Review on the Book Entitled *Syntactic Disambiguation of Modern Chinese with Experimental Prosodic*

ABSTRACT

One of the essential topics in human-computer communication is the disambiguation of the Chinese language. To understand the new progress in this field, this paper introduces the commonly used disambiguation methods, including the “three-domain theory” and the atomic theory used by linguists, and the neural network disambiguation and sliding semantic string matching disambiguation used by computer engineers. After pointing out the limitations of the existing disambiguation methods, the paper introduces a new disambiguation method, disambiguation with rhythm, which is studied in the book entitled *Syntactic Disambiguation of Modern Chinese with Experimental Prosodic*. The paper reveals that the book *Syntactic Disambiguation of Modern Chinese with Experimental Prosodic* combines theoretical discussion and social needs, and provides new ideas and methods for the study of Chinese disambiguation. In one word, the book, which has changed the direction of disambiguation studies, is a masterpiece in Chinese disambiguation research.

Keywords: Disambiguation of Chinese language, Disambiguation with rhythm, Human-computer communication,Syntactic Disambiguation of Modern Chinese with Experimental Prosodic

1. INTRODUCTION

Language is an essential communication tool for human beings and the carrier of human culture. It can be said that without language, human society would collapse. Because of the inseparability of humans and language, we often ignore some features of language. Language is ambiguous. The language we use is not accurate but expresses a vague concept. Sometimes an utterance has multiple meanings. In human communication, the language speaker and listener share the communication scene. The speech context can help both sides choose specific semantics, so language ambiguity will not obstacle to human communication. Sometimes speakers even deliberately create rhetorical methods such as "puns" to improve the effect of language expression, or to achieve some other communicative purposes with the ambiguity of language. For example, some advertisements prefer to attract audiences with the ambiguity of language.

With the development of science and technology, we have entered an information age. Artificial intelligence has entered every field of life. We hope that machines can be intelligent like human beings and can serve us better. To achieve this goal, the most basic requirement is that machines can understand human language and communicate with humans. Under the current technical conditions, machines cannot access context information, so it seems difficult for machines to eliminate language ambiguity with context information. The best way

to eliminate ambiguity in language is to study language grammar deeply and create some computer programs, which will increase the accuracy of human-machine communication.

Chinese is the lingua franca in China and one of the most important languages in the world. Like in other languages, there are ambiguities in Chinese. China is in the great journey of its second centenary and China cannot fall behind other countries in the competition of science and technology. Therefore, scholars in China have shouldered the great task of studying the ambiguity of Chinese deeply, finding out the way to eliminate Chinese ambiguity, and promoting the development of human-machine communications and artificial intelligence in China.

2. PREVIOUS STUDY OF ELIMINATING AMBIGUITY IN CHINESE

In order to distinguish the ambiguity in Chinese and help machines understand the Chinese language, linguists and algorithmic scientists have made efforts from many angles to eliminate Chinese ambiguity. Generally speaking, there are two approaches to eliminating ambiguity in Chinese. One is by linguistic knowledge, and the other by algorithm.

Linguists eliminate Chinese disambiguation with linguistic knowledge. Their research focuses on some basic theories, such as some questions about what ambiguity is, how many types of ambiguity there are, and what the triggering mechanism of ambiguity is. They study the grammatical structure, semantic features, and pragmatic environment of sentences, hoping to find ways to distinguish ambiguity by studying language ontology.

Zuo analyzed the ambiguity of negative structure with the “three domain theory” [1]. The “three domain theory” inherits the traditional ideas of cognitive linguistics, which believes that language exists in three domains: domain of action, domain of knowledge, and domain of speech. The domain of action, which expresses the basic meaning of language, reflects the objective world that people know. Usually, language speakers transform the language whose action domain is related to subjective cognition and reasoning into the domain of knowledge, and those related to speech act into the domain of speech with the help of metaphor [2]. After analyzing the negative sentences with the Chinese character “shi”, Zuo concludes that the pragmatic ambiguity of the negative sentences with the Chinese character “shi” does not exist in language usage or descriptive usage, but in the domain of action, domain of knowledge, and domain of speech [1]. The negation which is in the domain of knowledge and domain of speech is at the meta-representation level, without truth values. Therefore, when judging whether the negative structure is ambiguous, we should not rely on meta-negative or descriptive negative, but on the negative within the three domains.

Cai, Liu, and Wang analyzed the ambiguity of the V+NP+PP structure in Chinese with the Atom Theory [3], which is proposed by Erteschik-Shir and Rapoport [4]. This theory holds that any verb can be decomposed into one or more atomic components, which determine the body marker or argument of the verb in syntactic collocation. The verb’s mode, state, and place are determined by atoms that constitute the verb. If the atoms representing the state of the verb are highlighted, the verb will take on the characteristics of a noun and can be modified by an adjective; if the atoms representing the place of the verb are highlighted, the verb will take on the characteristics of a place noun; if the atoms representing the mode of the verb are highlighted, the verb will take on the characteristics of a verb and can be modified by an adverb. Based on the Atom Theory, scholars reveal the interactive relationship between the verb and other components in the syntactic configuration and explain the ambiguity of the V+NP+PP structure [5]. They believe that a verb structure whose components deconstructed by a verb act on the subject and does not interact with the prepositional phrase will become a single event structure, rather than an ambiguous

structure; if a verb structure whose components deconstructed by a verb interact with the prepositional phrase, will form a complex structure, and accordingly, ambiguity will occur. That is to say, the criteria for judging whether the V+NP+PP structure has ambiguity depends on whether and how the verb component interacts with the prepositional phrase.

The aforementioned two studies on Chinese ambiguity are conducted by linguists. Usually, such studies focus on the ambiguity of some language structures. The conclusions are some theoretical achievements of linguistic research, which have little direct guiding significance for helping computers understand language and promote human-machine communication. Computer engineers are concerned with solving practical problems. They refer to creating models or algorithms to help computers understand languages based on linguistic rules. Bayesian model [6], decision tree model [3], vector space model [7], and maximum entropy model [8-9] are these kinds of technologies for Chinese disambiguation by computer engineers. In recent years, with the development of science and technology, scholars have realized that Chinese disambiguation needs to integrate linguistic knowledge and various algorithms, which breeds disambiguation methods such as neural network disambiguation combining word form, part of speech and translation [10] and a word sense disambiguation model based on Sliding Match of Semantic String [11].

The neural network disambiguation combining word form, part of speech, and translation is proposed by Zhang, Zhao, and Gao [10]. This disambiguation system takes ambiguous words as the center and sets the left adjacent vocabulary window and the right adjacent vocabulary window on both sides of ambiguous words. When encountering ambiguity, the system will calculate the word form and part of speech of words in the left window and right window respectively with the model of neural network, to judge the semantic category of ambiguous words, and then achieve the purpose of disambiguation. According to some tests, the accuracy of this disambiguation method is significantly better than that of the traditional disambiguation method.

The word sense disambiguation model based on the Sliding Match of Semantic String is proposed by Wang and Huang [11]. The work principles of the model are as follows. Firstly, the synonym classification standard is established. Secondly, some N-gram semantic templates are extracted from the corpus annotated by semantic code and then one N-gram semantic template bank is built. Thirdly, each N-gram semantic code string is extracted from a test sentence to match with the N-gram templates in the N-gram semantic template bank, and one final semantic code is decided by choosing the one with the highest score that combines the votes of multiple matched templates. In this way, the computer can select a unique semantic for the ambiguous words, thus realizing the purpose of computer disambiguation.

3. DISAMBIGUATION WITH EXPERIMENTAL PHONETICS: A NEW APPROACH

Linguists and computer engineers have made various efforts to eliminate ambiguity in Chinese, and to realize human-machine conversation or automatic translation. After years of efforts, Chinese disambiguation has made some achievements. The models proposed by computer engineers after learning from the research results of linguists have improved the success rate of disambiguation much better than before. However, the current research results are not completely satisfactory. Until today, scholars have not found general and efficient disambiguation methods. Ambiguity is not a trouble in human-human conversation. In man-machine communication, however, because computers lack the same background knowledge as humans, ambiguity will affect the computer's processing of language, failing man-machine communication. Scholars have tried to eliminate ambiguity in Chinese through a variety of ways, but the effect has not reached our expectations for some reasons. One of

the essential reasons is that scholars have not found the most suitable path in disambiguation. As we all know, sound is the shell of language, which contains rich discourse information such as the speed of voice, the tone of voice, and pauses. Sound plays a vital role in communication. It is voice information that helps people eliminate most of the ambiguity in human-human communication. In the research on Chinese disambiguation, scholars regard language as written symbols [12] and pay more attention to grammatical information. It is precisely because of the lack of sufficient attention to voice information that the current disambiguation methods cannot reach a satisfactory level.

Scholars have paid attention to the importance of phonology in Chinese disambiguation, and have achieved some successes in disambiguating some ambiguous structures in Chinese with experimental prosodic [13-16]. However, these kinds of achievements are relatively scattered and have not yet formed a system. The book entitled *Syntactic Disambiguation of Modern Chinese with Experimental Prosodic* (ISBN978-7-5767-0755-7, CIP2023070823) is a monograph written to fill this gap in the academic community [17]. Its author is Dr. Yan Zhang, an influential young scholar in China. The book is the research achievement of two research projects, which are funded by the Ministry of Education of the People's Republic of China (*JiaoyubuRenwenShekeYanjiuXiangmu*) and Hebei Federation of Social Science Circles (*HebeishengShehuikexue Fazhan Yanjiu Keti*), respectively. One study gets both national funds and local funds, which reveals that the study is in the frontier of its field and meets the social requirements. This work is published by Harbin Institute of Technology Press, a well-known university press in China, showing that the research results are of high quality.

There are 23000 Chinese characters in the book *Syntactic Disambiguation of Modern Chinese with Experimental Prosodic*, including ten chapters in addition to the introduction and conclusion. Chapter 1 is a literature review, which summarizes the achievements on Chinese phonology and Chinese ambiguity, as well as some limitations of the relevant research. The second chapter introduces the specific research methods, providing the necessary conditions for repeated testing of the research process and results, which meet the requirements of contemporary academia. Chapter 3 and Chapter 4 introduce the phonetic characteristics of single-character sounds and double-character words in modern standard Chinese, respectively. This part can lay the foundation for non-professionals to read this book, and also make a perfect preparation for the following chapters. Chapter 5 to Chapter 10 are the core chapters of this book, which respectively introduce the prosodic features of "de" (得) sentence, "ye" (也) sentence, "de" (的) sentence, "he" (和) sentence, verb-object/partial structure, apposition/partial structure, and the methods to distinguish ambiguity by prosodic features. Each chapter introduces the experimental materials, experimental results, and the similarities and differences with the existing research results, showing in which aspects this study has achieved a breakthrough in the original research.

The topic of the book is novel, its content is rich, and its structure is reasonable. The most prominent feature of the book is its research methods. For a long time, studies of humanities and social sciences in China have been unable to get rid of traditional research methods. They are more accustomed to finding research materials through self-awareness or from previous studies, lacking rigorous experimental methods as natural sciences. Taking linguistic research as an example, the older generation of linguists usually investigate and record speech only by their mouths and ears. This traditional research method requires high-quality researchers' skills in distinguishing sounds and recording sounds. On the one hand, it will take a long time and a huge effort to train a qualified language researcher. On the other hand, due to the limitations of various factors, even an excellent recorder cannot guarantee the correctness of speech recording, let alone recording the subtle differences between

sounds with accurate data. When studying Chinese ambiguity, traditionally linguists and engineers often ignore phonetic factors and focus on grammatical factors. Part of the reason for the aforementioned phenomena is that it is difficult to record speech in traditional linguistic research. The book entitled *Syntactic Disambiguation of Modern Chinese with Experimental Prosodic* introduces some theories and research methods of experimental linguistics into the study of Chinese ambiguity, quantifies the phonological tone of Chinese, and records how speakers eliminate ambiguity in natural conversation through acoustic data such as pitch, intensity and duration ratio with accurate data. Such research methods are scientific and reasonable, and the research conclusions are convincing. The data obtained from experimental linguistics in the book can reveal the external manifestations of Chinese ambiguity and help computers distinguish ambiguity. In one sentence, the work entitled *Syntactic Disambiguation of Modern Chinese with Experimental Prosodic* gives directions for future studies on phonetic research and ambiguity research, perfectly combines theoretical study and social requirements, and provides new ideas and methods for Chinese disambiguation research. It is a rare masterpiece in the research of Chinese disambiguation in recent years.

4. CONCLUSION

Language is an important communication tool and thinking tool of human beings, and one of the essential characteristics that distinguish humans from animals. Language has been with us for a long time. It has become an important factor in uniting community members. Although there are ambiguities in language, speakers can eliminate ambiguities in language easily with speech intonation, background knowledge, and other information. The current society has entered the information age. Language's social function has strengthened. With the development of information technology, we hope that computers can do more for us, and expect that computers can communicate with us in human languages. Therefore, it is necessary for computers to be able to clearly distinguish ambiguities in language and find the most appropriate meaning among many ambiguities. Linguists and engineers have helped computers identify ambiguities in language in various ways. Although some achievements have been made, the achievements cannot meet our requirements. It is a new approach to eliminate ambiguities in languages with phonology information and intonation information. It is hoped that more scholars will be involved in relevant research to overcome the difficulty of Chinese disambiguation, so as to provide support and guarantee for the realization of human-machine communication.

REFERENCES

[1] Zuo B. Analysis of the ambiguity of negation from a cognitive perspective. Foreign Language Learning Theory and Practice. 2021(3): 47-55. Chinese.

[2] Sweetser E. From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure (Vol. 54). Cambridge: CUP, 1990.

[3] Cai Y, Zhang H, He Q, DuanJ. A novel framework of fuzzy oblique decision tree construction for pattern classification. Applied Intelligence. 2020, 50: 2959-2975. Chinese.

[4] Erteschik‐Shir N, RapoportT. Contact and other results. In HovavMR, Doron E,Sichel I. editors. Lexical Semantics, Syntax, and Event Structure. Oxford: Oxford University Press. 2010: 59-75.

[5] Cai M, Liu R, Wang C. The motivation of semantic ambiguity of the V+NP+PP special structure from the perspective of the Atom Theory. Modern Foreign Languages. 2022,45(5): 611-620. Chinese.

[6] Lu Z, Liu T, Zhang G, Li S. Word sense disambiguation based on dependency relationship analysis and Bayes model. Chinese High Technology Letters. 2003(5): 1-7. Chinese.

[7] Chen H, Cheng L, Zhang X. Unsupervised approach to word sense disambiguation based on vector space model. Computer Engineering and Design.2007(5): 1215-1218. Chinese.

[8] He J, Wang H. Chinese word sense disambiguation based on Maximum Entropy model with feature selection. Journal of Software.2010,21 (6): 1287-1295. Chinese

[9] Zhang Y, Huang G, Su W. Chinese word sense disambiguation based on Latent Maximum Entropy principle. Journal of Chinese Information Processing.2012,26(3): 72-78. Chinese.

[10] Zhang C, Zhao L, Gao X. Chinese word sense disambiguation based on word-translation and part-of-speech. Journal of Harbin University of Science and Technology.2020,25(3): 131-136. Chinese.

[11] Wang W, Huang D. Chinese word sense disambiguation based on Sliding Match of Semantic String (SMOSS). Journal of Chinese Computer Systems. 2020, 41(7): 1345-1350. Chinese.

[12] Li D, Zhang X, Fan Y, Wu X. Chinese syntactic parsing with word sense disambiguation. Acta ScientiarumNaturalium Universitatis Pekinensis. 2015,51(4): 577-584. Chinese.

[13] Huang C, Zhao Y. Prosodic realization of negative scope and negative focus in sentences with “Mei”. TCSOL Studies. 2023(1): 19-27. Chinese.

[14] Huang C. An experimental phonetic analysis on the ambiguous structure: “V double + N double”. Language Teaching and Linguistic Studies. 2012(3): 98-104. Chinese.

[15] Huang C. A phonetic experimental reanalysis of the ambiguous structures with the Chinese character “Dou”. Studies in Language and Linguistics. 2013,33(3): 52-57. Chinese.

[16] Huang C. The experimental phonetics analysis on ambiguity in semantic sentences of the Character “Dou”. Language Teaching and Linguistic Studies.2013(5): 91-96. Chinese.

[17] Zhang Y. Syntactic Disambiguation of Modern Chinese with Experimental Prosodic. Harbin: Harbin Institute of Technology Press, 2023. Chinese.