

Research Article

Xi Wang*

Translation correction of English phrases based on optimized GLR algorithm

<https://doi.org/10.1515/jisys-2020-0132>

received December 18, 2020; accepted April 14, 2021

Abstract: Basic syntactic analysis refers to sentence-level syntactic analysis. In the process of developing the Mat Link English–Chinese machine translation system, the Generalized Maximum Likelihood Ratio algorithm was improved, and a basic English syntax analyzer for English–Chinese translation was designed and implemented. The analyzer approves the structure of the analysis table with a variety of export products, introduces the character mapping function to realize the automatic recognition of the sentence boundary, uses the children of the same level to describe the grammatical structure of the sentence, and realizes the proverb from the original sentence to the target sentence stage conversion. Finally, through the analysis of example sentences, the design concept and working process of the basic grammar are explained.

Keywords: machine translation, syntactic analysis, GLR algorithm

1 Introduction

Basic syntax analysis has a wide range of applications in the field of natural language processing research, such as machine translation, information collection, information retrieval [1], natural language understanding, and speech analysis and synthesis. The Generalized Maximum Likelihood Ratio (GLR) algorithm is an improved algorithm based on the LR algorithm [2]. It uses graph structure stack (including Raph stack) data to solve multiple entry points of antiquities [3], which is a problem of syntactic ambiguity and has high analysis efficiency. When the GLR [4] algorithm is applied to shallow syntactic analysis, this article proposes a method for constructing a multi-exit analysis table, which reduces the complexity of the algorithm [5] and automatically recognizes phrases and boundaries by establishing symbol mapping [6]. Use child sibling trees instead of graph stacks to describe short syntax. In many automatic phrase analysis algorithms [6], some very simple structures in manual translation cannot be recognized [7]. There are few advanced automatic sentence analysis algorithms for English–Chinese machine translation [8]. Therefore, a new algorithm [9] is designed to use the sentences in the sentence, and the sentence function determines the position range in the translation result [10].

Zhou Yating analyzed the feature that the English [11] text machine translation conforms to the full stop office. The unit is NT clause. The part, translate, assemble (PTA) model in the translation unit system realizes the process of English–Chinese translation [12] and the construction of the text-oriented English–Chinese clause corpus. The PTA model is explained in detail, highlighting the importance of corpus [13]. Lu Rongwan improved the traditional rule-based machine translation model, using the English machine translation model based on the semantic network [14], in the specific implementation process, using the vector-based hybrid phrase synthesis semantic statistical English machine translation method [15], in the translation similarity model. In the measurement process, the cosine similarity calculation method is used

* **Corresponding author: Xi Wang**, School of Foreign Languages and International Education, Chengdu Technological University, 3-3-1201, Jinlanyuan, No.179, Jinkang Street, Wuhou District, Chengdu, Sichuan 610031, China, e-mail: x0a4wo@163.com

to obtain the semantic similarity of the two vectors [16], and the weighted vector method calculation rules are added to distinguish the differences between the two similar vectors, so as to obtain accurate translation results and ensure the quality of translation [17]. Huang Dengxian overcomes the use of pipeline layer-by-layer analysis technology to analyze machine translation, compares the segmented phrase words with the phrase corpus to analyze part of speech and syntax, and further obtains the syntactic structure of the English to be translated. The method errors are gradually transmitted and accumulated [18]. This leads to the disadvantage of lower translation accuracy. We designed a vocabulary semantic similarity and logarithmic linear model based on HowNet, saved the corresponding bilingual corpus in a Chinese-English dependency tree-to-string manner [19], and provided a structured processing of language dependency to ensure Correspondence between English and bilinguals, calculating the semantic similarity between the translated sentence and the vocabulary in the source language of the instance database for the calculation of HowNet operation input, which further improves the accuracy of translation and the translation results have higher accuracy [20].

2 LR and GLR algorithms

The LR analysis method was originally used to analyze programming languages. The algorithm for automatically creating the diagnosis table during the preparation process can be found in the “Chuan” file [21], so I won’t repeat it here. The action table basically describes the mapping of the binary group formed by the state S and the input symbol T in the analysis operation [22], that is, the function is the analysis operation, where T covers the sequence and origin of the analysis table; the new state S and the unstable N State, that is, GOTO5N [13] is a certain state. “Enter” means to push the new current state onto the stack and continue to read the next input symbol. State key status (1 relevant part of rule 0), which is a good reference! 0T0. The table shows the current state of the stack. The recognition process confirms that the input symbols do not conform to the given LR grammar [23]. In other words, the mapping described by the analysis table is two double actions. The regular CFG grammar ensures that there is no ambiguity [24].

Generally, because the GI_R algorithm has greater chance in the results of part-of-speech recognition [25], the identified data points have a high coincidence probability, which still cannot meet the existing part-of-speech recognition accuracy. This article improves the classic GLR algorithm and proposes to use the phrase center to analyze the structure of the phrase [19], which effectively reduces the probability of overlapping data points and improves the accuracy of part-of-speech recognition. The improved GI_R algorithm calculates the likelihood of the preamble of the phrase with the help of four-element clusters. The algorithm is shown in the formula:

$$G_E = (V_N, V_T, S, \alpha).$$

Assuming that P is any action in a and the corpse exists in V , after deduction, the formula:

$$P \rightarrow \{\theta, x, c, \delta\}.$$

To this end, Tomita has extended the LR analysis algorithm so that it can deal with changes in reduced ambiguity. This is a GLR algorithm. The GLR algorithm is a method of identifying general irrelevant languages. Each entry and exit stack is always a status icon; in the case of ambiguity, it is similar to the LR algorithm. Complete the analysis table operation separately, keep multiple options, create multiple results, and perform subsequent operations independently. Under this analysis, symbolic analysis when the analysis based on a specific analysis table fails, the analysis result will be discarded. Each result input can have a grammatical structure for diagnosis [19].

Although natural language is a large number of incoherent languages (CFL), phrases, as the basic unit of natural language meaning, are closer to CFL than natural languages. Therefore, in the basic syntactic analysis of natural language, that is, syntactic sentences, CFG grammar can be used to approximate

sentences. When applying the GLR algorithm to basic syntactic analysis, there are mainly the following problems:

- (1) There are many types of phrases, such as nouns, adjectives, adverbs, and various non-modal phrases and artificial phrases, to promote in-depth syntactic analysis. This requires its analysis table to have a multiple export structure.
- (2) There are many types of sentences, such as nouns, adjectives, adverbs, and various non-modal sentences and artificial sentences, to promote more in-depth syntactic analysis, which requires the analysis table to have multiple derived structures.
- (3) In the result of basic syntax analysis, each block should have stable properties. In other words, the sentence should have attributes such as subcategory, morphology, and semantics. These characteristics of a sentence are mainly composed of its main words, so it is necessary to specify the main words of the sentence while specifying the structure of the short grammar.
- (4) In the application of machine translation, pronouns should be formed in the basic stage of analysis. Therefore, the author proposes an enhanced CFG method.

2.1 Construction of phrase corpus

Table 1 describes the phrase information of the algorithm. The size of corpus phrases is about 580,000 English and Chinese words, which can roughly form 20,000 sentences and 10,000 sentences, while standard corpus sentences can only form 10,000 sentences. As shown in Table 1, the phrase corpus considers the scope and system of the corpus and can represent various communications in daily communication, business, technology, and machinery [26]. On the other hand, the corpus semantic method uses data, hierarchical structure, and processing methods to define sentences in text form, knows the various parts of short sentences to complete the sequence of sentences, and uses independent human–computer interaction to copy and calibrate English–Chinese translation sentences. The example illustrates the specific application of the oracle library.

Table 1: Phrase corpus information

Elements	Attribute	Information
Corpus marks the content	Size	Chinese has 300,000 words and English 280,000 words
	Application field	Daily communication, business, technology, machinery
	System	Words used in communication, proper nouns, written expressions
	Tense	Then, now. In the future
Expected notation	Data	Text format
	Level	Part of speech, alignment
	Processing	Man–machine autonomous communication
Expected application	Scope	English–Chinese machine translation

Original sentence: I know you are a machine factory of physics facilities.

Part-of-speech tag: I/Verb know/INC you/PRP are/NNSa/V BP machine/VBP factory/NNS of/INP physics/JJ facilities/NNS./

2.2 Part-of-speech recognition of phrase corpus

Part-of-speech recognition is an important processing step of the automatic phrase analysis algorithm for English–Chinese machine translation, which is suitable for large-scale sentences, sentence, and

grammatical ambiguities. According to the meaning of the phrase, divide the sentence into several words, arrange the words to form the relationship between the sentence and the person, and write independence in the phrase after acknowledging part of the sentence and forming a syntactic tree, which reduces English–Chinese machine translation. This is very useful for improving the efficiency of phrase processing. GLR is the basic algorithm for recognizing speech parts. It is a method of analyzing the probability of a phrase. The number of recognition results given by the standard GLR algorithm is uncertain, and there may be overlapping data points in different recognition results and the recognition accuracy is usually low. In the designed automatic sentence analysis algorithm for English–Chinese machine translation, the standard GLR algorithm is improved and a phrase structure centered on the phrase center is established to improve the accuracy of recognition.

2.3 Automatic phrase recognition algorithm correction process for English–Chinese machine translation

In the previous automatic algorithm of English machine translation proverbs, the discourse recognition result of the phrase was the final result, but the discourse recognition did not improve the organizational ambiguity between English and Chinese. It is necessary to correct the result of acknowledging the words. In the GLR algorithm, a linear analysis table is used to identify sentence functions. In addition, linear table analysis has another function, namely syntactic recognition, which uses progress, reduction, acceptance, termination, and errors as evidence to analyze part of the speech analysis. Figure 1 is a flowchart of automatic phrase correction algorithm correction for English–Chinese machine translation.

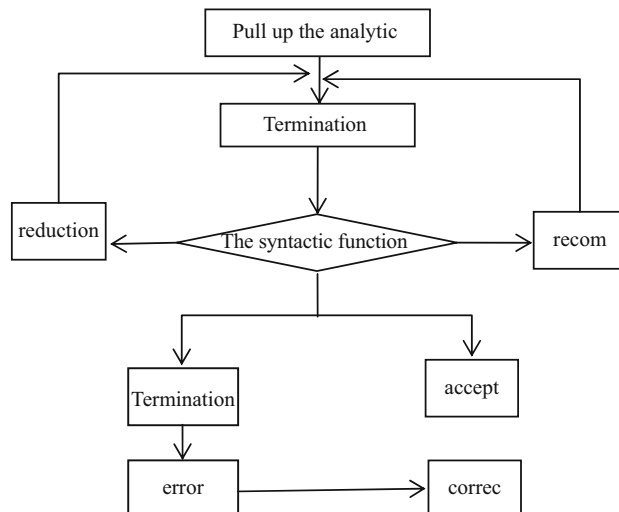


Figure 1: Algorithm correction flowchart.

It can be seen from Figure 1 that reduction is very similar to progress actions. Reduction refers to the reshaping of grammatical recognition constraints, which means that the previous constraints are invalid or there are errors in the cycle. Progress means that there is no error in the ambiguity structure in this syntactic recognition. The part of the speech analysis result of the sentence in front of the cover is correct, and then the receiving cursor is called and sent out for use. Normally, the approval cursor and indicator will appear at the same time. If the two are only one of the algorithm streams, it indicates a loop error or an error in the algorithm configuration. It is appropriate to review the analysis of the linear table to recover the acceptable

part of the speech analysis result. Before replacing the converter, check the type of cursor. If the search fails, enter and complete the cursor. The end point was created at the preparation site, and the structure of the site may be ambiguous. After displaying the endpoints, the algorithm will build a sentence tree, mark the symbol table, analyze whether the central symbol of the prepared point exists, and place it on the correct sentence type. If it does not exist or the position is incorrect, the algorithm immediately calls the wrong cursor to correct it. In the entire algorithm correction process, there are multiple sentence analysis output ports, and a receiving cursor can only send one recognition result at a time. When multiple confirmation results need to be sent at the same time (because the two phrases in the sentence are continuous), please write multiple confirmation results into the same node of the phrase tree, and the recipient cursor will automatically use them as the confirmation result.

3 Shallow syntactic analyzer

The English-based syntactic analysis tool designed in this article takes the utterance part of the speech after the meaning of the speech part as the analysis part and sends out the bulk order of the sentence. The Mat Link English–Chinese machine translation system contains 30 tags, that is, the system is closed. They are /noun,/verb,/auxiliary,/adj,/adv,/art,/det,/num,/prom/prep,/conj,/inter],/wh,/where,/is,/to,/Both,/and,/that,/pmark,/qmark,/emark,/comma./lsign./rsign./Dash./colon./scolon./sym and/donor; nine phrases can be identified in basic syntactic analysis, namely, *NP, *AjP, *AvP, *Ven, *Ving, *Toinf, *Tobe, *Vertu og, *PPY, and the other four symbols in the middle start from Hn-. Mn Det and Adj create symbols that are not related to the basic grammar, and the output of each block is a known phrase or a non-phrase label.

3.1 The structure of the analyzer

The structure of the entire basic equipment, it contains two diagnostic tables, one of which contains many export products. One is the TableA diagnostic table, which is closely related to noun phrases that can identify and publish noun phrases* NP, adjective phrases* AjP, and adverbs* AvP: the other is Tab1eB speech analysis table, which can recognize and send unlimited sentences** Tobe; currently, Word segmentation/type sentences *Ving and proverb sentences *Ven and *Be, and *PPY} are necessary for in-depth grammatical analysis. The reason for this division is to fully consider the amount of connection between different sentences and reduce the complexity of the analyzer. Since the analysis method structure is used in the two analysis tables, each part of the voice sequence number is no longer consistent with the sequence number of the sequence status symbol in each analysis table, so it is necessary to classify the flight symbols into system status symbols and connections table. The system cover refers to the input symbols of the analyzer, and the diagnosis table cover refers to the error diagnosis table. We have introduced character mapping function for each diagnosis table to realize the conversion from system station to diagnosis table station. For the diagnostic table, if a system interrupt is invalid in the diagnostic table, the symbol mapping function will map it to the diagnostic table of the workstation (Figure 2).

3.2 Analysis algorithm based on GLR

Each diagnostic table of the analyzer and the diagnostic process of each diagnostic tablet are similar to the GLR algorithm, but with the following differences:

- (1) There are five functions in the analysis table, including the meanings “1v in,” “decrease,” and “keep unchanged.” There are two types of errors: “dismissal” and “resolve errors.” For the service error of the

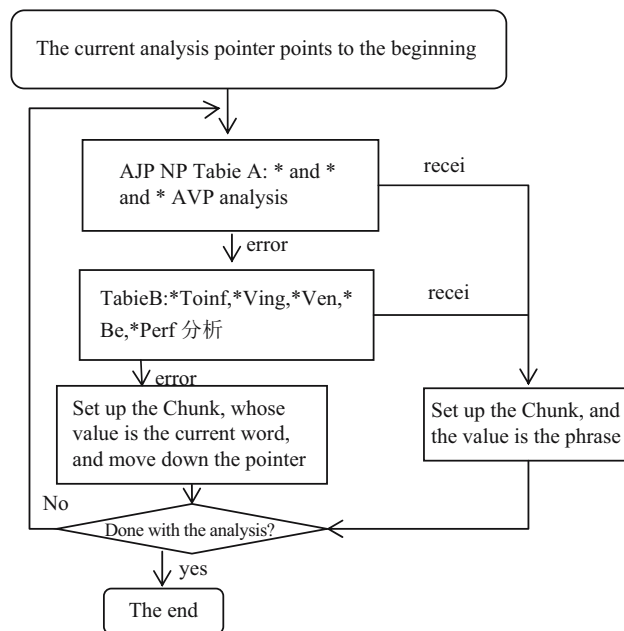


Figure 2: Shallow syntactic analyzer structure.

final investigation diagnosis table, “common error” refers to the service error of the stop diagnosis table. The “Finish” function does not regard the analysis as a failure, but keeps the current state unchanged, and returns the current winding system of the system to the suspension diagnosis table, and then continues the analysis; the “error” in the diagnosis table is considered a diagnosis failure. And restore the diagnosis cursor to the diagnosis position, and then end the analysis of the diagnosis table.

- (2) When trying to use rules for reduction, please check the restrictions of the rules; if the conditional Boolean activity value is true, then perform the reduction, otherwise perform the termination operation;
- (3) The analysis table is mostly exported, that is, there are many countries at the end of the analysis table.

Perform the “receive” function, each “receive” corresponds to a known phrase on the top of the symbol board. In the analysis, symbol keys and status keys are used, and they both perform push and pop operations at the same time. The token key holds a tree representing the sibling children of the token. Figure 3 shows the analysis algorithm based on GLR. The analysis steps are described as follows:

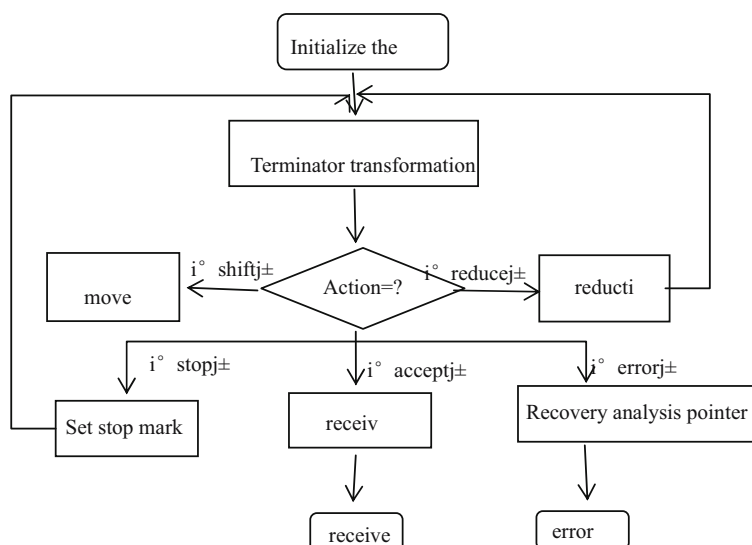


Figure 3: Syntactic analysis algorithm based on GLR.

Initialization. Push status 0 into the stack, analyze the status of the system station indicated by the cursor, and clear the termination flag.

- (1) Modification of the closing device. If the termination flag is not set, use the mapping function icon to map the current system cover to the diagnostic table or cover. If the termination flag is set, please directly map the current termination of the system on the diagnosis table.

Check the ACTION table and process each input function separately.

Input: Put the current state and current symbology (terminal and nonterminal) into the stack, and lower the diagnostic cursor.

Skip Code of Conduct: Call the precondition function to check the precondition. If the conditions are met, click the sub-node from the icon key to form the syntactic structure of the current icon tree. The grammatical structure is not closed and push the icon key; click the middle position state, check the GOTO table, and push the new state into the state pile. At the same time, point the middle word to the corresponding middle word and create a translation of the current irrelevant symbols according to the translation method. If the conditions are not met, the termination flag is set.

Termination: Set the termination flag.

Receive: Pop up the syntax of the tree at the top of the icon key, and then return.

Error: Restore the initial state and return, the instructions go to step 2.

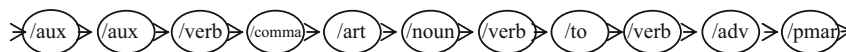
During the execution of the plot, we checked the limitations of the syntactic rules. Limitations mainly examine the four attributes of the word itself, the subparts of the language, and its lexical and semantic technology. After further generalization, the author believes that it is possible to gradually study methods of expanding discourse labels before analysis. For example, based on semantic and morphological characteristics, these nouns can be expanded into time nouns *n time*, address nouns *n-tle>*, noun possessive *epos*), and common nouns to bypass nouns. For example, the phrase “to be corrected}, the clock started to work well.” The phonetic sequence after the meaning of the voice part is *artificial/aux/verb} EN}/comma/art/noun/verb} ED}/to/verb/adv/pmark,*” as shown in Figure 3. Therefore, after a part of the case is expanded, part of the case will become *tracking/being/ven/comma/art/noun/ved./verb/adv/pmark*” as shown in Figure 3. The actions and their GOTO tables are shown in Tables 2 and 3. For each rule in the analysis table, an underscore is used to indicate the middle word of a word that does not end in the left part of the rule, and some rules do not exist because these irrelevant symbols are neither recognizable sentences, nor will they become The middle word of the known sentence:

Table 2: Evaluation results of statistical algorithms

Project	Closed beta	Develop test
Identification accuracy/point	78	74
Identify moderation/score	73	76
Update ability/score	61	58
Score/point	76.9	73.3

Table 3: Evaluation results of dynamic memory algorithm

Project	Closed beta	Develop test
Identification accuracy/point	94	93
Identify moderation/score	76	78
Update ability/score	68	60
Score/point	91.8	90.6



Part-of-speech sequence after part-of-speech tagging:



- (1) The word order after the part of speech is expanded, so the middle word of the term is meaningless. In addition, the meaning of each rule is not enumerated one by one. Now take the various rules of the TableA analysis table as an example. The third rule * NP -> /noun/number translation method can be described as “/number/noun,” produced by this rule. The translation of SNP is num across nouns). *NP -> /pron/The independent translation method is wool apron. It should be noted that the mapping shown in Figure 4 does not involve mapping. If the termination flag is set, each system shutter will be mapped to the termination analysis table.

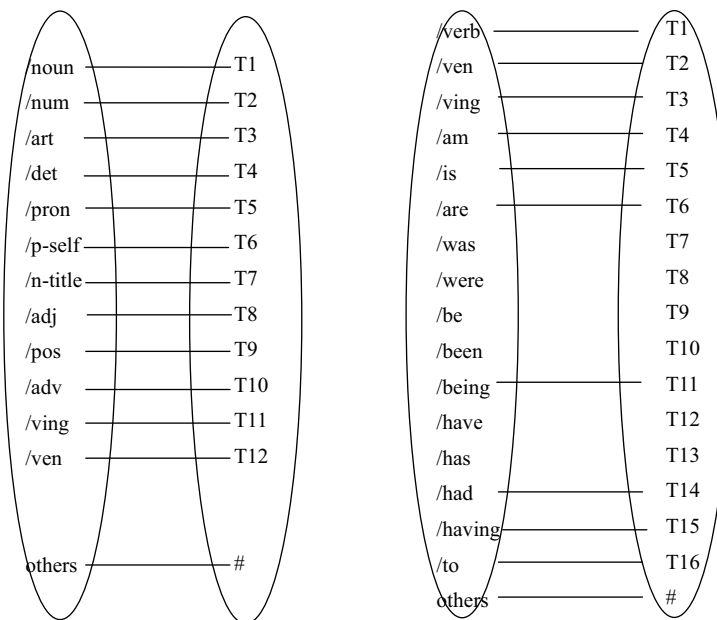


Figure 4: The symbolic mapping from the system terminator to the analysis table terminator.

4 Experimental evaluation

4.1 Evaluation method

The experiment organized an evaluation team to evaluate the performance of the English subtitle automatic sentence analysis algorithm designed in this article, including the accuracy of word analysis, recognition speed, and update ability. The evaluation team includes three English–Chinese machine translators, two English–Chinese translators, and two ratings. The three English–Chinese machine translations have the same specifications and are equipped with algorithms, statistical algorithms, and weaker algorithms after

initialization. The evaluation method adopts closed evaluation and development evaluation. Closed assessment refers to the automatic recognition of phrases in specific English–Chinese translation sentences; the English–Chinese translation sentences in the development assessment are randomly selected from the Internet. After using three automatic sentence analysis algorithms to identify and communicate the translation results, two English–Chinese translators communicated to translate. The scorer compares machine translation and manual translation and scores these three algorithms according to the scoring rules. The scoring rules are as follows:

- (1) The accuracy of the recognition algorithm, recognition speed, and update capability are, respectively, 90, 5, and 5% of the total level.
- (2) The certification accuracy scoring rules (not based on input errors) are as follows:
 - 100 points: The meaning of the translation structure is very accurate, and the grammatical structure does not need to be changed;
 - 80 points: The overall explanation is clear; there are minor grammatical errors; if necessary, some simple modifications should be made;
 - 60 points: The overall explanation is clear; there are many grammatical errors that must be changed, otherwise the meaning is not clear;
 - 40 points: Part of the explanation is clear, there is no obvious error in the expression, and the overall explanation is not continuous;
 - 20 points: All and part of the explanation is very confusing, and the sentence explanation is obviously wrong;
 - 0 points: The whole explanation and part of the explanation are very confusing and difficult to understand.
- (3) The recognition rules of recognition speed and update ability use weighted average method, that is, the total recognition time and total update time of the algorithm are multiplied by the weight, and then the sum is divided by the number of phrase analysis.

4.2 Evaluation results

The experiment performed phrase analysis on 60 sentences in closed assessment and development assessment. The evaluation results of the three algorithms are shown in Tables 2 and 4. The algorithm in this article has the highest score in the result evaluation, with an average of 92.3 points, while the lowest is the statistical algorithm with 75.1 points. The evaluation result of the dynamic memory algorithm is 91.2 points, which is not much different from the algorithm level in this article, but the update capability of the dynamic memory algorithm is seriously insufficient. From a long-term perspective, the algorithm in this article is more practical.

Table 4: Evaluation results of the algorithm in this article

Project	Closed beta	Develop test
Identification accuracy/point	95	93
Identify moderation/score	76	78
Update ability/score	78	76
Score/point	93.2	91.4

5 Examples

The following briefly introduces the analysis process of the phrase “clock has been repaired.”

Step 1: Identify the cursor pointing/having o4. First use Tab1eA diagnosis: According to Tab1eA symbol mapping function, /corresponds to the stop number, at the initial position 0, the corresponding function ERR. TableA analysis failed. Then use Tab1eB to analyze: according to tab1eB mapping function symbols, /have/be/ven correspond to stop T15, T10, and T2, respectively. In the initial position 0, after entering the functions S7, S8, and S10, the diagnostic cursor moves down to /comma; /comma corresponds to the shutter #, in state 10, the corresponding action is for R4, according to the law of rule 4, from the symbol key, take a picture of the symbols/have, /being and /ven, and use uncertain letters * to form a sub-sibling tree * screw in the symbol key, take the middle position 7, 8, and 10 in the status key, please refer to the current position and symbol of the stack. The state 0 and the symbol *Ving of the top cell of the stack are used to check the GOTO table Tab1eB and press the state 4 in the stack. Since the zoom out function does not move the diagnostic cursor down, the current diagnostic cursor still points to the symbol /comma, which corresponds to the stop sign. In state 4, the diagnostic function is ACC, so Tab1eB receives it. Create a block in the top cell of the symbol key *Wing to point to the child's siblings. The diagnosis result indicates/ has, has been, and/or shown in Figure 5. In the figure, the solid lines indicate signs of children and brothers, and the dotted lines indicate the words in the middle.

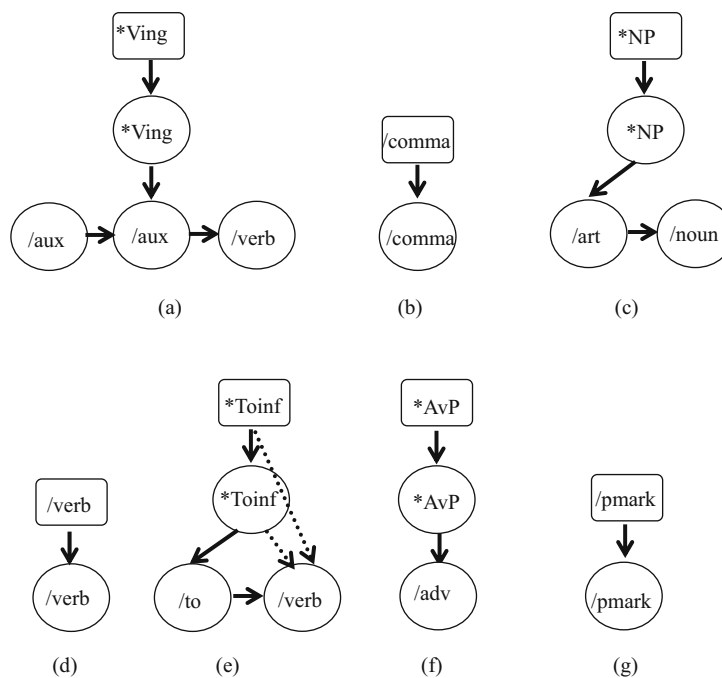


Figure 5: The analysis result of the symbol/comma (a–g).

Step 2: The current diagnostic cursor points to /comma. First use Tab1eA to analyze: According to the symbol mapping function, Tab1eA corresponds to the /comma stop sign. At the initial position 0, the corresponding function ERR TableA analysis failed; Q) Then use Tab1eB to analyze: According to the character mapping function, Tab1eB corresponds to the /comma stop sign. At the initial position 0, the corresponding function ERR, TableB analysis failed. Create a block to indicate the current diagnostic symbol /comma, the diagnostic cursor moves down, and the analysis result of the symbol/comma is shown in Figure 5.

Step 3. The current diagnostic cursor similarly points to /ae3, first use Table A to switch.

According to rules 13.9, 7, and 2, recognize/art/noun as *NP; then create a block in the top cell of the symbol key *NP to point to the child's siblings.

Step 4: The current diagnostic cursor points to “Wednesday.” Similarly, first use Table A and Table B for analysis, but both failed, and then create a block to indicate the current diagnostic symbol /ved, and then move the diagnostic cursor down, the symbol /ved diagnostic result is shown in Figure 5.

Step 5: The current diagnostic cursor points /points. Similarly, according to rule 6, first use Table A to diagnose and fail, then use Table B to diagnose and recognize /to/verb as *Toinf, and finally, create a block to point to the child’s *Toinf siblings. The symbol /t diagnosis result and /verb are shown in Figure 5.

Step 6: The current diagnostic cursor points to /adv. Similarly, according to Rule 29, using Table A analysis, /adv is recognized as “AvP; a block is created for “AvP.” The diagnostic result of the symbol /adv is shown in Figure 5f.

Step 7: The current diagnostic cursor points to /pmark. Similarly, the analysis using Table A and Table B fails. Create a block for /pmark and move the diagnostic cursor down. The diagnostic result of the symbol/signal is shown in Figure 5.

Step 8: After analyzing the entire sentence, stop. The analysis result of this sentence is shown in Figure 6.

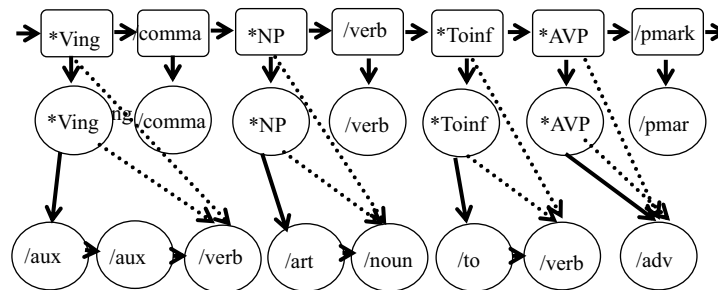


Figure 6: Sentence “Having been repaired, the clock began.”

6 Discussion

In the process of translating English–Chinese phrases, the ambiguity between English and Chinese grammars makes the task of analyzing sentences very difficult. Aiming at this problem, this article designed a translation system based on the extended CFG grammar and improved GLR algorithm. It was used for the translation correction of English phrases, and its usability was evaluated by the method of experimental evaluation.

First of all, in terms of the recognition accuracy, whether it was a closed test or a development test, the proposed algorithm was significantly better than the statistical algorithm, and its difference with the dynamic memory algorithm was small. From the perspective of the recognition speed, the score of the statistical algorithm was 73 and 76 points, respectively, which were lower than the dynamic memory algorithm and the algorithm proposed in this study. Finally, from the perspective of the update ability, the scores of the statistical algorithm and dynamic memory algorithm were both below 70 points, and the scores of the algorithm proposed in this study were 78 and 76 points, respectively, in the closed test and development test, showing significant advantages. Overall, the performance of the algorithm proposed in this study was better, and the final scores were 93.2 and 91.4 points, respectively. In the closed test, the score of the algorithm proposed in this study was 21.2% higher than the statistical algorithm and 1.53% higher than the dynamic memory algorithm. In the development test, the score of the algorithm proposed in this study was 24.69% higher than the statistical algorithm and 0.88% higher than the dynamic memory algorithm. The experimental results showed that the algorithm proposed in this study had better performance in the translation of English phrases and was more suitable for promotion and application in practice.

7 Conclusion

This article proposed an extended CFG grammar to describe English phrases and realized the translation correction of English phrases through the improvement of the GLR algorithm. It was found from the experimental results that compared with the statistical algorithm and dynamic memory algorithm, the method designed in this study had significant advantages in recognition accuracy, recognition speed, and update ability. In conclusion, the improved GLR algorithm has better performance in phrase analysis and is more suitable to be applied in actual translation systems to improve the efficiency and accuracy of syntactic analysis.

Conflict of interest: Author states no conflict of interest.

References

- [1] Dang SS, Gong XT. Design of intelligent recognition English translation model based on improved GLR algorithm. *Comput Meas Control*. 2020;28(259):166–9.
- [2] Guo YH, Wu BM, Wang BX. A shallow syntactic analyzer for English–Chinese machine translation based on GLR algorithm. *Comput Eng Appl*. 2004;(34):124–9.
- [3] Zheng YE, Zheng ZM. Discussion on the syntactic analysis of English long difficult sentences based on GLR algorithm. *J Anyang Norm Univ*. 2015;(2):32–5.
- [4] Zhu LQ. Automatic phrase recognition algorithm in English–Chinese machine translation. *Mod Electr Tech*. 2017;(15):126–8.
- [5] Guo YH. Research on key technologies of English–Chinese machine translation system [dissertation]. PLA Information Engineering University; 2006.
- [6] Huang DX. The algorithm design of English translation software translation accuracy correction. *Mod Electr Tech*. 2018;41(14):178–80+185.
- [7] Yuan SX. Research on Chinese and English cross-language plagiarism recognition technology based on translation features and content [dissertation]. Shanghai Jiaotong University; 2011.
- [8] Long W, Ta L, Jian L. Error Correction Algorithm for Chinese Speech Recognition Based on Phrase Translation Model [A]. Speech Information Committee of Chinese Information Society of China. [C]. Proceedings of the 14th National Conference on Human-Machine Speech Communication (NCMMSC '2017). Phonetic Information Committee of Chinese Information Society of China: Tsinghua National Laboratory of Information Science and Technology (Preparatory); 2017. p. 6.
- [9] Qiu N. A simultaneous interpretation box using BIRCH clustering algorithm to translate Mandarin and English; 2018.
- [10] Yao L. Research on computer intelligent proofreading system based on improved phrase translation model. *Electron Des Eng*. 2020;28(18):52–55+59.
- [11] Zhou Y. Advancing fundamentals and applications of X-ray birefringence imaging. Cardiff University; 2018.
- [12] Luo Y, Li M, Zhu J, Hu Guanlong. Research and implementation of machine translation decoding algorithm based on phrase statistics. *Comput Eng Appl*. 2007(30):171–173+178.
- [13] Liu S, Li S, Zhao TJ, Zhang CY, Wang B. Research on phrase machine translation algorithm based on pruning strategy. The Fifth National Machine Translation Conference, China; 2010. p. 187–93.
- [14] Rongwan L, Puspani I. Shifts in translation of English noun phrases into Indonesian with reference to the short story a scandal in Bohemia. *Humanis*. 2019;23(3):167.
- [15] Zheng LT, Shi HL, Yu WH, Dong YS, Fan QH. A method for manuscript translation optimization based on KNN algorithm. *J Chromatogr A*. 2015;1405:23–31.
- [16] Lu SX. Research on model training, adaptation and learning algorithms in statistical machine translation. *J arthroplasty*. 2014;29:923–8.
- [17] Miao HX, Cai DF, Song Y. Phrase-based statistical machine translation method. *J Shenyang Instit Aeronaut Tech*. 2007;(2):34–6.
- [18] Zhang H. Research on phrase statistics and machine translation decoding algorithm [dissertation]. Northeastern University; 2012.
- [19] Hao Z. Research on decoding algorithm for phrase statistical machine translation. Northeastern University; 2015.
- [20] Yutao Q. Research on decoding optimization technology of statistical machine translation based on phrase. Northeastern University; 2015.
- [21] Qiang J, Zhang J. Improvement of phrase extraction algorithm in phrase-based statistical machine translation. *Comput Eng Appl*. 2018;44(13):147–9.

- [22] Ren GJ, Turgen I, Aishan W. Improvement of Chinese-Uyghur phrase extraction algorithm in phrase-based statistical machine translation. *Mod Comput (Prof Ed)*. 2010;(5):11–3.
- [23] Su Q. Optimized simulation of English machine translation under weakened grammar rules. *Comput Simul*. 2016;57(11):64–76.
- [24] Lv CT. Research on Chinese-Vietnamese phrase machine translation method based on language characteristics. Kunming University of Science and Technology; 2016.
- [25] Wang X. Research on translation knowledge optimization method in statistical machine translation [dissertation]. Soochow University; 2014.
- [26] Chunxiang Z, Tiejun Z, Sheng L. Acquisition of translation equivalence pairs based on multiple linear regression model. *Comput Eng Appl*. 2006(04). 1–3+26 .