

Introduction to the Special Issue on Natural Language Generation

Robert Dale*
Macquarie University

Barbara Di Eugenio†
University of Pittsburgh

Donia Scott‡
University of Brighton

1. Introduction

There are two sides to natural language processing. On the one hand, work in natural language *understanding* is concerned with the mapping from some surface representation of linguistic material—expressed as speech or text—to an underlying representation of the meaning carried by that surface representation. But there is also the question of how one maps from some underlying representation of meaning into text or speech: this is the domain of natural language *generation*.

Whether our end-goal is the construction of artifacts that use natural languages intelligently, the formal characterization of phenomena in human languages, or the computational modeling of the human language processing mechanism, we cannot ignore the fact that language is both spoken (or written) and heard (or read). Both are equally large and important problems, but the literature contains much less work on natural language generation (NLG) than it does on natural language understanding (NLU). There are many reasons why this might be so, although clearly an important one is that researchers in natural language understanding in some sense start out with a more well-defined task: the input is known, and there is a lot of it around. This is not the case in natural language generation: there, it is the desired output that is known, but the input is an unknown; and while the world is awash with text waiting to be processed, there are fewer instances of what we might consider appropriate inputs for the process of natural language generation. For researchers in the field, this highlights the fundamental question that always has to be asked: What do we generate *from*?

Despite this problem, the natural language generation community is a thriving one, with a research base that has been developing steadily—although perhaps at a slower pace because of the smaller size of the community—for just as long as work in natural language understanding. It should not be forgotten that much of NLP has its origins in the early work on machine translation in the 1950s; and that to carry out machine translation, one has to not only analyze existing texts but also to generate new ones. The early machine translation experiments, however, did not recognize the problems that give modern work in NLG its particular character. The first significant pieces of work in the field appeared during the 1970s; in particular, Goldman's work on the problem of lexicalizing underlying conceptual material (Goldman 1974) and

* School of Mathematics, Physics, Computing and Electronics, Sydney NSW 2109, Australia

† Learning Research and Development Center, 3939 O'Hara Street, Pittsburgh, PA 15260, U.S.A.

‡ Information Technology Research Institute, Lewes Road, Brighton BN2 4GJ, UK

Davey's work on the generation of paragraph-long descriptions of tic-tac-toe games (Davey 1979) were among the first to focus on issues unique to NLG. The field really took off, however, in the 1980s; for those working in NLG, the decade began with a bang, and the Ph.D. theses of McDonald (1980), Appelt (1981), and McKeown (1982) have had a lasting impact on the shape of the field.¹

But what has happened in the last fifteen years since those major pieces of work first appeared? Although one does find articles on NLG in the pages of *Computational Linguistics* and other journals in the field, and papers on generation do appear at the major NLP conferences, the quantity and range of work being carried out in NLG tends to be underrepresented in these forums. Instead, the community has tended to present its results at the two biennial series of workshops—one European and one international—that have sprung up in the last ten years. Many of these workshops have led to books: see Kempen (1987); McDonald and Bolc (1988); Zock and Sabah (1988); Dale, Mellish, and Zock (1990); Paris, Swartout, and Mann (1991); Dale et al. (1992); Horacek and Zock (1993); and Adorni and Zock (1996). This special issue of *Computational Linguistics* was inspired by discussions at the International Workshop on Natural Language Generation held in Herstmonceux in 1996; the aim of the volume you are reading is to show the wider computational linguistics community something of the range of activities in NLG.

2. Some Perspectives on Natural Language Generation

What is natural language generation about? A definition offered by McDonald (1987, 983) over ten years ago has stood the test of time:

Natural language generation is the process of deliberately constructing a natural language text in order to meet specified communicative goals.

A more recent definition with a slightly different emphasis can be found in Reiter and Dale (1997, 57):

Natural language generation is the subfield of artificial intelligence and computational linguistics that is concerned with the construction of computer systems that can produce understandable texts in . . . human languages from some underlying non-linguistic representation of information.

Both definitions pick out some of the foci of interest that give work in NLG its distinctive flavor. From the first we note the emphasis on *deliberate choice* as the fundamental operation that underlies much work in the area, and on the generation of *texts* as opposed to single sentences; from the second, we note the emphasis on underlying representations of information that may be *nonlinguistic* in nature. Each of these points bears some elaboration:

- In work in NLG, a major concern is that of choosing between different ways of doing things, as the same content can often be expressed in many different ways. Although some of these choices may indeed be arbitrary, there is a view in NLG that a great many are not, and the choices between different ways to say things—different ways to structure a text, different ways to refer to objects, the use of different syntactic

¹ The latter two works are more widely available in revised form as Appelt (1985) and McKeown (1985).

constructions, and of different words to realize underlying concepts—need to be motivated in some way. Much research in NLG is oriented towards uncovering those motivations.

- There is a sense in which work in NLU tends to start with the sentence as the principal focus of inquiry. However, for much work in NLG, the primary focus is the text or discourse: although there are many important issues involved in the generation of sentential forms, those working in NLG research have long accepted that discourse-level issues are just as important, and probably more so. This relates to the previous point: it is often only by considering the context within which a sentence is being generated that the appropriate choice of surface form can be made.
- The input representation provided to an NLG system may be symbolic (for example, an expert system knowledge base) or numeric (for example, a database containing stock market prices) but it is generally nonlinguistic in nature. Early work in the field relied on the use of hand-crafted knowledge sources, which sometimes meant that the representations used embodied unspoken assumptions. More recent work has been able to take advantage of representations created for other purposes; using these as the input to the generation process reinforces the realization that the elements of the underlying representation may not correspond in a straightforward way to words and sentences.

Much work in NLG thus concerns itself with pragmatics and discourse-level considerations. Interestingly, these too have been somewhat underrepresented in the standard computational linguistics forums, where the bulk of the work carried out is often in the area of well-specified and rigorous formal treatments of sentential phenomena.

There is an important point here that bears emphasizing: natural language generation is *not* the inverse of the process of parsing.² Those working in NLG generally break down the process of generating a text into a number of stages, and it is only the last of these—generally referred to as surface realization—that corresponds to anything like the inverse of parsing. If we want to seek the mirror-image of work in NLG within research in natural language understanding, we have to consider the entire analysis process, all the way through to plan recognition in multisentential discourses or dialogues.

This is perhaps an appropriate place to review what the task of NLG is now commonly seen to involve:

- First, there is the question of **content determination**: deciding what to say. This impacts at both macro and micro levels. At the macro level, researchers in NLG are concerned with how the content of a multisentential text, or of a turn in a dialogue, can be determined. At a micro level, researchers are concerned with how the content of appropriate referring expressions can be worked out. In each case the problem is how to select the right information from that which is available; it is rarely appropriate to say everything we could say.

² It should be noted, however, that there is a body of work that looks at the use of bidirectional grammars, where a common declarative representation of grammatical knowledge is used both for parsing and for realization; see, for example, Shieber et al. (1990).

- There is also the question of **text structure**: texts are not just random collections of sentences; they exhibit a structure that plays a key role in conveying their meaning. Researchers in NLG are concerned with elucidating mechanisms for determining the most appropriate structures to use in particular circumstances, and with working out how the information to be conveyed can best be packaged into paragraph- and sentence-sized chunks.
- Closer to the kinds of issues that concern those working in parsing, there are the problems of **surface realization** and **lexicalization**: once the content of individual sentences has been determined, this still has to be mapped into morphologically and grammatically well-formed words and sentences. Where the underlying representation expresses informational elements at a granularity that does not map easily into words, decisions about how to lexicalize the conceptual material have to be taken.

These are the kinds of issues that have driven much research in NLG over the last 15 years. Our understanding of the issues has come a long way in that time. This issue of *Computational Linguistics* contains what is no more than a snapshot of work in the field at the current time; it should be read against the background of the broader picture we have attempted to sketch here, albeit briefly. In the next section, we provide short summaries of the papers collected together in this special issue.

3. An Overview of the Issue

From the 25 papers originally submitted to the special issue, our reviewers helped us eventually select five. There were many more papers that, given space, we would have included; we hope that some of these will appear in subsequent regular issues of *Computational Linguistics*.

3.1 Chu-Carroll and Carberry

As we mentioned earlier, “deciding what to say” is a key issue in NLG. Chu-Carroll and Carberry’s paper focuses on strategies for selecting the content of responses in collaborative planning dialogues. The authors concentrate on situations in which the system and the user have different beliefs that they attempt to reconcile, namely: cases when the system needs to gather further information in order to decide whether to accept a proposal from the user, and cases when the system must negotiate with the user to resolve a detected conflict in beliefs. In both cases, the implemented algorithms identify the subset of beliefs that the system believes will most effectively help solve either its uncertainty or the conflict in beliefs; further, the system chooses an appropriate strategy and produces a response that initiates a subdialogue addressing the impasse in conversation.

Two other points deserve special note. First, the computational model is based on a small but convincing corpus study. Second, the authors conducted a formal, even if limited, evaluation of their prototype implementation. The evaluation consists of human raters grading the system’s actual response and some distractors, obtained by selectively altering the system’s response generation strategies. The evaluation is suggestive that the proposed strategies and their implementation are effective.

3.2 Stede

Addressing the issue of “deciding how to say it,” Stede focuses on the role of the lexicon and of lexical choice within an NLG system. More specifically, Stede describes how his approach can generate verbal alternations that change the aspectual category of the verb. One such alternation is the causative, as in *The mechanic drained the oil from the engine*, the causative form of *The oil drained from the engine*. Stede takes the lexicon to be the central device for mapping between domain representations and intermediate semantic representations. Alternations are generated by applying one or more rules in a predetermined order to a basic lexical form; the choice of a specific alternation is determined by parameters such as salience. The intermediate semantic representation of a sentence is the input for the surface generator—in this case, Penman (Penman group 1989).

Interesting aspects of Stede’s approach are the capacity of his system to generate fine-grained distinctions of meaning, and the attention paid to both linguistic constraints and computational concerns. Although only English is discussed in this paper, Stede’s system uses the same mechanism to generate alternations in German as well.

3.3 Mittal, Moore, Carenini, and Roth

Generation technology is now increasingly finding a place in applied systems. One such application is described by Mittal, Moore, Carenini, and Roth, who have developed a system to generate captions to accompany the graphical presentations produced by SAGE (Roth et al. 1994). It is well known that the interpretation of even simple, conventional, graphics can be difficult without accompanying textual pointers (e.g., keys, labels of axes, and the like). SAGE is innovative in its ability to produce novel graphics for highly abstract and complex data. The comprehension of these presentations is often heavily reliant on captions: extended textual descriptions of the relation of the presentation to the data it depicts.

Mittal et al. show how a SAGE graphic, together with information about the perceptual complexity of its elements and the structure of its underlying data, can be used to generate an effective multisentential caption. This is demonstrated through examples in the domain of housing sales; however, with the exception of the lexicon, the caption generator is fully domain independent. Although the system has not yet been formally evaluated, we are told that users of SAGE report that the generated captions contribute positively to their understanding of complex graphical presentations.

3.4 Radev and McKeown

Automated text summarization is a practical problem of increasing interest, especially with the ever-widening dissemination of the World Wide Web; this is an obvious area where NLG can contribute. Radev and McKeown describe an application of NLG techniques towards the end of producing summaries of a kind that are, as the authors argue, beyond the scope of current statistical summarizers. There are two main elements to their approach. The first is the use of “summarization operators” that compare data structures containing information derived from different sources and thus allow the system to produce summaries of several input messages; the second is the use of a technique for identifying proper names and related descriptions from on-line text so that these can be used to extend the descriptions provided in summaries.

The data structures used as input to generate the summary texts are filled MUC-style templates; the task of identifying key information in source texts and extracting it has already been carried out. The paper thus provides an excellent application of established technologies, with new mechanisms being developed to complete the pic-

ture; the work shows well how NLG techniques can make a real difference to the important task area of summarization.

3.5 Oberlander

Texts are generated to be read, and while generators can provide a range of texts for a given context, the question of what expression is most appropriate remains nontrivial. Typically nowadays, the designers of NLG systems tune their generators to produce expressions compatible with those found in a corpus of “good” exemplars from the domain in question (see, e.g., Scott and Power [1994] and Paris et al. [1995]). But this approach is not always possible (e.g., for novel domains), and even in cases where there is an available corpus, judgments of quality can often only be made by appealing to convention. Psycholinguistic studies suggest themselves as a useful source of guidance but this too can be problematic: what speakers or writers typically produce often conflicts with their preferences as perceivers, as shown, for example, with regard to referring expressions by Stevenson and her colleagues (Stevenson, Crawley, and Kleinman 1994; Stevenson and Urbanowicz 1995).

Oberlander discusses this apparent paradox with reference to the generation of referring expressions—in particular, the suggestion by Dale and Reiter (1995) that generation algorithms for definite noun phrases should be based on observations about human language production rather than on a strict observation of the Gricean maxims (Grice 1975). Oberlander calls this the Spike Lee maxim: *Do the right thing*—where “right” is that which is human and simple. He shows that, when generating referring expressions, we can’t always tell whether the right thing is to mimic the preferences of language producers or language perceivers, since these preferences often conflict. He argues that until we develop a more sophisticated view of the expectations of speakers and hearers, developers of NLG systems should probably stick to the Spike Lee maxim: even with its known limitations it produces more natural results than are achieved by following a strict interpretation of the Gricean maxims—and we would all agree that even that is better than the Cole Porter maxim.³

4. Future Directions for Research in Natural Language Generation

We said earlier that NLG research has come a long way since its beginnings, but there is still a long way to go. What does the future hold? Crystal-ball gazing is always a risky business, but on the basis of our experience and some of the issues that arise both in the work presented here and in other submissions to the special issue, we would suggest the following aspects of NLG will be seen as important areas in the next five years.

Microplanning. Ever since Thompson (1977), there has been a tendency to see NLG as involving two problems, which Thompson characterized as being concerned with decisions of strategy and tactics: in short, questions about what to say and questions about how to say it. In the field, this translated into work in the two areas of text planning and linguistic realization, with researchers often declaring themselves as working on one or the other. In more recent years, there has been the realization that something is required in the middle; this was most notably expressed in Meteer’s work on what she called “the generation gap” (Meteer 1990). This has given rise to a body of work that explores questions of what is often referred to as microplanning: once a

³ This being: *Anything goes.*

text planning process has worked out the overall structure of a text and the content to be conveyed, how is this information packaged into sentences? Serious work here has only just begun: there are a great many unresolved issues, and in many cases the questions themselves are unclear.

Multimodal generation. Real text is not disembodied. It always appears in context, and in particular within some medium—for example, on a page, on a screen, or in a speech stream. As soon as we begin to consider the generation of text in context, we immediately have to countenance issues of typography and orthography (for the written form) and prosody (for the spoken form). These questions can rarely be dealt with as afterthoughts. This is perhaps most obvious in the case of systems that generate both text and graphics and attempt to combine these in sensible ways. We predict that the World Wide Web will be a major factor in forcing some of the issues here: if systems are to automatically generate the text on Web pages (see, for example, Milosavljevic and Dale [1996]), then they also need to consider other elements of that container.

Reusable resources. It may be an indication of a maturing of some subareas of NLG research that we are now in a position where there are reusable components for particular tasks. Specifically, three linguistic realization packages, FUF/SURGE (Elhadad 1993a, 1993b; Elhadad and Robin 1996), PENMAN/NIGEL (Penman group 1989), and its descendant KPML/NIGEL (Bateman 1997), are widely used in the field. For anything other than simple applications, it is now questionable whether it makes sense to build a linguistic realization component from scratch. We may expect other kinds of reusable components to be developed within the research community within the next 5–10 years; it is developments of this kind that signal significant progress, since being able to reuse the work of others obviously has the potential to increase research productivity. In related developments, there is a growing interest within the community in defining a reference architecture for NLG; if successful, this is likely to stimulate further research and development in NLG through the provision of a modular baseline for development, comparison, and evaluation.

Evaluation. Although there have been attempts at the evaluation of NLG techniques and systems in the past, formal evaluation has only recently come to the fore. For example, systems have been evaluated by using human judges to assess the quality of the texts produced (Lester and Porter 1997; Chu-Carroll and Carberry, this issue); by comparing the system's performance to that of humans (Yeh and Mellish 1997); by corpus-based evaluation (Robin and McKeown 1996); and indirectly through "task efficacy" measures (Young 1997). The major stumbling block for progress is determining what metrics and methods should be used: for example, how can the *quality* of an output text be measured? Because of the different nature of the task, it is unlikely that methods that have been used in NLU, such as the evaluation process adopted in the Message Understanding Conferences, can be carried over to generation. Dale and Mellish (1998) suggest that the NLG community could make progress by devising specific evaluation methods for NLG subtasks such as content determination, text structuring, and realization; this "glass box" approach is likely to result in a clearer understanding of how to evaluate NLG systems as a whole.

The particular foci we have just outlined are specific to work in NLG. However, just as corpus-based methods have become very important in NLU research, we may expect this to happen increasingly in work on NLG too. Raw or coded text has been used by researchers to investigate strategies in a number of different areas of NLG, as demonstrated in the papers by Radev and McKeown and by Chu-Carroll and Carberry

in this issue. Given the emphasis within NLG research on text-level issues, a major bottleneck for work here is the encoding of corpora with semantic and discourse structural features; see Di Eugenio, Moore, and Paolucci (1997). These are needed to uncover plausible text-structuring and microplanning strategies, but annotating corpora for such features will remain a laborious manual task at least for the foreseeable future. This effort may be alleviated if sharable corpora become available through the Discourse Resource Initiative (<http://www.georgetown.edu/luperfoy/Discourse-Treebank/dri-home.html>).

With so many rich seams to mine, natural language generation has a promising future. We mentioned at the outset that researchers in NLG face the unique problem of deciding what to generate *from*: Yorick Wilks is credited with pointing out that, while the problem of natural language understanding is somewhat like counting from one to infinity, researchers in natural language generation face the problem of counting from infinity to one. In order to make progress, researchers in NLG pick a reasonably high number and get to work; as researchers in NLU climb the numerical ladder from the other end, we can expect that some of the big numbers discovered in NLG will prove to be of use in NLU too.

Acknowledgments

We offer our grateful thanks to the body of reviewers who did so much work in helping us put this issue together. We also acknowledge the many fruitful discussions we have had with our colleagues, including especially Giuseppe Carenini for sharing his notes on evaluation in NLG and Ehud Reiter for his observations on the state of the field.

References

- Adorni, G. and M. Zock, editors. 1996. *Trends in Natural Language Generation*. Lecture Notes in Artificial Intelligence. Springer-Verlag, Berlin.
- Appelt, Douglas E. 1981. *Planning Natural Language Utterances to Satisfy Multiple Goals*. Ph.D. thesis, Stanford University, Stanford, CA. Available as SRI Technical Note 259.
- Appelt, Douglas E. 1985. *Planning English Sentences*. Cambridge University Press, Cambridge.
- Bateman, John A. 1997. Enabling technology for multilingual natural language generation: The KPML development environment. *Natural Language Engineering*, 3:15–55.
- Dale, Robert, Eduard H. Hovy, Dietmar Rösner, and Oliviero Stock, editors. 1992. *Aspects of Automated Natural Language Generation*. Lecture Notes in Artificial Intelligence. Springer-Verlag, Berlin.
- Dale, Robert and Christopher S. Mellish. 1998. Towards evaluation in natural language generation. In *Proceedings of the First International Conference on Language Resources and Evaluation*, Granada, Spain, May 28–30.
- Dale, Robert, Chris Mellish, and Michael Zock, editors. 1990. *Current Research in Natural Language Generation*. Academic Press, New York.
- Dale, Robert and Ehud Reiter. 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 18:233–263.
- Davey, Anthony C. 1979. *Discourse Production*. Edinburgh University Press, Edinburgh.
- Di Eugenio, Barbara, Johanna D. Moore, and Massimo Paolucci. 1997. Learning features that predict cue usage. In *Proceedings of the 35th Annual Meeting*, pages 80–87, Madrid, Spain. Association for Computational Linguistics.
- Elhadad, Michael. 1993a. FUF: The universal unifier—user manual version 5.2. Technical Report CUCS-038-91, Columbia University.
- Elhadad, Michael. 1993b. *Using Argumentation to Control Lexical Choice: A Unification-based Implementation*. Ph.D. thesis, Computer Science Department, Columbia University.
- Elhadad, Michael and Jacques Robin. 1996. An overview of SURGE: A reusable comprehensive syntactic realisation component. In *Proceedings of the 8th International Workshop on Natural Language Generation (Demos and Posters)*, Herstmonceux, Sussex, UK, June.
- Goldman, Neil M. 1974. *Computer Generation of Natural Language from a Deep Conceptual Base*. Ph.D. thesis, Stanford University.

- Available as Stanford AI Laboratory Memo AIM-247 or CS Technical Report CS-74-461.
- Grice, H. P. 1975. *Logic and Conversation*. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics 3: Speech Acts*. Academic Press.
- Horacek, H. and M. Zock, editors. 1993. *New Concepts in Natural Language Generation*. Pinter, London.
- Kempen, Gerard, editor. 1987. *Natural Language Generation: New Results in Artificial Intelligence, Psychology and Linguistics*. NATO ASI Series No. 135. Martinus Nijhoff Publishers, Boston, Dordrecht.
- Lester, James C. and Bruce W. Porter. 1997. Developing and empirically evaluating robust explanation generators: The KNIGHT experiments. *Computational Linguistics*, 23(1):65–102.
- McDonald, David D. 1980. *Natural Language Production as a Process of Decision Making under Constraint*. Ph.D. thesis, MIT, Cambridge, MA.
- McDonald, David D. 1987. Natural language generation. In Stuart C. Shapiro, editor, *Encyclopedia of Artificial Intelligence*. John Wiley and Sons, pages 642–655.
- McDonald, David D. and Leonard Bolc. 1988. *Natural Language Generation Systems*. Springer-Verlag, New York, NY.
- McKeown, Kathleen R. 1982. *Generating Natural Language Text in Response to Questions About Database Structure*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA, May. Available as Technical Report MS-CIS-82-05.
- McKeown, Kathleen R. 1985. *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Cambridge University Press, Cambridge.
- Meteer, Marie. 1990. *The Generation Gap: The Problem of Expressibility in Text Planning*. Ph.D. thesis, University of Massachusetts.
- Milosavljevic, Maria and Robert Dale. 1996. Strategies for comparison in encyclopedia descriptions. In *Proceedings of the Eighth International Natural Language Generation Workshop*, pages 161–170, Herstmonceux, Sussex, UK, June.
- Paris, Cécile, Keith Vander Linden, Markus Fischer, Anthony Hartley, Lyn Pemberton, Richard Power, and Donia Scott. 1995. A support tool for writing multilingual instructions. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1398–1404.
- Paris, Cécile L., William R. Swartout, and William C. Mann, editors. 1991. *Natural Language Generation in Artificial Intelligence and Computational Linguistics*. Kluwer Academic Publishers, Boston.
- Penman group. 1989. Documentation of the Penman Sentence Generation System. USC Information Sciences Institute, Marina del Rey, CA.
- Reiter, Ehud and Robert Dale. 1997. Building applied natural language generation systems. *Natural Language Engineering*, 3:57–87.
- Robin, Jacques and Kathleen McKeown. 1996. Empirically designing and evaluating a new revision-based model for summary generation. *Artificial Intelligence*, 85:135–179.
- Roth, Steven F., John Kolojechick, Joe Mattis, and Jade Goldstein. 1994. Interactive graphic design using automatic presentation knowledge. In *Proceedings of CHI'94: Human Factors in Computing Systems*, pages 193–200.
- Scott, Donia and Richard Power, editors. 1994. Characteristics of administrative forms in English, German and Italian. Deliverable EV-1, GIST project LRE 062-09. <http://ecate.itc.it:1024/projects/gist/gist-bibliography.html>.
- Shieber, Stuart, Gertjan van Noord, Fernando Pereira, and Robert Moore. 1990. Semantic head-driven generation. *Computational Linguistics*, 16(1):30–42.
- Stevenson, Rosemary, Rosalind Crawley, and David Kleinman. 1994. Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, 9:519–548.
- Stevenson, Rosemary and Agnieszka Urbanowicz. 1995. Structural focusing, thematic role focusing and the comprehension of pronouns. In *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, pages 328–332.
- Thompson, Henry S. 1977. Strategy and tactics: A model for language production. In W. A. Beach, S. E. Fox, and S. Philosoph, editors, *Papers from the 13th Regional Meeting of the Chicago Linguistics Society*, pages 651–668, Chicago, IL.
- Yeh, Ching-Long and Chris Mellish. 1997. An empirical study of the generation of anaphora in Chinese. *Computational Linguistics*, 23(1):169–190.
- Young, R. Michael. 1997. *Generating Descriptions of Complex Activities*. Ph.D. thesis, Intelligent Systems Program, University of Pittsburgh.
- Zock, Michael and Gérard Sabah, editors. 1988. *Advances in Natural Language Generation: An Interdisciplinary Perspective*. Ablex Publishing Corp., Norwood, NJ.

