Expressibility and the Problem of Efficient Text Planning

Marie Meteer
(Rensselaer Polytechnic Institute)

Reviewed by
Ruslan Mitkov
Bulgarian Academy of Sciences

In natural language generation, there is a gap between the stages of "what is to be said" and "when is it to be said" (text planning), and the stage of "how is it to be said" (text realization). This gap is due to the fact that the text-planning component usually chooses units and organizes them without taking into consideration what linguistic resources are available for expressing them. Expressibility is a requirement on the initial stages of NLG, the stages that take care of selecting and organizing of what is to be said. This principle has been clearly defined and paid special attention in Marie Meteer's work.

Meteer defines the problem of expressibility, and tackles it by means of *text structure representation* in an appropriate way. She looks at text structure from a linguistic perspective, demonstrating its adequacy for capturing the significant linguistic abstractions necessary for composing linguistic resources, and also at the representation of text structure in her Spokesman generation system, showing its importance for text planning. For the first time in generation, she presents a computationally motivated linguistic level of representation that is more abstract than surface structure and more concrete than conceptual structures, which have been taken a priori as sources for the generator.

Meteer's book *Expressibility and the Problem of Efficient Text Planning* contains eight chapters (188 pages), the first of which is introductory and presents the "generation gap," whereas the last is concluding, discussing the contributions of the research done by the author and envisaging future investigations. Chapter 2 is dedicated to terminology and issues in text generation, and Chapter 3 discusses motivations from language and applications focusing on the analysis of revisions. Chapter 4 is essential, and describes resources of the language and the text structure representation. Chapters 5, 6, and 7 are dedicated to practical issues and implementation of Meteer's approach as well as to alternative architectures.

According to Meteer, in order to exercise the full expressiveness of language, text planning needs to address the internal composition of clauses and not just their organization into larger structures. Clauses in actual texts reflect a combination of multiple atomic units. Systems that ignore this and begin with units that are inevitably realized have two major deficiencies:

1.   they are presuming that underlying programs have units of this size that

                    may be simply selected for inclusion in the message and then realized intact; and

2.   they are under-utilizing the power of natural language, which can use complex noun phrases, nominalizations, adverbial phrases, and other adjuncts to pack information from multiple units into one clause.

Meteer further claims that in order to plan complex utterances and ensure they are expressible in language, i.e., can be realized as a grammatical utterance, the text-planning process must know:

1.   what realizations are available to an element—that is, what resources are available,

2.   the constraints on the composition of the resources, and

3.   what has been committed to thus far in the utterance, which constrains the choice of resources.

Meteer's work provides the theoretical apparatus required to supply this information, centering on a new representation level, which she calls "text structure."

    It would probably have been a good idea if Meteer had considered also recent work on text revision outside the U.S.A. (e.g., Japan), which has not been cited in the references. The chapter "Alternative architectures" could also have been more complete if architectures intermediary with respect to the traditional two-component systems and interdependent architectures had also been mentioned.

    Meteer enriches the present research of text generation with valuable ideas. Her investigations on revision in text generation are a solid contribution to the improvement of the quality of the generated text. Meteer's work is an important milestone in generation research and it also provides an exciting ground for future research. I very much recommend reading this book, especially to people interested in natural language generation.

Ruslan Mitkov graduated from the Humboldt University in Berlin in 1979. He got his Ph.D. from the Technical University, Dresden, Germany. From 1991 to 1993 he worked as an associate professor at the Machine Translation Unit, Universiti Sains Malaysia, Penang, Malaysia. Currently Dr. Mitkov works at the Department of Mathematical Linguistics at the Institute of Mathematics, Sofia. Mitkov's address is: Institute of Mathematics, Acad. G. Bonchev str.bl.8, 1113 Sofia, Bulgaria; e-mail: ruslan@bgearn.bitnet.