The Modeling and Realization of Natural Speech Generation System

Chen Fang, Yuan Baozong
Institute Of Information Science, Northern Jiaotong University
Beijing, 100044, P. R. China

Abstract: The paper gives an overall discussion on problems in Chinese natural speech generation. We considered not only how to convert text into speech but also how to generate the necessary text in text-to-speech conversion.A Chinese Bi-directional Grammar is developed to suit for Chinese Language understanding and generation. The system gets the right text and generates speech which have good quality in naturalness and intelligibility using Chinese Text-to-Speech Conversion System.

Keywords: Grammar, Text generation, Text-To-Speech Conversion

1 Introduction

It's an ideal in dream and sleep that people can use natural language to communicate with machines. Natural speech generation system is researched to simulate people's capability of speaking. In speech generation systems, we consider not only how to convert text into speech but also how to generate the necessary text in text-to-speech conversion. According to the Systematic-Functional Grammar, a Chinese Bi-directional Grammar is developed to suit for Chinese Language understanding and generation. A comprehensive description about the structure of characteristic network of all ranks in language, include word, phrase, sentence and complex sentence. Our system can select topic that need to describe and built corresponding knowledge domain. Supported by the knowledge domain, it can complete text planning, text organization and grammar realization. We set up a series of strategy in text organization and grammar characteristic network. After these steps we obtained formal machine language. Using generation rules and generation correction rules appropriate text can be collected. At last we will get speech which have good quality in naturalness and intelligibility from Chinese Text-To-Speech Conversion System. The structure of the system is showed as following:

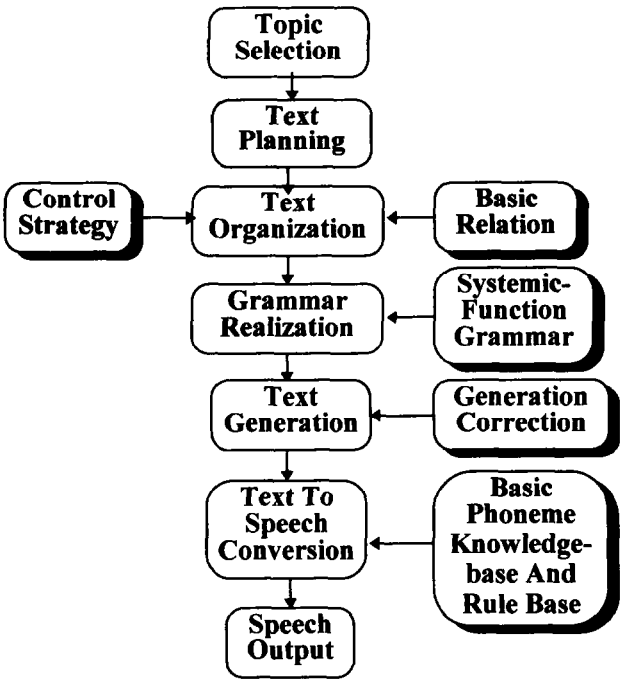


Fig 1 The Structure Of The System

2 Characteristic Network of Chinese Bi-directional Grammar

In the man-machine natural interactive system, natural language understanding and natural language generation are two important aspects. Many persons have done a large amount of work in language understanding with Transformational Generative Grammar, Augmented Transition Networks and etc. Natural language generating is a relatively new and increasingly active field of research. It haven't gotten satisfied results using the same grammar with understanding because those grammars is based on

formality ,which can't embody sentences function. Systemic-Functional Grammar avoids using transition rules. It need indicate not only relations between character strings and their meanings but also how to describe a certain meaning and how to understanding a certain representation.

Chinese grammar has many features which are different from English.We improve systemic-functional grammar according to the features in many aspects. We combine the systemic network and functional network into an integrate bi-directional grammar network that every correspondent items in systemic and functional grammar is revolved in it. We setup an characteristic grammar networks related to each rank of Chinese grammar, such as word, phrase, simple complicated sentence. We give a demonstration in the rank of phrase as Fig.2:

We set up three characteristic sub-networks of the first level in the phrase rank, which are Function Type, Structure Type and Fixed Type. In Function Type, the network gives several different choices of grammar characters, include Noun phrase, Verb phrase and Adjective phrase. In the sub-network there is sub-network of Function Intension in the second level, which gives out the role of three phrase with different grammar features. Structure Type indicates the structure in phrase composing. They are Coordinating Phrase, Modified Phrase, Predicate and Objective Phrase , Subject and Predicate Phrase. In the second level Structure Character sub-network shows the grammar characters of different structures while in the third level Grammar Function sub-network tells the function of those phrase in sentence. Another sun-network in the first level is Fixed Type. The structure of the phrase in the network is fixed. The second level sub-network Grammar Function also gives the function in sentence.

Item in the characteristic network of Chinese grammar is not limited in vertical order as an abstract form, so it is unmarkable and bi-directional suited for sentence generation. Based on the networks, it's easy to analyze the structure of input sentences and form the surface structure of output sentences. Assisted by the correspondent dictionary we can analysis and generate most of Chinese sentences and complex sentences.

3 Text Generation

A text generation system must have a lot of knowledge which can be selected so that machine can pick out different representation forms. We at first need to choose the generation topic , that is to select content related to generation. Text planning extracts the knowledge connected to concrete generation task. Then the Internal generation structure in machine is formed through text organization. After that the transformation from the internal structure to natural language text through grammar realization. At last natural language can be converted to speech. after generation correction.

(1) Topic Selection

Topic represents the knowledge related to a certain task in knowledge domain. The knowledge relation due to a topic forms a structure. When the text shaped, the generated text must be influenced by the structure, but the organizing order does not limited to the structure. The topic is defined for setting up global structure model and conform with the regular scope of that generation.

(2) Text Planning

Text planning is the process to determining the content and structure of generated text after topic selection. Text planning is hierarchical. It need at first give a global planning, which divided knowledge into direct domain knowledge and indirect knowledge. According to topic we can decide whether the indirect knowledge should be represented. The hierarchical planning method gives a global scheme that gives the detail step by step.

(3) Text Organization

Text organization decides how to form the sentence and how to organize complex paragraph through some

control strategies which are general appointment in text organization. We assigned the node to be described, then proceed strategy matching to get text organizing structure.

Control Strategy

We decided control strategies into three classes: basic strategy, multi-value strategy and null strategy.

Basic strategy includes describe strategy and find strategy. Describe strategies generate text associated with current node, and assign next current node.

Multi-value strategies are used to solve the contradiction between conflict strategies. It mainly includes select strategy, combine strategy and contradictory strategy. Select strategy picks out the best strategy in many candidates. Combine strategy is trigged when several strategy can be selected and have no priority. Contradictory strategy return to last node when several strategies conflict while combing.

Null Strategy: When no strategy matching null strategy will be adopted.

Many strategies circulate in generation.The machine records the executed process and text organization will be completed. Then we need to embody every strategy in organization. The implement of every strategy is different in language description according to grammar realization.

Basic Relation

Basic relation is the relation employed after control strategy. It ensures relation of things in topic domain, which combined as a whole through basic relation.

Basic relation include organization relation and increment relation. Organization relation is responsable for extraction information related to current node to form surface structure. Increment relation only increase information about the current node . It's an important kind of relation in generation, because it provides the methods to make text more and more to the same knowledge node.

The combing selection of several strategies can make text have more meaning. Only when every described node has a record and every selection of strategy on the basis of model defined by topic selection, we can get a continuos text.

(4) Grammar Realization

Grammar realization is consists of three step: they are grammar entry, grammar transmission and grammar embodying.

Grammar entry:

when the semantics need to be represent is in keeping with features in characteristic network, that is network entry condition, the grammar enter into every grammar rank to be generated.

Grammar transmission:

grammar network transmission means that system character is selected. Transmission activate the system characteristic network.

Grammar embodying:

After grammar entry and grammar transmission, the other task is to represent the selected result in form of declarative sentence.

(5) Text Planning

After we get surface grammar structure of text through grammar realization, connected text must be outputted. Except the assistance of dictionary, we set up some rules, They are:

- Insert rule: in characteristic network an item need to insert a certain component of language structure in grammar realization.
- Connect rule: in characteristic network an item need some component connected according to a certain order.
- Specialize rule: it is a common rule. In system characteristic network, we selected one or several item to realize.
- Function assign rule: it is a priory rule. When special rule acts to one item, some premise conditions must be possessed to define some functional components in language structure.
- Ordering rule: ordering rules determine the

priority of structure order.

Using the rules text will be generated after generation correction.

4 Text-To-Speech Conversion

The description from text-to-speech is based on its synthesis basic component. We build up a speech characteristic database with speech of 50 thousand phrases and hundreds of pronunciation rules. The rules includes pronunciation rules of words, phrase and sentence, semantic rules and rules for multi-sound words. The database gives completely a description from Chinese characters to speech. At first we recognize the structure of the input text in Chinese, abstract the rhythm characteristics in text through the analysis of semantic, grammar and lexicology. An efficient structure of the speech database is also designed. The whole Chinese character in GB2312-80(National standard for Chinese character)can be described to speech.. The analyzed text can be read in real time, the synthetic speech has good quality in naturalness and intelligibility.

5 Conclusion

Based on the research all above, a natural speech generation system is established. It can automatically plan and organize the output sentences in natural speech. The system can generate almost all types of sentences and complex sentences in Chinese.The analyzed text can be read in real time, the synthetic speech has good quality in naturalness and intelligibility.

Reference

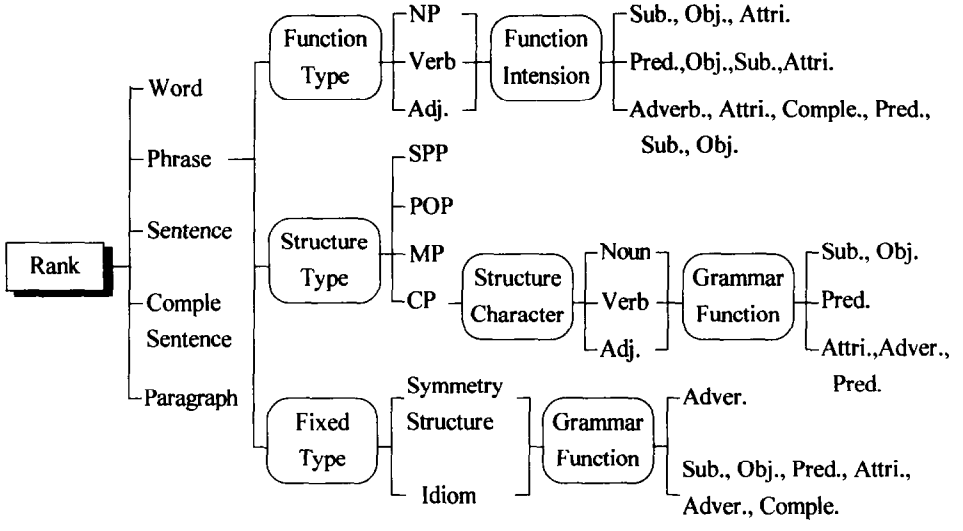
1.***T.Pattern***, Real-time generation of natural language, IEEE Expert, Oct.,1991.

2.***P.Sibun***, Generation text without trees, Computational Intelligence, Vol.8,No.1,Mar.,1992.

3.***George Ferguson, James F. Allen.*** Generic Plan Recognition for Dialogue Systems. 1993 DARPA Human Language Technology Workshop.

4.***Spoken Language Systems Group***, MIT. "Summary of Research". 1997.

5.***Chen Fang, Lin Biqin, Yuan Baozong,*** An improve-ment to BJD-92 Chinese Text to Speech System, ICSP'93,Oct.,1993.



CP: Coordinating Phrase MP: Modified Phrase

POP: Predicate and Object Phrase SPP: Subject and Predicate Phrase

Fig2 The Characteristic Network of Phrase Rank

The work was supported by National Science Foundation of China.