## Incremental Conceptualization for Language Production

**Markus Guhe**
(University of Edinburgh)

Mahwah, NJ: Lawrence Erlbaum Associates (distributed by Psychology Press), 2007, xii+260 pp; hardbound, ISBN 978-0-8058-5624-8, $75.00

*Reviewed by*
*Paul Piwek*
*The Open University*

For the past ten years or more, most work in the field of Natural Language Generation (NLG) has shied away from considerations regarding the processes underlying human language production. Rather, the focus has been on systems that automatically produce language—usually text—from non-linguistic representations, with the main objective being generation of a text that faithfully captures the meaning of those non-linguistic representations (see, e.g., Reiter and Dale's 2000 textbook on NLG). There is, however, also a different take on NLG "as not just competent performance by a computer but the development of a computational theory of the human capacity for language and processes that engage it" (McDonald 1987, page 642). Guhe's research monograph, based on his 2003 Ph.D. thesis, is firmly situated in the latter tradition. One of his main goals is to work out a computational architecture for Levelt's (1989) psycholinguistically motivated model of language production. According to Levelt's model, speaking involves three main activities: conceptualizing (deciding what to say), formulating (deciding how to say it), and articulating (saying it). Guhe's book focuses on the mental activity of conceptualizing.

Conceptualizing is a recalcitrant object of study, partly because of the problem of the "initial spark"; the decision to say something appears to be the result of volitional conscious decisions, which largely elude scientific study. Guhe avoids this problem by investigating conceptualization in settings where the main intention is already fixed: a speaker witnesses several events unfold and is instructed to *describe what happens* (while it happens). The research challenge then is to figure out how "subintentions" for individual speech acts come about. The benefit of using an on-line generation setting is that it provides information on both what a speaker says at a given point in time and what is being reported, that is, the data that drive the speaker's utterances.

The book consists of the usual preface and introduction, followed by four parts (A, B, C, and Results), a list of the book's theses, and an appendix that includes, among other things, a glossary, bibliography, name index, and subject index. Part A of the book is titled "Conceptualization." It starts with an introduction to the field of language production, with particular reference to Levelt's (1989) model. The notion of conceptualization as a "quasi-module," partly using Fodor's (1983) criteria, is presented and four subtasks of conceptualization are discussed:

1. construction (mainly mapping what is perceived to concepts from long-term memory)
2. selection (of events that are to be verbalized)
3. linearization (ordering selected events appropriate to the goal of the discourse)

4. generation of preverbal messages (mapping the conceptual representations that have been handled so far to semantic content that can interface with the linguistic formulator)

This chapter also introduces referential nets, the formalism that is used to represent conceptual content.

Part B ("Incrementality") traces the roots of the notion of incrementality in computer science, and provides an extensive overview of various notions of incrementality. Guhe settles on a definition of incrementality whose crux is the piecemeal processing of information and production of output before all input has been seen. He distinguishes between incremental processes, algorithms, and models; roughly speaking, incremental models contain a strictly uni-directional cascade of incremental processes that recursively call incremental algorithms. For Guhe, an essential characteristic of incremental algorithms is that they use only a local context, as opposed to all available knowledge, for their computations. He also adopts the common distinction between working memory and long-term memory. The former mediates the flow of information between incremental processes. "Increments," the small pieces of information that incremental processes operate with, can be read from it and written to it. It contains "situation and discourse knowledge," whereas long-term memory stores static "encyclopedic knowledge." This "blueprint for incrementality" is accompanied by a useful discussion of various dimensions of incrementality, such as monotonicity, lookahead, feedback, and discreteness.

Part C focuses on INC, the incremental conceptualizer, which is an implemented "working model" of the blueprint for incrementality. INC is offered as a framework, that is, a model which has been fleshed out in detail in some respects and left underspecified in others. A central role is played by four parameters of INC which influence its behavior. For example, two of these concern the storage of event representations in a buffer in working memory which mediates the flow of information between incremental processes. One parameter, length of traverse buffer (LOTB), concerns the size of this buffer, whereas the other, latency (LT), determines for how long an element is kept in the buffer until it is picked up by preverbal message generation. Small values for LOTB in combination with a large value for LT can lead to the "forgetting" of information: If the buffer has filled up and new information is added, the first element on the buffer is discarded and never reaches preverbal message generation. The book presents some evidence that variation of the parameter settings can account for some of the variation found among human speakers. This part of the book concludes with a discussion of the output of INC for two domains and output of human speakers for the same domains. It concerns a visual scene, from a bird's eye perspective, of two moving planes on a runway, and the replay of the drawing of a simple line drawing consisting of eight lines that represents a crossing.

The "Results" summarizes the main contributions of the book, makes some comparisons with Levelt's (1989) model, and proposes a number of future extensions, such as the addition of Levelt's monitor. The monitor takes as input the output of the speech-comprehension system and uses this to influence the processing of the conceptualizer. Finally, there are a good number of suggestions for further ways to parameterize INC.

The book is a rich source of information on language production, both from a computational and a cognitive point of view. It includes a good introduction to conceptualizing, and provides an insightful discussion of many varieties of incrementality. INC is an excellent starting point for others interested in on-line data-driven generation to both build on and respond to. The breadth of the work means that one gets a truly

holistic view of the problem and is given a good impression of the many debates that cross the boundaries of different disciplines. In this respect, the book goes against a recent trend in computational linguistics to show less interest in other language-related research communities (see Reiter 2007).

Although the wide scope of this book is in many ways what makes it attractive, it also leads to some of its weaknesses. In particular, the way INC is presented in this broad context did not feel optimal to me. Although the proper description of INC is delayed until Part C, there are numerous forward references to INC in the preceding parts. The reader will find several instances where a certain aspect of conceptualization or incrementality is discussed with reference to INC, only to find out later that this particular feature "is not implemented yet (apart from a dummy function)." It would have been fairer to the reader to separate a clear description of the current state of INC from the wider discussion surrounding it. Another presentational issue concerns the tight integration of locality and incrementality in the book's definitions. In particular, the virtual identification of incremental *algorithms* with computation on a local context makes one question why the book speaks of incremental rather than local algorithms. A more substantive point relates to Part C on INC. This part includes the description of two simulations that were run with INC. Somewhat frustratingly, both descriptions are incomplete. For instance, whereas for the first simulation the appendix contains the texts produced by human participants for the same task, there is no systematic analysis of the structural (dis)similarities between the output of INC and that of the human speakers. For the second simulation, there are some analyses of the similarities between the structure of INC's and the human speakers' output, but no transcripts of complete human outputs are provided. In both cases, there is also no detail about how long-term memory, referred to as the concept storage (CS), was populated for the relevant domains, even though the CS must have had a significant influence on the output that INC produces.

This book will be useful to research students and researchers in natural language generation who are interested in the study of generation systems as a computational model of human language production. Part B of the book, on incrementality, might also prove useful to those approaching NLG as an engineering problem. The main reason to consult this book is that it brings together in a single place information on conceptualization, incrementality, and various debates in philosophy, cognitive science, and computer science affecting these topics. INC, the incremental conceptualizer which is described in part C of the book, presents an ambitious attempt to implement a computational model of incremental conceptualization. The verdict on its adequacy is still out, given the limited empirical evaluation to which it has been subjected thus far.

Online generation, a central theme of this book, was adopted in 2007 at the International Conference on Intelligent Virtual Agents as a task (automated real-time reporting on simulated horse races) in the GALA competition for Embodied Lifelike Agents.[1] Work on embodiment and conceptualization, new insights into societal grounding of conceptual representations (e.g., DeVault, Oved, and Stone 2006), empirical and computational studies on generation (both incremental and non-incremental, from numerical data; e.g., van Deemter 2006), and recent experimental techniques for studying language production (see Roelofs 2004 for an overview) give a sense that this book could be part of an exciting revival of cognitively motivated NLG.

1 See `http://hmi.ewi.utwente.nl/gala/`.

## References

van Deemter, Kees. 2006. Generating referring expressions that involve gradable properties. *Computational Linguistics*, 32(2):195–222.

DeVault, David, Iris Oved, and Matthew Stone. 2006. Societal grounding is essential for meaningful language use. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06)*, Boston, MA, pages 747–754.

Fodor, Jerry A. 1983. *The modularity of mind*. MIT Press, Cambridge, MA.

Levelt, Willem J. M. 1989. *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA.

McDonald, David. 1987. Natural language generation. In Stuart C. Shapiro, editor, *Encyclopedia of Artificial Intelligence, Volume 1*, John Wiley & Sons, New York, pages 642–654.

Reiter, Ehud. 2007. The shrinking horizons of computational linguistics. *Computational Linguistics*, 33(2):283–287.

Reiter, Ehud and Robert Dale. 2000. *Building Natural Language Generation Systems*. Cambridge University Press, Cambridge, UK.

Roelofs, Ardi. 2004. The seduced speaker: Modeling of cognitive control. In Anja Belz, Roger Evans, and Paul Piwek, editors, *Natural Language Generation, Third International Conference*, LNCS 3123, Springer, Berlin, pages 1–10.

*Paul Piwek* is lecturer in computing at the Open University. His research interests are (multimodal) natural language generation and dialogue modeling. Piwek's address is: Centre for Research in Computing, The Open University, Walton Hall, Milton Keynes, UK; e-mail: p.piwek@open.ac.uk.