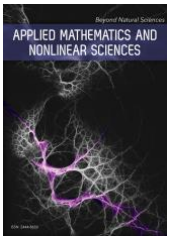

Applied Mathematics and Nonlinear Sciences

https://www.sciendo.com

Evaluation of Intelligent Proofreading Effect of English Translation of Lingnan Hakka Dialect Based on Dependency Syntactic Networks

Guiying Kong[1,†]

1. Xijiang River Valley Folk Literature Research Center, Wuzhou University, Wuzhou 543002, China.

Submission Info

Communicated by Z. Sabir
Received September 27, 2024
Accepted January 8, 2025
Available online February 27, 2025

Abstract

This study evaluates the effect of intelligent proofreading of English translation of Lingnan Hakka dialect based on dependent syntactic networks. Dependency syntactic networks help identify and correct common errors in translation by parsing the syntactic structure and semantic relations of dialect sentences. The evaluation results show that the use of dependent syntactic networks significantly improves the syntactic accuracy and semantic consistency of translation, especially when dealing with complex sentences and dialect-specific expressions. The method not only improves the quality of machine translation, but also enhances the natural fluency of the translation, providing effective technical support for intelligent English translation proofreading. The results of the study prove that the application of dependent syntactic network in dialect translation has a wide range of prospects.

Keywords: dependent syntax, Lingnan Hakka dialect, dialect English translation, intelligent English translation, proofreading effect
AMS 2020 codes: 91C05

† Corresponding Author.
Email address: camelliared1973@163.com

ISSN 2444-8656
sciendo
DOI: https://doi.org/10.2478/amns-2025-0106

1 Introduction

Lingnan Hakka dialect has attracted many linguists to study it for its unique language structure and rich cultural connotation, but the complex syntactic dependencies in it pose a challenge to the English translation of Lingnan Hakka dialect. The development of intelligent technology for English translation based on the dependency syntactic network can provide in-depth analysis of the complex syntactic structures in the dialect, which significantly improves the quality of English translation. However, there is a lack of systematic evaluation of the application effect comparison model for this technology. By constructing an experimental model and conducting an in-depth study and analysis of the intelligent effect of English translation in Lingnan Hakka dialect, this paper aims to explore the role of dependent syntactic networks in improving translation accuracy, fluency and semantic consistency, so as to provide scientific basis for the optimisation and promotion of related technologies.

In recent years, there have been more studies on dependent syntax at home and abroad. Aiming at the problem that existing deep learning models are difficult to extract the rich semantic information of online comments and thus difficult to accurately extract text sentiment, Xia Jiali et al [1] proposed a multi-feature multiple fusion sentiment classification model based on syntactic dependency and attention mechanism. Experimental results show that the model performs well for the task of online comment sentiment classification. Xie Xuemei et al [2] extracted product features and user sentiment viewpoints paired with them from reviews by manually setting extraction rules based on dependency syntactic analysis and a lexicon of product features, and quantified the viewpoints using sentiment computation in order to obtain user experience features. Dong Sven et al [3] extracted semantic words from the text around the expressions by dependent syntax based on FDS parsing expressions. The experimental results show that dependent syntax can effectively extract the semantic words of mathematical expressions. Yuling Zhang [4] proposed a relationship-aware graph-based convolutional network model to address the problem of ignoring the type of dependency relationships and the presence of noise in the text. Li Xiao [5] combines data augmentation strategy and corpus to train data for syntactic error generation, correction and detection model. The experimental results demonstrate that the syntax error model based on data augmentation strategy has high detection accuracy. Cui Xuran et al [6] proposed a sentiment analysis model based on BERT and dependent syntax to address the problem that existing models do not make sufficient use of syntactic trees, which leads to the inability to accurately understand the semantics of text.

With the depth of research, the potential of dependent syntax in domestic language annotation system and other aspects gradually attracted the attention of scholars. Zeren Drolma et al [7] formulated a set of Tibetan dependent syntax annotation system by referring to a variety of domestic and international dependent syntax annotation specifications and combining Tibetan grammar theory and language typology features. Guo Wenjing [8] investigated the dependency bias and syntactic complexity of Chinese-English relational clauses based on the dependency syntax theory, using the average dependency distance as a measurement index, and analysed the deeper motivation of the differences in the dependency bias and processing difficulty of Chinese-English relational clauses from the cognitive perspective. Qian Long et al [9] studied the syntactic complexity of Chinese learners' second language writing based on the dependency grammar as a theoretical guide and based on the dependency treebank of Chinese mediants.

Hakka dialect is a special dialect system in the Lingnan cultural system [10]. At present, there are fewer studies on intelligent proofreading of English translation for Lingnan Hakka dialect. Referring to the research results of intelligent proofreading system for machine translation [11-14], the research on intelligent proofreading of English translation for Lingnan Hakka dialect can not only fill the gaps at home and abroad, but also better serve the propaganda of Lingnan culture, which is conducive to the enhancement of cultural ties with overseas Hakka people. To address this issue, this study uses the dependency syntactic network to analyse the results of intelligent English translation of Lingnan Hakka dialect, to unravel the linguistic structure of the dialect sentences and the relationship between semantics, and to realise intelligent proofreading of Lingnan Hakka dialect by identifying the common grammatical errors and semantic deviations in the intelligent translation system.

2 The important embodiment of dependent syntactic network analysis in Lingnan Hakka dialect system

In dialectal English translation, language dependency is not only an important reference for the translation system to make translation decisions, but also an important index for readers to quickly understand the meaning of dialectal utterances and grasp the concepts of the text. Therefore, how to grasp and analyse the dependency syntax more accurately in dialectal English translation in order to dig out useful information is extremely important for the standardized English translation of Lingnan Hakka dialect.

In translation linguistics, the concept of dependent syntax refers to the relationship between connected words in the syntactic structure, which is the basis of language production and understanding. In the process of English translation of dialects, it is very important to understand and master the dependent syntactic relations of dialectal utterances, which determine the structural correctness of the translated sentences and the accuracy of the concepts expressed. However, the Lingnan Hakka dialect, as a language system with strong local and pronunciation uniqueness, has a syntactic structure that is very different from that of standard discourse. Therefore, it is particularly important to grasp and be familiar with the concept of dependent syntax in order to maintain a high level of accuracy in the English translation of Lingnan Hakka dialect. In the following, the importance of dependent syntax in the translation system of Lingnan Hakka dialect will be discussed and analysed in detail through specific examples.

2.1 Explanation of the concept of dependent syntax

Dependency syntax refers to the technique of revealing the interdependence between components within a language and its units. Dependency syntax provides a theoretical basis for the understanding of some unfamiliar languages and the processing of natural language through the in-depth analysis of some grammatical components in a sentence. In the process of analysing the language by using dependent syntax, it is not necessary to pay attention to some complicated phrase components in the language, but only need to pay attention to the meaning of the special words themselves and the dependency relationship between the words. An example of a typical dependency syntax is shown in Figure 1.

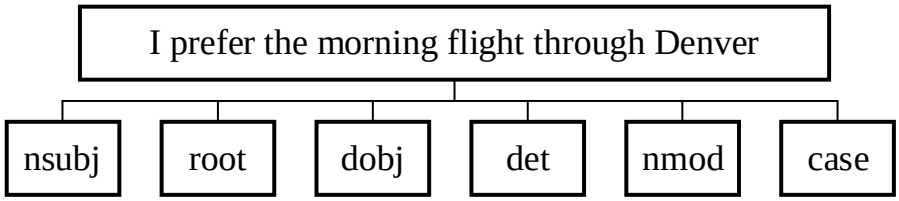

Figure 1. Dependency syntax case analysis

As can be seen from Figure 1, the central word of the example sentence is the verb prefer, and analyzing the overall sentence structure, the central verb is dependent on the subject word (nsub) "I" and the direct object structure of the sentence (dobj) flight. while the direct object structure of the sentence flight is dependent on the the definite article (det) "the" and (nmod) "morning", "Denvor" conjunction "through". The verb "prefer" exists independently and has no dependency on any word, but in the formed language it is necessary to construct a "root", which is dependent on the verb "prefer".

The syntactic structure of dependent syntax is shown in Figure 2.

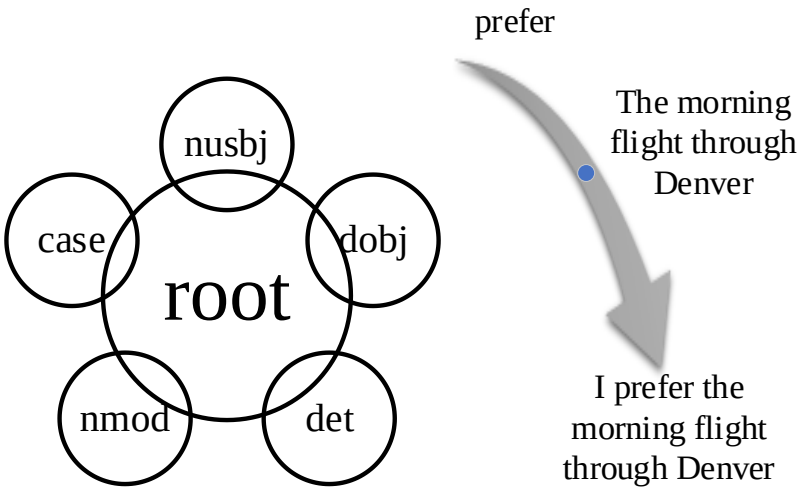

Figure 2. Syntactic structure composition diagram

In contrast to constituent syntax, dependent syntactic network analysis is a detailed analysis of sentence structure directly from the main components of the sentence such as subject and predicate. In addition, the results of the dependency syntactic network analysis are more direct in analyzing the relationship between words. As shown in Figure 2, the example of this dependency syntactic network analysis diagram, the direct object of the central verb prefer in the overall sentence structure is flight, and there is a strong dependency relationship between prefer and flight in the dependency syntactic network analysis tree.

The tree analysis of the sentence is shown in Figure 2. From Figure 2, it can be seen that there is a tertiary dependency, but this tertiary dependency between prefer and flight does not exist if analyzed by constituent syntax.

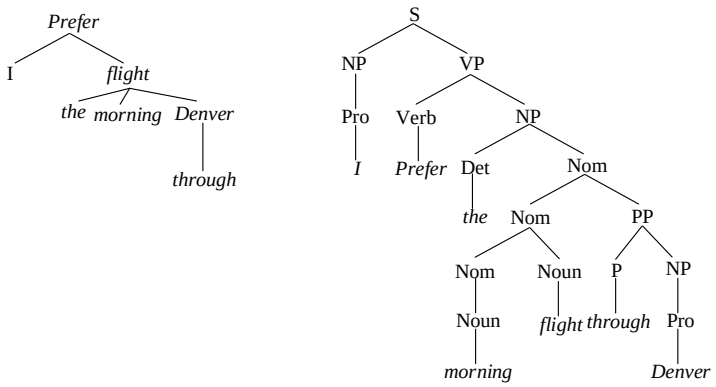

Figure 3. Dependency syntactic network analysis

When analyzing the English translation of Lingnan Hakka dialect based on the dependency syntactic network, the study of traditional grammatical relations is the basis for exploring the structure of linguistic dependency relations, and it is also the first step in learning the structure of the dependency syntactic network, but in the overall structure of the dependency syntactic network the existence of dependency relations is only binary simple relations. Each binary simple dependency is supported by a dependency and a centralizer, such as the SVP statistic line in the example above, which represents the direct dependency between I and refer, i.e., a dependency predicate is dependent on the centralizer subject. In linguistics, linguists have many definitions for the study of dependency in sentences, among which the definition of dependency syntax may vary from language to language.

2.2 Application of dependency syntax analysis in Lingnan Hakka dialects

The verbal meaning of “Lingnan” is divided into a broader and a narrower sense. Lingnan in the broad sense originally refers to the southern part of the Five Ridges in the south of China, while “Lingnan” in the narrow sense generally refers to Guangdong. Traditionally, the Hakka dialect is distributed over a wide range of areas and is not connected to each other, and in many areas south of the Yangtze River today there is a continuation of the Hakka dialect, including the Guangdong and Guangxi regions, Fujian, Sichuan, Taiwan and other places. It is even distributed abroad and is prevalent in some Southeast Asian regions. Among them, the Hakka dialect of the Lingnan region, due to the intersection of foreign cultures at that time and the later spread of a wider number of people, has a more pronounced structure of dependent syntax, which is analyzed in the following examples.

The Lingnan Hakka dialect is not very different from Putonghua in terms of linguistic structure, but it has obvious features in pronunciation, which are mainly as follows: firstly, in the development of the Lingnan Hakka dialect today, some of the ancient turbid consonants have been gradually eliminated, and most of them are pronounced as clear consonants regardless of their flat or oblique sounds, such as “Peach” (定平) and “Dao” (定仄) are pronounced as clear vowels today, even though one of them is flat and the other is oblique in the records. Secondly, in terms of consonants, the Lingnan Hakka dialect has preserved the full turbid consonants of ancient Chinese to a certain extent, for example, “床” (chuáng) is pronounced as [ch] in Putonghua, while in Hakka it is pronounced more similarly to the [z] sound. In addition, the Lingnan Hakka dialect is extremely rich in nasal consonants, such as [ŋai] for “milk” (nǎi). The Lingnan Hakka dialect has a more complex rhyme system, and maintains the incoming vowel endings in pronunciation, with common endings such as -p, -t, -k, etc. For example, “白” (bái) is pronounced as [bak] in Hakka. In terms of tones, Lingnan

Hakka dialect is very different from Mandarin. Lingnan Hakka dialect generally has six to eight tones, which is much richer than Mandarin. In Lingnan Hakka dialect, the pitch of the tones and the changes in the elevation of the tones obviously affect the meaning of the words, just as a word can have different meanings when read in different tones. Finally, in terms of overall pitch, Lingnan Hakka has very varied tones, which to a certain extent show a strong musicality, and changes in pitch are crucial to understanding and expression in Lingnan Hakka.

In the traditional Lingnan Hakka dialect, the analysis of dependent syntax can help us to simplify the understanding of some complex syntactic patterns and unique semantic structures of the dialect to a great extent. For example, in the Lingnan Hakka sentence "Granny cooks rice for Ah Sun", the dependency relations between sentences can help us understand the interdependence of the components in the sentence: "Granny" as the center word, i.e. subject, is dependent on the verb, i.e. the predicate "cook", and "rice" is dependent on the predicate verb "to cook" and on the prepositional phrase "give it to Ah Sun to eat". The word "rice" is dependent on the predicate verb "cook", and together with the prepositional phrase "畀阿孙食", it constitutes an extension of the predicate in the whole sentence structure. In this sentence, "畀" (to give) and "阿孙" (grandson) form a preposition-object combination, which is co-dependent on the predicate verb "to cook" and the other verb "to eat". Another verb "eat" (吃) further modifies "阿孙". In this example sentence, by analyzing the syntactic network of dependency, we can clearly see the dependency between the words, understand the complex semantic structure, clarify the relationship between the components, so as to understand and grasp the meaning and concept of the sentence as a whole, and preserve the unique expression of Lingnan Hakka dialect through the simple comprehension of the complex syntactic meaning and tone.

2.3 Modeling of dependency syntactic network analysis

Dependency syntactic network analysis is a method of analyzing complex linguistic structures based on dependency syntactic networks, which is applicable to many languages. In the process of dependency syntactic network analysis, the first step is to identify the dependency relationships among the words in various complex sentences, so as to construct syntactic trees. The basic modeling of dependency syntactic network analysis consists of two concepts, namely, dependency relation and syntactic tree. Each of the words in a sentence can be represented as a directed graph (G = (V, E)), where (V) is the set of complex words, which also represents nodes, and (E) is the set of dependency relations. Each dependency (e, in E) can be represented as (e = (h, m)), where (h) refers to the dominant word (head) and (m) refers to the dependent word (modifier), which clearly expresses the dependent syntactic relationships in it. The syntactic tree structure is shown in Figure 4.

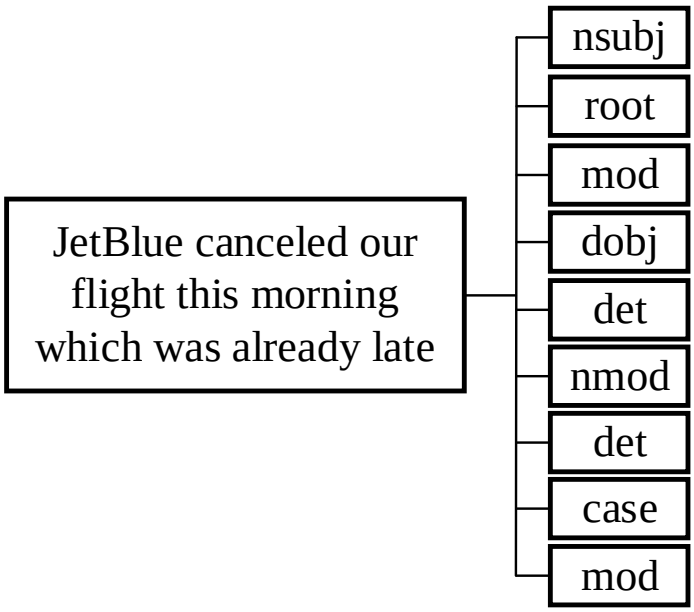

Figure 4. Non-projective dependency trees

In this example, we first set the definition of an edge (arc), which satisfies the Projective condition: suppose that this edge connects A1 and A2, where 1 < 2, and suppose that the central word of the sentence is head (which is one of A1 and A2). If there exists a path from the center word head to any complex word in that tree diagram relative to any of the complex words between 1 and 2, then it means that this line in the tree diagram satisfies the Projective property. On the contrary, if in a dendrogram representing dependency, there is a dependency tree in which there exists a path from the central word head to the word in each branch that satisfies the Objective property, where if there is any branch that does not exist a path from the central word head to the word, then it does not satisfy the Objective property, such as in the simple dependency dendrogram shown in the figure above, which The dendrogram has a branch (side arc) from flight, the center word (head), to was, which indicates that the individual word combinations contain more than one word between them, e.g., the center word (head), flight, has a direct path to which, (flight-was-which), and, conversely, there is no direct path from the center word (head) to the word in the words. On the contrary, there is no branch between the words “this” and “morning” and flight. Therefore, this edge (arc) does not satisfy Projective, i.e., it is non-Projective.

Based on the above operations, a simple analysis algorithm for dependent syntactic relations can be realized, as shown in Figure 5. In this algorithm, there is only one center word root in the initial state column, and all the words are arranged in a queue, and the algorithm loops until it reaches the end state (i.e., only the center word root is left in the state column, and the other queues are empty).

function DEPENDENCYPARSE(*words*) **returns** dependency tree

state← {[root],[*words*],[]} ;initial configuration
while *state* **not final**
t← ORACLE(*state*) ;choose a transition operator to apply
state← APPLY(*t*,*state*) ;apply it,creating a new state
return *state*

Figure 5. Dependency syntax analysis algorithm based on transformation

To perform the arithmetic, the following arithmetic steps are constructed

$$X_1 = \sum_{n=1}^{N}(1 + voc + a_n)\, f_n \tag{1}$$

X is the base model import structure

Voc is the root word in the syntax

an is the nth dependency occurrence rate

fn is the nth syntactic factor in the input process

$$X_2 = \sum_{n=1}^{N}(1 + X_{n+voc} + bn)\, f_n \tag{2}$$

$$S_n = max\left[S_n + \sum_{n=1}^{N} \underset{b}{\overset{M}{E}}{}_{a}^{n}\,(1 + X_{n+voc} + bn + 1 + voc + a_n)\right] \tag{3}$$

$$S_n = \frac{n(1+n)^x}{(1+n)^x - 1}\left(\sum_a X_1\,(Sa + Sb) + \sum_b X_{2(1+X_{n+voc}+bn)}\, S_{(Sa+Sb),}\right) \tag{4}$$

$$S_n = max[X_1 + \sum_{s=1}^{S} p_s\,(S_1 + S_2)] + \beta R_r \tag{5}$$

$$\mathrm{p} = 1 - \sqrt{\mathrm{In}\left(\frac{1}{1-\beta}\right)\frac{1}{\mathrm{S}}} \tag{6}$$

$$WL_{\underset{a}{n}} = \sum_{n=1}^{W}\sum_{a=1}^{L}(A_n R_a + R_n A_a) \tag{7}$$

$$WL_n = S(X_1 - X_2) \tag{8}$$

S_n represents the number of Stacks entered into the monitoring system, which includes different categories such as root, book, the, me, etc., and WL is the Word List entered. To optimize the system operation, the duplicate items in the two sets are processed as follows

Input function to find duplicate items R, single duplicate Ro, multiple duplicate Rm, βis the operation constant

$$R_m = R_o + \frac{1}{1-\beta}\sum_{s=1}^{S} X_s\,\beta_s \tag{9}$$

$$R_o \geq 0 \tag{10}$$

$$R_m \geq \ 0 \tag{11}$$

$$\beta_s \geq X_s - R_{o+m} \tag{12}$$

$$R_{m\beta} = max\left[R_o + \frac{1}{1-\beta}\sum_{s=1}^{S} X_s\,\beta_s\right] + \beta R_o \tag{13}$$

$$Ac = \sum_a X_1\,(Sa + Sb) \tag{14}$$

$$Re = \sum_b X_{2(1+X_{n+voc}+bn)}\,S_{(Sa+Sb),} \tag{15}$$

The simple analysis algorithm for dependent syntactic relations is shown in Table 1. The results of the algorithm analysis are shown in Table 2, Figure 6 and Figure 7, respectively. The results show that the algorithm maintains high accuracy at any step with an error of no more than 10%.

Table 1. Dependency syntax analysis algorithm

Step	Stack	Word list	Action	Relation added
0	[root]	[book, me, the, morning, flight]	SHIFT	(book→me)
1	[root, book]	[me, the, morning, flight]	SHIFT	
2	[root, book, me]	[the, morning, flight]	RIGHTARC	
3	[root, book]	[the, morning, flight]	SHIFT	
4	[root, book, the]	[morning, flight]	SHIFT	
5	[root, book, the, morning]	[flight]	SHIFT	
6	[root, book, the, morning, flight]	[]	LEFTARC	(morning←flight)
7	[root, book, the, flight]	[]	LEFTARC	(the←flight)
8	[root, book, flight]	[]	RIGHTARC	(book→flight)
9	[root, book]	[]	RIGHTARC	(root→book)
10	[root]	[]	Done	

Table 2. Algorithm analysis data

Step	Stack (Detailed data)	Algorithm accuracy	Algorithmic error
0	0.2536	0.0000	9.3625
1	0.2453	0.0000	8.1237
2	1.2785	0.0000	9.2351
3	9.3625	0.0000	3.2617
4	1.2635	0.0000	3.6210
5	/	0.0000	/
6	2.3627	0.0000	7.9215
7	/	0.0000	/
8	23.0124	0.0000	/
9	9.2621	0.0000	/
10	7.3624	0.0000	/

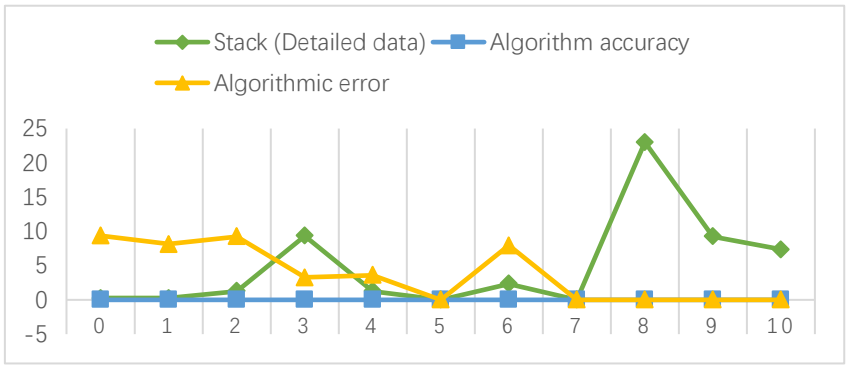

Figure 6. Algorithm analysis data graph

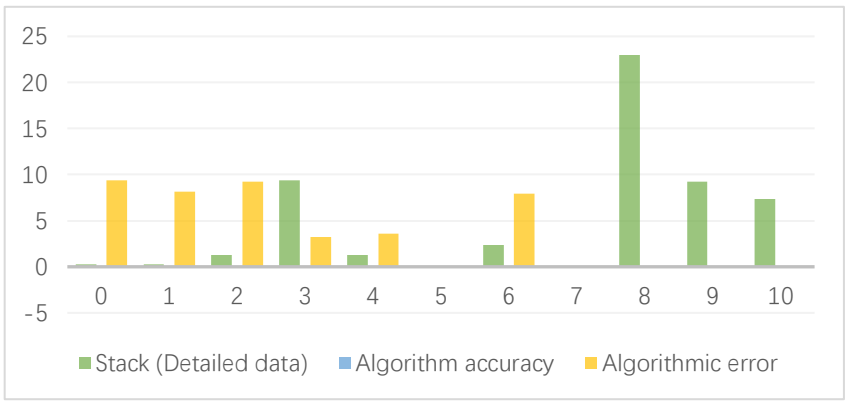

Figure 7. Algorithm analysis data graph

In the loop of the above dependent syntactic network algorithm, each step of the loop is automatically analyzed by the system according to the current state, and then the Oracle function is used to perform calculations, and finally the best algorithmic step is selected to perform the operation task. However, when the following situations occur, it may lead to an algorithmic error in the Oracle function, which ultimately leads to an error in the calculation result. If the final parse result is wrong, but it can be proved that the current dependency dendrogram satisfies the Projective nature according to the above algorithm, then at least one operation sequence can be performed in the algorithm. In the following, we take a typical sentence in Lingnan Hakka dialect as an example: “Granny cooks rice for the grandchildren”. We use the dependency syntactic network method to set up a dependency syntactic tree diagram for step-by-step parsing: first, we step-by-step parsed the dependency relationship between the main verb (predicate) and the subject, and the subject center word “阿婆” serves as a dependency of the predicate verb “煮”. In this sentence, the subject center word “阿婆” serves as the dependency of the predicate verb “煮”, and its dependency relation can be expressed as (eA1 = (text{boil}, text{Annie})), where “煮” is the dominant word and “阿婆” is the dependency word. Analyzing the sentence structure, it is clear that “granny” is the executor of the action “cook”. The next step is to analyze the dependency relationship between the object and the verb. The object “rice” is directly dependent on the predicate verb “cook”, and the algorithm is expressed as (e A2= (text{cook}, text{rice})), where the object “rice” is the subject of the predicate verb “cook”. “rice” is the action object of the predicate verb “cook”, i.e., the predicate verb “cook” is executed in order to form the object “rice”. The object of the predicate verb “cook” is the object of the action of the

predicate verb “cook”. The next step is to analyze the network dependency structure of the prepositional phrase. The purpose of the prepositional phrase “畀阿孙食” in the sentence structure is to modify the predicate verb “煮”, and the algorithm is expressed as (e A3= (text{cook}, text{畀})), where “畀” is the object of the action of the predicate verb “cook”, i.e., the predicate verb “cook” is executed in order to form the object “meal”. where “give” is the head verb of the prepositional phrase in the sentence structure. In the sentence structure, the preposition “畀” (i.e. the head verb) and the object “阿孙” form a dichotomous dependency, which is represented by the algorithm (e A4 = (text{畀}, text{阿孙})), i.e., it represents the action referred to by the predicate verb “the recipient of the action referred to by the predicate verb. Finally, the dependency between the complements is analyzed. In this example sentence, the complement “food” is not dependent on the subject-centered clause, but on the structure of the prepositional phrase, which is represented by the algorithm (e A5 = (text{畀}, text{food})), and the structure of the sentence “food” as a complement to the action “cook” is not dependent on the subject-centered clause. The sentence structure in which “food” is used as a complement to further explain the action of “Ah Sun” makes the meaning of the sentence clearer. After the complete algorithm operation and analysis of the above departments, finally, the dependency syntactic structure of the whole sentence is clearly shown by constructing a dependency syntactic tree diagram. When the dependent syntax analysis is carried out, the algorithm is keyed into the analysis, and this intelligent analysis method can help people to analyze the dependent syntactic structure of the sentence more clearly, and at the same time, it can also accurately convey the unique syntactic logic of Lingnan Hakka dialect, and translate foreign languages with the systematic thinking of Lingnan Hakka dialect, which can make the communication more precise and provide a powerful support to the grammatical research and translation precision of the Lingnan Hakka dialect. It provides a strong support for the grammatical research of Lingnan Hakka dialect and the precision of translation.

3 Intelligent development of English translation of Lingnan Hakka dialect

Lingnan Hakka dialect has been spreading widely in Lingnan area and even in the whole country with its unique language structure and profound cultural background, and its unique language structure, complicated vocabulary structure and diversified comprehension meanings have been troubling the translators for many years, especially when the tide of intelligent translation is coming, the intelligent English translation of Lingnan Hakka dialect is a great challenge faced by the technicians. With the rapid development of artificial intelligent translation technology, especially the breakthroughs in the field of Natural Language Processing (NLP) and machine translation, intelligent English translation of Lingnan Hakka dialect has gradually become possible. The following section will discuss the current situation and future outlook of the intelligent development of English translation in Lingnan Hakka dialect.

3.1 Construction of Lingnan Hakka dialect corpus

The core of building a modern multilingual intelligent translation system lies in creating a rich and comprehensive dialect corpus. In the process of developing intelligent English translation of Lingnan Hakka dialect, the difficulty that still needs to be faced is the construction of Lingnan Hakka dialect corpus, the cost of which is still being raised, and the construction of highly intelligent corpus requires large-scale high-quality parallelisms, which are being accumulated and sorted out at the moment, and

the slow process of constructing high-quality corpus of Lingnan Hakka dialect leads to the limited data for the training of the intelligent translation model. Intelligent English translation based on the existing Lingnan Hakka dialect corpus still needs to rely on manual collection and some traditional methods of collation, because of its narrow coverage, so it can not go to a comprehensive reflection of the diversity and complexity of the Lingnan Hakka dialect.

The construction of Lingnan Hakka dialect corpus is crucial to the development of intelligent English translation of Lingnan Hakka dialect, and the process of its construction is complicated and critical, involving a large number of linguistic and phonetic data collection and data organization. At present, the construction of Lingnan Hakka dialect corpus is mainly in the stage of collecting Lingnan Hakka dialect text and special pronunciation audio data. The text collection of Lingnan dialect mainly includes the collection of some literary works with Lingnan Hakka dialect as the main material, as well as some oral histories and folktales from local villagers, etc. The special audio collection of Lingnan Hakka dialect focuses on the recording of natural dialogues and some reading materials of Lingnan Hakka dialect speakers in different regions and at different age stages.

In the process of corpus data organization, firstly, we should carry out detailed word division, labeling and annotation of the collected materials, and then we should construct grammatical structures, analyze the dependent syntactic structures, and carry out detailed organization and summarization of the phonological system of the vocabulary list. In addition, it is also necessary to add rich contextual features and cultural background information according to the language habits of Lingnan Hakka dialect. The core purpose of constructing the Lingnan Hakka dialect corpus is to enhance the practicality of the corpus, and focusing on the organization of linguistic and cultural details is the most convenient way to achieve its purpose. At present, although the construction of Lingnan Hakka Corpus has made some progress based on the efforts of all parties in collecting and organizing information, the scale and coverage of the Lingnan Hakka Corpus still need to be further expanded in order to satisfy the needs of research and intelligent translation of the Lingnan Hakka dialect.

3.2 Contextual understanding and cultural translation in intelligent English translation of Lingnan Hakka dialects

In the process of intelligent English translation of Lingnan Hakka dialect, the factors of local cultural characteristics and the unique meanings of some vocabularies in the context should not be neglected. Lingnan Hakka dialect contains a large number of local cultural connotations, which need to be dealt with through the contextual understanding and some cultural translation strategies in the process of language translation. In the process of upgrading the Lingnan Hakka dialect translation system, we need to update the following two models: the context-aware model, the principle of which is to make the system training model have the ability to perceive the context and differentiate between different words and sentences in the dialect, so as to translate more accurately. Cultural Translation Database Model: By constructing a database that covers the entire Hakka cultural background in Lingnan, it can help the AI system to make accurate cultural conversions when translating. The intelligent translation system for Lingnan Hakka dialect relies on multi-modal learning by combining various data types such as dialectal speech, text, and images to achieve more accurate recognition and translation of Lingnan Hakka dialect. At the same time, the cross-language modeling of the intelligent translation system will also help the system to have stronger adaptability and learning ability when dealing with complex language environments.

3.3 Intelligent English translation process of Lingnan Hakka dialect

The intelligent English translation system for Lingnan Hakka dialect mainly relies on NLP and deep learning technologies. In the intelligent English translation of Lingnan Hakka dialect, the detailed operation steps are shown in Figure 8.

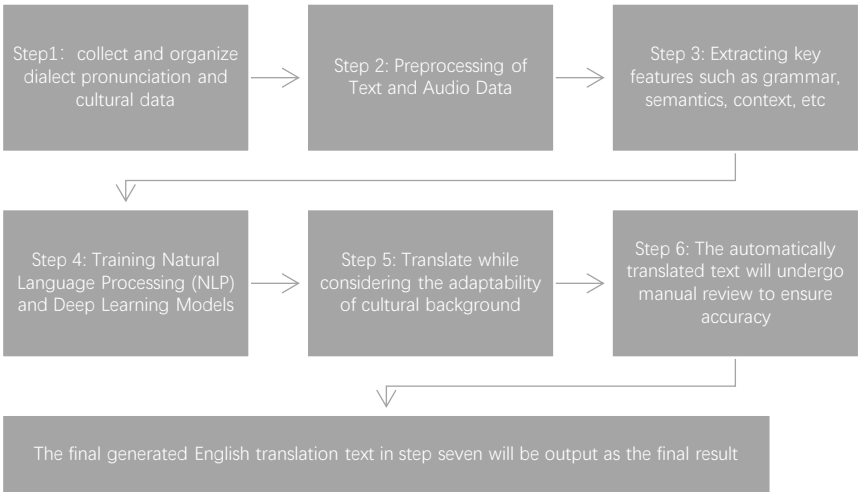

Figure 8. Intelligent English translation process of Lingnan Hakka dialect

As can be seen from Figure 8, a comprehensive corpus containing contextual understanding and cultural translation of Lingnan Hakka dialect needs to be constructed first. Then, through the preprocessing step, the pronunciation correction and the sub-sentence and phonetic labeling are performed to ensure that the intelligent Lingnan Hakka dialect translation system can correctly identify and understand the features of the Lingnan Hakka dialect. Finally, in the model training phase, a deep learning algorithm is used to train the translation model so that the intelligent system can accurately handle the unique grammar and vocabulary of the dialect. In the actual translation, the intelligent system will translate the Hakka dialect into English by analyzing the dependent syntactic structure through the automatic translation tool, and ensure the accuracy and naturalness of the English translation through the contextual analysis and cultural adaptation of artificial intelligence.

4 Evaluation of the effect of intelligent proofreading on English translation of Lingnan Hakka dialect

In order to improve the quality of intelligent English translation of Lingnan Hakka dialect, it is a general trend to construct an intelligent translation proofreading system for the widely popular dialectal English intelligent translation system. By using the object class library ADO.NET, we design the statement relationship table, analyze the syntactic structure in detail, and collect the database icon structure such as the translation decision data table based on the key qualitative factors such as dependent syntax. The application is written in Java language, and the interface program is run, which is combined with the comprehensively constructed homepage, editor, etc., and the display of view class, so as to obtain the proofreading client of the intelligent translation system. By using the IaaS cloud platform of Open Nebula and computer algorithms, the coding module of the server side of the intelligent translation proofreading of Lingnan Hakka dialect is constructed. The word vector information diversification fusion module is extracted from the established comprehensive corpus, and the database icon structure such as the translation decision data table is input again to

obtain the intelligent translation result proofreading platform by structuring the word proofreading unit module of the corpus and the cultural context and semantic proofreading unit module.

4.1 Constructing the database of intelligent proofreading system for Lingnan Hakka dialect English translation

Intelligent translation proofreading database, as the main storage location of translation data and the means of connecting diversified languages, plays a decisive role in the data structure of Lingnan Hakka dialect intelligent dialect translation system and the correctness of translation. By using the object class ADO.NET, the core Lingnan Hakka dialect intelligent translation proofreading database table structure is designed, as shown in Table 3-Table 5. Among them, the statement relationship table between each language structure determines the corresponding words, complex texts, and sentences to be extracted from the statements generated by the intelligent translation system. The translation decision table is used to detect whether there is a direct translation relationship between the source language of the natural translation system and the target language of the Lingnan Hakka dialect intelligent translation system, i.e., to analyze whether the Lingnan Hakka dialect intelligent translation system supports the translation processing between the languages, and the structure of the table is important for strengthening the expandability, plasticity and flexibility of the Lingnan Hakka dialect intelligent translation system, and for increasing the number of intelligent translations of complex dialect language varieties in the future. The data table structure lays a material foundation for strengthening the expandability, plasticity and flexibility of the Lingnan Hakka dialect intelligent translation system, and for increasing the number of intelligent translations of complex dialects and language types in the future, and provides the function of storing the classic phrases and cultural and historical traditions of Lingnan Hakka dialect as well as storing the initial texts and corrected pairs of phrases of the natural source language and the target language of the intelligent translation.

Table 3 Dialect sentence relationship table

Field name	Data type	Attribute	Attribute
Yjgx_id	8.4632	0.0000	Relationship table encoding
English	9.2365	0.0000	Source Language
Chinese	3.2320	0.0030	Target Language
E_id	/	0.0250	Source language encoding
C_id	9.2356	0.0000	Target language encoding
fragment_Eid	7.2361	0.0030	Source language statement encoding
fragment_Cid	1.0216	0.0050	Target language statement encoding

Table 4 Translation judgment table

Field name	Data type	Attribute	Definition
fypd_id	/	0.0001	Judgment table code
English	20.3622	0.0013	Source language
Chinese	32.0325	0.0008	Target language
interpretable	17.0025	0.0000	Is there a translation relationship

Table 5 Storage table of Lingnan Hakka dialect sentences

Field name	Data type	Data type	Definition
yjcc_id	0.2315	0.0000	Storage table encoding
English_y	0.2951	0.0002	Stored source language statements
Chinese_y	0.3062	0.0000	Stored target language statements
English_j	0.0214	0.0000	Stored target language statements
Chinese_j	0.1961	0.0010	Stored target language statements

4.2 User side of Lingnan Hakka dialect intelligent translation proofreading system

The user side of Lingnan Hakka dialect intelligent translation and proofreading system is located in the front end of the whole Lingnan Hakka dialect translation and proofreading system. The main interface of the application program of this translation and proofreading system consists of the continuation relationship between the view and the view group objects, and it is written in Java language through the running of the application program. The main interface function of this intelligent English translation and correction system is to complete the transmission of language text and media information and the jumping function of the main page of English translation and correction by applying the interrelationship between the view and view group objects, and the main principle of the system is to complete the system response and processing by using the NLP and the deep learning technology.

The proofreading of the translation effect of the intelligent system for the Lingnan Hakka dialect is the key link to ensure the translation quality of the intelligent translation system. After the Lingnan Hakka dialect intelligent translation system completes the preliminary search and analysis of the comprehensive corpus constructed and makes reference to the initial text storing the natural source language and the target language of the intelligent translation, the correction system evaluates and corrects the output translation results, and the proofreading stage audits and optimizes the resulting translation by the various means described above. The intelligent translation correction system analyzes the translation results to automatically detect grammatical errors in the translation results, conducts a comprehensive analysis in conjunction with the grammatical structure of the dependent syntax, eliminates unnatural linguistic expressions as well as errors in cultural connotations and natural contexts, provided that appropriate expressions in cultural connotations and natural contexts were typed into the language and writing system during the previous construction of the corpus, and corrects potential errors in the translated structure output by the intelligent translation system. potential misunderstandings of semantic comprehension in the translation structure output by the system. The final manual proofreading is carried out, which mainly emphasizes in-depth analysis of the culture and some special expressions in the regional context, so as to correct the subtle differences that the intelligent language system is unable to accurately analyze and deal with, as well as the unique expression habits of the Lingnan Hakka dialect. The final output of the translation should reach the standards of accurate semantic expression, natural emotion and proper transmission of Lingnan Hakka culture after several rounds of proofreading. The intelligent translation correction system for Lingnan Hakka dialect not only improves the translation efficiency, but also improves the translation quality through the above correction steps.

5 Conclusion

Through the study and analysis of the comprehensive data, the evaluation of the effect of intelligent proofreading of Lingnan Hakka dialect English translation based on dependent syntactic network shows that the method has significant advantages in improving the quality of the results of Lingnan Hakka dialect English translation. The intelligent English translation results of Lingnan Hakka dialect are analyzed in depth using the dependency syntactic network to unravel the linguistic structure of the dialect sentences as well as the relationship between the semantics, and the common grammatical errors and semantic deviations in the intelligent translation system are identified so that they can be corrected efficiently in the proofreading process. This method largely improves the grammatical accuracy of the intelligent translation of Lingnan Hakka dialect, and also enhances the precise expression of complex sentences and cultural connotations in the dialect. Overall, the dependent syntactic network provides a strong technical support for the proofreading of intelligent English translations, making the final translations more accurate and natural.

References

[1] Xia, J., Yu, Z., Deng, Q., et al. (2024). Multi-feature multiple fusion sentiment classification model based on syntactic dependency and attention mechanism. *Application on Research of Computers, 41(11),* 3295-3302. https://doi.org/10.19734/j.issn.1001-3695.2024.04.0104

[2] Xie, X., & Wang, M. (2024). User experience of intelligent human-computer interaction products based on dependency parsing. *Digital Library Forum, 20(6),* 44-53.

[3] Dong, S., Li, X., Yang, F., et al. (2024). Mathematical expression query expansion method based on dependency parsing. *Computer Applications and Software, 41(4),* 251-255, 290.

[4] Zhang, Y. (2024). Research on aspect-level sentiment analysis method based on dependency syntax [Doctoral dissertation, Shandong Normal University]. https://doi.org/10.27280/d.cnki.gsdsu.2024.000394

[5] Li, X. (2024). Research on data augmentation strategies and grammar error analysis techniques in AI translation. *Automation and Instrumentation, (07),* 243-246. https://doi.org/10.14016/j.cnki.1001-9227.2024.07.243

[6] Cui, X., Wang, R., & Lie, K. (2024). Sentiment classification model based on BERT and dependency syntax. *Modern Computer, 30(18),* 66-70.

[7] Zerenzhuoma, Q., & Xia, W. (2024). Construction and application of Tibetan dependency syntax annotation system. *Journal of Northwest University for Nationalities (Natural Science), 45(3),* 80-88.

[8] Guo, W. (2024). A study on attachment preference of Chinese and English relative clauses based on dependency syntactic annotated corpus [Doctoral dissertation, North Minzu University].

[9] Qian, L., & Gao, S. (2024). A study of syntactic complexity in CFL learners' writing based on dependency treebank. *Language Teaching and Linguistic Studies, 2024(06),* 14-26.

[10] Situ, S. (2014). Immigration and formation of Dongjiang Hakka culture. *Journal of Huizhou University, 34(05),* 1-5.

[11] Li, Y. (2021). Application of machine translation intelligent proofreading system in translation practice. *Computer Programming Skills and Maintenance, (11),* 104-105. https://doi.org/10.16184/j.cnki.comprg.2021.11.039

[12] Liu, L. (2016). Dialect translation strategies in English literature. *Journal of Suzhou University, 31(06),* 52-55.

[13] Cui, D., & Li, S. (2024). Design of natural language information extraction translation proofreading system based on AI algorithm. *Modern Electronic Technology, 47(10),* 111-116. https://doi.org/10.16652/j.issn.1004-373x.2024.10.021

[14] He, S. (2023). Design of an intelligent proofreading system for English translation based on LSTM algorithm. *Information Technology, (07),* 118-124. https://doi.org/10.13274/j.cnki.hdzj.2023.07.021

Acknowledgements

This work was sponsored in part by General project of the National Social Science Fund "Character Rhyme Group Books Dialect Phonetic Research of Lingnan and Literature Pedigree Sorting " (Project number: 21BYY130)

About the Author

Guiying Kong was born in Guangdong, China, in 1973. From 2004 to 2007, she studied in Yunnan Normal University of School of Foreign Languages & Literature and received her Master's degree of Foreign Linguistics and Applied Linguistics in 2007. From 2021 to 2023, she studied in University of Perpetual Help System DALTA in the Philippines and received her PhD of English in 2023. Currently, she works in College of Foreign Languages in Wuzhou University , Guangxi, She is also a researchmember of the Research Center of Xijiang River Basin Folk Literature in Wuzhou University. She has published 36 papers, her research interests are included English teaching, local culture, language comparison, etc.

E-mail:camelliared1973@163.com