

MLWFA: A Multilingual Weather Forecast Text Generation System¹

Tianfang YAO Dongmo ZHANG Qian WANG

Department of Computer Science and Engineering
Shanghai Jiao Tong University

1954 Hua Shan Road

Shanghai, China, 200030

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

Area: Application systems using language technology components or Natural language processing systems

Keywords: Multilingual text generation; Weather forecast generation system

Abstract

In this demonstration, we present a system for multilingual text generation in the weather forecast domain. Multilingual Weather Forecast Assistant (MLWFA) is the first practical application system of the project Applied Chinese Natural Language Generation (ACNLG) that is a cooperation research project between German Research Center for Artificial Intelligence (DFKI²) and Shanghai Jiao Tong University (SJTU) (Huang, 1996).

The Department of Computer Science and Engineering, Shanghai Jiao Tong University developed MLWFA. Shanghai Meteorological Center (SMC) is a cooperation partner and provides the instructions for the meteorological techniques.

We used pipeline architecture (macroplanner, microplanner and surface generator) as kernel architecture in our system. The weather chart processor (WCP) (Yao et al., 1998) is an important component in the macroplanner. Depending on the requirement of forecasters, we provided the processing environment, in which the forecasters can intuitively adjust the weather chart and modify weather forecast data. Thus, the numerical weather forecast and the experience of forecasters are organically

combined for further enhancing the accurate rate of weather forecast. For the purpose of strengthening the flexibility of the macroplanner, we proposed an input mode with three layers, which is an open interface for user to define application objects and relationships. We also improved the schema approach (Mckeown, 1985; Huang, Yao and Gao, 1997; Wang and Yao, 1997) and integrated template technique with schema technique. The domain adaptability and the expression flexibility of text content are enhanced. In the microplanner, the sentence-optimizing technique (Li, Zhang and Yao, 1998) and the mapping technique of the linguistic resource class are implemented. The redundant of the sentences that are independent of languages is removed. Moreover, we adopted Text Structure (TS) (Meteer, 1991) to organize the linguistic resource class that defines the structure of sentences and phrases for different languages. Based on TS, we defined the formalism of the linguistic resource class, as well as proposed a match algorithm for the mapping of the linguistic resource class. The linguistic resource expression and the sentence construction have flexibility in the mapping of the linguistic resource class. In addition, the formalism of the linguistic resource class and the match algorithm are also suitable to extend the linguistic constructions of different complexity

¹ The project is financially supported by the VW-Foundation, Germany, the National Natural Science Foundation, China, and the Science and Technology Commission of Shanghai, China

² Deutsches Forschungszentrum für Künstliche Intelligenz

(like sentence and phrase), and even to add a new language. We developed FB-LTAG (Feature Based-Lexical Tree Associate Grammar) (Joshi, 1985; Joshi and Schabes, 1992; Vijay-Schanker, 1987; The XTAG Research Group, 1995) based multilingual surface generator as well as the corresponding lexicon and syntactic and semantic rule library for Chinese, English and German. The syntactic and semantic rule is general for languages and domains, so that it can be to adapt new languages and new application domains. Because the rule formalism used is declarative, it is supposed to be modified and complemented for all languages. Moreover, it can be reused for integrating the surface generator in other systems, for example, information retrieval system, machine translation system, etc.

We have implemented the system MLWFA in VC++ under Windows 95/98 or NT. It can generate multi-city, marine and precipitation forecasts in Chinese, English and German. The system has shown that the generated texts conform to the readable requirement and the domain expression custom. At the same time, the flexibility and the extensibility in the system are also enhanced.

System Features:

- The system can be to adapt the variation of application domains
- Users can redefine text structure and content according to their requirement
- Redundancy between sentences can be removed to minimal range
- The microplanner can be extensible to add different linguistic constructions and even a new language
- There are various sentence structures for same content included in the generated text
- The grammar is general and has a big coverage for the linguistic phenomena
- Generate Chinese, English and German weather forecast texts at real-time
- Good readability
- User-friendly interface

Acknowledgements

Our thanks go to the VW-Foundation, Germany, the National Natural Science Foundation, China,

and the Science and Technology Commission of Shanghai, China.

References

- Huang X. R. (1996) *The Project ACNLG*. In: Proc. of the International Symposium on Natural Language Generation and the Processing of the Chinese Language, INP (C)-96, Shanghai, China, pp. 5-10.
- Huang X. R., Yao T. F. and Gao G. D. (1997) *Generating Chinese Weather Forecast with Stylistic Variations*. In: Proc. of 17th International Conference on Computer Processing of Oriental Language, Hong Kong, China, pp. 733-738.
- Joshi A. K. (1985) *An Introduction to Tree Adjoining Grammar*. Technical Report MS-CIS-86-64, LINCLAB-31, Department of Computer and Information Science, Moore School, University of Pennsylvania.
- Joshi A. K. and Schabes Y. (1992) *Tree-adjoining Grammar and Lexicalized Grammars*. In "Tree Automata and Languages", Nivat M. & Podelski A. eds., Elsevier Science, pp. 409-431.
- Li J. Q., Zhang D. M. and Yao T. F. (1998) *The Study of Sentence Optimizing for Multilingual Text Generation*. Application Research of Computer, Vol. 15, No. 1., Cheng Du, China, pp. 52-55.
- Mckeown K. R. (1985) *Using Discourse Strategies and Focus Constraints to Generation Natural Language*. Cambridge University Press, Cambridge, U.K.
- Meteer M. W. (1991) *Bridge the generation gap between text planning and linguistic realization*. Computational Intelligence, Vol. 7, No. 4, pp. 296-304.
- 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, pp. 5-10.
- Vijay-Shanker K. (1987) *A Study of Tree Adjoining Grammars*. Ph.D. Thesis. Department of Computer and Information Science, University of Pennsylvania.
- Wang Q. and Yao T. F. (1997) *Design and Implementation of Macroplanner for Chinese Weather Forecast Automated Generation System*. In "Language Engineering", Chen L. W. & Yuan Q. eds., Tsinghua University Press, Beijing, China, pp. 176-181.
- Yao T.F. et al. (1998) *An Editing Method of Weather Chart for Weather Forecast Text Generation System*. Journal of Chengdu Institute of Meteorology, Vol.13, No.3, Chengdu, China, pp. 222-230