4.4 **METEOCOGENT: A KNOWLEDGE-BASED TOOL FOR GENERATING WEATHER FORECAST TEXTS**

Richard Kittredge* and Benoit Lavoie
CoGenTex, Inc., Ithaca, NY

1. INTRODUCTION

The production of textual forecasts by professional meteorologists is a time-constrained activity requiring the interaction of knowledge about meteorology, knowledge about language and knowledge about the conventions used to communicate the right amount of weather information to meet the needs of diverse user communities. In situations where forecasts are written manually, the time spent composing text is usually time lost for analyzing the ever-increasing amount of guidance information. The production of written draft forecasts from forecast data has long appeared to be an excellent application for knowledge-based techniques, given the high volume of text produced and the circumscribed and recurrent nature of the task. Today, however, the issue is not *whether* to build a rule-based system, but rather *how much* knowledge representation is required, *how* text production should be integrated into the user's authoring process, and *how flexible and user-accessible* to make the knowledge components of such a system. These issues are closely related to the life-cycle costs and benefits of the software, but more importantly, to user acceptance.

Attempts to automate forecast composition over the past two decades have met with partial success and fall into two basic categories, depending on the amount of linguistic knowledge used: Computer-worded forecasting (CWF), and Knowledge-based text generation (KBTG). In both CWF and KBTG, text production usually starts from "explicit weather elements" that are derived from the gridded data output of numerical models by various "summarizing techniques", which work to pick a representative characterization of weather elements across the grid points in a zone of interest, cf. Ruth and Peroutka (1993). Neither CWF nor KBTG aims to replace the forecaster as final arbiter of output text, but rather to compose instantaneously formatted draft texts that the forecaster can revise and release quickly.

1.1 Computer-Worded Forecasting (CWF)

In CWF, short phrases inserted into slots in a fixed template structure to form forecast sentences. The template corresponding to the whole sentence is chosen based on the general type of information to be conveyed, including the presence of mixed or unusual weather conditions. Then within the template, the string values of slots to be filled are calculated using conditions on the numerical or symbolic values of weather elements. Basically, for each slot filler in a template, there is a direct mapping from a boolean combination of conditions to a piece of text which will appear in the output. The piece of output text has no internal structure, and is not given any linguistic marking as to its syntax or semantics. Rather, it has a single information category in its role of slot filler in the template for the full sentence.

CWF approaches have been developed in several weather offices around the world, cf. Glahn (1979), but until recently have been used mostly as guidance. Following a period of development at NOAA's Techniques Development Laboratory, an interactive CWF approach (ICWF) has begun to be used operationally to produce public zone forecasts, cf. Ruth and Peroutka (1993) and Calkins and Peroutka (1997). CWF techniques are clearly meeting a need, especially in producing texts with simple, short sentences. Nevertheless, there are several limitations in CWF approaches: (1) as the length and complexity of sentences and texts increase, the number of templates and the complexity of their usage conditions can become very large and hard to manage, (2) the lack of a linguistic basis in the rules makes it difficult to guarantee smooth and grammatical text under all conditions, and (3) CWF systems are relatively hard to adapt to local need, or extend to new

* *Corresponding author's address*: Richard Kittredge, CoGenTex, Inc., 840 Hanshaw Road, Ithaca, NY 14850-1589; authors' e-mail: {richard|benoit}@cogentex.com.

forecast varieties. Despite the shortcomings, CWF systems are relatively easy to get up and running initially, when the rule base is small. Their popularity in large organizations seems also to be due partly to the fact that they are "home grown" and hence more economical and controllable by in-house staff.

1.2 *Knowledge-Based Text Generation (KBTG)*

KBTG uses a more modular approach to text construction, typically broken into three major stages, (1) content planning, (2) text planning and (3) text realization. *Content planning* involves building a small number of significant weather event "messages" from the time series of input weather elements. The concepts (e.g., "significant wind change") used in describing events may be defined in accordance with user community interests and priorities. The process of *text planning* orders and groups the messages, and sets up the future clause structure and conjunctions between clauses. Text planning feeds directly into grammar-based *text realization*, in which details such as word order and subject-verb agreement are handled by general rules which access an application-dependent lexicon. The first KBTG system for operational forecasts has been Environment Canada's FoG system, cf. Goldberg et al. (1988, 1994), running on the Forecast Production Assistant (FPA) workstation, cf. Paterson et al., (1993). Similar systems are being developed by national weather services in a few other countries.

KBTG has potential advantages of expressive power and modularity over CWF. Texts are planned at a conceptual level, which delays purely linguistic operations until after the major sentence structures have been determined. Single concepts tend to be mapped to single words within a KBTG system. By contrast, a given word may appear on the right hand side of dozens of rules (which map complex conditions to phrases) in a CWF system. In principle, a KBTG system which incorporates linguistic knowledge gives (1) higher quality text when sentences are long or complex, (2) a better basis for speech synthesis, (3) a more natural framework for multi-lingual output, (4) easier tailoring of text to local needs, and (5) greater perspicuity for maintenance and reusability. The advantages of KBTG over CWF approaches become especially clear in the domain of synoptic forecasts, where the number of templates required for a CWF system is very high. The drawbacks of KBTG in the past have been the higher initial costs of building a knowledge-based system with the required linguistic knowledge, and the scarcity of specialized linguistic programming skills.

Having seen first-hand the reluctance of potential users to embrace KBTG, CoGenTex has addressed some of their concerns in MeteoCogent, a knowledge-based tool for building weather forecast text generators. MeteoCogent facilitates implementation, customization and maintenance of forecast generators without the need for extensive linguistic training. The next section discusses MeteoCogent's features and operation, while section 3 addresses some of the current limitations and plans for future improvement.

2. METEOCOGENT

2.1 *MeteoCogent Design Features*

MeteoCogent is a tool which aims to help meteorologists build KB generators for texual forecasting. First of all, it helps set up the knowledge resources needed to carry out KBTG for a particular text product. It builds in linguistic features that can give an application generator the capability to synthesize complex sentences, and if required, produce multilingual output, etc. MeteoCogent also includes several new features to speed the process of KBTG system building, facilitate the configuration of forecasts to meet local needs, support generator maintenance and integration with other applications:

- MeteoCogent supports generator configuration by means of several resource files in which the user specifies concept definitions, global text organization and output format of the resulting forecasts;
- MeteoCogent enhances access to linguistic resources, so that users can, for example, add or change terminology by editing word entries in the lexicon;
- MeteoCogent uses RealPro, cf. Lavoie and Rambow (1997), a fast new syntactic realizer based on the same Meaning-Text language models used in FoG;
- User-configured generators are set up to produce hypertext documents that can combine forecast text, tables and graphics;
- MeteoCogent is provided as server with a programmatic interface which allows integration with existing applications; It is also provided with a web interface allowing its use as a standalone application;
- MeteoCogent components are implemented in C++ and/or Java for optimal performance and portability.

2.2 *MeteoCogent System Components*

MeteoCogent runs as a server communicating with other applications through sockets. Figure 1 illustrates the MeteoCogent architecture. The system is composed of three components using several customizable resource files:

- *Server Interface.* This component interprets the incoming requests and redirects them to the Weather Data Manager or to the Text Generator.
- *Weather Data Manager*. This component contains the functions used to build concepts from weather element data.
- *Text Generator*. This component generates the weather forecasts using resource files to control content, organization and output format. It is composed itself of several subcomponents not represented in Figure 1 and which represent the different stages of the text generation process: text planning, sentence planning, realization and formatting.

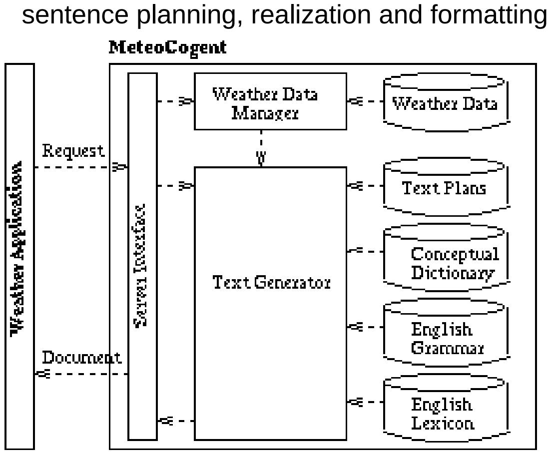

Figure 1. MeteoCogent Design

2.3 *How MeteoCogent Works*

In order to illustrate how MeteoCogent works and how its resource files can be used to customize the application product generator, let us assume that the user is building a generator for public zone forecasts, which he in turn uses to generate the simple forecast (here, from simulated data) illustrated in Figure 2.

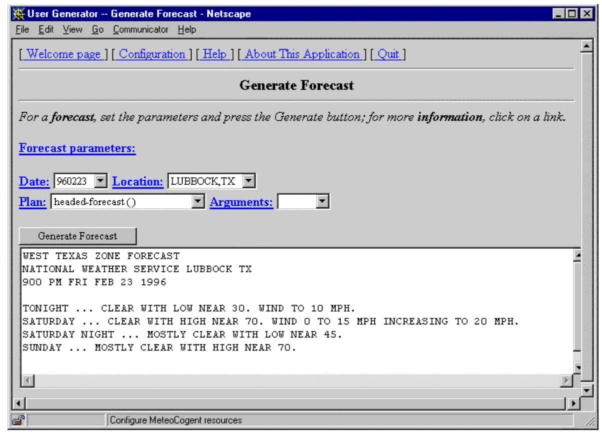

Figure 2: Output from User-configured Generator

To generate text of a particular type, a user must first configure a global plan for the text. A text plan is a tree or graph structure defining the possible content, organization, and format of a text and specifying under which constraints this information can occur. A partial text plan corresponding to the zone forecast with header shown in Figure 2 is illustrated in Figure 3. The text plan components have hypertext links to their definitions, which can then be edited. Text plan components can also be edited at generator run time.

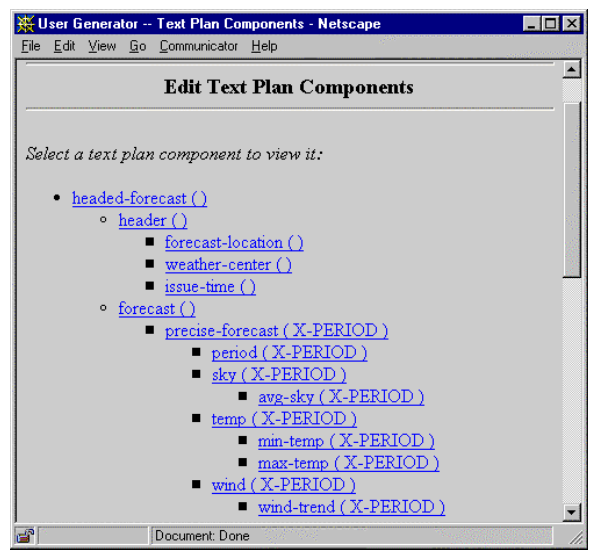

Figure 3. Partial Text Plan

The definition of a text plan constituent includes a specification of the conceptual representation of the corresponding text segment. This specification can have variable parts which are instantiated by requesting the missing information to the Weather Data Manager. For example, *(#SKY #category ?)* is an uninstantiated specification of the state (category) of the sky. Following Ruth and Peroutka (1993), the category is instantiated in a range of *1* to *5* where a category of *1* indicates a *clear sky* while a category of *5* indicates an *overcast sky*. Note that the purpose of a conceptual representation is to be language independent, allowing for possible multi-lingual generation, and that the use of English labels at the conceptual level is simply to facilitate readability.

MeteoCogent uses a conceptual dictionary to associate conceptual representations with linguistic representations at the (deep) syntactic level. Figure 4 shows how five distinct degrees of sky coverage are lexicalized. Lexicalization can be controlled by simply modifying the appropriate entries in the conceptual dictionary: for example although in figure 4, *(#SKY #category 1)* is being lexicalized as *(CLEAR*), it is also possible for the user to substitute *(CLOUDLESS)*.

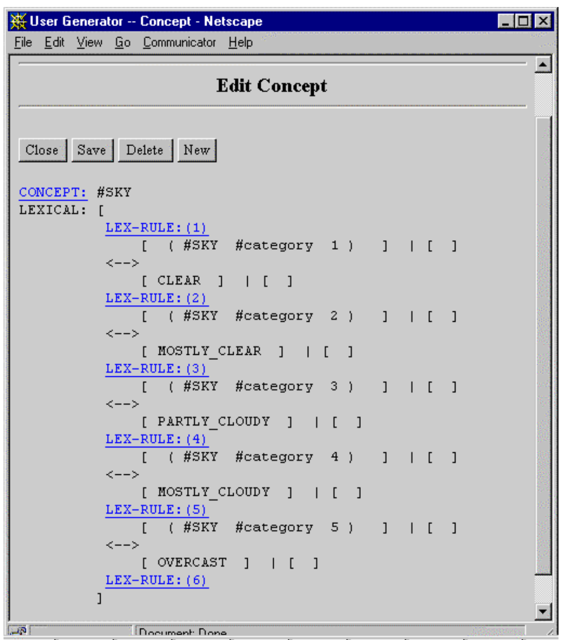

Figure 4. Sample of Conceptual Dictionary Entry

By modifying one or more resource files used by MeteoCogent, a user can control the content and the organization of output weather forecasts. Figure 5 illustrates a simple variation of the text shown in Figure 2 obtained with only a few changes to the resource files.

WEST TEXAS ZONE FORECAST
NATIONAL WEATHER SERVICE LUBBOCK TX
900 PM FRI FEB 23 1996

TONIGHT ... CLOUDLESS. MINIMUM NEAR 30 WITH WIND UP TO 10 MPH.
SATURDAY ... CLOUDLESS. MAXIMUM NEAR 70 WITH WIND 0 TO 15 MPH INCREASING TO 20 MPH.
SATURDAY NIGHT ... MOSTLY CLEAR. MINIMUM NEAR 45.
SUNDAY ... MOSTLY CLEAR. MAXIMUM NEAR 70.

Figure 5. Output from Modified Generator

3. DISCUSSION

These simple examples cannot do full justice to the potential power of MeteoCogent to generate much more complex texts. The RealPro realizer, in conjunction with a variety of text plans and planning techniques, has already been used to produce complex text in several different domains. RealPro contains a broad-coverage English grammar of the kind needed to generate synopses. What remains for building an actual synopsis generation system, is for meteorologists to use MeteoCogent's framework to specify text plans and concept definitions which are appropriate for these texts. At present, users can define concepts (e.g., 'rapid drop in temperature') in terms of a library of C++ functions which access weather element data to compute differences, averages, maxima, etc. Clearly, this set needs to be extended as MeteoCogent moves from its current alpha status to beta test sites. These extensions would best be done by the users themselves, as a part of their site customization.

REFERENCES

Calkins, J., and M. Peroutka, 1997: Local Effect Phrases in the Interactive Computer Worded Forecast. *Preprints Thirteenth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology*, Long Beach, Amer. Meteor. Soc., 120-123.

Glahn, H., 1979: Computer worded forecasts. *Bul. Amer. Meteor. Soc.*, **60**, 4-11.

Goldberg, E., N. Driedger and R. Kittredge, 1994: Using Natural Language Processing to Produce Weather Forecasts. *IEEE Expert*, **9**(2), 45-53.

Goldberg, E., R. Kittredge and A. Polguère, 1988: Computer Generation of Marine Weather

Forecast Text. *J. Atmos. Oceanic Tech.*, **5**(4), 473-483.

Kittredge, R., T. Korelsky and O. Rambow, 1991: On the Need for Domain Communication Knowledge. *Canadian Computational Intelligence Journal*, **7**(4),305-314.

Lavoie, B., and O. Rambow, 1997: A Fast and Portable Realizer for Text Generation Systems. *Proc. 5th Conference on Applied Natural Language Processing*, Washington, 265-268.

Paterson, R., B. de Lorenzis, N. Driedger, E. Goldberg, B. Greaves, and R. Trafford, 1993: The Forecast Production Assistant. *Preprints Ninth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology*, Anaheim, Amer. Meteor. Soc., 129-133.

Ruth, D., and M. Peroutka, 1993: The Interactive Computer Worded Forecast. *Preprints Ninth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology*, Anaheim, Amer. Meteor. Soc., 321-326.