

Retraction

Retracted: Collocation Features in Translated Texts Based on English Analogy Corpus

Scientific Programming

Received 1 August 2023; Accepted 1 August 2023; Published 2 August 2023

Copyright © 2023 Scientific Programming. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] B. Liu and J. Wang, "Collocation Features in Translated Texts Based on English Analogy Corpus," *Scientific Programming*, vol. 2022, Article ID 8294254, 7 pages, 2022.

Research Article

Collocation Features in Translated Texts Based on English Analogy Corpus

Bin Liu  and **Jing Wang**

School of Foreign Studies, Suqian University, Suqian 223800, Jiangsu, China

Correspondence should be addressed to Bin Liu; binliuandy@sq.edu.cn

Received 8 December 2021; Accepted 9 March 2022; Published 30 March 2022

Academic Editor: Le Sun

Copyright © 2022 Bin Liu and Jing Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The aim of this paper is to explore the characteristics of the use of verbal collocations in English, to compare the use of verbal collocations in the English translation and the original English text, and then to compare and analyse the characteristics of the choice of verbal collocations in the English text. In this paper, we take bilateral marked causative complex sentences as the object of study and use deep learning methods to automatically explore the implied features of complex sentences while incorporating the significant knowledge of relational words in linguistic research. The experimental results achieved an F1 value of 92.13%, which is better than that of the existing comparison models, demonstrating the effectiveness of the method.

1. Introduction

A corpus is a large collection of natural language materials, both written and spoken, collected systematically and scientifically for research purpose. A corpus is a large collection of authentic and reliable linguistic materials that provides a comprehensive and accurate representation of a language or an aspect of language, providing a wide range of verbal material for language research and revolutionising the way language research is conducted [1]. Since the 1980s, corpus-based translation research has become a new research paradigm in the field of translation research at home and abroad. Corpus translation studies take translated texts as the object of study and adopt the mode of combining intra-linguistic and inter-linguistic comparisons to describe and explain translation phenomena from a large-scale translated text or translated language as a whole, so as to explore the essence of translation [2, 3]. The corpus provides a new tool for translation studies, opening up new ideas and expanding the scope of translation research. Baker classifies translation corpora designed for different research purposes into three categories: parallel corpora, multilingual corpora, and analogical corpora, of which Baker considers the analogical corpus to be the most significant for translation research [4].

Through the comparative analysis of two texts in the analogical corpus, the researcher can explore the norms of translation in a particular historical and cultural context and discover some specific patterns of translated texts, i.e., the universality of translation. The salient features of the translated language are in the area of vocabulary, mainly in the conventionalisation of words used in the translated texts and the emergence of new word combinations [5, 6]. This new combination of words is a reflection of the lexical collocation characteristic of translated texts [7]. The linguistic features of the translated text are therefore highlighted in the lexical aspect, especially in the collocation of words, where the differences in collocation patterns reflect the differences between the original text and the translated text. The lexical collocation features reflect the specific meaning of the linguistic forms realised in the context and truly reflect the frequently used, habitual collocation forms of words in linguistic communication [8].

In recent years, with the continuous development of corpus translation science, studies on the lexical collocation characteristics of translated texts based on corpus have emerged at home and abroad, but empirical studies on lexical collocations in English translations of Chinese medicine texts using corpus are not common [9, 10].

Therefore, in this paper, the authors use the corpus to conduct a statistical study on the use of verb-name collocations in the English translation of TCM texts and the original medical English texts and then compare and analyse the patterns of verb-name collocations in medical texts, with the aim of providing some reference for the English translation of TCM texts, i.e., how to select suitable words for collocation in the English translation of TCM texts, and discovering lexical collocations.

2. Related Studies

Complex sentences are classified as marked or unmarked according to whether they contain relational words or not. At present, automatic recognition of marked compound sentence categories is mainly based on rule-based methods and machine learning methods. Wang et al. [11] combined the syntactic theory of Chinese compound sentences and the theory of relational tagging collocation to automatically identify the relational category of biological non-full-state marked compound sentences; the calculation method of semantic relatedness was used to calculate the relatedness of two words, so as to identify the relational category of compound sentences [12]. Igaab and Abdulhasan [13] used decision tree algorithms to extract features such as lexical properties to identify causal and juxtaposition relations between Chinese sentences.

For the recognition of relationship categories of unlabeled compound sentences, deep learning methods are mostly used for the recognition of relationship categories of unlabeled compound sentences due to the lack of relationship words and the absence of obvious manual recognition features [14]. Li et al. [15] used an attention-based mechanism of convolutional neural network on a Chinese chapter book library [16] for the recognition of unlabeled compound sentence relations. Algburi and Igaab [17] combined word vectors with lexical features as the input of the model and used CNN to classify unlabeled complex sentence relations. The study of unlabeled complex sentences still faces some difficulties, namely, the difficulty of data annotation, the relatively small amount of training data, and the uneven distribution of data among categories, which easily leads to overfitting of the model and makes the model's generalisation ability insufficient. Among the many deep learning models, the transformer model [18] has a simple network architecture with an attention mechanism as its main structure [19]. In this paper, we explore the fusion of relational word features in deep learning models to enable automatic recognition of biological marked causal complex sentence relations.

3. Data Statistics and Analysis

3.1. Word Frequency Statistics. Word frequency is the total number of occurrences of a word item or a class of words in the corpus, and counting word frequency can provide certain reference information about the stylistic or linguistic features of a discourse. The study of word collocation should centre on the behaviour of real words, and collocation studies should be conducted mainly by selecting real words as nodal words; the behaviour of functional or grammatical

words has mostly been described in detail by grammarians [20, 21]. Therefore, the first criterion for choosing nodal words in this paper is real words. Moreover, of the four main categories of real words (nouns, verbs, adjectives, and adverbs), nouns and verbs have the highest collocation power, thus further defining the nodal words studied in this paper as verbs. The following statistics are commonly used in English text bases: the number of tokens, number of the types, type/token ratio, word length, average word length, and so on [22]. In this paper, WordSmith 5.0 was used to obtain statistics on the common parameters of the self-built TCM English corpus and then rank the verb morphology of the corpus in descending order of frequency [23].

3.2. Extraction of Collocations for Verbal Structures. The collocation of these three verbs (influences, caused, and treating) in the self-built TCM English corpus was searched using AntConc software [24], which requires that the collocation must be in the lexical noun form and that they serve as an object of the sentence. Therefore, collocation that did not meet the requirements of the study was eliminated, leaving 200 significant collocations for each of the three verbs. Since the BNC has its own search data analysis function, which allows the selection of collocations in different genres of texts from different generations, the three verbs (influences, caused, and treating) were directly entered into the BNC and the collocations that met the requirements of the study were selected and then analysed quantitatively and qualitatively with the previous data [25, 26].

3.3. Analysis of Data. In Figure 1, the verb influence is most often found in the native English corpus with sphere, followed by decisions, and in the translated English corpus with factors, followed by range; in Figure 2, the verb cause is most often found in the native English corpus with trouble, followed by harm, and in the translated English corpus with damage, followed by problems; in Figure 3, the verb treat is most often found in the passive form with damage, followed by problems. The verb cause tends to occur in the passive form with damage, followed by problems in the translated English corpus; from Figure 3, the verb treat tends to occur most often with patients in the native English corpus, followed by symptoms, while treating tends to occur in the progressive tense in the translated English corpus. It is often found in the progressive tense, most often with disease, followed by pain [27].

The number of collocations of nodal words selected in the native English corpus is significantly higher than that in the translated English corpus, which indicates that the medical English native texts are more varied in terms of word usage and use a larger vocabulary than the translated texts, which is not beyond expectation since, after all, English translations of Chinese medical texts are mostly done by translators and are not as rich in terms of word usage as the native English texts. Once widely accepted, some high-frequency word combinations in translated texts may enter the target language and become the translation counterparts of several near-sense expressions, thus partially confirming the tendency of translated languages towards lexical simplification.

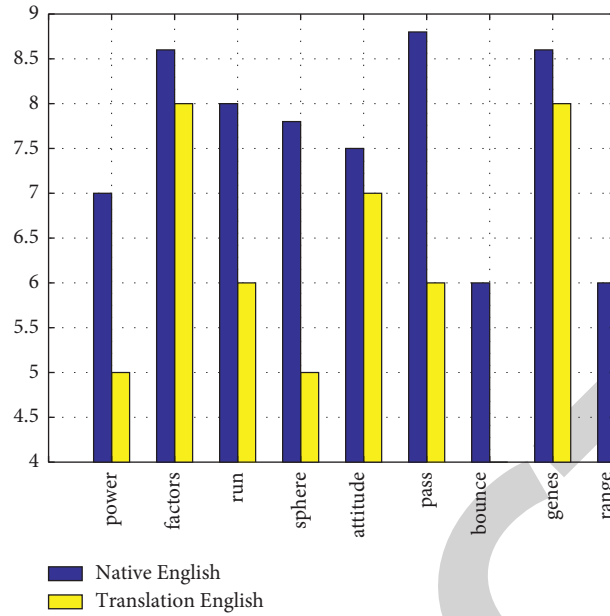


FIGURE 1: Comparison of the number of verb-noun collocations of nodal words.

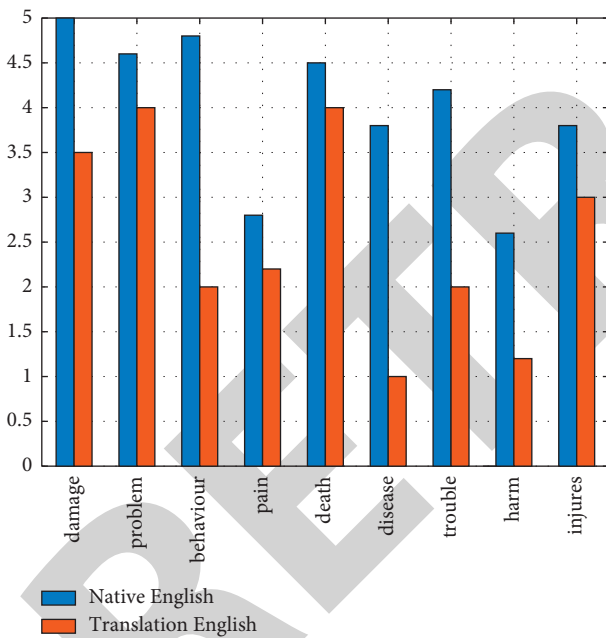


FIGURE 2: A comparative figure of the number of verb-noun collocations of the nodal word caused.

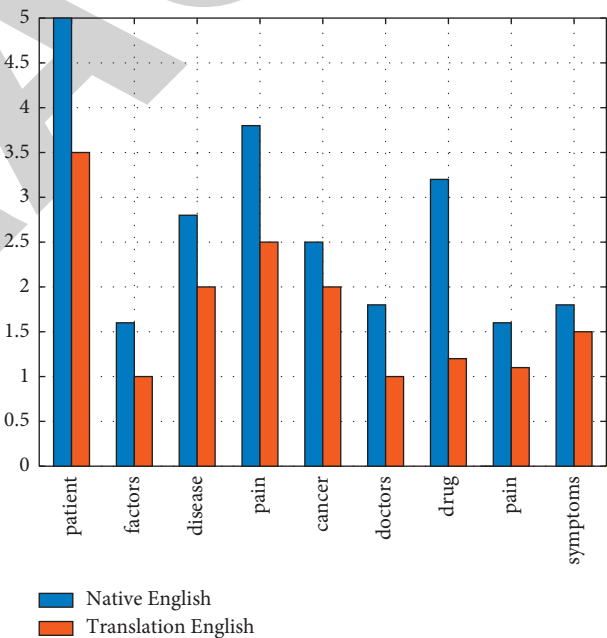


FIGURE 3: Number pairs of verb names of node word "treating."

4. The Transformer Model

4.1. Model Structure. Transformer is essentially an encoder decoder structure, which is composed of multiattention mechanism and feedforward neural network [28]. The multiheaded attention mechanism combines the context with the distant words and processes all words in parallel, thus achieving parallel computation and capturing the global semantic information. The structure of the RM-transformer model used in this paper is shown in Figure 4.

4.2. Model Input. In this paper, a pretrained word2vec word vector [29] is stitched with relational word features as model input. The 6-dimensional one-hot encoding is used for the relational word features, and all words in the word list are represented by the 6-dimensional relational word features. The first dimension uses 1 and 0 to indicate the presence or absence of a relation, and the next 5 dimensions correspond to the 5 relations of cause and effect, hypothesis, condition, inference, and purpose. Gensim’s word2vec model is used to train a 122 dimensional word vector, which is then stitched with the 6-dimensional relational word feature vector to

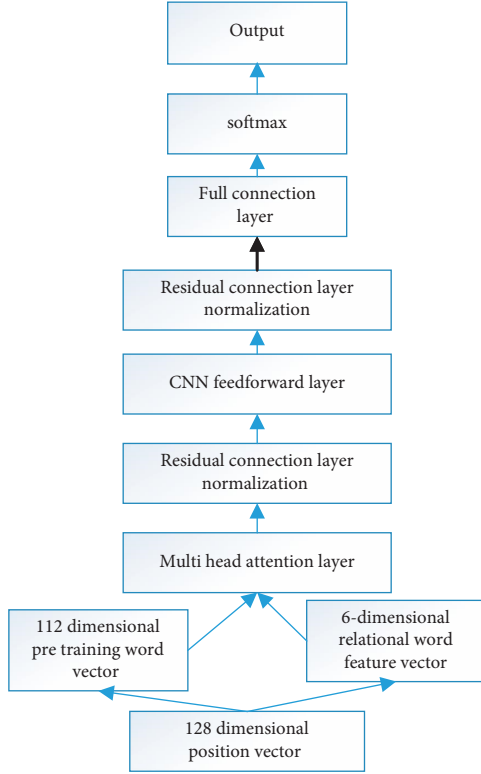


FIGURE 4: Structure of the RM-transformer.

obtain a 128-dimensional vector. If an input sentence is of length n , W_j ($j=1, \dots, n$) denotes the pretrained word vector for the j th word and R_j ($j=1, \dots, n$) denotes the relational word vector for the j th word. Then, the vector WR for each word is represented as follows:

$$WR_j = W_j \oplus R_j, \quad (1)$$

where \oplus indicates a splicing operation.

The multiheaded attention mechanism can obtain the information of long-distance features and can also perform parallel calculation, but it cannot represent the position information of the input sentence. Here, using Position Embedding proposed by Google [30], each position is encoded so that the multiheaded attention mechanism can obtain the position information of each word. The equation for the position vector is shown in equations (2) and (3).

$$P_{(j,2i)} = \sin\left(\frac{\text{pos}}{10000^{(2i/d_{\text{model}})}}, \right), \quad (2)$$

$$P_{(j,2i+1)} = \cos\left(\frac{\text{pos}}{10000^{(2i/d_{\text{model}})}}, \right), \quad (3)$$

where j ($j=1, \dots, n$) is used to represent the position information of the word, P_j (odd and even positions) represents the position vector at the j th position, and i is the index of each value in the vector. $d_{\text{model}} = 128$ is consistent with the dimensionality of the word vector after adding features. At even positions, sine coding is used; at odd positions, cosine coding is used. The vector representation of the input model is as follows:

$$\text{Vector}_j = WR_j + P_j, \quad (4)$$

where $+$ indicates direct summation of word vectors.

4.3. Transformer Feature Extraction. Self-attention is the calculation of the weight of each word vector on the input, which randomly initializes a set of weight matrices Q, K, V . Q, K, V refer to the result of multiplying the word vectors of the input model with a random initialization matrix, and d_k is the dimension of the Q, K vectors [31] and is calculated as follows:

$$\text{attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (5)$$

The self-attention layer is used to obtain the global semantic information of the input sentence, and after the self-attention layer, a feedforward layer is connected. NN uses a one-dimensional convolution operation and first performs an inner layer convolution operation, with the number of inner layer filters using the parameters set by oneself, using the relu activation function. The inner convolution operation is then performed, with the number of outer filtrators being the same as the dimensionality of the word vector, ensuring that the dimensionality of the input CNN is consistent with the dimensionality of the output. After the above process of transformer feature extraction, the output is fed into the next transformer encoder. Once the feature extraction is complete, a fully connected lawyer and a software layer are used to output the probability distribution for each category [32].

5. Experimentation and Analysis

5.1. Experimental Data. In this paper, we identify the relational categories of biological marked causal compound sentences, and the datasets are the Corpus of Chinese Compound Sentences (CCCS), an annotated corpus from Huazhong Normal University, and THUCNews, Tsinghua News Classification Corpus [33]. The CCCS is a special corpus of 658 Chinese compound sentences (447 items) from the People's Daily and the Yangtze River Daily [15]. The Tsinghua News Corpus THUCNews is a filtered corpus of 14 types of short-text news items, based on historical data from the Sina News RSS feed from 2005 to 2011 [19]. A total of 91,646 two-sentence marked causal compound sentences were annotated, forming a corpus abbreviated as CTCCCS (the Corpus of Two-Sentence Causal Chinese Compound Sentences), and the data distribution of each relationship category in the dataset is listed in Table 1. In the experiment, 75% of the data were selected as the training set and 25% of the data were selected as the test set, and the data were divided as listed in Table 2.

5.2. Experimental Comparison and Analysis. The experiment compares each of the same hyperparameter settings of the model and the values of different weights in CNN [3, 4, 5]. It can be known that convolution kernels of different sizes can

TABLE 1: Data distribution of the CTCCCS corpus by relationship category.

Relationship category	Number of complex sentence instances	Scale (%)
Causal class	30180	32.93
Hypothetical class	31444	34.31
Condition class	11576	12.63
Target class	12438	13.57
Inference class	6008	6.56

capture features of different sizes, which are more effective than data fitting using a single convolution kernel.

The number of layers in the LSTM and BiLSTM is set to 1, and the hidden layer is set to 128. The experimental results are listed in Table 3. The accuracy of the model was improved by 3.27%, 0.98%, and 0.3%, respectively. Compared to the transformer model without the addition of relational features, accuracy improved by 13.74%. Using a fixed sequence length of 100 in the model, the RM-transformer improved the precision, recall, and F1 values more significantly compared to the CNN, by 3.38%, 2.83%, and 3%, respectively. The learning of long-range features may be difficult, although multiple convolutional kernels of different sizes are used to capture features of different sizes.

The RM-transformer performs parallel computation through a multiheaded attention mechanism while learning global feature information and then learning sequential local features through the CNN feedforward layer, thus achieving better results than CNN.

The RM-transformer has an improvement of 1.26% and 0.18% in F1 values compared to the LSTM and BiLSTM, respectively, which is not particularly significant compared to the CNN. The recall of RM-transformer is 0.12% lower than that of LSTM. LSTM and BiLSTM are relatively mature in dealing with input text sequences and can learn the long-range features of the sequences, but LSTM relies on the above information and BiLSTM relies on the contextual information. The self-attention mechanism in the RM-transformer enables the direct correlation of long-range features to obtain global features, and the RM-transformer can achieve similar effects as the LSTM and BiLSTM. This paper has limited manually annotated data, and the parallel computational power of the transformer may be able to have better results when experimenting on a more data-rich complex sentence dataset. When comparing the RM-transformer with the transformer model without adding relational features, the F1 value increased by 11.63%, which is a significant improvement, indicating that relational words play a very important role in determining the relationship of causal complex sentences. Although the deep learning model can automatically mine the text for some semantic and other feature information, it can be made more effective by adding some obvious manually identified features. The results of the classification experiments for each category of cause-effect complex sentences are listed in Table 4.

TABLE 2: CTCCCS dataset division.

Relationship category	Number of training sets	Number of test sets
Causal class	22635	7545
Hypothetical class	31444	7861
Condition class	8682	2894
Target class	93287	3110
Inference class	4506	1502

TABLE 3: Comparison of experimental results for each model.

Algorithm	Accuracy	Precision	Recall	F1
Transformer	81.35	82.33	79.05	80.50
CNN	89.79	88.84	89.44	89.13
LSTM	92.08	89.61	92.39	90.87
BiLSTM	92.76	91.99	91.96	91.95
RM-transformer	93.06	92.22	92.27	92.13

TABLE 4: Correct classification rates for each category of causal complex sentences.

Relationship category	Accuracy (%)
Causal class	93.25
Hypothetical class	92.41
Condition class	95.20
Target class	96.85
Inference class	83.69

From Table 4, it can be seen that the recognition rate of inferred compound sentences is significantly lower than that of other categories. The possible reasons for the classification errors in the experiment are as follows. (1) The corpus of the experiment is mostly from the news corpus, and inferred compound sentences are used infrequently, so the collected corpus is on the low side, and overfitting occurs during the training process. (2) There are multiple quasi-relatives (words that can act as relatives) in the sentences, corresponding to different categories. In this sentence, “since” indicates an inferred relationship, but “that” can indicate both a hypothetical and an inferred relationship, which should have been judged as a hypothetical relationship [34, 35].

Next, the dependent syntactic tree [18] can be used as input to incorporate richer syntactic information, and the graphical CNN model [19] can be used for the recognition of relational categories of marked compound sentences to further improve the accuracy of compound sentence category recognition.

6. Conclusions

In summary, an analogous study of collocation characteristics of verbal names in English translations based on the corpus found the following.

- (1) Compared with other texts, verbal-name collocations in the English language are more concise, passive forms are used more frequently, verbal-name

collocation patterns are relatively fixed, and the choice of English vocabulary reflects the professional and concise nature of the medical language, with a simple and logical collocation structure.

- (2) The choice of words in the English translation is somewhat narrower than that in the original English text, and the nodal words have far less influence on the collocations than the target language, reflecting the fact that the translator is influenced by the source language when translating and has certain limitations in word choice, which differs from the target language in the use of verbs.

A corpus can provide a large amount of authentic and natural linguistic data for text translation and provide a more objective and comprehensive picture of the characteristics and intrinsic patterns of Chinese medical English. The use of English corpus to study lexical collocation features can help explore the universal laws of Chinese medical English translation, grasp the characteristics of the translated text itself, and provide a basis for the standardisation of English terminology. At the same time, by searching the corpus of native speakers, exploring the constitutive rules of medical English vocabulary as well as customary collocation, and digging deeper into the meanings of English words, new translation ideas and translation methods can be provided for English translations.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] S.-Ho Han, "A review of research on restaurant brand personality: a focus on the hospitality and tourism journals listed at Korea research foundation," *Journal of Tourism Sciences*, vol. 35, no. 2, pp. 337–353, 2011.
- [2] K. Ahrens, H. Zeng, and S. H. R. Wong, "Using a corpus of English and Chinese political speeches for metaphor analysis," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 2018.
- [3] F. Mohamad and N. Abdul Malik, "Metaphor, religion, and gender: a case study of metaphor analysis in Islamic motivational speech corpus/Norasyikin Abdul Malik and Faizah Mohamad," *International Journal of Modern Languages and Applied Linguistics (IJMAL)*, vol. 5, no. 3, pp. 95–121, 2021.
- [4] T. Wang and S. Ge, "Corpus-based semantic prosody study of English-Chinese translation: taking trump's popular saying "it is what it is" as an example," in *Learning Technologies and Systems*, pp. 420–429, Springer, New York, NY, USA, 2020.
- [5] P. Boontam, "A corpus-based study of food-related metaphors in English and Thai," *The Academic Journal: Faculty of Humanities and Social Sciences Nakhonsawan Rajabhat University*, vol. 6, no. 1, pp. 1–43, 2019.
- [6] M.-L. Chen, "A corpus-based study on imagery and symbolism in Goldblatt's translation of Red Sorghum," *Babel. Revue internationale de la traduction/International Journal of Translation*, vol. 65, no. 3, pp. 399–423, 2019.
- [7] G. Liu and X. Zhang, "The meaning-shift unit of Chinese intensifiers: a corpus-based study," *Journal of Language Teaching and Research*, vol. 12, no. 1, pp. 111–119, 2021.
- [8] Y. Wang, "Book review: understanding metaphor through corpora: a case study of metaphors in nineteenth century writing," *International Journal of Translation, Interpretation, and Applied Linguistics (IJTIAL)*, vol. 3, no. 2, pp. 50–55, 2021.
- [9] H. Feng, I. Crezee, and L. Grant, "Form and meaning in collocations: a corpus-driven study on translation universals in Chinese-to-English business translation," *Perspectives*, vol. 26, no. 5, pp. 677–690, 2018.
- [10] E. F. Nurhidayat, E. Apriani, and S. Edy, "The analysis of cohesive devices used by tertiary English students in writing English paragraphs," *International Journal of Multicultural and Multireligious Understanding*, vol. 8, no. 4, pp. 70–81, 2021.
- [11] L. Wang, C. Zhang, Q. Chen et al., "A communication strategy of proactive nodes based on loop theorem in wireless sensor networks," in *Proceedings of the 2018 Ninth International Conference on Intelligent Control and Information Processing (ICICIP)*, pp. 160–167, IEEE, Wanzhou, China, November 2018.
- [12] S. Perdana, S. Fajaruddin, G. E. Kiswaga, A. Ariyanto, and A. Fujiastuti, "Experiential meaning breadth variations of the English-bahasa Indonesia Alice in Wonderland movie texts," *Script Journal: Journal of Linguistics and English Teaching*, vol. 6, no. 1, pp. 44–55, 2021.
- [13] Z. K. Igaab and H. Abdulhasan, "Collocation in English and Arabic: a contrastive study," *English Language and Literature Studies*, vol. 8, no. 4, pp. 89–103, 2018.
- [14] Z. K. Igaab, "The pragmatics of blackmail in English and Iraqi-Arabic," *International Linguistics Research*, vol. 4, no. 3, 72 pages, 2021.
- [15] B. Li, Y. Li, C. Xu et al., "The NiuTrans machine translation systems for WMT19," in *Proceedings of the Fourth Conference on Machine Translation*, vol. 2, pp. 257–266, Florence, Italy, August 2019.
- [16] D. Suparno, M. Farkhan, and M. A. Wahab, "Word formation in modern standard Arabic Palestinian and Tunisian," *AIUA Journal of Islamic Education*, vol. 1, no. 1, pp. 65–88, 2019.
- [17] B. Y. J. Algburi and Z. K. Igaab, "Defamation in English and Arabic: a pragmatic contrastive study," *International Linguistics Research*, vol. 4, no. 2, 31 pages, 2021.
- [18] H. Feng, *Form, Meaning and Function in Collocation: A Corpus Study on Commercial Chinese-to-English Translation*, Routledge, England, UK, 2020.
- [19] K. Saito, "Multi- or single-word units? The role of collocation use in comprehensible and contextually appropriate second language speech," *Language Learning*, vol. 70, no. 2, pp. 548–588, 2020.
- [20] R. Raksangob Wijitsopon, "Collocations and local textual functions of quantifiers in learner English essays," *Linguistic Research*, vol. 34, no. 1, pp. 1–49, 2017.
- [21] H. Li, D. Zeng, L. Chen, Q. Chen, M. Wang, and C. Zhang, "Immune multipath reliable transmission with fault tolerance in wireless sensor networks," in *Proceedings of the International Conference on Bio-Inspired Computing: Theories and Applications*, pp. 513–517, Springer, Singapore, October 2016.
- [22] M. Jafar Jabbari and N. Kavooosi, "An investigation into the collocations used in the translation of official documents from

- Persian into English,” *Communication and Linguistics Studies*, vol. 3, no. 2, pp. 15–21, 2017.
- [23] E. L. Jiménez-Navarro, “Nominal collocations in scientific English: a frame-semantic approach,” in *Proceedings of the International Conference on Computational and Corpus-Based Phraseology*, pp. 187–199, Springer, Malaga, Spain, September 2019.
- [24] P. An, Z. Wang, and C. Zhang, “Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection,” *Information Processing & Management*, vol. 59, no. 2, Article ID 102844, 2022.
- [25] L. Lei and D. Liu, “The academic English collocation list,” *International Journal of Corpus Linguistics*, vol. 23, no. 2, pp. 216–243, 2018.
- [26] K. Mroczyńska, “A dictionary of legal English collocations as an aid for mastering the legal English genre,” *Linguistics*, vol. 6, pp. 130–141, 2020.
- [27] S. Choi, “Processing and learning of enhanced English collocations: an eye movement study,” *Language Teaching Research*, vol. 21, no. 3, pp. 403–426, 2017.
- [28] L. Chen, “Corpus-aided business English collocation pedagogy: an empirical study in Chinese EFL learners,” *English Language Teaching*, vol. 10, no. 9, pp. 181–197, 2017.
- [29] A. Frankenberg-Garcia, R. Lew, J. C. Roberts, G. P. Rees, and N. Sharma, “Developing a writing assistant to help EAP writers with collocations in real time,” *ReCALL*, vol. 31, no. 1, pp. 23–39, 2019.
- [30] D. Wu, C. Zhang, L. Ji, R. Ran, H. Wu, and Y. Xu, “Forest fire recognition based on feature extraction from multi-view images,” *Traitement du Signal*, vol. 38, no. 3, pp. 775–783, 2021.
- [31] A. Kotelnikova and E. Kotelnikov, “SentiRusColl: Russian collocation lexicon for sentiment analysis,” in *Proceedings of the Conference on Artificial Intelligence and Natural Language*, pp. 18–32, Springer, Tartu, Estonia, November 2019.
- [32] P. Saliminejad and G. Karimkhanlooei, “A study on the type and frequency of unacceptable collocations in the English-Persian translations of Hemingway’s Masterpiece: for Whom the Bell Tolls,” *Journal of Language and Cultural Education*, vol. 6, no. 3, pp. 85–100, 2018.
- [33] Q. Zhang, H. Lu, H. Sak et al., “Transformer transducer: a streamable speech recognition model with transformer encoders and rnn-t loss,” in *Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7829–7833, IEEE, Barcelona, Spain, May 2020.
- [34] S. Singh and A. Mahmood, “The NLP cookbook: modern recipes for transformer based deep learning architectures,” *IEEE Access*, vol. 9, pp. 68675–68702, 2021.
- [35] L. Cai, K. Janowicz, G. Mai, B. Yan, and R. Zhu, “Traffic transformer: capturing the continuity and periodicity of time series for traffic forecasting,” *Transactions in GIS*, vol. 24, no. 3, pp. 736–755, 2020.