

System Demonstration

Multilingual Weather Forecast Generation System

Tianfang Yao Dongmo Zhang Qian Wang

Department of Computer Science and Engineering, Shanghai Jiao Tong University

1954 Hua Shan Road, Shanghai 200030, Shanghai, China

E-mail: yao-tf{zhang-dm, wangqian}@cs.sjtu.edu.cn

Abstract

The MLWFA (Multilingual Weather Forecasts Assistant) system will be demonstrated. It is developed to generate the multilingual text of the weather forecasts automatically. The raw data from the weather observation can be used to generate the weather element chart. According to the weather change trend, the forecasters can directly modify the value of the element on the chart, such as the center point value, the isoline and the isocircle. After that, the modified data are stored as the input for the system. The system can select a schema depending on the input or the requirement from the users. The schema library can be conveniently maintained, such as the schema modification or extension. Through optimizing and mapping the schema tree, the microplanner constructs the brief and coherent internal text structure for the surface generator. After the processing of the generator, the multilingual weather forecasts used for the broadcast program are generated.

Keywords: Multilingual Generation, Weather Forecast Assistant.

1 Introduction

The MLWFA system is developed as the first application of the project ACNLG [Huang et al. 97a & b] which is an international cooperation between the German Research Center for Artificial Intelligence (DFKI) and Shanghai Jiao Tong University (SJTU). The system mainly consists of four components: the graphic processor, the macroplanner, the microplanner and the surface generator. The graphic processor is used to adjust weather forecasts data by the forecasters. The technique adopted for the macroplanner is based on the schema approach [McKeown 85], but we expand the operator of schema. The microplanner is based on the sentence structure optimizing which is independent of the language and language resource mapping which is associated with the language. On the basis of the FB-LTAG (Feature-based Lexicalized Tree Adjoining Grammar) [Joshi 85, XTAGRG 95], the surface generator identifies the feature of the nodes, compounds the grammar-trees and finally generates Chinese, English and German weather forecasts.

2 Architecture

The architecture of the system is shown in figure 1. The macroplanner obtains original predicted weather data either from the interface or from the users. These data mainly include weather status, wind direction, wind force, temperature and so on.

In the stage of macroplanning, we adopt the schema approach and introduce the "||" operator which indicates that the order of the predicates can be exchanged. The macroplanning procedure

contains five major steps: inputting meteorological data, building knowledgebase dynamically, initializing predicate, selecting schema and filling schema [Wang et al. 97].

The users can dynamically define objects and input their property value. Then the system will dynamically build a hierarchical knowledgebase. With the help of that, the system transforms original data to the conceptual value. After producing the conceptual value, the system initializes the predicates with the conceptual value. The process of initialization is to fill the conceptual value in the predicate by matching the semantics of the conceptual value and the argument. We add the layout information on the predicate and define some valid argument combinations that enhance the expressible ability of the predicate. In the macroplanner, system selects schema according to three conditions. There are selecting information carried by the schema item, big grain information that is deduced by several conceptual values according to the domain conditions and the selecting priority.

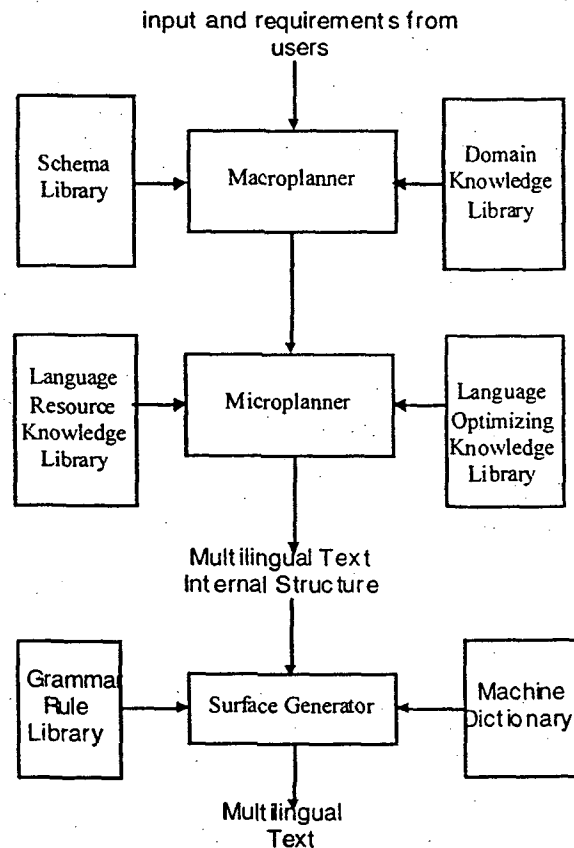


Fig. 1 MLWFA System Architecture

The work of the microplanner can be divided into two closely related parts: the sentence structure optimizing (SSO) and the language resources mapping (LRM). SSO aims at removing redundant information in the schema tree so as to make the sentence natural, fluent, and readable, and a semantic-based rule driven method is adopted in this component. LRM adds the function words into it in conformity to the specific language syntax, and enriches them with the specific language resources so that every expression in the output is flexible and grammatically correct.

In accordance with the weather forecast domain, 5 types of optimizing tasks are classified: they are the aggregation of the proposition with the common parts, the constitute conjoining, the

domain restriction, the constitute omission and the reference planning. Each stands for a group of the optimizing operations. In this idea SSO can be confined within the semantic level which is irrelevant with any specific language, so a single operation can be done for all languages.

As to LRM, the sentence structure classes are used to represent the detailed sentence structures as well as the syntax and semantic information for every composition of the sentence. There are several sentence structure classes defined for every predicate. Considering text requirements, one of the classes above is chosen and determined to construct the corresponding text structure tree. Furthermore, according to the characteristics of the different language, the dynamic information is added to the relevant nodes of the tree. As the definition differs widely for different target languages, we implement this module for different languages respectively.

The input of surface generator is the output of the microplanner, which is a dependent tree structure including not only the information of the text content, the sentence syntax and the sentence semantics, but also the result of the optimizing.

We have used the FB-LTAG as a generation grammar for our generator [Yao et al. 97, Wang 97]. The algorithm of the surface generator can be summarized as:

- Instantiate the trees and compute all legal subtree compositions;
- Pick out the first tree from the set of acceptable derived trees, which satisfies the feature constraints and has a complete structure;
- Traverse the selected tree and output the sentences (do some morphology according to the feature values, if necessary).

3 Demonstration

The main window of the system during the macroplanning is shown in figure 2.



Fig. 2 The Main Window

At the top of the screen is the main menu of the system. In the system submenu the user can maintain all the database in the system. In the data submenu the system can generate Chinese, English or German text, through executing the macroplanning, the microplanning and the surface generation procedure in turn. In the prediction submenu the user can change options of the system, either produce forecasts step by step or do it uninterruptedly. The left window on the middle of the screen shows the weather element chart, the user can modify the element value on the chart of that window directly. The right window shows the result of the macroplanning, we split that window into two parts: One shows the tree structure of the macroplanning result; the

other shows the detail information of the focusing node.

4 Conclusion

We have implemented a prototype of MLWFA system consisting of four components in VC++ under Windows 95. Currently we are replenishing more optimizing rules and sentence structure classes, and further modifying some existing ones. After doing that, we will plan to put this prototype into test use.

Acknowledgements We want to thank sincerely Dr. Xiaorong Huang for his concrete instruction and valuable suggestion. Thanks also to the people who took part in the project or made valuable suggestions.

This work is supported by the VW-Stiftung, Germany, the National Science Foundation, China, and the Science and Technology Commission of Shanghai, China.

References

- [Huang et al. 97a] Xiaorong Huang, Tianfang Yao and Guodong Gao. 1997. Generating Chinese Weather Forecast with Stylistic Variations. In Proc. 17th International Conference on Computer Processing of Oriental Language, Hongkong, 733-738
- [Huang et al. 97b] Xiaorong Huang, Tianfang Yao, Huanye Sheng. 1997. The Project *ACNLG*: Applying Natural Language Generation in China. In Proc. DFKI Workshop on Natural Language Generation, Saarbruecken, Germany.
- [Joshi 85] Aravind K. Joshi. 1985. An Introduction to Tree Adjoining Grammars. Technical Report MS-CIS-86-64, LINCLAB-31, Department of Computer and Information Science, Moore School, University of Pennsylvania.
- [McKeown 85] Kathleen R. McKeown. 1985. Text Generation. Cambridge Press, Cambridge, UK.
- [Wang et al. 97] Qian Wang and Tianfang Yao. 1997. Design and Implementation of Macroplanner for Chinese Weather Forecast Automated Generation System, In Proc. of the Joint Symposium Computational Linguistics '97, Beijing, China, 176-181
- [Wang 97] Yufang Wang. 1997. A Tree Adjoining Grammar for Chinese Weather Forecasts. In Proc. of the Joint Symposium Computational Linguistics '97, Beijing, China, 207-213
- [XTARG 95] The XTAG Research Group. 1995. A Lexicalized Tree Adjoining Grammar for English. Technical Report IRCS 95-03, Institute for Research in Cognitive Science, University of Pennsylvania.
- [Yao et al. 97] Tianfang Yao, Xiaorong Huang and Huanye Sheng. 1997. The MLWFA System. In Proc. of the Natural Language Processing Pacific Rim Symposium 1997. Phuket, Thailand, 609-612.