweets, number of tweets that contain figurative language, and number of tweets that do not contain figurative language.

The test dataset used in the experiment consisted of metaphors categorized into three

Table 3. Training dataset

Dataset	Literal	Non-Literal	Total
Train0	11, 007	699	11,585
Train1	10, 992	716	11,585
Train2	10, 989	721	11, 586
Train3	11, 009	706	11, 586
Train4	10, 999	706	11, 586

Table 4. Test dataset

Dataset	Literal	Non-Literal	Total
Test0	2, 742	188	2, 915
Test1	2, 758	170	2, 915
Test2	2, 760	166	2, 914
Test3	2, 740	181	2, 914
Test4	2, 750	181	2, 914

different types: Single, Multiple, and Multi-word. The total count of the metaphors in each category per dataset is presented in Table 5. The Single type of metaphors had the highest frequency across all test cases, with an average count of 152 metaphors. This shows that the majority of metaphors in the dataset were composed of a single word per sentence. The Multiple type of metaphors followed with an average count of 12, this shows that there were a few instances where multiple occurences of metaphors within a single sentence. The Multiword type of metaphors had an average count of 18, indicating that there were more instances of multiple word metaphors than multiple single word metaphors in a sentence.

Table 5. Metaphor Categories

Dataset	Single	Multiple	Multiword
Test0	160	13	23
Test1	144	13	18
Test2	142	12	16
Test3	167	7	13
Test4	149	15	20

Performance Evaluation

Table 6 presents the results of correctly predicted metaphor categories, namely Single, Multiple, and Multiword. The table indicates that the model demonstrates a higher accuracy in predicting Single metaphors compared to Multiple and Multiword metaphors. Notably, the model was unable to detect any instances of Multiword metaphors across all datasets.

Table 6. Correctly Predicted Metaphor Categories

Dataset	Single	%	Multiple	%	Multiword
Test0	31	19.37%	4	30.76%	0
Test1	32	22.22%	5	38.46%	0
Test2	34	23.94%	2	16.66%	0
Test3	39	23.35%	0	0%	0
Test4	15	10.06%	1	6.67%	0

The performance evaluation results for the test dataset are shown in Figure 7. The evaluation metrics include precision, recall, and F1 score. Precision measures the proportion of correctly predicted positive instances out of the total instances predicted as positive. Recall calculates the proportion of correctly predicted positive instances out of the actual positive instances. F1 score is the harmonic mean of precision and recall, providing a balanced measure of the model's performance.

The experiment results (Figure 7) show varied performance across the test datasets. F1 scores ranged from 18.26% to 41.92%, with Test0 achieving the highest and Test4 the lowest. Precision values ranged from 70.37% to 93.33%, with Test3 achieving the highest and Test4 the lowest. Recall values ranged from 10.49% to 28.23%, with Test1 achieving the highest and Test4 the lowest.

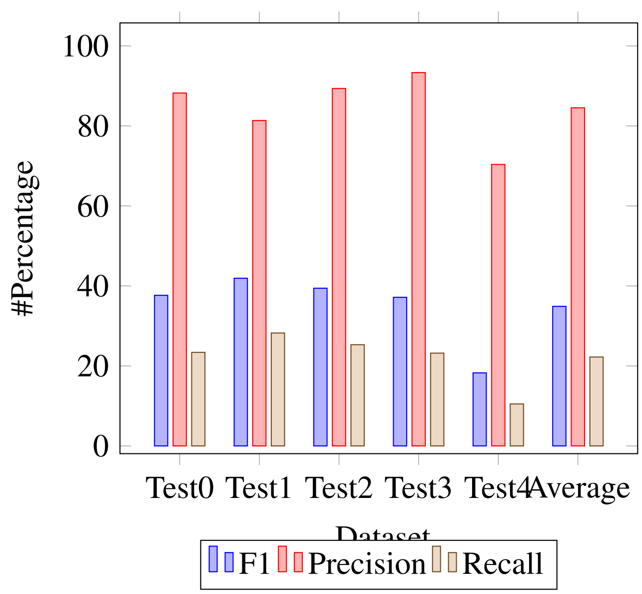

Figure 7. Test Dataset Performance

SUMMARY AND CONCLUSION

In conclusion, this study successfully addressed its objectives by creating a specialized corpus for Filipino metaphors and developing an algorithm based on the SPV and MIP approaches to detect metaphors in Filipino sentences. This is further enhanced by the introduction of an automatic definition selector and a modified metaphor counting method. The performance algorithm was evaluated using precision, recall, and F1 score metrics resulting in an average precision of 84.53%, recall of 22.23%, and F1 score of 34.89%.

For future works, the size of the dataset can be increased to improve the performance of the detection algorithm. Exploring alternative stemmers and dictionaries or wordnets can be used to derive the definitions of the target words used. Another recommendation is exploring and comparing the performance of other metaphor identification algorithms.

LITERATURE CITED

ARIAS, L. H., & SENCIL, M. (2017). Reflecting filipino identities through hugot in three philippine films: A contextual and prosodic analysis. *Undergraduate thesis, Department of English: Mindanao State University-Iligan Institute of Technology Iligan City. Search in.*

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), 2081.

CATAYNA, C. R., & CLARIÑO, M. A. A. D. (2022). Sentiment analysis and topic classification of twitter-based data set on the face-to-face classes resumption in the philippines during the covid-19 pandemic. In *2022 2nd international conference in information and computing research (icore)* (p. 39-44). doi: 10.1109/iCORE58172.2022.00027

CHOI, M., LEE, S., CHOI, E., PARK, H., LEE, J., LEE, D., & LEE, J. (2021). Melbert: metaphor detection via contextualized late interaction using metaphorical identification theories. *arXiv preprint arXiv:2104.13615*.

CONNEAU, A., KHANDELWAL, K., GOYAL, N., CHAUDHARY, V., WENZEK, G., GUZMÁN, F., ... STOYANOV, V. (2019). Unsupervised cross-lingual representation learning at scale. *CoRR*, *abs/1911.02116*. Retrieved from `http://arxiv.org/abs/1911.02116`

DEL PILAR SALAS-ZÁRATE, M., ALOR-HERNÁNDEZ, G., SÁNCHEZ-CERVANTES, J. L., PAREDES-VALVERDE, M. A., GARCÍA-ALCARAZ, J. L., & VALENCIA-GARCÍA, R. (2020). Review of english literature on figurative language applied to social networks. *Knowledge and Information Systems*, *62*(6), 2105–2137.

DEVLIN, J., CHANG, M.-W., LEE, K., & TOUTANOVA, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

FERNANDEZ, J. L., & ADLAON, K. M. M. (2022). Exploring word alignment towards an efficient sentence aligner for filipino and cebuano languages. In *Proceedings of the fifth workshop on technologies for machine translation of low-resource languages (loresmt 2022)* (pp. 99–106).

GAITAN-BACOLOD, M. A. (2010). Ang metapora sa wikang filipino. *Philippine Social Sciences Review*, *62*(1), 1.

GROUP, P. (2007). Mip: A method for identifying metaphorically used words in discourse. *Metaphor and symbol*, *22*(1), 1–39.

HAAGSMA, H., & BJERVA, J. (2016). Detecting novel metaphor using selectional preference information. In *Proceedings of the fourth workshop on metaphor in nlp* (pp. 10–17).

HEIN, A. M. (2010). Identification and bridging of semantic gaps in the context of multi-domain engineering. In *Forum on philosophy, engineering & technology* (pp. 58–57).

JURAFSKY, D., & MARTIN, J. H. (2014). *Speech and language processing*. Pearson Education.

KRISHNAKUMARAN, S., & ZHU, X. (2007). Hunting elusive metaphors using lexical resources. In *Proceedings of the workshop on computational approaches to figurative language* (pp. 13–20).

LAKOFF, G., & JOHNSON, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive science*, *4*(2), 195–208.

LAPITAN, F. R., BATISTA-NAVARRO, R. T., & ALBACEA, E. (2016). Crowdsourcing-based annotation of emotions in filipino and english tweets. In *Proceedings of the 6th workshop on south and southeast asian natural language processing (wssanlp2016)* (pp. 74–82).

LIU, J., O'HARA, N., RUBIN, A., DRAELOS, R., & RUDIN, C. (2020). Metaphor detection using contextual word embeddings from transformers. In *Proceedings of the second workshop on figurative language processing* (pp. 250–255).

LIU, Y., OTT, M., GOYAL, N., DU, J., JOSHI, M., CHEN, D., ... STOYANOV, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

MAO, R., LIN, C., & GUERIN, F. (2018). Word embedding and wordnet based metaphor identification and interpretation. In *Proceedings of the 56th annual meeting of the association for computational linguistics.*

MARCOT, B. G., & HANEA, A. M. (2020, jun). What is an optimal value of k in k-fold cross-validation in discrete bayesian network analysis? *Computational Statistics*, *36*(3), 2009–2031. Retrieved from `https://doi.org/10.1007%2Fs00180-020-00999-9` doi: 10.1007/s00180-020-00999-9

MESIONA, R. M. B., AGRABIO, J. E., DIANASAS, L., AMBRAD, R. C., & DIONES, L. L. (2022). Therese marie villarante select songs: A textual analysis on cultural influences of cebu. *International Journal of Modern Developments in Engineering and Science*, *1*(11), 48–55.

MYKOWIECKA, A., WAWER, A., & MARCINIAK, M. (2018). Detecting figurative word occurrences using recurrent neural networks. In *Proceedings of the workshop on figurative language processing* (pp. 124–127).

PANICHEVA, P. V., MAMAEV, I. D., & LITVINOVA, T. A. (2023). Towards automatic conceptual metaphor detection for psychological tasks. *Information Processing & Management*, *60*(2), 103191.

POWERS, D. M. (2020). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.

SHUTOVA, E., KIELA, D., & MAILLARD, J. (2016). Black holes and white rabbits: Metaphor identification with visual features. In *Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 160–170).

STEEN, G., DORST, A. G., HERRMANN, J. B., KAAL, A., KRENNMAYR, T., & PASMA, T. (2010). A method for linguistic metaphor identification. *Amsterdam: Benjamins*.

TITONG, A. M. (2014, May). *Research about figures of speech.* Retrieved from [[https://www.academia.edu/4022976/Research_about_Figures_of_Speech]]

VELASCO, D. J., ALBA, A., PELAGIO, T. G., RAMIREZ, B. A., CRUZ, J. C. B., & CHENG, C. (2022). Automatic wordnet construction using word sense induction through sentence embeddings. *arXiv preprint arXiv:2204.03251*.

WAN, H., LIN, J., DU, J., SHEN, D., & ZHANG, M. (2021). Enhancing metaphor detection by gloss-based interpretations. In *Findings of the association for computational linguistics: Acl-ijcnlp 2021* (pp. 1971–1981).

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

WISE, I. (2013, Oct). *High- and low-context cultures.* Wordpress. Retrieved from https://blogonlinguistics.wordpress.com/2013/10/22/high-and-low-context-cultu

ZHANG, S., & LIU, Y. (2022). Metaphor detection via linguistics enhanced siamese network. In *Proceedings of the 29th international conference on computational linguistics* (pp. 4149–4159).