AI Magazine Volume 16 Number 3 (1995) (© AAAI)

The Seventh International Workshop on Natural Language Generation

Koenraad De Smedt, Eduard Hovy, David McDonald, and Marie Meteer

■ The Seventh International Workshop on Natural Language Generation was held from 21 to 24 June 1994 in Kennebunkport, Maine. Sixty-seven people from 13 countries attended this 4-day meeting on the study of natural language generation in computational linguistics and AI. The goal of the workshop was to introduce new, cutting-edge work to the community and provide an atmosphere in which discussion and exchange would flourish.

The Seventh International Workshop on Natural Language Generation was held from 21 to 24 June 1994 at the Nonantum Inn on the seacoast in Kennebunkport, Maine. Sixty-seven people from 13 countries attended this successful 4-day meeting, coming from as far away as Japan, Australia, and Europe.

The study of language generation in computational linguistics and AI is still overshadowed by the study of parsing and analysis. By all ways of measuring, natural language generation has received less attention: fewer conferences, fewer dissertations and books, considerably less space in textbooks, and so on. Over the past 15 years, however, natural language generation has started forming an identity as a separate field of research. The first workshop was held in 1983 and has been followed every other year by an international workshop. An additional European workshop series, held in the alternate years, reflects the importance given to generation in Europe that is not given in the United States. Other related workshops devoted to specialized subtopics such as evaluation, explanation generation, and summarization occur with increasing frequency in conjunction with major meetings, such as that sponsored by the Association for Computational Linguistics.

The international workshops have traditionally been the premier gatherings for the presentation of research by this community, bringing together researchers in what is an increasingly more consolidated field. Several of the workshops have led to the production of proceedings and carefully edited books on the state of the art in the field (Dale et al. 1992; Paris, Swartout, and Mann 1991; Kempen 1987).

The goal of this latest workshop was to introduce new, cutting-edge work to the community through a mixture of presented papers, invited talks, posters, a panel, and computer demonstrations and provide an atmosphere in which discussion and exchange would flourish. Although the days were long, containing long and short paper presentations followed by discussions, joint meals and social events in a relaxed setting provided opportunities for informal conversations. The choice of a relaxing, casual site contributed greatly to the success of the workshop in stimulating the exchange of ideas.

The goal of this latest workshop was to introduce new, cutting-edge work to the community.

Although much work of recent vintage has focused on either the automated planning and organization of multisentence text structure or the problem of multilingual generation (these areas were prominently represented at the workshop), the workshop included several new themes with considerable promise. Of notable interest was a set of papers and an invited talk dealing with lexical aspects—not only the problem of lexical choice but the nature and content of lexicons and aspects of multilinguality.

Presentations at the workshop also addressed the following questions: What is the relation between strategic generation (what to say) and tactical generation (how to say it)? What kind of planning and what kind of representations are involved in creating a substantial (even multiparagraph) discourse? What specific issues are involved in creating instructional texts? At what levels of generation is information processed with respect to time, event structure, and so on? How can we generate multilingual texts efficiently?

The topics presented at the workshop included overall structural issues (generator architecture, input, and so on); the capture of the relationship between representations of the world and language in various ways (ontologies, viewpoints, simulations, cognitive modeling, and so on); lexical choice and lexicons; content selection and planning (including the use of text plans and discourse coherence relations as well as planning techniques from AI); the bridging of the gap between text planners and sentence realizers, including reference; multilingual generation (instructions and other text); grammatical issues and the realization process; dialogue (including digressions and responses); generation in different styles and modes (from sign language to genre characterization); the use of generators in a specific task environment; and system descriptions.

Two invited speakers described their perspectives on two areas outside the field that might become an important part of language generation research in the future. James

Pustejovsky (Brandeis University) presented his views on the richness of what can be encoded in what he calls a generative lexicon and the versatility that such a lexicon would provide. He revived the idea of an *enthymeme* as a way to control what is omitted from texts because of inferences implicit in the words that are used. Mari Ostendorf (Boston University) presented the current picture of the state of the art in speech generation (synthesis). She pointed out that research in prosody, for example, is badly needed to improve present-day speech synthesis, as is a richer model of the structure underlying the text. Currently, this structure is invariably determined by parsing an already finished written text, and as a result, most of the structure is missed because it cannot be recovered by state-of-the-art parsers. If, however, the text were produced by a generation system, the structure could easily be provided because generators need to plan their multisentence text.

Results and progress in language generation have always been difficult to measure largely because the customary testing paradigm used in text or speech understanding is impossible to apply. In the understanding (input) direction, there is total and uncontroversial agreement on what you start with—the characters of the input text or the recorded waveforms of the speech—and (in both cases) the output is specified as needed for some task or application. In contrast, in the generation (output) direction, although the nature of the final result is clear, the nature of the starting point of the process is itself one of the prime research areas. Differing assumptions about the source form lead to different choices of module and different distributions of the knowledge sources used. Because no knowledge base or expert system can provide enough of the information required to adequately specify the exact form of output text, all nontrivial generators have to include numerous default settings; deciding which aspects to control, which to set by default, and which simply to remove from the generator itself is another little understood facet of generation. Different generator designers make different choices, and the resulting systems are hard to compare.

For this reason, the possibilities for meaningful cross-system glass-box evaluation have always seemed remote. The only hope for success seems to lie in focusing on some particular methodology and function, such as incremental generation (talking while thinking) versus offline text production, dialogue versus monologue, and single language versus multilingual production of the same information, and on some particular task, such as the generation of tutorials, explanations, reports, and summaries.

In an effort to come to grips with how research on language generation might be evaluated, a panel was included in the program, with invited presentations by people who have had extensive experience in evaluating natural language systems. Madelyn Bates (BBN Inc.) and Stephanie Seneff (Massachusetts Institute of Technology) spoke on their experiences in organizing and defining the evaluation procedures for the Advanced Research Projects Agency's (ARPA) Air Traffic Information System (ATIS) speech-understanding endeavor and on the changes to the procedures that are being contemplated. Eduard Hovy (USC Information Sciences Institute) described the evaluations in the recent ARPA machine-translation effort. Although arguing strongly that evaluation was a necessary component of research, all three speakers were sensitive to the problem of the tail wagging the dog, where funding-sensitive, multisite competitive evaluations tend to unreasonably shape the character of the research.

In the discussion following the panel presentations, an unexpected consensus emerged. Competitive evaluation in generation was rejected as premature and likely to stifle research. At the same time, the need for evaluation as a means of measuring progress was universally accepted along with a collective need to work out the particulars of what any evaluation might concretely involve. To this end, there was common agreement that any individual generation project should define—in its own terms—what it would choose to measure from year to year so that quantitative factors can be incorporated into what to date has been an exclusively qualitative process. By sharing and comparing measurement definitions, the community could, over time, accumulate a palette of techniques that groups could adopt to the degree that they were appropriate to what they did. All agreed that a competitive evaluation was sensible only as an end-to-end black box in which competitors would take on virtually all the components of the task, permitting them to freely choose the source representation and distribution of processing in a way that fits their theoretical principles.

It was generally agreed that this workshop was one of the most enjoyable and productive workshops held recently in the generation community. Preparations for the next international generation workshop have already begun. The meeting will be held in the United Kingdom in 1996. For further details as they become available, contact either Donia Scott (Brighton University, e-mail: donia.scott@itri.bton.ac.uk) or David McDonald (Brandeis University, email: mcdonald@cs.brandeis.edu).

References

Dale, R.; Hovy, E.; Roesner, D.; and Stock, O., eds. (1992). *Aspects of Automated Natural Language Generation.* (Lecture Notes in Artificial Intelligence No. 587). Berlin: Springer.

Kempen, G., ed. 1987. *Natural Language Generation: New Results in Artificial Intelligence, Psychology, and Linguistics*. Dordrecht, The Netherlands: Martinus Nijhoff Publishers.

Paris, C. L.; Swartout, W. R.; and Mann, W. C., eds. 1991. *Natural Language Generation in Artificial Intelligence and Computational Linguistics.* Boston: Kluwer Academic.