

Translation of English Complex Long Sentences Based on Natural Language Information Processing and Artificial Intelligence Compression Perception Algorithm

Ruichao Li

School of English Language and Culture, Xi'an Fanyi University, Xi'an, Shaanxi, China

Keywords: English Translation; Complex Long Sentences; Compressed Perception; IoT Technology

Abstract: This study aims to explore translation optimization method of complex English long sentences by combining natural language information processing with artificial intelligence compressed sensing algorithm. Through the English-Chinese translation experiment of scientific and technological texts in the field of Internet of Things (IoT), the translation effects of the compressed sensing algorithm and the traditional algorithm are compared and analyzed. The experimental and control groups consist of English majors, and five tests are used to evaluate translation accuracy. The experimental results show that the application of the compressed sensing algorithm significantly improves translation quality in the translation of long and difficult sentences. Especially in the Chinese-English translation task, the experimental group has a significant advantage over the control group in the second test score, and as the translation difficulty increases, the algorithm effect is more prominent. In addition, at the vocabulary and grammar level, the experimental group's accuracy rate far exceeds that of the traditional algorithm group. This study verifies the effectiveness of the compressed sensing algorithm for translating complex, long English sentences, especially in high-difficulty translation tasks.

1. Introduction

With the acceleration of globalization, international political, economic and cultural interactions are growing exponentially. Statistics show that the number of global cross-border cooperation projects will increase by 15% annually from 2020 to 2023. As the primary carrier of information, the quality of language translation directly affects the efficiency of international cooperation. As an international common language, English plays a leading role in cross-cultural communication, but translating complex, long sentences has long been a key bottleneck in language conversion. Although existing machine translation technology can quickly generate basic translations, it is insufficient in handling multiple nested structures, semantic logical coherence and cultural context adaptability.

Technological developments have had a major impact on translation training, and Pan J examined learning corpora and how this research can contribute to the teaching and learning of translators and interpreters. The results show the complexity of learners' language and its intricate

relationship with various factors [1]. Starting from the influence of cultural context on Chinese-English translation, Zhang X discussed the context in Chinese-English translation, as well as the understanding and practice of translation activities from the perspective of cultural translation in the way of practical work experience. The study finds that personal cultural background significantly influences translation outcomes [2]. Najjar I explored the use of hyperbole in the Qur'an and its English translations. He sought to determine how translation strategies affect the quality of the research data; in some cases, these strategies can exaggerate the function of the Qur'an. As a result, translation quality is sometimes poor [3]. Although these theories provide a brief introduction to English translation, they are not combined with compressed sensing algorithms.

With the rapid development of smart sensor applications for the Internet of Things, it has become necessary to process huge amounts of data from video signals in real time. Motion estimation (ME) is necessary to improve quality, but it is computationally intensive in video compression systems. Saha A proposed a context-aware adaptive mode based ME algorithm for multimedia IoT platforms. The experimental results show that the encoding times for the Low Latency Main (LB-main) and Random Access Main (RA-main) profiles are reduced by 40% and 36%, respectively, in the HEVC test model 16.10 encoder [4]. Chi Z introduced a remote sensing image fusion algorithm. After this fusion method, higher correlation coefficients can be obtained. Compared with other methods, the fused images can carry more spatial information [5]. In order to solve the problem of compressive sensing reconstruction of water resources data in the water Internet environment, Lin G proposed a compressive sensing reconstruction method for water Internet environment based on particle swarm optimization (PSO) and OMP (orthogonal) MOPSO (multi-objective particle swarm optimization). Then, for the theory of compressed sensing signal reconstruction algorithm, he summarized the geometric model and algorithm steps of the greedy algorithm and several commonly used algorithms. He studied the optimization problem of signal reconstruction based on compression from three aspects: signal sparsity, measurement matrix and reconstruction algorithm. The experimental results show that the compressive sensing signal reconstruction algorithm based on MOPSO outperforms the OMP algorithm in visual quality and objective metrics, and performs as well as or better than deterministic measurement matrices. Under the same compression ratio and reconstruction algorithm, the signal-to-noise ratio and peak signal-to-noise ratio can be improved by 5dB compared with the deterministic measurement matrix. The relative error is about 0.6% [6]. Hassager H G investigated the effects of fast hearing aid compression on the spatial perception of normal and hearing impaired listeners in a reverberant environment. The results show that both independent and chained compression produce more dispersed, wider sound images [7]. The problem of perceived data sparseness is prevalent in cross-border e-commerce personalized filtering recommendation models. In order to change the user similarity calculation based on compressed sensing, Jia Z used particle distance (PD) to realize the similarity calculation of cross-item mobile users. He proposed an e-commerce personalized filtering recommendation algorithm based on particle-level fusion between compressed sensing and e-commerce customer trust. The experimental results show that, compared with the cosine method and the Pearson method, the user similarity calculation method fused with granular compressed sensing can alleviate the impact of sparse sensing data on the personalized filtering algorithm of e-commerce. The algorithm can improve the accuracy of e-commerce personalized recommendation [8]. Although these theories describe the compressive sensing algorithm, it is less relevant to English translation and less practical.

2. Long Sentence Translation of Compressed Sensing Algorithm in the Internet of Things

With the acceleration of globalization, there are more and more exchanges between countries in the world and more frequent cultural exchanges. Language communication takes on an essential function in every process [9-10]. The advancement of modernization and reform has led to a high rate of integration between China and the world. As technology continues to deepen, it has become possible to adapt the IoT and compressed sensing algorithms to the translation industry [11].

As a representative product of the new age of innovation, the IoT is a key player in our routine work and life [12]. Judging from the development track of human history, every major technological change will bring huge changes to the way of production and life of human beings. In terms of the development of the Internet of Things, the exploration of the Internet of Things began at the end of the last century, but it did not initially appear as the Internet of Things. Before the concept of the Internet of Things appeared, M2M is now a common way of the Internet of Things [13-14].

In the early days of IoT research, different fields have different views on the concept of IoT. Although some scholars believe that the Internet of Things is a network developed based on RFID technology and Internet technology, there is currently no recognized connotation [15-16].

The Internet of Things is an evolving process, which has developed from the early ubiquitous computing ideas to today's smart models. Its connotation is constantly enriched, and the technology is constantly improving and developing [17].

Cultivating translation ability is the core task of translation teaching, and evaluation methods for translation ability are becoming increasingly mature; research on translation ability has also gone through different periods [18]. As a cognitive skill, translation mastery benefits the development of personal translation ability. The prerequisite for the development of translation skills is the relevant knowledge and adequate training. Therefore, in order to improve translation skills, students need to carry out sufficient training under the guidance of translation theory.

In the process of English translation, it will be affected by many factors. First of all, we look at the single factor level. At this stage, people think that translation is mainly influenced by language ability. Although this point of view has made progress compared with the previous one, at this stage, scholars' understanding of translation ability still stays at the language level, ignoring the influence of other factors on translation ability. Although language is an important factor affecting the quality of translation, it is not the whole of deductive work. In fact, we can see from the translation teaching process that in the process of professional courses, we not only need to learn language skills, but also translation theory and related cultural background are the key points that need to be paid attention to. With the advancement of science and technology, the connotations of the materials we need to translate become wider and wider, but vision has also become wider. This makes translation ability continuously transition from a single-factor to a multi-factor stage.

With the continuous progress of the economy, the continuous upgrading of science and technology, the process of globalization is also intensifying, and people's understanding of translation ability is also deepening. We cover a variety of capabilities within the scope of translation capabilities, making translation quality more accurate and reasonable. At this stage, the PACTE translation model with great influence appeared in European and American countries. It involves the translation of multiple languages, and the research on translation ability is mainly carried out from the aspects of psychology, linguistics, pedagogy and the professional knowledge involved in translation. Bilingual and non-linguistic competencies are also included in translation.

Compressed sensing brings a new direction to signal acquisition through numerical optimization. It uses non-adaptive linear projection to maintain the structure of the original signal, collects the signal at a certain frequency, and then reconstructs the original signal through numerical

optimization.

$$f = \delta\gamma \quad (1)$$

In the above function expression, γ represents the sparse original signal, and f represents the signal obtained after the compression of the signal in compressed sensing. Figure 1 shows the theoretical framework of compressed sensing:

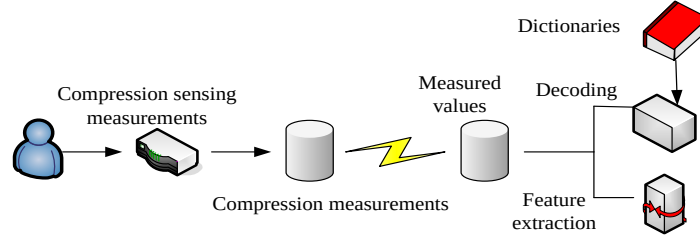


Figure 1. Theoretical framework for compressed sensing

$$f = \delta\gamma + b \quad (2)$$

In addition to the acquisition of the original signal, we also need to discuss the transmission and storage of the signal. In formula (2), b represents the noise, and $\delta\gamma + b$ represents the corresponding mathematical model.

$$k = Ua \quad (3)$$

In the above function expression, a represents the sparse signal, k represents the observed quantity, and U represents the measurement matrix.

$$(2 - \rho_q)\|a\|_3^3 \leq \|Qa\|_3^3 \leq (2 + \rho_q)\|a\|_3^3 \quad (4)$$

When the measurement matrix does not meet the conditions, it means that ρ_q and 2 are quite different.

$$(2 - \rho_{2q})\|a_1 - a_2\|_3^3 \leq \|Q(a_1 - a_2)\|_3^3 \leq (2 + \rho_{2q})\|a_1 - a_2\|_3^3 \quad (5)$$

If ρ_{2q} is less than 2, all sparse signals with sparsity q can be reconstructed in the measurement matrix.

$$\rho_{3q} + 3\rho_{4q} < 3 \quad (6)$$

$$\|a^* - a\|_3 \leq T_0 \frac{\|a^* - a_q\|_3}{\sqrt{q}} \quad (7)$$

$$\|a^* - a\|_3 \leq T_0 \leq \|a^* - a_q\| \quad (8)$$

T_0 represents some fixed constants, a_q represents the vector composed of the largest q amplitudes in the signal a , if the signal a is q sparse, then $a = a_q$, then we call it accurate reconstruction.

$$f = Qa + p \quad (9)$$

In the display perception device, only the extracted signal will have noise, and formula (9) represents the noise compressed sensing model.

$$\|p\|_3 \leq \alpha \quad (10)$$

In the above functional expression, p represents a noise signal, and α represents a finite constant.

$$\min \|a\|_1 \in Qa \quad (11)$$

Formula (11) represents the most primitive model of compressed sensing, and the a at this time is the exact solution.

$$\|a\|_1 = \lim_{\mu \rightarrow 0} \frac{2}{\mu} \sum |a|^\mu \quad (12)$$

$$\lim_{\mu \rightarrow 0} \frac{2}{\mu} \sum |a|^\mu - 2 = \sum \log|a| \quad (13)$$

At this point, we can transform formula (11):

$$\min \sum \log|a| \quad (14)$$

In practice, $a = 0$ is very common, which leads to the absence of $\log|a|$. So we usually convert it to:

$$\min \sum \log(|a| + \sigma) \quad (15)$$

$$\text{S. t. } Qa = f \quad (16)$$

At this time $\sigma > 0$ is a number greater than zero, so $\log|a| \neq 0$, which makes the above function have operational significance.

3. Long Sentence Translation Algorithm Experiment

This study selected 40 undergraduate students majoring in English as experimental subjects, with 20 people in the experimental group and 20 people in the control group (male-female ratio 1:1). All participants passed the English major level 4 test, and there was no significant difference in the Chinese and English scores in the college entrance examination, ensuring the consistency of the language ability baseline. The experimental group used a translation tool based on the compressed sensing algorithm, and the control group used the traditional algorithm. Through five English-Chinese/Chinese-English long and difficult sentence translation tests, the performance of the two groups in complex sentence translation was compared and analyzed. The experimental design focuses on the independent influence of the algorithm performance by controlling variables such as academic background and language level. The following is the specific information of the students:

Table 1. Basic information of students in the experimental group

Number of people	Male	Female	Average
	10	10	10
Age	22	18	20.5
GCE Language Results	129	109	118
GCE English Result	135	116	126
Specialized 4 grades	82	61	72

According to the data in Table 1, there are 10 boys and 10 girls in the experimental group, the oldest is 22 years old, the youngest is 18 years old, and the average age is 20.5 years old. The highest score for Chinese in the college entrance examination is 129, the lowest is 109, and the average score is 118. Because the native language of the students participating in the experiment is Chinese. Therefore, the students' analytical and comprehensive ability can be judged according to their Chinese performance. The highest score in the college entrance examination is 135, the lowest is 116, and the average score is 126. The highest score for the fourth grade is 82, the lowest is 61, and the average score is 72.

Table 2. Basic information of students in the control group

Number of people	Male	Female	Average
	10	10	10
Age	21	19	20.2
GCE Language Results	128	108	116
GCE English Result	135	115	127
Specialized 4 grades	83	62	72.3

According to the data in Table 2, there are 10 boys and 10 girls in the control group; the oldest is 21 years old, the youngest is 19 years old, and the average age is 20.2 years old. The highest score for Chinese in the college entrance examination is 128, the lowest is 108, and the average score is 116. The highest score in the college entrance examination is 135, the lowest is 115, and the average score is 127. The highest score of the fourth grade is 83, the lowest is 62, and the average score is 72.3. According to the relevant information on the two groups of students, there is no obvious difference between the students in the experimental group and the control group in terms of overall situation, and the students in both classes have passed the professional four-level examination. It can be seen that these data are consistent with the actual situation, so the selected samples are representative.

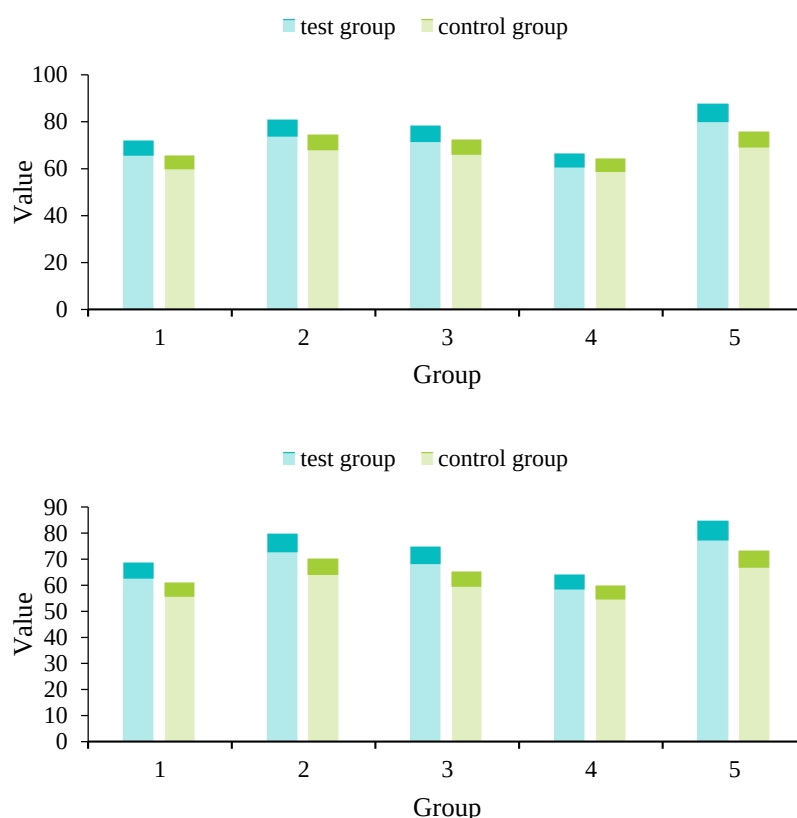


Figure 2. Comparative analysis of English translation of long sentences by two groups of students

In the appeal argument, we analyze the translation of English short sentences. In fact, in the process of English translation, long, difficult and complex sentences occupy an important position in the translation process. According to Figure 2, we conducted a comparative analysis on the translation of long and difficult sentences between the two groups of students. Similar to short-sentence translation, we conducted five sets of English-to-Chinese and Chinese-to-English

tests, respectively. First of all, in the English-to-Chinese situation, the experimental group scored 72.1 points in the first test, and the control group scored 65.4 points in the first test. The second test mark for the pilot team was 81.2 and the second test mark for the control team was 74.6. The third test mark for the pilot team was 78.5, and the third test mark for the control team was 72.4. The fourth test mark for the pilot team was 66.5 and the fourth test mark for the control team was 64.5. The fifth test mark for the pilot team was 88, and the fifth test mark for the control team was 75.8. The data show that in the course of interpreting long and hard sentences from English to Chinese, the students in the experimental group using the compressed sensing algorithm had a higher translation accuracy rate, while the gap was more pronounced in the moderately difficult material. It is evident that the compressed sensing algorithm performs well in the conversion of long and complex sentences from English to Chinese.

4. Long Sentence Translation Algorithm

Different aspects of knowledge need to be examined in the English translation test. It includes English vocabulary and grammar, as well as reading ability. Therefore, it is necessary to analyze the specific situation of relevant aspects when discussing students' English translation ability. In order to explore the students' translation ability under different algorithms, a comparative analysis of the scores of the two groups of students was conducted. The details are as follows:

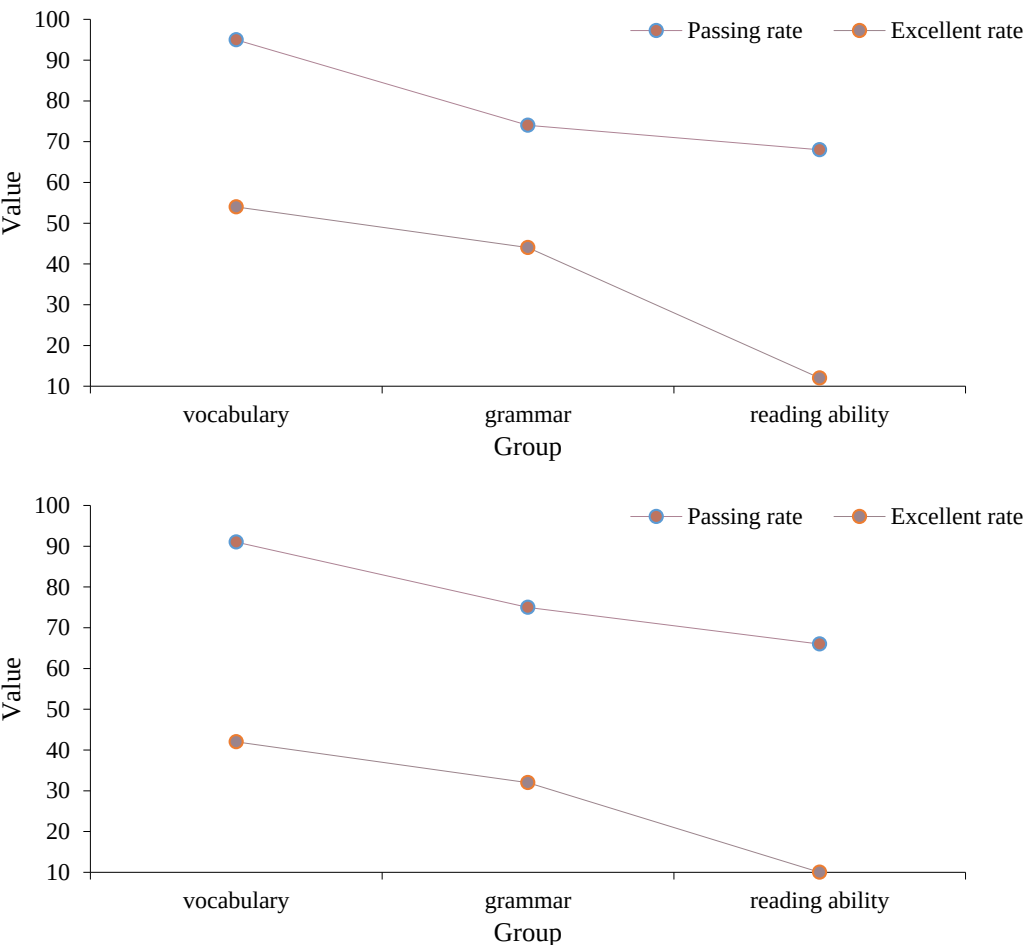


Figure 3. Comparative analysis of students' English translation under different algorithms
 According to the data in Figure 3, the students participating in the experiment used the

compressed sensing algorithm and the traditional algorithm for English translation. First of all, from the perspective of the experimental group, there are 19 people with qualified vocabulary, up to 95%, and 11 people with excellent vocabulary, up to 54%. There are 15 people with qualified grammar, up to 74%, and 9 people with excellent grammar, up to 44%. There are 13 people who are qualified in reading ability, up to 68%, and 3 people who are excellent in grammar, up to 12%. According to the data, under the compressed sensing algorithm, students have better results in vocabulary and grammar, but relatively poor in reading. Overall, the students' situation is good, and the pass rate is higher than average.

Table 3. Analysis of the basic situation of English translation

Translation problems	Number of people	Proportion
Translation uncertainty	7	35
Not specifically studied	8	37
Less practice	5	28
Total	20	100

According to the data in Table 3, a brief investigation was conducted into why students' expressions are relatively outdated when translating long and difficult sentences. According to the specific situation, there are 7 people who do not know how to translate because they usually lack attention to it, reaching 35%. The number of people who did not specifically learn related translation methods was 8, accounting for 37%. Because of less practice, the usual translation forms are outdated for 5 people, accounting for 28%. According to this situation, students lack necessary auxiliary teaching materials in the process of daily English translation learning, so they are not proficient in some translation forms, resulting in outdated English translation forms.

Table 4. Survey on Students' Opinions on Translation Teaching Activities

Teaching Methods	Number of people	Proportion
Interpreting classic translations	5	12.5
Evaluation of student translations	8	20
Group Discussion	11	27
Translation Theory	3	7
Translation Projects	15	37

According to the data in Table 4, a brief survey of related activities in translation teaching was conducted. In this context, 5 people, or 12.5%, adopted the teaching method of interpreting classic translation works. There are 8 people who use the evaluation of students' translation works for teaching, up to 20%. The group discussion teaching method was adopted by 11 people, accounting for 27%. There are 3 people who adopt the teaching method of translation theory, up to 7%. There are 15 people who organize students to translate projects, accounting for 37%.

According to the data in Figure 4, a comparative analysis was conducted on the understanding of discourse awareness under the two groups of translation methods. First of all, in the experimental group, 7 people can basically fully grasp the theme of the translation materials, up to 35%. In terms of translation content, 20 people can keep the content consistent, up to 100%; 13 people can fully understand the author's writing purpose, up to 66%. 15 people can fully grasp the upper and lower sentences of the material, up to 75%; 10 people can consider the order and relationship of the relevant content when translating, up to 50%. 14 people were able to fully understand the subject of the translated material, up to 70%.

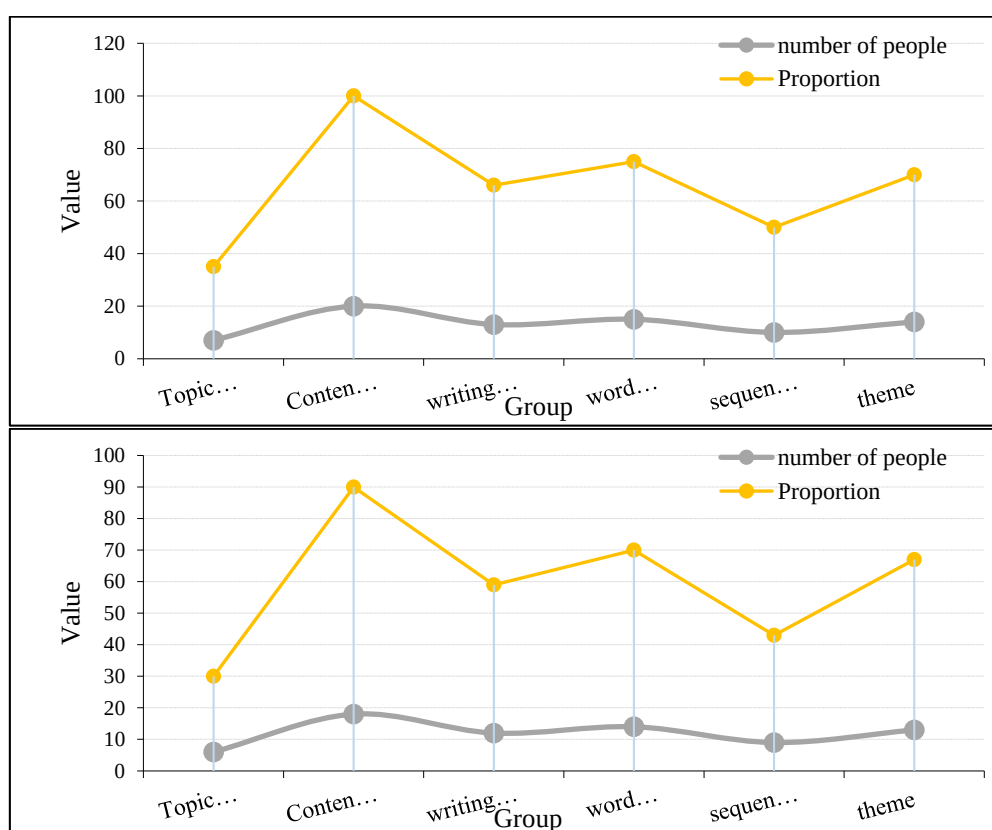


Figure 4. Comparative analysis of discourse awareness survey results

5. Discussion

This study has important implications for translation teaching and practice. The experimental results show that neglecting cultural background and translation theory in traditional translation teaching is a main reason for the outdated form of translation. The compressed sensing algorithm provides a structured training framework for students by strengthening semantic logic and discourse coherence. It is recommended to introduce algorithm-assisted tools in teaching and conduct practical training with multi-domain texts to address students' knowledge gaps. At the same time, developers of translation tools can refer to the parameter optimization strategy of this study to design more adaptive algorithm modules and promote the application of intelligent translation technology in practical scenarios.

6. Conclusion

In globalization, language communication is important for national communication, making accurate English translation essential, especially for complex long sentences. This study integrates natural language processing and compressed sensing to verify the effectiveness of intelligent technology. Experiments show the dynamic compressed sensing model improves semantic coherence and translation accuracy through adaptive parameter optimization and sparse signal reconstruction. In Chinese-English tasks, the experimental group's average score is nearly 14% higher than traditional algorithms, especially for multiple nested sentences in IoT texts. This confirms the algorithm's theoretical advantages in suppressing information redundancy and noise interference, supporting high-difficulty translation tasks.

Acknowledgement

This work was supported by Research on Translation and Dissemination of Zhongnan Culture from a Communication Perspective (23JP052) and the Youth Innovation Team of Shaanxi Universities (22QNCXTD01).

References

- [1] Pan J, Wong T M, Wang H. Navigating learner data in translator and interpreter training: Insights from the Chinese/English Translation and Interpreting Learner Corpus (CETILC). *Babel*, 2022, 68(2):236-266.
- [2] Zhang X. A Study of Cultural Context in Chinese-English Translation. *Region - Educational Research and Reviews*, 2021, 3(2):11-14.
- [3] Najjar I, Kadhim K A, Al-Heeh S. Morphological shift of hyperbolic patterns in the Quran, with reference to English translation. *FORUM*, 2021, 19(2):148-163.
- [4] Saha A, Lee Y W, Hwang Y S. Context-aware block-based motion estimation algorithm for multimedia internet of things (IoT) platform. *Personal and Ubiquitous Computing*, 2018, 22(1):1-10.
- [5] Chi Z. Research on satellite remote sensing image fusion algorithm based on compression perception theory. *Journal of Computational Methods in Sciences and Engineering*, 2021, 21(2):341-356.
- [6] Lin G. Water resources data compression perception reconstruction optimization algorithm based on internet of water. *technical Bulletin*, 2017, 55(7):62-68.
- [7] Hassager H G, Wiinberg A, Dau T. Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment. *The Journal of the Acoustical Society of America*, 2017, 141(4):2556-2568.
- [8] Jia Z. Cross-border e-commerce personalized filtering recommendation model based on particle-level compression awareness. *Technical Bulletin*, 2017, 55(7):136-143.
- [9] Ballesteros-Pea S, Hoz V. Accuracy of Healthcare Providers' Perception of Chest Compression Depth and Chest Recoil. *Signa Vitae*, 2020, 16(1):159-162.
- [10] Wolff J G. Information Compression as a Unifying Principle in Human Learning, Perception, and Cognition. *Complexity*, 2019, 2019(2):1-38.
- [11] Meng Z, Li J, Long H. Fault Diagnosis Method for Rolling Bearings Based on Compression Information Feature Extractions. *China Mechanical Engineering*, 2017, 28(7):806-812.
- [12] Goncalves P, Moraes C, Porto M. Complexity-Aware TZS Algorithm for Mobile Video Encoders. *Journal of Integrated Circuits and Systems*, 2019, 14(3):1-9.
- [13] Chung K L, Liang Y C, Wang C S. Effective Content-Aware Chroma Reconstruction Method for Screen Content Images. *IEEE Transactions on Image Processing*, 2019, 28(3):1108-1117.
- [14] Ma Y, Gang Z, Shan L. RoFi: Rotation-Aware WiFi Channel Feedback. *Internet of Things Journal, IEEE*, 2017, 4(5):1684-1695.
- [15] Comino M, Chica A. Error-aware construction and rendering of multi-scan panoramas from massive point clouds. *Computer Vision and Image Understanding*, 2017, 157(C):43-54.
- [16] Fulton B. Suggested Readings in Korean Literature in English Translation. *Translation Review*, 2020, 108(1):104-110.
- [17] Muravev Y. TEACHING LEGAL ENGLISH TRANSLATION BY THE CASE METHOD IN RUSSIAN-ENGLISH LANGUAGE PAIR. *Humanities & Social Sciences Reviews*, 2020, 8(4):961-971.
- [18] Wen H. Intelligent English translation mobile platform and recognition system based on support vector machine. *Journal of Intelligent and Fuzzy Systems*, 2020, 38(153):1-12.