Research Article

Chinese Cross-Language Definition Recognition Method Based on DE-BP Model

Yanxia Shen,1 Yi Zhang,2 and Mingwei Zhu3

1College of Liberal Arts and International Education, Xi’an Peihua University, Xi’an, Shaanxi 710125, China
2College of Liberal Arts, Huaibei Normal University, Huaibei, Anhui 235000, China
3College of Computer Science and Technology, Xi’an University of Technology, Xi’an, Shaanxi 710054, China

Correspondence should be addressed to Yanxia Shen; shenshen@peihua.edu.cn

Received 22 May 2022; Revised 27 June 2022; Accepted 1 July 2022; Published 1 August 2022

Academic Editor: Lianhui Li

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

As an important linguistic phenomenon in verbal communication, relayed speech exists in a large number of news texts and is also one of the most prominent features of the Chinese language. However, at present, there are few systematic comparative studies on the recognition of the relayed problems in Chinese language crossover, and the existing methods are highly subjective. This paper makes a qualitative comparative analysis of the reported speech in the Chinese language by the DE-BP model, which combines differential evolution (DE) algorithm and BP (backpropagation) neural network to recognize the Chinese cross-language paraphrase. After that, we obtained some meaningful findings as follows. In the Chinese language, the frequency of indirect paraphrase is the highest, followed by direct paraphrase, while other categories, namely, free indirect paraphrase, free direct paraphrase, and narrative paraphrase of speech acts, are relatively rare. Through the identification and manual labeling of reported verbs and then the word frequency statistics, it is found that the number of reported verbs in English newspapers is dominant in general, and there is a significant difference between them.

1. Introduction

Relaying speech is a basic form of human language. When people want to convey what has happened in the past, what is happening now, or what may happen in the future, either in their own words or in the mouths of others, they need to communicate through relaying speech [1]. Only in this way has human civilization survived; without speech, language has been reduced to mere communication tools. The most basic feature of reported speech is reflexivity (i.e., people use language to refer to language itself). People report not only what others say but also what others comment, criticize, and so on. Accordingly, the definition of relayed speech generally involves two features, namely, reflexive feature and two-layer structure [2]. The reported speech is defined as the speech of the words, words of words, and also the words of the words and words about words. The uniqueness is embodied in paraphrased words. Two different contexts or styles can also appear in a single syntactic structure and in the process of interaction between speakers and hearers [3].

Paraphrased words have common features in all human languages. Now, we know that all languages have been paraphrased in one way or another; even those innate configurations of bad language also have their own paraphrased way and sometimes may have to rely on contextual characteristics; paraphrased words have a certain universality in human behavior. Most languages have their own paraphrased way; as said, paraphrased words as a kind of double-layer structure in the process of the formation and development of human language are indispensable and play an important role. Every conversation is filled with numerous paraphrases and interpretations of what others have said. According to this line of thinking, every speech is actually a free translation because every speech is a patchwork of quotes, assimilation, and conversion of others’ words [4, 5].

As for the terms of paraphrased speech, the usage is more complicated. For the paraphrased speech (perhaps the most well-known term so far), there is no unified opinion, and there is no consistent view among the various expressions of the verbal paraphrase phenomenon. By telling a story,
considering the ability of human memory and expression, most of the words may not be original but transferred from other people or other media [6].

The research on the recognition of Chinese language paraphrase will further enrich the research on paraphrase, which is a common phenomenon in communication, no matter in daily conversation or written discourse, which is full of a large number of other speeches. On the one hand, it can explain the ideographic meaning of public discourse, and on the other hand, it can deepen the understanding of the discourse reported by people in the news. As a language form, reported speech not only contains two speech forms and contexts but also reflects the system and essence of social communication and cognition. Analysis of relayed speech helps to show the conversational nature of language: constructive studies of dialogue, especially of relayed forms, because they reveal the basic and constant active acceptance of other speakers that constitutes the essential feature of dialogue. Human personality is only a part of the society as a whole, which is realistic from the perspective of history but productive from the perspective of culture [7, 8]. The study of reported speech is helpful to further understand the sociological essence behind these phenomena. In addition, the study of reported speech also helps to reveal the complex cognitive mechanism of human beings, including perception, storage, memory, and verbal expression, and can further enrich the different perspectives on paraphrased speech [9].

It can further verify the feasibility of the critical linguistics and corpus-combined technology. Critical discourse analysis has made great achievements in revealing the relationship between language and ideology, but criticism and doubts about it have been constantly heard. Although many scholars put forward criticisms, they also put forward many constructive suggestions, such as proposing that only reliable conclusions can be drawn based on a large corpus to strengthen critical discourse analysis. Contrastive analysis is an important means of language research [10]. It is of great academic and practical value to reveal the common and special laws between one's own language and other's language by studying other languages or by contrastive analysis of different languages. This research is faced with two main difficulties: First, the research on free translation of Chinese is not enough, and there is still a big gap in theory and category. Second, there are few direct references. Because of the above difficulties, it is of great significance to strengthen the research on the recognition methods of Chinese free translation [11].

From the perspective of application, it will help improve the level of Chinese journalists’ news reporting, distinguish foreign news pitfalls, maintain the national image, improve the ability of critical thinking and reading, and enlighten the field of foreign language teaching and translation to a certain extent. From the point of communication, the success of external publicity is largely determined by the effect of information dissemination. In fact, China’s current external communication capacity is far from satisfying the image of a rapidly rising power. This study is helpful in enhancing the mutual understanding between China and foreign countries, improving the effectiveness of communication, and avoiding misunderstanding and has a strong realistic and contemporary significance [12, 13]. It gives a qualitative comparative analysis of the reported speech in the Chinese language from the types of reported speech by the DE-BP model and obtains some meaningful findings.

This section mainly introduces the research topics of discourse paraphrase in existing literature, the related research on Chinese language paraphrase recognition, and the specific application of recognition methods in news discourse, summarizes the achievements and shortcomings of previous studies, and determines the direction of this research. Discourse paraphrase is a universal phenomenon in human speech communication, which has been widely concerned by linguists, philosophers, and literary critics. Various terms have appeared successively, including quotation and paraphrase [14]. The study of language expression has always been troubled by the confusion of terms and concepts. Different terms represent different aspects of the process of language translation, and different theoretical perspectives reflect the complexity and diversity of discourse translation itself [15].

Quotation is widely used in literature, reflecting scholars’ discussions on the reflexivity of language and the nature of paraphrase. In semantic and philosophical studies, quotation is a general term, and its classification is mainly based on the use and mention of language [16, 17]. From the perspective of projection structure, the juxtaposition is a quotation, while the subject-subordinate is a report. Thus, it can be seen that quotation is not very clear in terms of conceptual significance. Sometimes it refers to the quotation of abstract language expression, and sometimes it only refers to the direct quotation of concrete discourse. Quoted speech is a term widely used in linguistic research of speech paraphrase, mainly used to refer to different ways of paraphrasing another speech. It mainly focuses on the grammatical characteristics of different types of relayed speech and the rules of phase five conversion [18]. Most of the researches on discourse relayed use direct speech and indirect speech to refer to relayed speech. A relatively general term is used when discussing paraphrasing in various styles. According to the stylistic attributes of the original discourse, three direct forms of discourse are distinguished: direct speech, direct thinking, and direct writing. Thus, relayed speech is a term mainly used in linguistic research of discourse relaying [19].

As far as English is concerned, the differences between direct and indirect discourse lie in tenses, person, and the form of expression of the original discourse. In terms of morphology, direct discourse retains the original tense of the verb, while indirect discourse needs to refer to the time in the clause to move the time of the relayed discourse. Textual free translation mainly focuses on various free translations in literary works, especially the analysis of the relationship between direct speech and indirect speech. In addition to direct speech and indirect speech, there are also free direct speech, free indirect speech, and narrative paraphrase of speech act in the Chinese language [20, 21]. Direct speech has two features that can show the presence of the narrator:
quotes and paraphrases clauses; free direct speech is to remove one or two of these characteristics; The narration of speech act only conveys the occurrence of the act, but does not convey the content and form of speech, so the content of speech is not so important in this form [22].

The pragmatic study of discourse retransmission tends to explore the mode and function of discourse retransmission in context. Speech connects two discourse events. The main term used is relayed speech, and direct speech and indirect speech are used when referring to various types of relayed speech. From the perspective of linguistic function, the paraphrase is referred to as the signal sound in the text: if the speaker or writer in some way signals the presence of another voice in the text, then any paragraph derived from that voice [23, 24]. Free translation is a misleading term that means that one can freely translate the words of another while retaining the nature of the original speech. The term used is constructed dialogue: framing discourse as dialogue is not paraphrasing but a recontextualization of words within the current discourse [25]. In the study of discourse free translation, free translation is a topic that scholars pay most attention to. The early studies mainly describe the grammatical characteristics of various types of reporting and the rules of conversion from direct speech to indirect speech under the influence of structuralism. When people want to paraphrase what they have said before or what someone else has said, there are two ways: give the exact expression of direct speech used by the original speaker; modify indirect discourse according to the relayed context [26].

The function of discourse retransmission is closely related to various ways of discourse retransmission [27]. This section mainly introduces the attitude of the narrator and the pragmatic functions of discourse retransmission in the specific discourse. In the sequence of relayed speech, direct speech is the conventional pattern, and from direct speech to the right, the sense of freedom gradually strengthened. Free indirect speech is on the more indirect side of the speech expression gradient, allowing the narrator to take two perspectives, creating a distance between the reader and the character’s words and inserting the author’s voice, resulting in a satirical effect. The sound effect of direct speech is a kind of guidance and quotation, not indirect speech with direct speech mixed with narrative language. It provides the narrator with a summary of the character’s language, and can speed up the contrast between the narrator’s direct speech and the interrogative speech and control the dialogue [28].

The narrative of speech act has the function of highly checking and concealing and turning the human and object discourse into speech act to narrate, and the author exercises the biggest intervention right. The recognition of discourse paraphrase focuses on the stylistic effects of various paraphrase types in literary works, and its advantage lies in that these categories are regarded as a conscious paraphrase of any words or thoughts by the author through the choice of structure. Attitude refers to the current relayed on the original information or speaker evaluation, including neutral, positive, and negative [29]. Therefore, no matter how precise the paraphrased words may be and no matter how pure the motives of the paraphrase may be, removing a passage from its original context and recontextualizing it in a new network of relationships are bound to interfere with its original effect [30]. Functional research of discourse paraphrase is reflected in two aspects: the intervention and paraphrase of the relayed discourse and the pragmatic functions of discourse in the whole discourse. Stylistics studies have shown that different ways of paraphrases and narrator paraphrases are related to the degree of involvement as reported speech. Recent empirical studies have fully demonstrated that discourse paraphrasing appears in different types of discourse and has multiple functions. On the one hand, citing the discourse of others can provide an objective basis for the current discourse. And different free translation methods reflect the differences between subjects. From the perspective of the existing research on speech paraphrase, the research on speech paraphrase recognition has gradually moved from static grammatical description to dynamic pragmatic study, and significant findings have been obtained [31].

This paper mainly uses the DE-BP method to study the recognition problem of Chinese cross-language paraphrase, which is beneficial to verbal communication. This research mainly consists of four parts. The first part introduces the research background and significance of comments on the framework of this paper. The second part mainly introduces the DE-BP-based Chinese cross-language paraphrase recognition. The third part introduces the experimental results and analysis. The fourth part is the summary part of this research.

Based on the above analysis, the main contributions of the paper are shown as follows:

1. It makes a qualitative comparative analysis of the reported speech in the Chinese language by the DE-BP model
2. It is found that other categories, namely, free indirect paraphrase, free direct paraphrase, and narrative paraphrase of speech acts, are relatively rare
3. It is found that the number of reported verbs in English newspapers is absolutely dominant in general, and there is a significant difference between them

2. DE-BP-Based Chinese Cross-Language Paraphrase Recognition

2.1. Chinese Cross-Language Paraphrase Recognition. Chinese cross-language paraphrase recognition is the restriction of the various contextual factors, making it impossible to accurately reproduce the original discourse, and this research fully reveals that the speech paraphrased the complex characteristics of a speech act; however, the research on fidelity of speech transmission is also lacking of detailed description, because the specific forms of speech reporting are closely related to fidelity. In a specific context (e.g., news context), what is the similarity or arbitrariness between the paraphrase and the original discourse? In addition, the relationship between various forms of retransmission and the authenticity of the content of retransmission needs further investigation. The typical Chinese cross-
2.2. DE-BP Model. Based on the above discussion, the BP (backpropagation) neural network is used here to solve the problem. By simulating the neural function of the human brain, the data samples are learned, and the link weights and thresholds between units are established to deal with complex nonlinear problems without specific functional forms. According to the error between the actual value and the expected value, from the output layer through the hidden layer to the input layer, the link weight between each layer is revised layer by layer. By iteratively correcting the weights, the gap between the observer and the predicted value becomes smaller and smaller, and the model performance becomes better and better. When the error is less than a certain value, it indicates that the network training is completed.

\[ E = \sum_{i=1}^{m} (x_i - c_i)^2. \]  

(1)

Then, we expand the above error definition to the hidden layer:

\[ E = \frac{1}{2} \sum_{k=1}^{l} [d_k - f(\text{net}_k)]^2 = \frac{1}{2} \sum_{k=1}^{l} d_k - f \left( \sum_{j=0}^{m} \omega_{jk} y_j \right)^2. \]  

(2)

We expand further to the input layer; there is

\[ E = \frac{1}{2} \sum_{k=1}^{l} d_k - f \left( \sum_{j=0}^{m} \omega_{jk} f \left( \sum_{i=0}^{n} \chi_i x_i \right) \right)^2. \]  

(3)

It can be seen from the above formula that the network input error is a function of the weights of each layer, so the error can be changed by adjusting the weights. Obviously, the principle of adjusting weights is to reduce errors continuously, so the weights should be proportional to the gradient descent of errors:

\[ \Delta \omega_{jk} = -\eta \frac{\partial E}{\partial \omega_{jk}}, \quad j = 0, 1, 2, \ldots, m; \quad k = 1, 2, \ldots, l, \]  

(4)

\[ \Delta v_{ij} = -\eta \frac{\partial E}{\partial v_{ij}}, \quad i = 0, 1, 2, \ldots, n; \quad j = 1, 2, \ldots, m. \]

Then, the weight adjustment formula of each layer is

\[ \Delta \omega_{jk}^{h+1} = \eta \delta_k y_j = n(d_k - \omega_k) o_k. \]  

(5)

According to the above rule layer-by-layer analogy, the weight adjustment formula of the first hidden layer is as follows:

\[ \Delta \omega_{pq}^1 = \eta \delta_q x_p = \eta \left( \sum_{r=1}^{m} \omega_{qr} \delta_r \right) y_q^1. \]  

(6)

There are some problems such as slow convergence speed, poor performance, uncertain learning rate, and easy to fall into local minimum. A proposed optimization algorithm for the traditional BP neural network, in the whole calculation process, generally exists in the local extremely small convergence speed and sometimes even oscillation and divergence problems to overcome its limitations in the traditional BP neural network. The differential evolution (DE) algorithm is introduced into the neural network to optimize the initial weights and thresholds of the network, which shows better nonlinear mapping ability and improves its prediction accuracy. Population initialization is as follows:

\[ x_{i,1} = x_{i}^L + \text{rand}(x_{i}^U - x_{i}^L), \quad i = 1, 2, \cdots, NP. \]  

(7)

The mutation operation formula is as follows:

\[ v_{i,G+1} = x_{i,G} + F(x_{i,2G} - x_{i,3G}). \]  

(8)

Then the interlace operation is as follows:

\[ u_{j,G+1} = \begin{cases} v_{j,G+1}, & r_j \leq CR \text{ or } j = \text{rand}(i), \\ x_{j,G}, & r_j \geq CR \text{ or } j \neq \text{rand}(i). \end{cases} \]  

(9)
Formulas (7)–(9) mainly study the mutation operation formula and interlace operation. Accordingly, the selection operations are as follows:

\[ x_{i,G+1} = \begin{cases} u_{i,G+1}, & f(u_{i,G+1}) \leq f(x_{i,G}) , \\ x_{i,G}, & f(u_{i,G+1}) > f(x_{i,G}). \end{cases} \quad (10) \]

The fitness function is

\[ f(X) = \frac{1}{N} \sum_{i=1}^{N} (Y_i^0 - Y_i)^2. \quad (11) \]

Based on (1)–(11), Figure 2 gives DE-BP-based Chinese cross-language paraphrase recognition proposed in this paper, which mainly includes the original sentence, the neural machine translation, the interlanguage sentence, the neural translation, and the paraphrase sentence.

3. Experimental Results and Analysis

3.1. Introduction to Experimental Dataset. This study mainly adopts the method of corpus research and combines quantitative statistics with qualitative analysis. First of all, we collected news texts from People’s Daily for consecutive months and established a large database. Then, according to the research purposes of different chapters, we extracted samples from the main corpus and built a secondary corpus. For some chapters, we also classified and collected them according to the subject needs. Then, we sorted all the
Corpora in each chapter of the newspaper into text documents to establish a pair of subcorpora.

In corpus collection, we mainly consider the following issues: the collected newspaper texts are all issued in years to reflect the recent situation. The establishment of the Chinese newspaper corpus should not only pursue the size but also consider whether it can fully explain and fit the research objectives of this paper.

3.2. Experimental Result Analysis. After establishing the cross-language recognition method based on the DE-BP algorithm, the established method is verified based on the data of 3.1. The specific simulation results are described as follows: after the corpus is established, we will mark the corpus of each section. Indirect reporting is judged by the absence of quotation marks and the use of the third person; free direct paraphrase is mainly in quotes without leading sentences. Among them, the indirect paraphrase is the residual form of indirect speech, which needs to be judged according to the specific context.

The main characteristic of narrative paraphrase of speech act is that it usually centers on a paraphrase verb or is used alone or forms a verb-object structure to indicate who performs a speech act. The specific results of similarity calculation between corpus are shown in Figure 3. The main purpose of the results of the graph is for the subsequent Chinese language paraphrase recognition service.

Finally, we use two sets of classification models to analyze the sources of information: the former takes the proper nouns and pronouns that appear in the discourse to represent people or organizations as the judging criteria. According to the categorization of definite degree, the definite source of information generally has a surname or a given name, uses a definite noun phrase, or uses a definite anaphora or anaphora pronoun. Ambiguous sources are represented by indefinite noun phrases or by coreferential pronouns with no definite identity in the context. Sources
that cannot be identified are more complex, use the passive voice of the hidden actor, or use an unidentifiable noun structure, prepositional phrase construction that conceals the source, or some other vague expressions. Since the similarity is calculated in Figure 1, based on the obtained similarity, the relationship between Chinese paraphrase accuracy and corpus similarity is shown in Figure 4, as can be seen from the figure above. The higher the similarity is, the higher the recognition accuracy will be, and vice versa.

The recognition accuracy of Chinese cross-language paraphrase by different algorithms is shown in Figure 5. The curves are drawn from 2,000 samples, with each dot representing 50 samples. Then, we analyze the extent to which all methods of free translation are related to content word variation extraction. The bottom half of Figure 5 shows the results of the evaluation (curves labeled with CWV). PosedScore’s sample size was significantly reduced compared to English and Japanese, but accuracy remained high. This is mainly due to the linguistic crossover caused by globalization.

In addition, the graphical representation of the size of the set of definitions for each language is shown in Figure 6. It can be seen from the figure that we extracted a significantly different number of identifying definitions for each language. The number of free translations generally varies greatly in size from the collection of French, Arabic, and Chinese free translations and other free translations used to extract free translations for that language. This is because we used a relatively large bilingual corpus for these three languages and a smaller recognition scale for the other languages, resulting in similar recognition results.

In the traditional sense, indirect speech is characterized by the use of the third person as the subject of the introductory sentence, and then the related sentence usually uses an object clause, as well as the moving of the verb tense, such as the present tense to the past tense. In addition, the words and phrases about time and place should be changed accordingly, such as today to yesterday. In addition, it should be pointed out that those paralinguistic features originally directly reported are also lost; for example, pronunciation and intonation, omission of address forms, and gestures and expressions are simplified or lost. Because individual words are separated from the original discourse, they have been broken away from the original grammatical structure and lost the original context, and the semantics have changed, or even the opposite.

Figure 7 gives the example of a paraphrase process for an input Chinese sentence. Figure 2 shows nodes (double circles) and edges (solid lines) from the input sentence by the original word, and at the bottom, Figure 2 shows the last lattice of new nodes (single-line circles) and new edges (dashed lines) from the paraphrase. It can be seen that the proposed method increases the diversity of source phrases and thus provides more flexible translation options in the decoding process.

4. Conclusions

The use of various paraphrasing strategies shows that the news paraphrase controls and manipulates the paraphrase; even if the paraphrase is highly similar to the original language, it will deviate from the original speaker’s intention in a specific context.

On the whole, it is the way in which the Chinese language paraphrases behavior. The diversity of choice of diction and its influence on similarity reflects the conventional choice of paraphrase under the constraints of institutional context. However, the strategic regulation of news relaying behavior in the process of original discourse event itself (such as the adjustment of information type, information state, and amount of information) and relaying expression fully reflects the active choice of the relaying behavior under the influence of various institutional context factors. On the one
hand, it reflects the adaptability of news discourse reporting behavior, and on the other hand, it also reflects the subjectivity of news discourse reporting behavior. However, the research in this paper is only limited to small Chinese language paraphrases data. Big data and multiple kinds of data will be the focus of future research [32].

Data Availability
The dataset can be accessed upon request.

Conflicts of Interest
The authors declare that there are no conflicts of interest.

References
[1] H. Asghari, O. Fatemi, S. Mohtaj, and P. FailiRosso, "On the use of word embedding for cross language plagiarism detection," Intelligent Data Analysis, vol. 23, no. 3, pp. 661–680, 2019.
[2] L. Li, B. B. Lai, and J. P. Huang, "Paraphrase identification based on interpretable mechanism," in Proceedings of the 2013 3rd international conference on artificial intelligence and advanced manufacture (AIAM), IEEE, pp. 416–421, Manchester, United Kingdom, October 2021.
[3] B. Sun, F. Zhang, J. Yuan, Z. Wei, and S. Ting, "Chinese medical paraphrase generation: based on neural machine translation," Journal of European Economy, vol. 47, no. 1, pp. 1–6, 2021.
[4] B. Sun, Y. Wu, Y. Zhao, Z. Hao, L. Yu, and J. He, "Cross-language multimodal scene semantic guidance and leap sampling for video captioning," The Visual Computer, pp. 1–17, 2022.
[5] V. Kumar, A. Verma, and N. Mittal, Anatomy of preprocessing of big data for monolingual corpora paraphrase extraction: source language sentence selection//Emerging Advances in Intelligent Systems and Computing, vol. 5pp. 495–505, Springer, Singapore Singapore 2019.
[6] Y. Wang, M. Liu, Y. Zhang, Z. u. Jin’an, and C. Yufeng, "Research on the construction and application of paraphrase parallel corpus," Beijing Da Xue Xue Bao, vol. 57, no. 1, pp. 68–74, 2021.
[7] Y. Zhai, P. Safari, G. Illouz, A. Allauzen, and A. Vilnat, "Towards Recognizing Phrase Translation Processes: Experiments on English–French," 2019, http://arxiv.org/abs/1904.12213.
[8] M. Nagata, C. Katsuki, and M. Nishino, "A Supervised Word Alignment Method Based on Cross-Language Span Prediction Using Multilingual BEKT," 2020, http://arxiv.org/abs/2004.14516.
[9] X. I. N. Bin and G. A. O. Xiaoli, Reported Speech in Chinese and English Newspapers: Textual and Pragmatic Functions, Routledge, England, 2021.
[10] J. Li, Z. L. Zha, and A. Wulamu, International Journal of Legal Medicine, vol. 135, no. 2, pp. 455–456, 2021.
[11] C. Mi, L. Xie, and Y. Zhang, "Improving data augmentation for low resource speech-to-text translation with diverse paraphrasing," Neural Networks, vol. 148, pp. 194–205, 2022.
[12] C. Federmann, O. Elachqar, and C. Quirk, Multilingual Whispers: Generating Paraphrases with translation, W-NUT, in Proceedings of the 5th Workshop on Noisy User-Generated Text, pp. 17–26, W-NUT, Hong Kong, November 2019.
[13] L. Shi and Y. Dong, "Chinese graduate students paraphrasing in English and Chinese contexts," Journal of English for Academic Purposes, vol. 34, pp. 46–56, 2018.
[14] G. Li, F. Liu, A. Sharma et al., "Research on the natural language recognition method based on cluster analysis using neural network," Mathematical Problems in Engineering, vol. 105, no. 2, pp. 583–597, 2021.
[15] P. Zhu, D. Cheng, F. Yang et al., "Improving Chinese named entity recognition by large-scale syntactic dependency graph," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 12, no. 5, pp. 55–68, 2022.
[16] M. Byas, N. Malik, A. Bilal, and S. Razzaq, "Plagiarism detection using natural language processing techniques," Technical Journal, vol. 26, no. 01, pp. 90–101, 2021.
[17] J. Jun, "Paraphrasing Chinese idioms: paraphrase acquisition, rewording and scoring[!]" Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, no. 3, pp. 1999–2005, 2021.
[18] G. Rocha and H. Lopes Cardoso, "Recognizing textual entailment: challenges in the Portuguese language," Information, vol. 9, no. 4, p. 76, 2018.
[19] H. Le, L. Vial, J. Frey et al., "Flaubert: Unsupervised language model pre-training for french," 2019, http://arxiv.org/pdf/1912.05372.
[20] X. Jia, J. Wang, Z. Zhang, N. Cheng, and J. Xiao, "Large-scale Transfer Learning for Low-Resource Spoken Language Understanding," 2021, http://arxiv.org/pdf/2008.05671.
[21] N. T. T. Thuy, N. X. Bach, and T. M. Phuong, "Cross-language Aspect Extraction for Opinion mining," in Proceedings of the 2018 10th International Conference on Knowledge and Systems Engineering (KSE), pp. 67–72, IEEE, Ho Chi Minh City, Vietnam, November 2018.
[22] M. Abdelhamid, F. Azouaou, and S. Batata, A Survey of Plagiarism Detection Systems: Case of Use with English, French and Arabic Languages[ ], arXiv preprint arXiv:2201.03423, 2022.
[23] L. Feng, J. Yu, D. Cai, S. Liu, H. Zheng, and Y. Wang, "ASR-GLUE: A New Multi-Task Benchmark for ASR-Robust Natural Language Understanding," 2021, http://arxiv.org/pdf/2108.13048.
[24] J. P. Corbeil and H. A. Ghadivel, "Bet: A Backtranslation Approach for Easy Data Augmentation in Transformer-Based Paraphrase Identification context," 2020, http://arxiv.org/abs/2009.12452.
[25] S. Alzahrani and H. Aljuaid, "Identifying cross-lingual plagiarism using rich semantic features and deep neural networks: a study on Arabic-English plagiarism cases," Journal of King Saud University-Computer and Information Sciences, vol. 3, no. 3, pp. 15–29, 2021.
[26] J. W. Lin and R. G. Chang, "Chinese story generation of sentence format control based on multi-channel word embedding and novel data format," Soft Computing, vol. 26, no. 5, pp. 2179–2196, 2022.
[27] J. A. Alzubi, R. Jain, A. Kathuria, and A. KhandelwalSaxenaSingh, "Paraphrase identification using collaborative adversarial networks," Journal of Intelligent and Fuzzy Systems, vol. 39, no. 1, pp. 1021–1032, 2020.
[28] L. Shi, "Application of big data language recognition technology and GPU parallel computing in English teaching visualization system," International Journal of Speech Technology, vol. 12, pp. 452–462, 2021.
[29] B. Thompson and M. Post, "Paraphrase Generation as Zero-Shot Multilingual Translation: Disentangling Semantic
[30] M. Zhang, L. Tan, Z. Tu et al., "Don’t change me! User-controllable selective paraphrase generation," vol. 28, no. 107, pp. 3667–3676, 09290, arxiv.org/abs/2008.10064.

[31] Q. Du and Y. Liu, "Foregrounding learner voice: Chinese undergraduate students’ understanding of paraphrasing and source use conventions for English research paper writing," Language Teaching Research, Article ID 136216882110270, 2021.

[32] H. Inaguma, T. Kawahara, and S. Watanabe, "Source and Target Bidirectional Knowledge Distillation for End-To-End Speech translation," 2021, http://arxiv.org/pdf/2104.06457.