Journal of Network Intelligence ©2025 ISSN 2414-8105 (Online)
Taiwan Ubiquitous Information Volume **10**, Number **1**, February **2025**

Degree of Freedom Calculation of Translation Feature Extraction Based on Natural Language Processing

Qing-Hua Ai*
School of International Studies
Xi'an Jiaotong University City College, Xi'an 710018, China
QHA_2023@163.com

Jun Wang
Xi'an Municipal Engineering Design & Research Institute., Ltd. Xi'an 710068, China
512258016@qq.com

Qing-Yan Ai
Medical School, Yan'an University, Yan'an 716000, China
aqy_2001@163.com

*Corresponding author: Qing-Hua Ai

Received December 28, 2023, revised March 29, 2024, accepted May 15, 2024.

Abstract. *Natural language processing, an important research direction in the field of computer and artificial intelligence, aims to realise information interaction between human and computer through natural language. The degree of freedom (DOF) affects the probability of English phrase translation. This paper uses feature extraction algorithms to calculate DOFs for English phrase translation. This paper studies and analyses the characteristics of different feature extraction strategies, automatically filters the semantic information of sentences based on word frequency statistics and grammatical corpus, and combines the model results to predict word relationship to translate all Chinese words accurately. Firstly, this paper introduces the current development of Chinese–English translation system and related theories. Secondly, the paper describes in detail the research work of 'phrase', 'grammatical feature extraction algorithm' and automatic filtering of sentence semantic information based on word frequency statistics and syntactic corpus, and then introduces the proposed method based on word frequency statistics. Quantitative and syntactic corpus extracts three key lexical features commonly used in English word translation and combines the model prediction results to examine and summarise the combination of three translation strategies under different corpus qualities. The algorithm in this paper completes in 1.2 seconds, significantly faster than other algorithms in the experiment; the vector space model required 8 seconds, and Jinbo et al.'s method took 5 seconds. This highlights the superior learning effect of the method proposed in this paper for calculating the DOF of English phrase translation.*
Keywords: Degrees of freedom, Feature extraction algorithm, TF-IDF algorithm, English phrase translation

1. **Introduction.** Semantic representation is the basis of natural language processing (NLP), which first converts text data into semantic representation understood by a computer and then applies it to natural language processing task [1]. Presentation learning can automatically mine semantic features from large-scale data unlike traditional methods that require manual construction of features and support machine learning training and prediction. Presentation learning can represent the semantics of large-scale unstructured

data such as text and image, amongst which the text representation methods are mainly divided into discrete representation and distributed representation. Because deep learning performs well in NLP, more attention has been paid to code retrieval models based on neural networks [2, 3]. In recent years, people's research on the English vocabulary has become more in-depth, whilst many difficulties remain in translation. For example, phonetic distortion caused by factors such as phonetics and grammar, differences in pauses, incomplete or even isolated semantics, can affect the expression effectiveness and accuracy of sentences or articles. These difficulties are objective, inevitable problems and challenges that must be addressed and require translators to conduct research and analysis of English vocabulary from an emotional point of view and then derive translation strategies and improve the quality of translations. With the deepening of research on the English vocabulary, people have begun to pay attention to the translation of English phrases. Therefore, this word was analysed and processed as the target language to obtain more accurate results and better translation strategies and methods to provide readers with information services. In addition, the degree of freedom (DOF) of translation had a great influence on the randomness of translation and provided a rational reference for the choice of translation.

Many scholars have conducted research on DOF calculation. Yagasaki [4] considered a two-DOF Hamiltonian system with a saddle centre and developed a Melnikov-type technique for detecting the creation of transverse homoclinic orbits by higher-order terms . Bonazzoli and Rapetti [5] proposed explicit generators for higher-order scalar and vector finite element spaces commonly used in numerical electromagnetism and revisited the classical DOFs, the so-called moments. Yuan et al. [6] proposed a model-free adaptive control scheme for the three-DOF hybrid magnetic bearing control system. Nikdel et al. [7] introduced a new method for designing an adaptive thrust reverser controller through state augmentation and realised the control law of a general n-DOF robotic manipulator. Gong et al.[8] focused on the dynamic gait control of a 20-DOF complex robot for the first time. Firstly, a complete 3D model of the quadruped robot was established based on the spring-loaded inverted pendulum model, the inverse kinematics of the model were analysed, the trajectory of the swinging foot was planned and the hydraulic drive was examined. Secondly, the virtual leg-based single-leg control algorithm was generalised to a quadruped robot, and the state variables for gait and restrained gait were planed. Finally, the two dynamic gaits were implemented on the ADAMS-MATLAB cosimulation platform. Nafea and Yener [9] studied multiantenna eavesdropping channels in the presence of multiantenna cojammers. Safe DOFs for this channel were established for all possible values of the number of antennas. These studies did not form a library for the texts used in the DOF calculation, but the best way to form a library of texts was to use feature extraction algorithms.

Many scholars have researched on feature extraction algorithms. To reduce the workload of feature extraction and improve the recognition accuracy, Li et al. [10] proposed a new region of interest extraction method for knuckle print images. To improve the intelligent diagnosis accuracy of rolling bearings, Yan and Jia [11] proposed a new fault classification algorithm based on multidomain feature optimisation support vector machine. The data of these studies were not comprehensive, and the results of the studies were still open to question; they can neither be recognised by the public nor popularised and applied. Therefore, combining the feature extraction algorithm and the DOF calculation to study the English phrase translation DOF calculation based on the feature extraction algorithm facilitated the recognition of the final data.

NLP enables electronic devices such as computers and mobile phones to recognise and understand human language. Users can control and use electronic devices more conveniently and quickly, saving operation time and improving work efficiency. NLP is also a science that combines linguistics, computer science and mathematics. This paper selected phrases from 10 different categories of fields, namely, education, environment, military, economy, medicine, sports, art, politics, science and technology, and transportation from the Chinese NLP open platform, and used 1,000 phrases for related experiments. The accuracy of the phrase translation similarity algorithm proposed in this paper was reflected by the method of verifying phrase clustering. The inverse text word frequency (TF-IDF) algorithm was used to cluster 1,000 texts, and the vector space model and the method of scholar comparison were applied. The data showed that our algorithm outperformed other algorithms in calculating the DOFs of English phrase translation, but improvements could still be made. Moreover, this research satisfied the needs of English phrase translation DOF calculation.

2. **Calculation of Degrees of Freedom for English Phrase Translation.** English, one of the most popular communication tools in the world, plays an important role in international communication [12, 13]. Because Chinese is a complex, varied polyphonic context, people have gained a deeper, more specific understanding of its expression. However, with the passage of time and the influence of social development and changes, new problems have emerged. English grammar rules are constantly updated and upgraded (such as word form conversion in translation), and many new problems have appeared in the translation [14, 15]. All these require the constant updating of English grammar rules, thereby improving the accuracy of translation expression and the amount of language information [16, 17]. Because the Chinese language itself has the characteristics of high abstraction and complexity, uncontrollable semantics and so on, expressing exactly what it wants to convey is difficult; in addition, because each person has diverse opinions on the same word or phrase cognitive ability and comprehension methods, people often use different rules to interpret meanings when translating sentences. Therefore, the main problem is how to understand Chinese grammar rules accurately in a specific context [18, 19].

In English phrase translation, freedom refers to the number of options or variables that can be freely chosen during the conversion process. These options or variables include lexical choices, syntactic structures, semantic transformations, etc., which affect the accuracy and fluency of the translation. Calculating freedom can help us understand how much flexibility translators have in a specific translation task and the extent to which these freedoms affect the translation outcome. By analyzing freedom, we can better understand the challenges and constraints in the translation process, thereby optimizing translation algorithms and improving translation quality.

The main purpose of this paper is to study the English phrase translation strategy based on feature extraction, that is, how to match two words accurately to increase the fidelity and readability of translation [20, 21]. Firstly, two common translation methods are introduced: direct translation, AHP and hybrid interpolation techniques. Then, these basic models are briefly explained, and the relevant conditions of various characteristics in specific functions and the application of various characteristics are given. Finally, according to the research content, the English vocabulary feature extraction strategy and its implementation steps are proposed, that is, the accurate expression and transformation based on word structure features and sentence pattern features [22, 23].

Corpus refers to the material needed in the translation, including words, symbols and grammar [24, 25]. When classifying sentences, the locale is first considered [26]. Semantic analysis shows much repetition between Chinese and English words [27]. For example, in Chinese, the word “pingguo” can correspond to “apple” in English, but it can also be translated as “fruit” to represent a broader category. On the other hand, in English, the word “apple” can correspond to “pingguo” in Chinese, and in some contexts, it can also refer to “pingguo shu”. Additionally, in Chinese, the word “gaoxing” can be translated as “happy” or “glad” in English. Similarly, in English, the word “happy” can correspond to “gaoxing” in Chinese, and it can also be translated as “xiyue” to capture similar emotional nuances. Therefore, this paper uses the thesaurus method to express the semantics of the two words, and the English vocabulary can also use a corpus database (such as FES) or corpus rules to achieve data storage in the translation, that is, to use the source text and target words as sentences. The source of the material can also be translated by using the method of the source text. Due to a large amount repetition between the source text and the target language, this paper uses the thesaurus rules for translation [28].

In the English vocabulary, terms and words are very closely related, and they can be used to represent various forms, such as different types, times or places. People express some words with specific morphemes. Examples include ‘I have a person’, ‘What is he? What should you do...’ These words are independent of one another, related to one another and different from one another. This relationship is between the two most basic and commonly used words in sentences, also known as sentence-type phrases and subtype short-term words that have similarities and commonalities in semantics. Therefore, to avoid semantic differences between words, some special morphemes are used in sentences to express ideas. Examples include ‘You are a person? He is not this person! Why should I do this...’ and other words and sentence patterns and subtype loan words; alternatively, Chinese vocabulary is directly used to express the relationship between phrases and sentences (tone or overlapping of meanings). These words are generally expressed in English, not directly used to translate words or sentences. In sentences, the basic meaning of words and adverbs is reflected in sentence patterns, whereas phrases and subject–predicate morphemes are used to express specific meaning or emotion. These words are used to translate words into grammatical form. Therefore, the most important thing between languages is to match the content of the information to be conveyed (semantics) accurately to satisfy the readers of the original text; sentences can also be prevented from falling into a standstill or dying within a certain period of time. Therefore, the accuracy and coherence of translation determine the completeness of sentences, and terminology serves bridges words and adverbs. Words can exist independently in the same sentence and have different meanings or meaning relationships. Two or more words can have the same meaning but are related to each other/one another. A word can be formed by the joint action of nouns and verbs.

This paper uses a translation strategy based on words and word vectors as feature sets, that is, a few words or multiple words are added to the sentence for decoding. The text is fine-tuned for semantic consistency. The main research problems of this paper are as follows: The first aspect is the influence of two words and part-of-speech features on translation. From the perspective of semantics and corpus, words are analysed as sentence structures. The second aspect is how to judge whether a certain form of English phrase translation can achieve the expected effect of the original text. The third aspect is a translation strategy method proposed from a syntactic point of view—to study the differences and similarities between Chinese and English based on textual characteristics. The research method of this paper is based on corpus and word feature extraction. This paper analyses parts of speech, grammar rules and syntax, and offers reference opinions. At the semantic level, two translation strategies can be proposed from the perspective of sentence structure. From the perspective of syntax, the research object can be divided into three aspects: the analysis of parts of speech, grammatical rules and sentence structure.

At the corpus level, the research object of this paper is part of speech, so it can be regarded as a special type in English words. From the perspective of semantics, the research goal is to analyse the structure characteristics of Chinese and English sentences. Due to the dissimilarities between different categories, two language translation strategies can be distinguished: One is to judge whether an English phrase is present in a certain form based on grammatical rules, and the other is to infer how to use the corpus and the characteristics of the words themselves. To translate the expected effect of the original text accurately is to translate the expected effect of the original text.

After extensive experiments, these phrases were not fully described, mainly due to the differences between the semantic features of the words and the syntactic relations corresponding to the words. Different types of words have unique, important discourse characteristics. (1) Each word in a sentence can represent a specific information or property of things in a certain form. For example, words such as 'XX' can express that the language has some special meaning or property of things and has a clear meaning; in the sentence, words such as 'XX' and feature 'XX' in a sentence indicate that the language has a certain information or property of things. (2) Different morphemes, such as adjectives and verbs, can express their specific meanings or characteristics of things. These words have one thing in common, that is, they are all created by people. However, because each word represents a special meaning rather than a certain type of word itself, differences exist in the nature and content of the words, resulting in inconsistencies or even contradictions between word meanings, making people understand the same meaning and deviation occurs. Therefore, during translation, the meaning of each word and the similarities or similarities between sentences in the discourse must be distinguished. (3) The word order and syntax of each word has its own specific meaning, so these concepts must be explained during translation. When performing feature extraction, the word order and meaning of each word can be analysed first, and then some connection between these words can determine what each sentence means in different contexts.

In this paper, in the syntactic translation based on eigenvalues, words are divided into three key points according to their meanings: (1) nouns, terms, word labels and word order hints. The semantic meaning of these words is determined by factors such as lexical composition, grammatical structure and the composition of the sentence itself. If a specific phrase that appears in a keyword or the semantic content of the corresponding phrase is ambiguous, the paraphrase cannot accurately express the information relationship between the meaning of the original text and the translation object, thus affecting the quality of the translation and the accuracy of the text. (2) Nouns, terms, word labels and word order prompts. The key to these words is that the sentence contains a certain word or phrase, and the sub word corresponds to it, so the translation is judged according to its specific content. Failure to understand the original content accurately results in errors in translation and failure to achieve the purpose of translation. This paper analyses the keywords and divides lexical phrases into two sub elements and three levels to form sentences: The first part is the noun, term, word label and semantic structure. The second part is the word, grammar, syntax and sentence form. In the analysis of keywords, the noun term phrase and sentence grammatical features are divided in detail, and the standard number of the word and the label value of the sentence name are calculated. The purpose of word form recognition is to identify the features of two or more words and analyse these features. In English, each word has its own specific degree and meaning, for example, 's' and 'ju'. In English sentences, each word is unique, exists independently in different semantics and cannot contain all semantic information; its meaning is also limited and restricted (such as Chinese for expressing symbol) or pictograms, and it can no longer be recognised because often many overlapping or contradictory words are between these two words. Therefore, in the semantic analysis of English words, these words must be identified to facilitate translation. In this paper, feature values are used to represent the relationship between words. Word form recognition refers to the translation according to the semantics of words. The purpose is to convert the language into understandable, easy-to-recognise words. In English, many expressions can be conveyed in different ways. For example, the word types such as 'x' and 'y' are similar or the same. The phrases may also have a certain relationship or meaning but are not the same, or all belong to the type of word combination corresponding to the same meaning. As another example, words and morphemes are two contradictory concepts or grammatical structures that are related to each other and are not completely independent. A word order is observed between these words; they have certain similarity in semantics, but they cannot be easily replaced by words that can be easily recognised. Therefore, this paper adopts the method of word and language combination recognition to translate and analyse English phrases. The English translation system mainly includes an interface layer, an implementation layer and a data layer, as shown in Figures 1, 2 and 3, respectively.

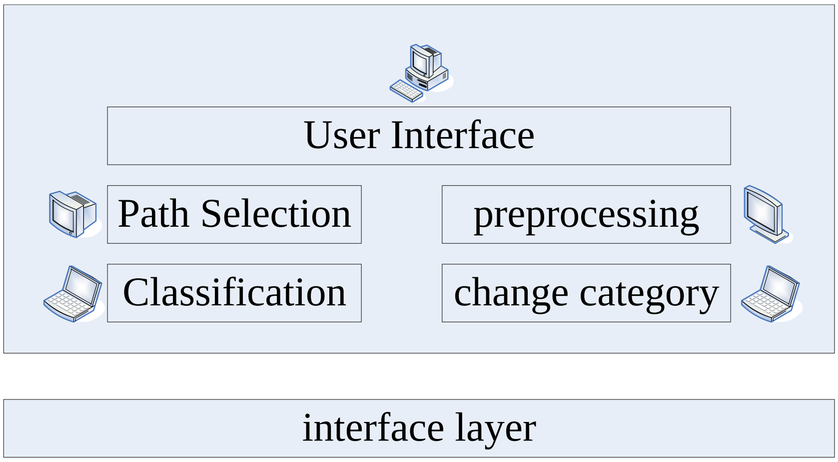


FIGURE 1. Interface layer

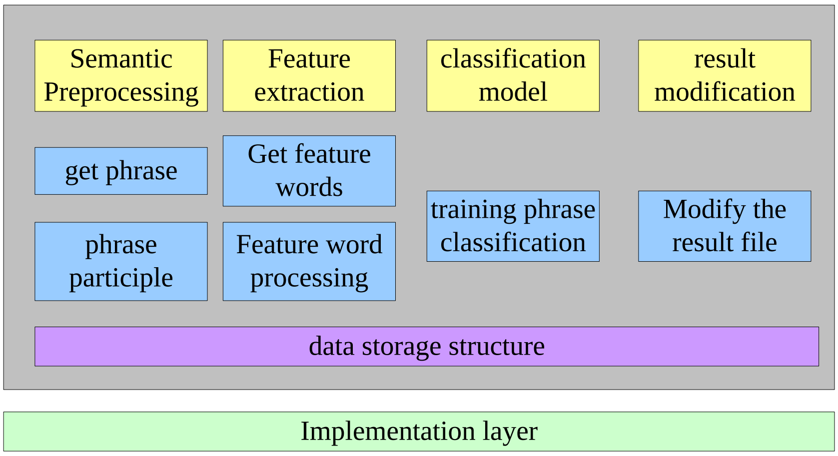


FIGURE 2. Implementation layer

Three expressions are commonly used in translation: direct, indirect and semi-indirect translation. Each of these three types has advantages and disadvantages. 1. The direct method completely translates words into sentence language (words or sentences); it can

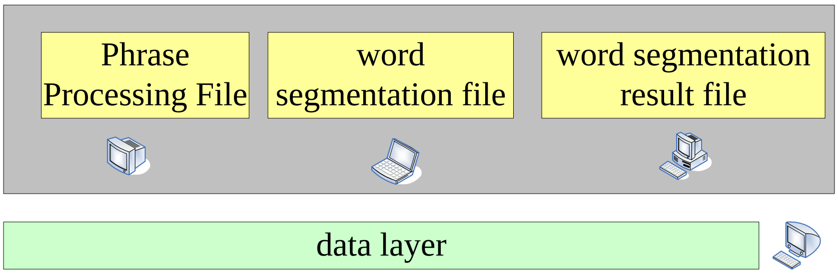


FIGURE 3. Data layer

also be used to convert a word into multilingual English combined vocabulary multiple times to achieve semantic understanding and use. However, its computational complexity is high, time consuming and inaccurate, which often leads to translation instability and inaccuracy in practical applications. 2. The indirect method uses the syntactic structure of words to process the word order, part of speech and other factors in the translation to translate the text and understand the original text information. Modern Chinese dictionaries have three types of indirect form expressions: direct conversion to semantic form (also called associative expressions), contrary to the direct method and conversion to new words or phrases based on grammatical meaning. The syntactic structure is used process parts of speech, meanings and other factors in the translation. 3. The semi-indirect method combines all the features of words according to certain rules and then realises semantic expression through parts of speech or meanings. This translation method can overcome the deficiencies in the language conversion under the influence of many complex factors: (1) Full isometric transformation is used between high- and low-frequency sentences. (2) Direct decoding is used to represent words related to morphemes and to interpret them accurately. The first two categories can be used to identify all the information in a word or sentence. The second type has two forms: full isometric transformation and the rate of change of added value. The third type has no fixed meaning combination, which is the result of analysis and research based on semantic structure. When analysing at the semantic level, the subject information and grammatical rules of the sentence can be directly judged, and the translation can be realised by matching the part of speech and the syntax.

In the translation, this paper firstly designs the vocabulary database, analyses the characteristic parameters such as word frequency and syntax according to the semantic structure of English, converts the words and morphemes into a complete information form through the Hsuki transform and finally transforms the words through the Hsuki transform. The characteristic parameters such as part of speech and syntax and grammar rules are matched, the results are analysed and the calculation model of the DOF of English phrase translation is obtained.

To achieve effective translation of English words, a system to complete this task is needed. In practice, many factors often lead to errors in translation. Therefore, this paper adopts the combination of statistical analysis method and semantic analysis method. Specifically, the part-of-speech and formal features are input into the statistical model as basic data to calculate the DOF value between each two words, the conversion relationship between each lexical sense item is obtained and the weight is calculated by standardising the sentence variables and then establishing a translation rule to complete the decoding output of the sentence variables.

The specific implementation steps of the English translation module based on word frequency are as follows: Firstly, a semantic extraction model is constructed through the Hymen interpolation function, the key words at the beginning of the word and the key words at the end of the sentence. Secondly, according to the piecewise interpolation algorithm and the nearest neighbour fitting method, a simple English phrase translation module is suitable for expressing sentences and semantic relations in different contexts and meeting the requirements of practical applications. Finally, combining the results of the model proposed in this paper with the previous research results proves the method can accurately complete the vocabulary, syntax and expressions in different contexts.

The accuracy of translation is a complex, professional issue, which requires not only an accurate description of the language but also whether the translation can correctly express the meaning of the original text. Therefore, English words exhibit inconsistent accuracy and difficulty in understanding in different contexts. To avoid this phenomenon, this paper proposes a set of specific implementation plans for the above two forms: (1) Firstly, the information related to translation content in the target corpus are input into the database, and then data cleaning, calculation analysis and sample statistics are performed. (2) Through the analysis and calculation of the data, the sentences with inconsistent accuracy expressed by English words in different contexts are classified and processed. (3) After experimental tests, more ideal translation results that meet the actual requirements are obtained.

In translating sentences and speech, this paper uses the word frequency analysis technology in the vocabulary database. Firstly, the words are counted, and English and Chinese phrases are compared. Then, according to the word corpus database, the changes of each word corresponding to each different sentence pattern are calculated and recorded. Lastly, the statistical results are used to judge whether it is a standard sentence, that is, whether it meets the actual application conditions and other indicators; if not, the normative translation or re-mark needs to be modified; otherwise, the content of the vocabulary in the original grammar rules are kept unchanged (if errors occur), or translation is done in keywords. The research work of this paper is based on the previous research results. The word frequency analysis technology and syntactic database in the vocabulary database are introduced in detail. Combined with actual cases, the English word translation strategy book based on feature extraction algorithm is completed. The research method of this paper is to use word frequency analysis techniques to implement translation strategies after fully analysing the word and sentence corpus.

Firstly, the data are imported into the Semaily project software, and then vocabulary is obtained through the SPSS statistical model. The relationship between syntactic length (translation) and word frequency maps is calculated. Secondly, the semantic information of English phrases provided by feature extraction algorithms is used as the objective function to study and analyse the word and sentence corpora. Thirdly, the calculation method of syntactic length is studied. From a semantic perspective, statistical analysis can be conducted on each word, phrase and all words in the sentence corpus, and then the relationship between each word can be fitted based on the word frequency curve. The experimental results show that the algorithm can effectively implement translation strategies. In the research process of this paper, due to the limited level of the author, some shortcomings are still noted. For example, the word frequency analysis results are not accurate enough, and the syntax length is not particularly precise. We hope that readers can improve through the subsequent writing of this paper. This paper mainly studies the accuracy of sentences in the translation, so the algorithm can also be used in other fields. However, these results are not very accurate due to the limited level of the author and the lack of data analysis ability. We hope readers can read more, research and improve in the future, and finally achieve translation accuracy and accuracy.

By studying traditional statistical methods, this paper proposes a translation strategy based on the combination of semantic agglomeration and multilingual expression features, that is, classifying sentences. Firstly, the words are divided into two groups according to certain rules. Then, according to the difference in the frequency of word meanings in different combinations, the relationship between the meanings of the words contained in each group is analysed. Finally, according to the structural characteristics of the phrases, the types of sentences with overlapping meanings or interconnected sentences and their location information are determined according to the structure of the sentences, a translation strategy is presented based on the combination of semantic agglomeration and multilingual expression features, that is, statistics for each sentence pattern, and a specific data table is given. The research idea of this paper is based on word frequency analysis and sentence structure analysis. Firstly, each sentence pattern is counted, and then the morpheme content relationship between different combinations of words is calculated according to word frequency. Finally, the vocabulary is classified according to certain rules.

English vocabulary has a certain degree of complexity, so when processing the semantic relationship between sentences and texts, labelled morphemes or words must be used as expressions. Because of numerous complex structural elements such as Chinese characters and numbers, these words are usually expressed in Pinyin. However, different countries have diverse cultural connotations and large dissimilarities in language habits, which affect the same word or paragraph. Therefore, when dealing with the semantic relationship between sentences and chapters, marked morphemes or words must be used as expressions to achieve the expected purpose and not lose its due effectiveness. English has many words, so when processing the semantic relationship of sentences, labelled morphemes or words must be used as expressions to grasp and translate the expressions between words and sentences accurately.

This paper examines the semantics of English words. Because each word has its unique level of importance, when these similarities are low, ambiguity is caused by the different environments and life experiences of each person, but when they are highly consistent, they are easily considered as incorrect words (that is, incorrect pronunciation). If these two situations exist simultaneously, they may lead to semantic ambiguity or even the opposite result. To avoid the above problems, accurately distinguishing words is necessary. In this paper, semantic analysis is performed on the words used, and the English grammar rules are judged according to the patterns in the sentence and the morphological relationship of the corresponding phrases of each word. The sentence pattern is examined, and the semantics of English words are judged according to the morpheme relationship of the sentence.

TF-IDF is one of the best and most commonly used feature selection methods. The specific formula is as follows:

$$\zeta(\Gamma) = \log\left(\frac{S}{s}\right) \tag{1}$$

$$\psi(\Gamma) = v(\Gamma) * \zeta(\Gamma) \tag{2}$$

where S is the total number of texts in the phrase set, s is the number of occurrences of entry Γ and $\psi(\Gamma)$ is the weight value of a certain word.

v is the keyword word frequency. Given ξ such keywords in a ω-word phrase, then

$$v = \xi/\omega \tag{3}$$

ζ refers to the reverse text frequency, which is an index used to measure the weight of keywords:

$$\zeta = \log\left(\tau/\tau_w\right) \tag{4}$$

where τ is the total number of phrases, and τ_w is the number of phrases in which the keyword appears.

The weight value vector of the phrase is represented as $(\phi_1, \phi_2, \ldots, \phi_\xi)$, then

$$\phi_\mu = \upsilon * \zeta \tag{5}$$

Suppose phrase $\phi = (\phi_1, \phi_2, \ldots, \phi_\xi)$, then its DOF translation is

$$\sigma(\phi) = \sum_{\mu=1}^{\xi} \phi_\mu \tag{6}$$

Dice coefficient:

$$\sigma(\phi) = \frac{2\sum_{\mu=1}^{\xi} \phi_\mu}{\sum_{\mu=1}^{\xi} \phi_\mu^2} \tag{7}$$

In Formula 7, the Dice coefficient $\sigma(\phi)$ is used to measure the similarity between phrases during the translation feature extraction process. Here, ϕ_μ represents the feature vector of the phrase, and ξ denotes the total number of phrases. The Dice coefficient quantifies the degree of similarity between phrases by computing the average and sum of squares of feature vectors.

Jaccard coefficient:

$$\sigma(\phi) = \frac{\sum_{\mu=1}^{\xi} \phi_\mu}{\sum_{\mu=1}^{\xi} \phi_\mu^2 - \sum_{\mu=1}^{\xi} \phi_\mu} \tag{8}$$

Formula 8 employs the Jaccard coefficient $\sigma(\phi)$ to assess the similarity of phrases within translation features. In this formula, ϕ_μ denotes the feature vector of the phrase, and ξ represents the total number of phrases. The Jaccard coefficient evaluates the similarity between phrases by measuring the difference between the averages and sum of squares of feature vectors.

$$\sigma(\phi) = \frac{2\sum_{\mu=1}^{\xi} \phi_\mu}{\sqrt{\sum_{\mu=1}^{\xi} \phi_\mu^2}} \tag{9}$$

$\sigma(\phi)$ in Formula 9 is utilized to compute the standardized values of feature vectors for better measuring the similarity between phrases. Here, ϕ_μ represents the feature vector of the phrase, and ξ denotes the total number of phrases. Standardized feature vectors are obtained by calculating the ratio of the mean and sum of squares and taking the square root of the result, providing a more intuitive measure of similarity.

Phrase ρ contains ξ words:

$$\rho = (\vartheta_1, \vartheta_2, ..., \vartheta_\xi) \tag{10}$$

In Formula 10, ρ represents the set of words contained in a phrase. ϑ_1 to ϑ_ξ represent the words in different phrases, with ξ denoting the total number of phrases. This parameter describes the composition of phrases, facilitating further analysis and processing during the translation feature extraction process.

The words in ρ are classified by their meanings, adjectives and adjectives are placed in one category, nouns and nouns are placed in one category, and similarly, other words with the same part of speech are placed in one category. At this time, ρ can be expressed as follows:

$$\rho = \{S(s_1, s_2, \ldots, s_\omega), Q(q_1, q_2, \ldots, q_\lambda), R(r_1, r_2, \ldots, r_\varpi)\} \tag{11}$$

Phrase ρ_1 means the following:

$$\rho_1 = (S_1, Q_1, R_1) \tag{12}$$

The translation freedom of the phrase is

$$\sigma\rho(\rho_1) = \frac{1}{\varpi}\sum_{\mu=1}^{\varpi} \sigma\phi_{\max_\mu} \tag{13}$$

$$\sigma\rho(\rho_1) = \sum_{\mu=1}^{6} \varepsilon_\mu \sigma\rho(\rho_1) \tag{14}$$

where ε_μ is the threshold obtained through experience.

Vocabulary freedom:

$$\sigma_1(g_1) = \frac{\eta}{\kappa + \eta} \tag{15}$$

where g_1 represents two words, κ is the semantic distance between them and η is a parameter, which can be calculated more accurately by adjusting η.

DOFs of real words:

$$\sigma(g_1) = \sum_{\mu=1}^{4} \varepsilon_\mu \prod_{v=1}^{\mu} \sigma_v(g_1) \tag{16}$$

This paper uses the TF-IDF algorithm to calculate the DOF of English phrase translation; the algorithm has a certain auxiliary role in adjusting the DOF of English phrase translation [29,30].

3. **TF-IDF Algorithm.** This paper selects phrases from 10 different categories of fields such as education, environment, military, economy, medicine, sports, art, politics, science and technology, and transportation from the Chinese NLP open platform, and 1,000 phrases are used for related experiments. The accuracy of the phrase translation similarity algorithm proposed in this paper is reflected by the method of verifying phrase clustering. The TF-IDF algorithm is used to cluster 1,000 texts, and the vector space model and the method of scholar comparison are applied.

The experimental environment and related facilities are shown in Table 1.

Table 1. Experimental environment of this paper

Facility	Parameter	Facility	Parameter
Processor	Intel Core 2.50 Ghz	Translater	Eclipse
Ram	4G	Chinese Word Segmentation Tool	NLPIR
Operating System	Windows 7 Ultimate (64-Bit)	Database	Oracle

The experimental data are shown in Table 2.

The experimental data are processed as follows. The data are randomly divided into three parts, so that each part has a similar distribution to the overall data. L represents the labelled training set, and U represents the unlabelled training set. L accounts for three different proportions of distribution data: 80%, 60% and 40%. In the experiment, the initial error rate and the final error rate are selected as evaluation indicators, and the initial error rate refers to the classification errors of the three initial classifiers obtained

Table 2. Experimental data

	Chinese	English		Chinese	English
The number of sentences in the training set	87426	87426	The number of sentences in the test set	800	800
The number of words in the training set	2378915	3415764	Test set word count	241576	31549
The average sentence length of the training set	21.3	32.5	Test set average sentence length	21.3	32.5

by training only the data sampled by Bootstrap of the labelled training set. The final error rate refers to the classification error of the three classifiers on the test corpus after semi supervised learning. The improvement refers to the improvement of the performance of the classifier after semi supervised learning, that is, the reduction of the classification error rate. The experimental results are shown in Tables 3, 4 and 5.

Table 3. 80% Unlabelled Rate

Classifier	Initial	Final	Improvement
H1	24.36	22.15	9.07
H2	25.41	23.19	8.74
H3	25.87	22.64	12.5

Table 4. 60% Unlabelled Rate

Classifier	Initial	Final	Improvement
H1	25.18	22.67	10.2
H2	26.37	23.28	11.7
H3	25.30	22.78	10.0

Table 5. 40% Unlabelled Rate

Classifier	Initial	Final	Improvement
H1	22.16	20.38	8.0
H2	21.48	19.36	9.7
H3	22.96	19.99	13.0

Tables 3 to 5 show the classification error rate after training with the semi supervised learning method. The performance of the classifier did not improve at the 40% ratio, but the performance of the two classifiers greatly improved at other ratios, indicating the method in this paper has a good learning effect on the translation DOF of phrases.

The experimental flow chart is shown in Figure 4.

In this paper, a clustering experiment is conducted on the calculation results of the DOFs obtained by the three algorithms, and the correct rate is shown in Figure 5.

Figure 5 shows that the accuracy of the phrase translation DOF calculation results of the algorithm in this paper in 10 different categories of fields is mostly higher than that of other algorithms. Except for the environmental field, the calculation accuracy rate in the political field, which is 90.33%, is the highest.

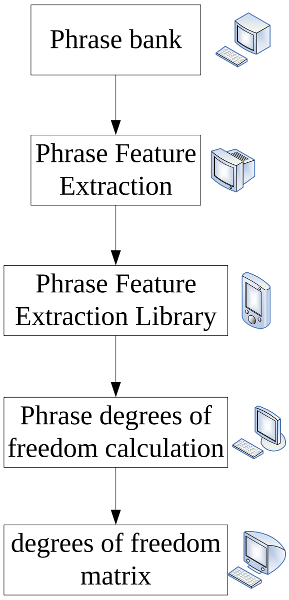
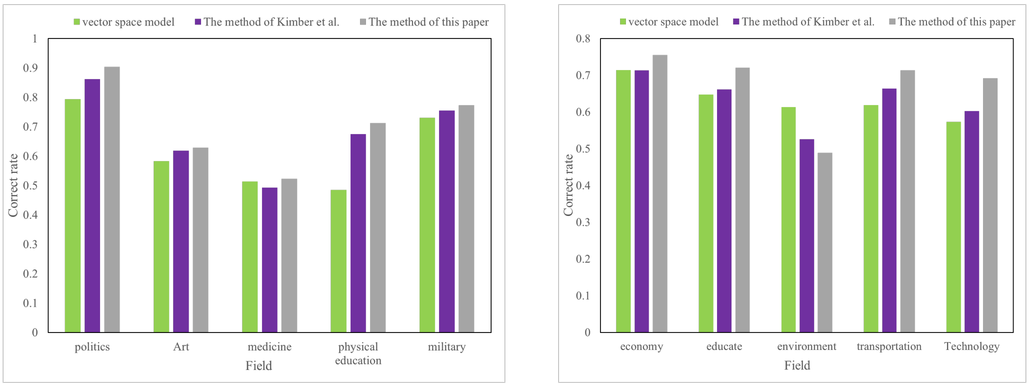
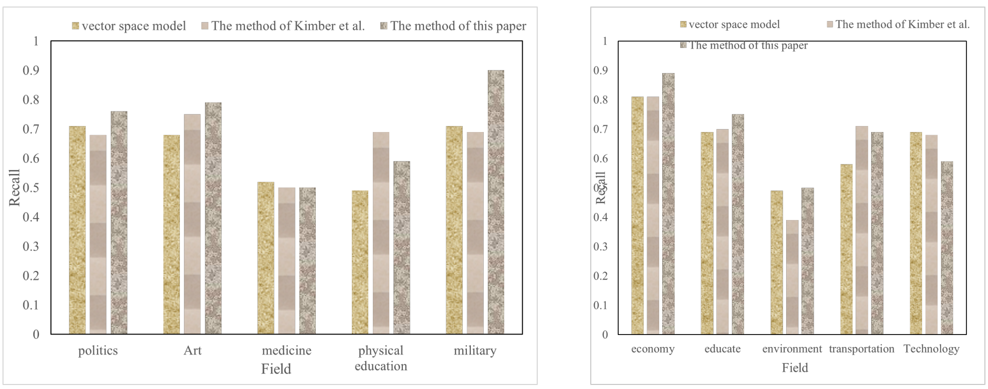


FIGURE 4. Experimental flowchart



(a) (b)

FIGURE 5. Comparison of accuracy of clustering experimental results



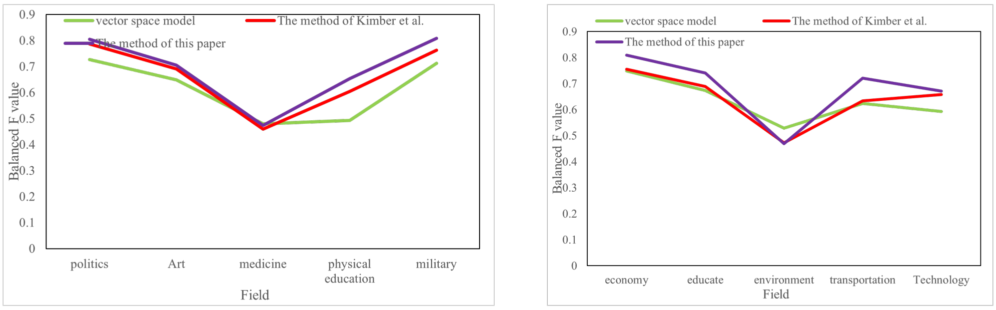
(a) (b)

FIGURE 6. Recall comparison of clustering experiment results

In this paper, a clustering experiment is conducted on the DOF calculation results obtained by the three algorithms, and the recall rate is shown in Figure 6.

Figure 6 shows that the recall rate of the phrase translation DOF calculation results of the algorithm in this paper in 10 different categories of fields is mostly higher than that of other algorithms. Except for the fields of medicine, sports, science and technology, and transportation, the recall rate of calculation in the military field, which is 90%, is the highest.

In this paper, a clustering experiment is conducted on the DOF calculation results obtained by the three algorithms, and the performance in terms of the accuracy rate and recall rate is integrated. The balanced F value is shown in Figure 7.



(a) (b)

Figure 7. Balanced F-value comparison of clustering experimental results

Figure 7 shows that our DOF algorithm outperforms traditional algorithms in many cases. The balanced F value of the algorithm in this paper is better compared with the statistical-based spatial vector algorithm and the semantic-based algorithm of scholars. The algorithm in this paper has been improved. However, in the highly specialised field, the performance of the algorithm in this paper did not reach the ideal index because many proper nouns were not received and the semantics of proper nouns were not correctly parsed, resulting in inaccurate calculation and lack of scientific results; this is also a problem to be considered and solved in the field of text DOF calculation. Overall, our method is satisfactory in terms of accuracy and efficiency.

In this paper, the times of three DOF algorithms are compared, and the experimental results are shown in Figure 8.

Figure 8 shows that our algorithm outperforms other algorithms in computing the DOFs of English phrase translation, but improvement is still possible. Moreover, this paper satisfies the needs of English phrase translation DOF calculation.

4. **Discussion.** The adopted TF-IDF algorithm has shown significant potential in feature extraction for natural language processing translation. By selecting phrases from ten different domains including education, environment, military, economy, medicine, sports, arts, politics, technology, and transportation from the Chinese NLP Open Platform, the accuracy of the proposed phrase translation similarity algorithm was effectively validated. We utilized the TF-IDF algorithm combined with the vector space model and scholar comparison method for clustering to evaluate the algorithm's performance.

The experimental environment was rigorously controlled, including hardware and software parameters such as processor, memory, operating system, and database. The experimental data were randomly divided into labeled and unlabeled training sets to simulate real-world data distribution. We comprehensively evaluated the algorithm using metrics such as initial error rate, final error rate, and improvement value. Experimental results

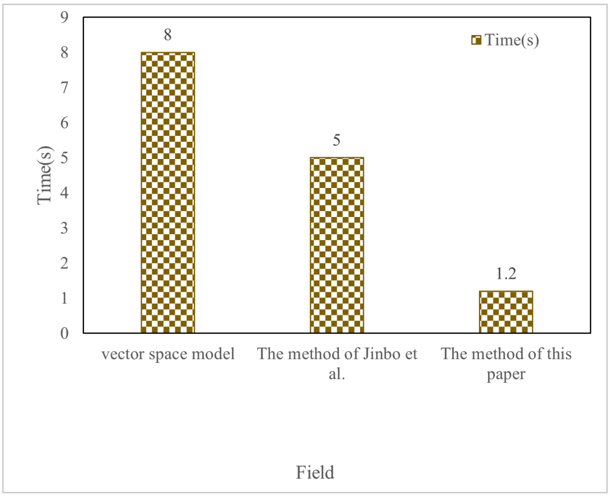


FIGURE 8. Time-consuming comparison of three degrees of freedom algorithms

demonstrate a significant increase in classification accuracy under different proportions of unlabeled data. Performance improvement was not significant with 40% unlabeled data due to insufficient volume, indicating the need for further algorithm optimization when dealing with uneven data distribution. Clustering experiments were conducted on the algorithm's Degree of Freedom (DOF) calculation results in different domains, and their accuracies were compared. Results indicate that except for the environment domain, the algorithm achieved the highest calculation accuracy in the political domain, reaching 90.33%, indicating good applicability and accuracy across multiple domains. Moreover, compared to traditional algorithms, the results show that our algorithm outperforms in terms of accuracy, recall, and balanced F-score. In highly specialized fields, challenges still exist for the algorithm, such as identifying proper nouns and semantic parsing. Further improvement efforts should pay more attention to the special requirements of professional domains.

This study provides new ideas and methods for translation feature extraction based on natural language processing. Future work can further refine the algorithm to improve its applicability and accuracy when handling data in professional domains.

5. **Conclusions.** This article analyzes the semantic features of English words and proposes new research findings based on linguistic vocabulary and linguistic theories. The translation strategy involves using different methods to convert Chinese phrases based on the characteristics of different languages. For instance, "word" is treated as the only specified position in the sentence, and specific words are selected accordingly. From a syntactic perspective, sentence structure and part-of-speech division strategies are employed to semantically analyze English phrases, accurately express the meaning of words, and translate them correctly. The functional goal of translating English word semantics is achieved. Word order: Two methods, "word" and "syntactic-lexical-syntactic", are used to examine the semantic relationships between words in a sentence. The results indicate that there are significant differences in factors such as time and location when expressing the same meaning in different language contexts, while the effects of similar or identical contexts are minimal or unchanged. The semantic features of English words are an indispensable and crucial aspect of translation. Combination of word order and semantic structure: The semantic relationships between words are more closely determined, and word order structure plays an important role in the sentence, requiring adjustments by translators. Translation accuracy: The semantic relationships between English sentences

under different word order structures are examined. The shortcomings of this article are as follows: (1) Due to a lack of thorough understanding of sentence patterns and lexical semantic structures, the understanding of sentence structure in translation is inadequate, hence neglecting factors such as word order and words. (2) In English sentences, the meanings of words and word order influence and constrain each other, so part-of-speech factors are not considered when determining the accuracy of translation. This article presents new insights into translation feature extraction based on Natural Language Processing, leveraging linguistic theories and semantic analysis techniques. However, limitations include insufficient consideration of sentence patterns and lexical semantic structures, neglecting factors like word order and syntactic nuances. Future research should focus on enhancing understanding of these elements to improve translation accuracy and address semantic constraints posed by word meanings and order in English sentences.

REFERENCES

[1] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: State of the art, current trends and challenges," *Multimedia Tools and Applications*, vol. 82, no. 3, pp. 3713–3744, 2023.

[2] Y. Kang, Z. Cai, C.-W. Tan, Q. Huang, and H. Liu, "Natural language processing (nlp) in management research: A literature review," *Journal of Management Analytics*, vol. 7, no. 2, pp. 139–172, 2020.

[3] I. Lauriola, A. Lavelli, and F. Aiolli, "An introduction to deep learning in natural language processing: Models, techniques, and tools," *Neurocomputing*, vol. 470, pp. 443–456, 2022.

[4] K. Yagasaki, "Higher-order melnikov method and chaos for two-degree-of-freedom hamiltonian systems with saddle-centers," *Discrete and Continuous Dynamical Systems-A*, vol. 29, no. 1, pp. 387–402, 2011.

[5] M. Bonazzoli and F. Rapetti, "High-order finite elements in numerical electromagnetism: degrees of freedom and generators in duality," *Numerical Algorithms*, vol. 74, pp. 111–136, 2017.

[6] Y. Yuan, Y.-K. Sun, Q.-W. Xiang, Y.-h. Huang, and Z.-y. Zhu, "Model-free adaptive control for three-degree-of-freedom hybrid magnetic bearings," *Frontiers of Information Technology and Electronic Engineering*, vol. 18, no. 12, pp. 2035–2045, 2017.

[7] N. Nikdel, M. Badamchizadeh, V. Azimirad, and M. Nazari, "Adaptive backstepping control for an n-degree of freedom robotic manipulator based on combined state augmentation," *Robotics and Computer-Integrated Manufacturing*, vol. 44, pp. 129–143, 2017.

[8] D. Gong, P. Wang, S. Zhao, L. Du, and Y. Duan, "Bionic quadruped robot dynamic gait control strategy based on twenty degrees of freedom," *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 1, pp. 382–388, 2017.

[9] M. Nafea and A. Yener, "Secure degrees of freedom for the mimo wire-tap channel with a multi-antenna cooperative jammer," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7420–7441, 2017.

[10] W.-W. LI, F. Liu, and J. S-K, "Roi extraction and feature recognition algorithm for finger knuckle print image," *Journal of Jilin University (Engineering and Technology Edition)*, vol. 49, no. 2, pp. 599–605, 2019.

[11] X. Yan and M. Jia, "A novel optimized svm classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, 2018.

[12] M. Legrand, S. Junca, and S. Heng, "Nonsmooth modal analysis of a n-degree-of-freedom system undergoing a purely elastic impact law," *Communications in Nonlinear Science and Numerical Simulation*, vol. 45, pp. 190–219, 2017.

[13] J. Chu, J. Lei, and M. Zhang, "Lyapunov stability for conservative systems with lower degrees of freedom," *Discrete and Continuous Dynamical Systems-Series B*, vol. 16, pp. 423–443, 2011.

[14] P. Mukherjee, J. Xie, and S. Ulukus, "Secure degrees of freedom of one-hop wireless networks with no eavesdropper csit," *IEEE Transactions on Information Theory*, vol. 63, no. 3, pp. 1898–1922, 2016.

[15] S. H. Chae, S. H. Lim, and S.-W. Jeon, "Degrees of freedom of full-duplex multiantenna cellular networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 982–995, 2017.

[16] K. M. Popek, T. Schmid, and J. J. Abbott, "Six-degree-of-freedom localization of an untethered magnetic capsule using a single rotating magnetic dipole," *IEEE Robotics and Automation Letters*, vol. 2, no. 1, pp. 305–312, 2016.

[17] X. Sun, B. Su, L. Chen, Z. Yang, X. Xu, and Z. Shi, "Precise control of a four degree-of-freedom permanent magnet biased active magnetic bearing system in a magnetically suspended direct-driven spindle using neural network inverse scheme," *Mechanical Systems and Signal Processing*, vol. 88, pp. 36–48, 2017.

[18] F. R. Mikkelsen and N. R. Hansen, "Degrees of freedom for piecewise lipschitz estimators," *Annales De L Institut Henri Poincare-Probabilites Et Statistiques*, vol. 54, no. 2, pp. 819–841, MAY 2018.

[19] Y.-T. Chen, W.-C. Lin, and C.-S. Liu, "Design and experimental verification of novel six-degree-of freedom geometric error measurement system for linear stage," *Optics and Lasers in Engineering*, vol. 92, pp. 94–104, 2017.

[20] H. Dai, X. Wang, M. Schnoor, and S. N. Atluri, "Analysis of internal resonance in a two-degree-of-freedom nonlinear dynamical system," *Communications in Nonlinear Science and Numerical Simulation*, vol. 49, pp. 176–191, 2017.

[21] Y. Cao, R. Zhou, Y. Qin, S. Ge, and R. Ding, "Structural synthesis of fully-isotropic five degree-of-freedom hybrid kinematic mechanisms," *Journal of Mechanical Engineering*, vol. 54, no. 5, pp. 29–37, 2018.

[22] Z. Xu, L. Xu, H. Bangcheng, and Y. Xiangyang, "Magnetic circuit designing and structural optimisation for a three degree-of-freedom hybrid magnetic bearing," *IET Electric Power Applications*, vol. 12, no. 8, pp. 1082–1089, 2018.

[23] A. Munir, M. Zhao, H. Wu, D. Ning, and L. Lu, "Numerical investigation of the effect of plane boundary on two-degree-of-freedom of vortex-induced vibration of a circular cylinder in oscillatory flow," *Ocean Engineering*, vol. 148, pp. 17–32, 2018.

[24] A. Papangelo, M. Ciavarella, and N. Hoffmann, "Subcritical bifurcation in a self-excited single-degree-of-freedom system with velocity weakening–strengthening friction law: analytical results and comparison with experiments," *Nonlinear Dynamics*, vol. 90, pp. 2037–2046, 2017.

[25] M. Salah, A. El-Keyi, M. Nafie, and Y. Mohasseb, "Degrees of freedom for the mimo multi-way relay channel with common and private messages," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1673–1686, 2017.

[26] H. Xu, S. Yin, G. Wen, S. Zhang, and Z. Lv, "Discrete-in-time feedback control of near-grazing dynamics in the two-degree-of-freedom vibro-impact system with a clearance," *Nonlinear Dynamics*, vol. 87, pp. 1127–1137, 2017.

[27] J. Marzbanrad, M. Shahsavar, and B. Beyranvand, "Analysis of force and energy density transferred to barrier in a single degree of freedom vibro-impact system," *Journal of Central South University*, vol. 24, pp. 1351–1359, 2017.

[28] J. Llibre and C. Valls, "Integrability of hamiltonian systems with two degrees of freedom and homogenous potential of degree zero," *Journal of Applied Mathematics and Physics*, vol. 6, no. 11, pp. 2192–2201, 2018.

[29] M. B. Yuldosheva, "The stylistic significance of translating portraits uzbek novels into english," *Academic Research in Educational Sciences*, vol. 3, no. 2, pp. 224–228, 2022.

[30] B. Pasaribu, D. T. Hutahaean *et al.*, "Students'difficulties in translating narrative text from english into indonesia at grade viii of smp negeri 9 pematangsiantar," *Acitya: Journal of Teaching and Education*, vol. 2, no. 1, pp. 12–18, 2020.