Lecture Notes on Language and Literature (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/langl.2023.061705
ISSN 2523-5869 Vol. 6 Num. 17

Investigation on Chinese Function Words and Construction of Generalized Function Word Knowledge Base in Natural Language Processing

Lin Yang[1,a,*], Mengyang Yu[1,b]

[1]*School of Computer Science and Technology, North University of China, Taiyuan, Shanxi, China*
[a]*ggg758670@gmail.com*, [b]*yumychn@163.com*
[*]*Corresponding author*

Keywords: Knowledge Base Construction, Generalized Function Words, Chinese Function Words, Natural Language Processing, ELMo Model, NNLM Model

Abstract: The study of function words in Chinese has a long history and has achieved fruitful results. However, most of the existing research on function words is human-centered, and it is difficult to avoid subjective and vague description of the character characteristics of function words. Therefore, it is difficult to apply it directly to the research of Natural Language Processing (NLP). From the perspective of computational linguistics, and combined with the existing literature and the People's Daily of part of speech tagging and its corpus, this paper aimed to provide some scientific basis for the use of function words in Chinese. This paper mainly explains the theory and method of conjunctions and phrases in Chinese function words. This paper integrated, proofread and identified Chinese function words through various kinds of dictionary corpus, and the knowledge base was constructed. By comparing the two preprocessing models in NLP, this paper mainly compared the function word recognition rate, text, grammar, semantics, lexical resolution, and the amount of training text data. The results showed that the ELMo model improved the recognition rate of function words by 14.5% compared with NNLM model in the constructed knowledge base, and had a higher understanding of all kinds of grammar and morphology. There would also be more text data for training. However, the NNLM model would be more stable in the amount of training text.

1. Introduction

NLP is an important research field in artificial intelligence technology. It is a method that enables people to communicate with people in natural language. NLP is a kind of computer technology, which can analyze the naturally occurring characters on the level of one or more language analysis and show a series of behaviors similar to human language. Computers use NLP methods to reveal the secrets of human language, so that machines can communicate with people. "Natural language" is a language that evolves naturally in culture. It is an important means for people to communicate and think, and also a concentrated embodiment of human intelligence.

2. Related Work

Different regions have different languages and cultures, which requires the study of natural languages in different regions. In the process of NLP, many linguists have launched discussions. In recent years, deep learning has been widely used in NLP. Otter Daniel W described its architecture method, and made in-depth exploration in other fields, thus finding the core problem of language processing [1]. In order to represent the model of text data, Galassi Andrea designed a definition model of NLP. According to the representation of input, compatibility function, distribution function, and the diversity of input and output, he showed how to use prior information in the model, and also explained the problems that might be faced by subsequent research [2]. Pre-Training Model (PTM) processing of natural language indicated that a new stage was entered in NLP. Qi Xipeng conducted a comprehensive review of the PTM of NLP. At first, he briefly introduced the language representation learning and its research progress, and systematically classified the existing PTM by category from four different perspectives. In addition, Qi Xipeng also introduced how PTM knowledge was combined with downstream work, and looked forward to the future development prospects of PTM research [3]. Li Hang summarized the application of deep learning in NLP, and discussed its advantages and problems. He believed that natural language had five major functions, such as classification, matching, translation, structure prediction and sequence judgment, and found that the effect of deep learning was better or significantly better than traditional methods [4].

In this regard, Zhang Wei Emma proposed various methods to attack a wide range of NLP applications, and also discussed some outstanding issues to bridge the gap between the current progress and the more powerful adversarial attacks on NLP and DNN [5]. The existing universal clinical NLP system has been successfully applied to extract information from clinical texts. However, users usually have to customize the existing system for their personal tasks, which may require a lot of NLP skills. In this regard, Soysal Ergin introduced CLAMP (Clinical Language Annotation, Modeling and Processing), which was a newly developed clinical NLP toolkit. It not only provided the most advanced NLP components, but also offered a user-friendly graphical user interface that could help users quickly build customized NLP pipelines for their personal applications [6]. Hu Yingjie proposed a computing framework for obtaining local place names from house advertisements with geographical markers. The proposed framework included two stages: NLP and geospatial clustering. The NLP stage checked the text content of the housing advertisement and extracted the candidate place names [7].

3. Investigation on Chinese Function Words and Construction of Generalized Function Word Knowledge Base

3.1 Characteristics and Current Situation of Chinese Function Words

The conjunctions and auxiliary words in function words also have many language functions. Each different function word has its own language characteristics [8-9]. By starting from the language characteristics of Chinese function words, their functions and uses are not many, but their application scope and frequency are very high. At the same time, their uses vary widely among different types. At present, Chinese function words lack systematicness and similarity. Therefore, in order to fully and thoroughly reveal the grammatical rules of Chinese and scientifically and reasonably explain them, it is necessary to carefully analyze various types of function words. In the past 20 years, function words have become a hot topic in the field of Chinese grammar. From different angles, different levels, different theories and methods, the paper discusses the function words in Chinese from different angles, and makes beneficial exploration at different levels. At

present, Chinese function words are mainly classified into adverbs, prepositions, conjunctions, auxiliary words, modal words, and locative words, as shown in Figure 1.

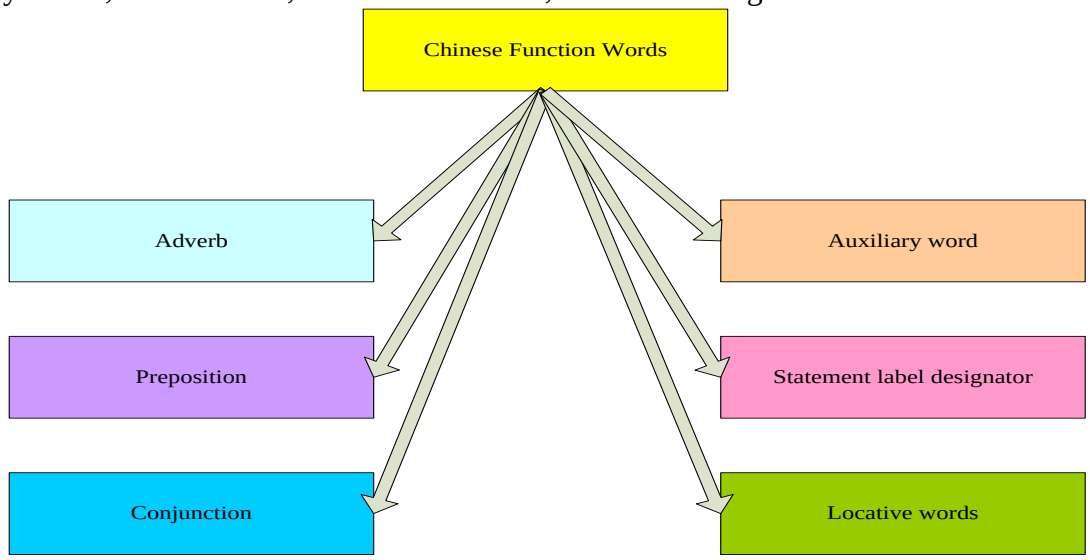


Figure 1: Main classification of Chinese function words

Looking at the history of Chinese grammar, we can see that in the past 20 years, the study of conjunctions in function words has reached its peak. There are many original ideas and achievements in the research of function words in academic circles. Most scholars will not explain the usage of conjunctive structural phrases entirely. Their main concern is the comprehensiveness of all kinds of function words. At the same time, it can not only describe and analyze conjunctive structural phrases in a synchronic and static way, but also conduct a horizontal and dynamic investigation and discussion of conjunctive structural phrases in terms of time. The greatest progress in the study of conjunctive structural phrases is to break through the shackles of traditional concepts, to be all-inclusive in theory, and to make new breakthroughs in methods. The study of conjunctive structural phrases also shows a trend of vigorous development [10-11]. In the study of conjunctive structure phrases, various monographs on function words emerge in endlessly, and numerous professional journals, academic journals, papers and doctoral theses.

Language is a dynamic process, which is not static and fixed. Function word is the most basic grammatical form of Chinese, and its mode, speed, function and result are constantly changing. Due to the widespread popularity of network communication, many specific and even temporary grammar skills can be widely copied and used in a very short time, which has led to changes in many forms of Chinese expression. Among them, the use of many function words has also undergone great changes. Over time, these small changes would change from quantitative to qualitative. Especially after the new century, in the past decade, a series of features have emerged in the development of function words in Chinese, which are quite different from those in the past.

Through in-depth investigation and comparison of modern Chinese function words, it can be seen as follows: (1) the expansion and neutralization of grammatical functions; (2) refining, transforming and subjectifying performance functions; (3) the structure and mark of common function words; (4) the appearance and popularity of function words make them fast and simple. However, in the existing function word dictionaries, grammar works, and even in the current textbooks for teaching Chinese as a foreign language, the use of function words is mostly based on previous experience, and rarely reflects some phenomena, usages and laws. This situation has great contradiction with the research, teaching and application of Chinese function words.

3.2 Theories and Methods of the Study of Conjunction Structure Phrases

Grammaticalization theory: In a specific language environment, lexical items and syntax constitute a process of expressing grammatical functions, which is also a new grammatical function [12]. Unlike Indo-European languages, Chinese entries and syntactic structures do not evolve into inflectional affixes after grammaticalization. Instead, they are combined with neighboring words to form a new phrase, which makes the adverbial group extremely grammaticalized and moves backward in the semantic center of "extreme+VP". This makes it change from "predicate object" to "adverb", and thus completely becoming an adverb. Most function words in Chinese are generated from the original words or phrases. In Chinese, the study of function words has been greatly developed.

The theory of subjectivity: The function of function word method inevitably involves subjectivity. Subjectivity refers to its corresponding structural form or corresponding evolution process. In other words, when a person speaks a word, there are his position, attitude and ideas in the words, and leave his own brand in the words. Among adverbs, the formation of commentary adverbs is a grammaticalization process, which is also a subjective process [13-14]. Therefore, the theory of subjectivity plays an irreplaceable role in the interpretation of evaluative adverbs and modal particles. Therefore, the subjectivity of Chinese function words is the most important aspect in current research.

Lexicalization theory: The lexicalization can be divided into synchronic and diachronic, which mainly refers to diachronic lexicalization. However, people are concerned about the lexicalization of cross-hierarchy structure in lexicalization. That is to say, in the initial stage of lexicalization, the two components have no direct connection. The "double tone" tendency of function words in Chinese makes some grammatical elements appear "empty" in the process of "lexicalization".

Constructional theory: Constructive grammar is a grammar theory proposed by linguists. The definition of construction is as follows: C is a construction. If C is a pairing of form and semantics, some forms or semantics of C cannot be completely predicted by the components of C or other existing constructions. Construction refers to the overall structure of part of speech, meaning and phrase composition. The grammaticalization of construction is defined as a process of reanalysis, which would lead to grammaticalization and lexical fusion. In fact, function words in Chinese are composed of syntactic components, which can also be analyzed through the grammar of construction. Therefore, it has irreplaceable advantages to study the development and evolution of Chinese structure and function words from the perspective of construction.

Markup theory: Under certain conditions, words would further "empty" and "solidify", thus forming particles and prepositions. For example, the tenses of "zhe, le, and guo", "de and di"; another example is "look, and don't say". For example, the word "zhong" was originally a noun. However, as time goes on, its position changes and its part of speech becomes a function word, and its further virtualization becomes a continuous symbol; the "gei" was originally a verb, and it became a preposition and a special focus symbol finally.

Metalinguistic theory: The "metalanguage" is a new concept originated from modern logic and philosophy of language, which is greatly different from the basic language. Metalanguage has four major functions: calibration of subject structure; organizational conversation theory; evaluation proposition demonstration; communication channels. Some parentheses and conjunctions in Chinese are abundant in the use of metalanguage. In particular, Chinese commentative adverbs have a variety of meta-language usages, including marking, evaluation and relevance.

Semantic map model: This model is a widely used method at present, which is mainly used in language types and Chinese language semantics. Semantic mapping can also effectively segment the polysemy and homophone features of the divided morphemes, so as to intuitively reflect the

close and distant relationship of each function and describe the change path of language in a sense. The semantic map model is used to study the function and usage of Chinese function words, as shown in Figure 2.

Affixation theory: Affix refers to the loss of phonetic independence and dependence on an independent word. However, it still retains the position of the word (not the morpheme within the word) in grammar, that is, “affixes, appendixes, and endings”. At present, the word “affixation” is rarely mentioned in the Chinese grammar circle, and affixation is a common grammar in people’s language at present. Its basic principles are as follows: strict syntax, lenient tone, syntactic emphasis on word order change, structural dislocation, pronunciation, etc.

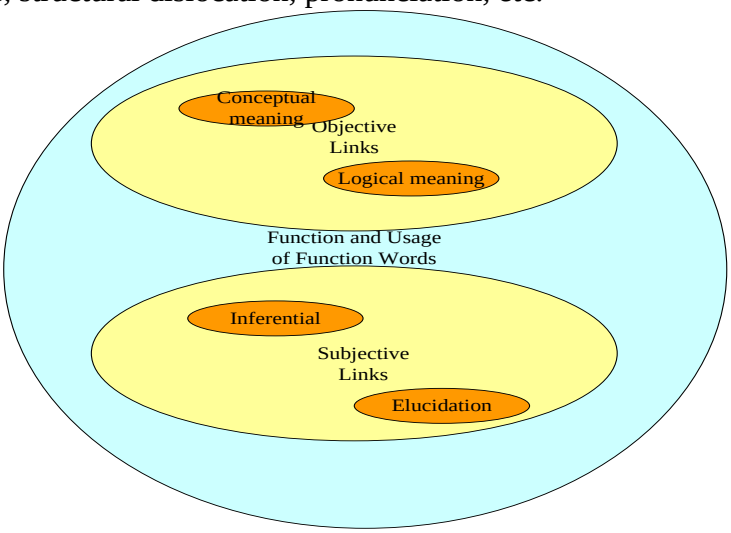


Figure 2: Function and usage of function words

3.3 Construction of Function Word Knowledge Base

The knowledge base construction of Chinese function words mainly includes the usage dictionary, usage rules and Chinese corpus of Chinese function words [15]. The construction of function word knowledge base is shown in Figure 3.

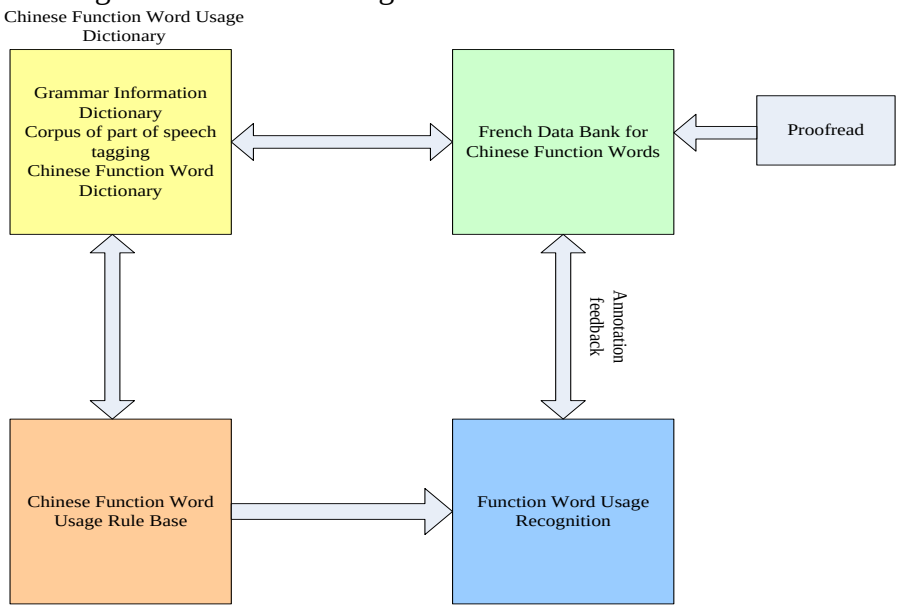


Figure 3: Function word knowledge base construction

The construction of function word knowledge base needs to build and standardize the function word dictionary first. Through tagging corpus, additional processing is needed in proofreading. First of all, the grammatical features of various parts of speech are designed. The entries are classified according to various dictionaries, and the functional descriptions, definitions, example sentences and other supplements are provided; the description of the usage specification is redesigned, and a basic usage rule is manually constructed according to the usage instructions of the dictionary; a rule-based automatic recognition method is designed to realize the automatic annotation of the usage of function words in the corpus. In automatic annotation, if the annotation usage is inconsistent, the final annotation result would be determined after the third party's negotiation, and the annotation result, proofreading dictionary and usage would be standardized. The error rate and the word order of the content are reduced by manual detection and automatic labeling in the rule base [16].

In the usage dictionary of function words, the dictionary needs to design the structure framework, supplement and revise the chapters and paragraphs. In the framework design, the main content includes four categories: sentence identification, syntactic function description, category description and usage description. Among them, the recognition category is the same for the six categories. It is to assign a unique code to the use of each function word. The whole coding process are [tn], [m], [x], and [y]. On the whole, the structures of various parts of speech have the same nature, but they also have their own characteristics. This makes full use of the knowledge base of function word usage in the process of processing.

Content fill: According to the views of different scholars, this dictionary decomposes and differentiates the descriptions of various usages according to the needs of natural language, and extracts operable usage features to fill in.

Feedback modification: In the proofreading of the corpus, the function word usage dictionary has also been feedback and screening, and some words in the dictionary have been added or deleted, which is prone to different meanings. The construction of the function word usage dictionary is a process of continuous improvement and improvement. The distribution of various function word entries and usage in the Chinese generalized function word dictionary is shown in Table 1.

At present, the function word usage dictionary has formed a certain embryonic form. However, due to the frequent use of function words (which appear more frequently in the corpus), linguists have made a more detailed classification of their semantics and usage. The use of function words is less and more simple, and the distinction in semantics and usage is also scattered. The usage of function words in modern Chinese is based on rules. On the basis of a preliminary discussion of the usage dictionary of Chinese function words, this paper puts forward four stages of establishing rules: format and specification, establishing rules, content annotation, and feedback modification. The operable discriminant conditions are extracted from different use characteristics of function words, and the rule base of "function words" in modern Chinese is established [17].

Table 1: Entry and usage of function words

Usage of entries	Adverb	Conjunction	Preposition	Auxiliary word	Statement label designator	Locative words
1	744	210	65	44	33	144
2	251	152	25	12	13	38
3	126	46	16	9	4	14
4	78	12	15	7	2	11
5	56	12	7	0	1	7
6	23	9	7	1	0	4
7	12	0	2	0	1	8
Total usage	1290	441	137	73	54	226

Rule building: The use rules of function words are constructed manually based on the instructions of the existing function words dictionary, with high automatic recognition accuracy as the starting point. For more complex use, a use can be described by multiple rules.

In the aspect of feedback correction, for the content of rules, the corpus after manual correction is compared with the corpus automatically marked by rules, and the correct rate is automatically marked according to different usage rules. The sentences that are marked with errors or cannot be understood are analyzed, and the rules that meet the specifications are extracted from them. In terms of rule ordering, the automatic recognition of function words is not determined according to the sequence, but according to their folding in the dictionary and the probability of their occurrence in the corpus. Although all the formal things are described as much as possible when defining rules, it can only deal with formal instructions.

The virtual word corpus is a basic standard corpus for word segmentation and character classification of nearly nine million parts of speech. It introduces rule-based part of speech tagging, and the corresponding manual tagging is also completed by researchers in computer and linguistics. At the same time, it also tests the use of pseudonyms in phonetic materials through the usage instructions in the function word usage dictionary, and takes the case that the two words are different into the third party's argument in order to get the final result. In order to ensure the accuracy of the materials used, it is also necessary to form a criterion according to the annotation procedure and the results of the research process, which is the basic criterion for the division of the use of words. In the process of determining and applying the scope of use of function words, there are three methods of automatic recognition of the use of function words: Automatic recognition of the use of function words based on rules, automatic recognition of the use of statistical function words, and automatic recognition of the use of rules and statistical function words.

According to the rules, the use of function words is labeled. First, the corpus is read out from the corpus, and the content of the corpus is divided into several sentences. According to the function word search rules to be labeled, the corresponding matching of the six types (F, M, L, R, N, and E) and the specific frame matcher are called respectively to annotate their use rules. In the specification, a lot of modifications and improvements have been made to the rule base, which has significantly improved the accuracy of its description and use of rules in the specification. Due to some difficulties in describing rules, automatic recognition of function words cannot be well realized when using rules.

3.4 Natural Language Pre-Training Model

There are many limitations in the relationship modeling between words. Therefore, this paper focuses on NNLM and ELMo models. NNLM is a shallow model based on forward learning, which introduces the concept of vocabulary vector into the model. The NNLM training model can be expressed as follows:

$$N = f(w_t, w_{t-1}, \ldots w_{t-n+2}, w_{t-n+1}) \tag{1}$$

$$f(w_t, w_{t-1}, \ldots w_{t-n+2}, w_{t-n+1}) = p(w_t | w_1^{t-1}) \tag{2}$$

Among them, w_t represents the t word in the sequence, and w_1^{t-1} represents the sequence between the first column of words and (t-1) words. The conditions that need to be met for the establishment of the model are as follows:

$$f(w_t, w_{t-1}, \ldots w_{t-n+2}, w_{t-n+1}) > 0 \tag{3}$$

$$\sum_{i=1}^{|V|} f(w_t, w_{t-1}, \ldots w_{t-n+2}, w_{t-n+1}) = 1 \tag{4}$$

Among them, V is the probability value of each component.

ELMo is a feature-based language model that can be combined with context. In ELMo, the word vector is an internal state function of a network, and the word vector expressed by it is dynamic. The model consists of forward LSTM and reverse LSTM.

The model formula of forward LSTM can be expressed as follows:

$$p(t_1, t_2, \ldots t_M) = \prod_{k=1}^{M} p(t_k | t_1, t_2, \ldots t_{k-1}) \quad (5)$$

The model formula of reverse LSTM can be expressed as follows:

$$p(t_1, t_2, \ldots t_M) = \prod_{k=1}^{M} p(t_k | t_{k+1}, t_{k+2}, \ldots t_M) \quad (6)$$

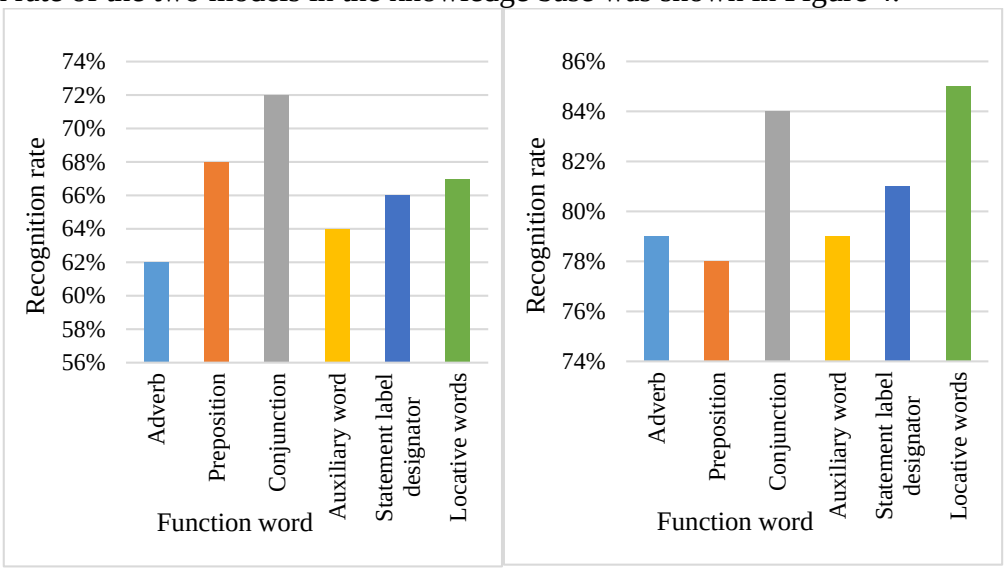
The formula realizes the probability prediction of the k-th word by knowing the following modeling. The maximum likelihood probability of logarithmic forward and reverse can be expressed as follows:

$$P = \sum_{k=1}^{M} (\log p(t_1, t_2, \ldots t_{k-1}; \Psi x, \Psi_s) + \log p(t_k | t_{k+1}, t_{k+2}, \ldots t_M; \Psi x, \Psi_s) \quad (7)$$

Among them, Ψx is the input of the initial language vector, and Ψ_s is the parameter of the software layer. Both have a certain weight and cannot be fully interdependent. This paper also compares the two models.

4. Comparison of Two Training Models in NLP

This paper mainly compared the NNLM model and ELMo model in NLP, which was mainly based on the research of Chinese function words and the knowledge base of generalized function words. Through the recognition rate of various function words in the knowledge base, text, grammar, semantics, lexical resolution, and the amount of training text data, this paper conducted model processing in the knowledge base of Chinese function word research and generalized function word construction. In terms of the recognition rate of various function words, the experimental training in this paper mainly analyzed the six kinds of adverbs, prepositions, conjunctions, auxiliary words, modal words and locative words. After the model pre-training, the recognition rate of the two models in the knowledge base was shown in Figure 4.



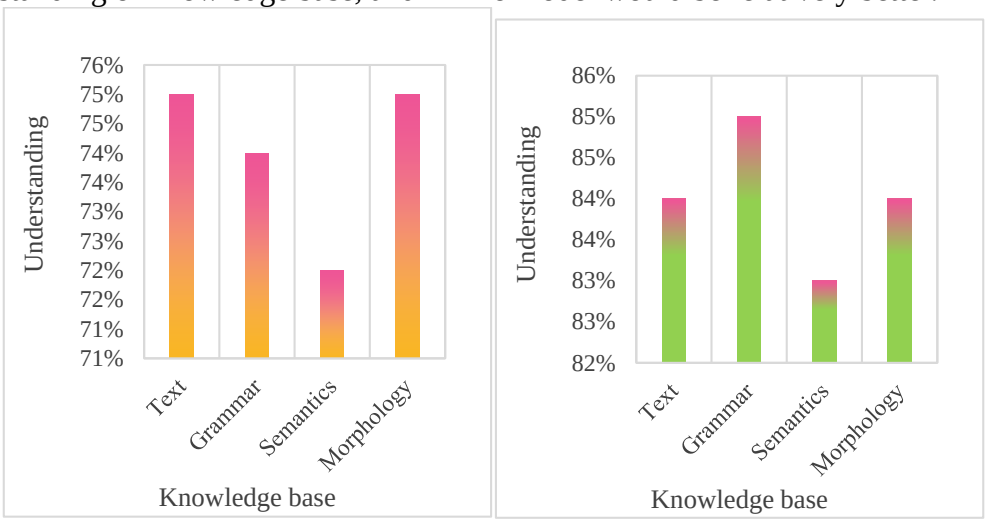
4A: Recognition rate of NNLM model in knowledge base
4B: Recognition rate of ELMo model in knowledge base

Figure 4: Recognition rate of the two models in the knowledge base

It could be seen from Figure 4A that the recognition rate of the NNLM model for adverbs, prepositions, conjunctions, auxiliary words, modal words and locative words were 62%, 68%, 72%, 64%, 66% and 67% respectively, and the average recognition rate of the model for function words was 66.5%. In Figure 4B, the recognition rate of ELMo model for various function words were 79%, 78%, 84%, 79%, 81%, 85% respectively, and the average recognition rate of ELMo model for function words was 81%. It could be seen that ELMo model improved the recognition rate of function words by 14.5% compared with NNLM model, and ELMo model was more suitable for the pre-training of knowledge base.

The text, grammar, semantics and morphology in the Chinese knowledge base were understood by two models. The understanding degree of the two training models to the Chinese knowledge base was shown in Figure 5.

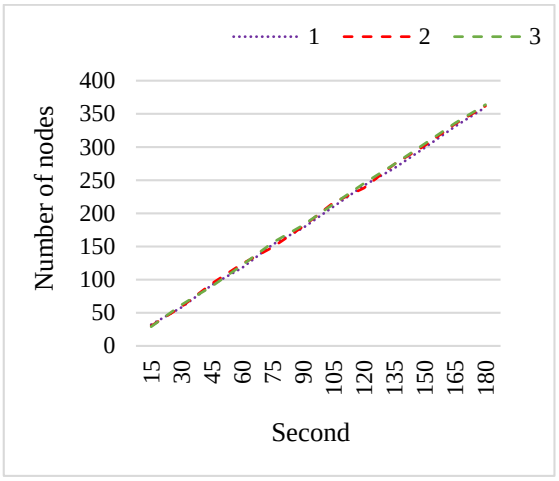
It could be seen from Figure 5A that the NNLM model had 75%, 74%, 72% and 75% understanding of the text, grammar, semantics and morphology in the knowledge base. In Figure 5B, the ELMo model's understanding of the knowledge base were 84%, 85%, 83% and 84% respectively. It could be found that the NNLM model had a lower understanding of Chinese in the knowledge base than ELMo model, but the overall understanding was still good. Both models had a good understanding of knowledge base, and ELMo model would be relatively better.

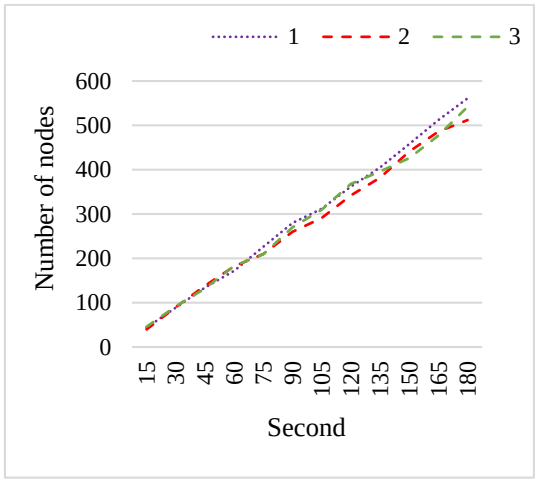


5A: NNLM model's understanding of knowledge base
5B: ELMo model's understanding of knowledge base

Figure 5: Understanding of Chinese knowledge base by two training models

In NLP, the data volume of model training also played an important role. The more data the model trains, the higher the understanding of the corpus and words in the knowledge base would be, and the selection of function words in the text would be more in line with the context. In this paper, the number of bytes trained by the two models was compared at different times. In order to verify its accuracy, three experiments were carried out. The amount of training text data of the two models was shown in Figure 6.





6A: NNLM model training text data volume 6B: ELMo model training text data volume

Figure 6: Training text data volume of two models

As could be seen from Figure 6A, the NNLM model training text data grew with time, and the word node data showed a linear growth. At 180 seconds, the number of bytes increased to more than 360. In Figure 6B, when the ELMo model training text data was 180 seconds, the number of bytes exceeded 500. Through three experiments, it could be found that the ELMo model trained more text data, but the number of bytes in training was unstable; the bytes trained by the NNLM model were relatively stable and were in a relatively stable growth range.

5. Conclusions

In this paper, we mainly build a knowledge base of conjunctions in Chinese function words, through natural language processing. It mainly describes the six main types of function words, elaborates the steps of building a knowledge base, and explains the usage dictionary, usage specification, and automatic tagging of the corpus of conjunctive structure phrases, thus forming a preliminary modern function word use dictionary, function word application standard database, and a small-scale Chinese generalized function word information database of modern function word use corpus, Some of these problems are studied in detail in order to improve the accuracy of the automatic standard for the use of virtual words. At the same time, the cooperation and development with various fields such as the Chinese generalized function word use dictionary, function word use rules and function word use French database were further strengthened, and a more accurate and comprehensive modern Chinese universal function word information database was gradually formed to meet the needs of natural language. However, the inadequacy of this paper was that, due to some objective reasons and the constraints of human resources and materials, this paper was not able to communicate face to face with the main creators of the corpus for research and analysis used in the paper, and not fully understood the problems and difficulties actually encountered in the specific process of building the corpus. It was impossible to build a corpus by one person in a short time. It needed the cooperation of corpus linguistics experts, computer programming experts, second language acquisition research experts, interlanguage theory research experts and front-line teachers for teaching Chinese as a foreign language.

References

[1] Otter, Daniel W., Julian R. Medina, and Jugal K. Kalita. "A survey of the usages of deep learning for natural language processing." IEEE transactions on neural networks and learning systems 32.2 (2020): 604-624.

[2] Galassi, Andrea, Marco Lippi, and Paolo Torroni. "Attention in natural language processing." IEEE Transactions on Neural Networks and Learning Systems 32.10 (2020): 4291-4308.

[3] Qiu, Xipeng. "Pre-trained models for natural language processing: A survey." Science China Technological

Sciences 63.10 (2020): 1872-1897.
[4] Li, Hang. "Deep learning for natural language processing: advantages and challenges." National Science Review 5.1 (2018): 24-26.
[5] Zhang, Wei. "Adversarial attacks on deep-learning models in natural language processing: A survey." ACM Transactions on Intelligent Systems and Technology (TIST) 11.3 (2020): 1-41.
[6] Soysal, Ergin. "CLAMP–a toolkit for efficiently building customized clinical natural language processing pipelines." Journal of the American Medical Informatics Association 25.3 (2018): 331-336.
[7] Hu, Yingjie, Huina Mao, and Grant McKenzie. "A natural language processing and geospatial clustering framework for harvesting local place names from geotagged housing advertisements." International Journal of Geographical Information Science 33.4 (2019): 714-738.
[8] Zhang, Yi. "Chinese adverbs: A Cognitive Grammar exploration." Review of Cognitive Linguistics. Published under the auspices of the Spanish Cognitive Linguistics Association 20.2 (2022): 330-356.
[9] Daojia, C. "The analysis of the errors in acquisition of chinese adverbs by Uzbek speakers." Journal of Advanced Research in Dynamical and Control Systems 12.S2 (2020): 171-181.
[10] Zhang, Ning. "Sentence-final aspect particles as finite markers in Mandarin Chinese." Linguistics 57.5 (2019): 967-1023.
[11] Liu, Chensheng. "Projecting adjectives in Chinese." Journal of East Asian Linguistics 27.1 (2018): 67-109.
[12] Lu, Xiaolong. "The constructionalization of antonymous compounds: Evidence from the adverb chi-zao in Mandarin Chinese." Concentric 48.1 (2022): 30-69.
[13] Li, Long, and Canzhong Wu. "DEGREE OF INTENSITY in English-Chinese translation: a corpus-based approach." Functional Linguistics 6.1 (2019): 1-19.
[14] Hsieh, Miao-Ling, and Yu-Fang Wang. "Acquiring the polysemous adverb HAI in Chinese by English-speaking, Japanese-speaking, and Korean-speaking CSL learners." Concentric 46.2 (2020): 173-205.
[15] Xu, Zheng. "The word status of Chinese adjective-noun combinations." Linguistics 56.1 (2018): 207-256.
[16] Shen, S., Zhu, C., Fan, C. Wu, C., Huang, X., Zhou, L. Research on the evolution and driving forces of the manufacturing industry during the "13th five-year plan" period in Jiangsu province of China based on natural language processing. PLoS ONE .2021, 16(8): e0256162.
[17] C. Zhang, C. -N. Hsu, Y. Katsis, H. -C. Kim and Y. Vázquez-Baeza, Theoretical Rule-based Knowledge Graph Reasoning by Connectivity Dependency Discovery. 2022 International Joint Conference on Neural Networks (IJCNN), 2022, pp. 1-9