

Bilingual Generation of Weather Forecasts in an Operations Environment

L. Bourbeau, D. Carcagno, E. Goldberg,
R. Kittredge and A. Polguère

1 Introduction

In 1986 the first experiments in text generation applied to weather forecasts resulted in a prototype system (RAREAS[6,3]) for producing English marine bulletins from forecast data. Subsequent work in 1987 added French output to make the initial system bilingual (RAREAS-2[11]). During 1988-1989 a full-scale operational system was created to meet the needs of daily marine forecast production for three regional centres in the Canadian Atmospheric Environment Service¹. In contrast to the earlier systems, the most recent one uses general models for both text planning and sentence realization (see sections 4 and 5 below).

This new implementation, dubbed FoG for Forecast Generator, may constitute one of the first "industrial" uses of text generation. FoG is of interest to computational linguists for three additional reasons:

- the conceptual input to the text generation process is derived from data that also drive a graphic display on a workstation for forecasters; this determination of text from a selected subset of graphically displayed data represents an important paradigm for the transformation of information;
- conceptual processing results in an "interlingual" representation, a kind of deep syntactic structure for both English and French in this sublanguage;
- sentence generation is carried out using a "streamlined" version of the Meaning-Text linguistic model; this may represent the first time that such a general model has been adapted to the descriptive problems arising in telegraphic sublanguages.

¹This work was carried out at Odyssey Research Associates, Montreal, under contract with Environment Canada. Current affiliations of the authors are: Bourbeau, Kittredge and Polguère at Université de Montréal [e-mail: bourbeau/kittredg/polguere@iro.umontreal.ca]; Carcagno at CGE-Marcoussis, France [carcagno@crge1.cge.fr]; Goldberg at Environment Canada, Toronto [eli@aeshq.uucp].

2 The Graphical Environment of Weather Forecasting

Operational meteorologists normally work with graphical representations of the information available to them. "Charts" are used to display the large volumes of observational data and also the results of global simulations of the atmosphere. The graphical entities displayed on these charts (such as weather fronts, and low pressure systems) are manipulated to adjust for more recent data, and for perceived errors in the simulations. This results in manually created weather depictions which are valid at some future time (24 to 36 hours in the future).

The weather situation is always being monitored and updated as new information is received and assimilated. During the normal course of events, much of the communication between forecasters is done using these charts. When it is time to write a forecast for some user community, the forecaster has to extract the pertinent information from these charts and recast them into a structured text. In addition, the 'primary' information taken from the charts has to be modified for local geographic effects. The forecaster appears to do this while the text is being composed. This mental transposition of meteorological information from graphical to text form is believed to be open to a number of subjective errors. In addition, the pressure to compose text often conflicts with the scientific demands of analyzing an emerging weather situation.

FoG is part of the recently implemented Forecast Production Assistant (FPA)[7], which uses interactive computer graphics to allow the meteorologist to view and edit a display of the weather situation. All of the fields produced by the large scale computation are directly available on the FPA together with any manually produced products. This makes it possible to obtain numerical values directly from the charts and to use them in other applications such as FoG. All of the fields required to produce forecast text can be obtained from the computer graphics.

3 From Data to Concepts

A sampling procedure is used to determine values of these fields at specific latitude and longitudes which have been pre-selected as being representative of weather conditions over a specified geographic area. Computer animation techniques are used to interpolate between the standard chart times (normally every 12 hours) to whatever time resolution is required for the text product. Currently, charts are available at intervals of three hours through the forecast period. The problem is that this yields nine values for a 24 hour forecast. Practical considerations limit the number of events (e.g. shifts in wind speed or direction) in a forecast to three or four, depending on the severity of the weather. The conceptual phase of the processing treats the sampled data so that only the significant events in time and space are passed on to textual encoding.

Conceptual processing involves several stages. 1) Events requiring "weather warnings" are identified and stored before any data smoothing is done, 2) Sampled data is smoothed with respect to time so that only the significant weather changes are retained, 3) Spatial smoothing is done so that areas sharing similar weather conditions can be grouped together in the text. We have noticed, however, that the notion of "significance" is partly dependent on the ability of the lexicon of the forecast language to make semantic distinctions. Thus, a wind change of 30 degrees is more likely to be judged significant when it crosses the boundary between, say, *north-easterly* and *easterly*, than when it stays entirely within the range of one of these terms. The semantic granularity of temporal adverbs has a similar "anticipatory" effect on the way generalizations are made over time. This constitutes a kind of filter on content determination that precedes formal text planning.

4 Text Planning

Text planning in FoG consists of three stages: content determination, text structuring and interlingua production. Content determination covers the problems of (1) converting the smoothed data on significant meteorological events into complex objects appropriate for inferencing, one object for each meteorological event of interest, and (2) using the structured data objects to compute additional concepts needed to talk about transitions between weather events. The output of content determination is, for each forecast area, an enriched data object called a "text content representation".

Text structuring consists basically of finding the optimal way of cutting each text content representa-

tion (the conceptual representation corresponding to one future text) into sentence-sized chunks of information ("sentence partitioning") within the complex text structure. The chunks are then linearly ordered according to principles that are sometimes domain-specific, but often more general (e.g., temporal sequence). There is a subsidiary problem of making full or partial copies of certain concepts to assure continuity of reference between consecutive sentences. The output of the text structuring process gives, for each forecast area, a partitioned and possibly enriched structure called the "text representation".

The final stage in text planning involves converting the single partitioned text representation into an actual sequence of conceptual representations for individual sentences. The strong similarity between forecast styles and structures used in Canadian marine forecasts in the two official languages makes it possible to formulate a single interlingual structure, which can map quite directly to the "deep" syntactic structure of the corresponding sentence in either English or French. The primary issue here is the identity of information conveyed in the two parallel sublanguages, and the fact that sentence scoping may be performed in identical ways on the text content representations used for English/French. There is no guarantee that such an interlingua would suffice for a language using a very different conceptual system or communication style for weather phenomena (e.g., Inuktitut).

5 Meaning-Text Realization Component

The last part of forecast generation involves the relatively well-developed technique of sentence realization. By this we mean the conversion of interlingual representations of English/French sentences into acceptable word strings in one or the other of these two languages. To guarantee generality and long-term flexibility of linguistic modelling, we have chosen to use the Meaning-Text linguistic theory of Mel'čuk[8], which has also served as the framework for text generation in other technical sublanguages[5,4]. Because of the lack of semantic paraphrase in the forecasting sublanguage considered, however, we have eliminated the semantic net representations from the processing stages, passing directly from interlingual representations to deep syntactic dependency trees[10]. We have implemented a fragment of an existing Meaning-Text model for English[9] and adapted this model for French.

6 Implementation

FoG is written in Quintus Prolog and runs on a Hewlett-Packard 9000 workstation as part of the FPA system. The graphics software on the FPA workstation is programmed mostly in C. As of April 1990, the entire FPA is undergoing testing by three regional weather centres in Eastern Canada and is being phased into daily production, initially during one of the three daily work shifts.

7 Future Plans

Since FoG is now configured only to produce marine forecasts for the Halifax, Gander and Great Lakes regions of Canada, an early priority is to adapt the software (mostly the text planner) to the different content and style of forecasting found in Pacific Canada and other marine regions, and to specialized marine forecasts (e.g., for small craft). Concurrently, investigation should continue into extending the system to other forecast types, including agricultural and general public forecasts. We expect that our linguistic model will also facilitate the addition of high-quality voice output as an option at some future time.

References

- [1] Carcagno D. (1989) *Text Planning for Generation of Weather Forecasts*, Technical report, Odyssey Research Associates, Montreal.
- [2] Carcagno D. and Iordanskaja L. (1989) "Content Determination and Text Structuring in GOSSIP", *Extended Abstracts of the Second European Workshop on Natural Language Generation*, Edinburgh.
- [3] Goldberg E., Kittredge R. and Polguère A. (1988) "Computer Generation of Marine Weather Forecast Text", *Journal of Atmospheric and Oceanic Technology*, vol.5, no.4, pp.473-483.
- [4] Iordanskaja L., Kittredge R. and Polguère A. (1988) "Implementing a Meaning-Text Model for Language Generation" supplement to *Proc. of the 12th Intl. Conf. on Computational Linguistics*, Budapest.
- [5] Kittredge R., Iordanskaja L. and Polguère A. (1988) "Multi-Lingual Text Generation and the Meaning-Text Theory", *Proc. of the 2nd International Conf. on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Carnegie-Mellon University.
- [6] Kittredge R., Polguère A. and Goldberg E. (1986) "Synthesizing Weather Forecasts from Formatted Data", *Proc. of the 11th Intl. Conf. on Computational Linguistics*, Bonn.
- [7] McLeod J.C. (1990) "The Forecast Production Assistant" *Preprints Sixth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology*, Anaheim CA, Amer. Meteor. Soc.
- [8] Mel'čuk I. (1981) "Meaning-Text Models", *Annual Review of Anthropology*, vol.10, pp.27-62.
- [9] Mel'čuk I. and Percov N. (1987) *Surface Syntax of English*, Benjamins.
- [10] Polguère A. (1988) *A Dependency Grammar for Marine Weather Forecasts*, Technical report, Odyssey Research Associates, Montreal.
- [11] Polguère A., Bourbeau L. and Kittredge R. (1987) *RAREAS-2: Bilingual Synthesis of Arctic Marine Forecasts*. Technical report, Odyssey Research Associates, Montreal.