Computational phonology: A constraint-based approach

Steven Bird
(University of Edinburgh)

Cambridge University Press (Studies in natural language processing, edited by Branimir K. Boguraev), 1995, xv + 203 pp.
Hardbound, ISBN 0-521-47496-5, $44.95

Reviewed by
Deirdre Wheeler and Bob Carpenter
Carnegie Mellon University

1. Introduction

This book is a revised and expanded version of the author's Ph.D. thesis (Edinburgh University, 1990), entitled *Constraint-based phonology*. The field of computational phonology is rather immature as a whole and there have been only a handful of publications on constraint-based theories of phonology. Bird's thesis, and its revision, stand as conceptual and computational milestones; we believe this work will act as a focal point for further investigation into constraint-based phonology.

The intended audience for this book includes computational linguists, phonologists and computer scientists. We briefly describe the contents of the book and then evaluate the strengths and weaknesses we see from the perspective of theoretical phonologists and computational linguists. Overall, our evaluation is that there is certainly something of interest for members of each of these audiences, but that none will be left completely satisfied. The theory of constraint-based phonology as presented in this book is suggestive, but its ultimate success or failure will depend on thorough phonological analyses of particular languages, broader coverage of typologically diverse languages, and a better understanding of the computational properties of constraint-based systems.

2. Summary of the book

The stated goal of the book is "to clarify the role that constraints play in phonology and then—drawing on insights from constraint-based grammar and constraint-programming—to formalize and implement a *constraint-based phonology*" (p. xiii). The first chapter consists of an introduction to phonology, a discussion of the formal adequacy of autosegmental notation, a brief overview of alternative approaches to computational phonology, a discussion of the methodology underlying constraint-based phonologies, and a historical survey of constraints in phonology. The chapter serves to highlight previous work as it relates to the defining properties of constraint-based phonology, which Bird takes to be intensionality, compositionality, monostratality, and lexicalism.

The second chapter lays out the logical foundation that Bird adopts in the rest of the monograph. A grammar is taken to be a set of constraints on well-formed representations. Bird adopts feature-structure representations of prosodic and segmental structure, thus allowing an easy interface with feature-based syntactic theories such

as HPSG (Pollard and Sag 1994). Constraints are expressed in a fragment of first-order logic. Issues touched on include dominance, re-entrancy, and the licensing of structure. More attention is paid to the temporal organization of autosegments, for which Bird adopts a standard logic based on the relations of overlap and precedence. The relations between hierarchical and temporal structure in autosegmental relationships can be constrained by precedence and overlap to produce a so-called "gestural score," which indicates the temporal restrictions on articulatory realizations of a phonological form. In the last section, the way in which transformational "rules" can be realized as constraints on a monostratal representation is discussed. The chapter ends with a distressingly brief notational introduction of Reiter's default logic, with no semantics to back it up. The computational power of the default system stands in stark contrast to the limited fragment of first-order logic employed elsewhere; this is reinforced by the omission of defaults in both the implementation discussed in Chapter 5 and the alternative logic discussed in Appendix A.

In Chapter 3, a more detailed study is undertaken of the consequences of assuming a monostratal view of phonology and the most serious potential criticism: that it is not powerful enough to express phonological generalizations. It begins with a discussion of allophonic variation, and then considers various deletion phenomena, which Bird claims can either be described as "alternations with zero" or explained away as phonetic processes. Restructuring processes, such as resyllabification, are then discussed, followed by feature-changing harmony rules. In each case, the challenge is to offer an analysis in a constraint-based approach.

In Chapter 4, Bird focuses on developing a theory of segmental structure. He begins with a survey of the literature on feature geometry, reviewing phonological evidence for the hierarchical organization of features, including the laryngeal, supralaryngeal, manner, and place nodes (see Clements 1985 and Sagey 1986, for example). Bird tends to favor Sagey's theory because of its phonetic basis, and he goes on to try to tie it together with the articulatory model of Browman and Goldstein (1989). The final section of the chapter presents a formalization of this theory of segmental structure.

Chapter 5 focuses on further formalization and implementation of the constraint-based theory. The motivation for an implementation appears to be that of testing linguistic theories rather than on applications such as speech recognition. C is used for the construction of models to satisfy the constraint system and Prolog is used to drive the search for solutions; this strategy is becoming more popular with the need for coroutining constraints, and will be further facilitated by SICStus Prolog 3.0, which builds in the architecture described by Bird. While there is a limited discussion of the complexity of model-building in C (non-polynomial), there is no evaluation of the behavior of the entire system. Purely constraint-based implementations of systems such as HPSG have been so fraught with nondeterminism that standard parsing mechanisms have been overlayed in almost every system currently being developed.

The conclusion, Chapter 6, provides a modest summary of the book. Bird claims that theoretical phonologists will benefit from the succinctness of the representations, the monostratal nature of the theory, and its formal precision. We believe that it is the monostratal, constraint-based nature of the theory that sets it apart from other theories in principle; the representations are similar to those in other theories, at least on the surface, and other theories could, in principle, be made more precise. Bird also claims that descriptive phonologists should benefit from the fact that systems can be automated to check their analyses. He further highlights the straightforward interactions with phonetics in his system, although this relationship is not precisely specified in the book. The conclusions for computational linguists are less clear, although presum-

ably the mere fact that the system can be implemented and integrated with popular syntactic and semantic theories is supposed to be appealing enough.

Appendix A contains an alternative modal-logic formalization of the theory, which Bird used in his thesis. Bird provides no motivation for the alternative formalization, and most readers will be more comfortable with the first-order presentation in the text.

3. Audiences

This section is organized according to the potential audiences for whom the book was written; we have taken the liberty of combining the perspectives of computer scientists and computational linguists.

3.1 For the theoretical phonologist

For the theoretical phonologist, there are several aspects of the book that should prove interesting. While some of the representations may at first seem unfamiliar and overly complicated (contrary to Bird's claims), it is significant to recognize that by formalizing existing graphical notations used in phonology, we come to a better understanding of the nature of the objects and relations encoded in these diagrams (*cf* Chapter 2). The need for such clarification in theoretical phonology is made clear by Bird and Ladd's (1991) review of Goldsmith's (1990) textbook; it is unfortunate that more of this kind of argumentation did not make it into the book. Bird's approach is much more rigorous than the standard in theoretical linguistics, and we believe that the field as a whole would benefit by greater formalization of the assumptions behind, and mechanisms involved in, a particular analysis.

While numerous phonological "processes" are discussed throughout the book, there is never an extended discussion of the phonology of a single language. In the end, this has several consequences. First and foremost, it leaves open the question of whether or not constraint-based phonology, as Bird envisions it, offers a viable phonological theory in the sense of allowing a phonologist to capture all significant generalizations in a language. Given the monostratal nature of his theory, all generalizations that are expressed through constraints on representations must hold of a single representation, the surface representation. Can phonologists be content with a theory which allows for no derivation? Chapter 3 is intended to address this issue both by criticizing destructive, transformational accounts and by providing alternatives. For instance, consider a deletion rule that operates on one representation with a segment in a specified context, returning another representation with that segment deleted. Bird replaces this notion by an "alternation with zero": the context of the deletion disallows the appearance of a phoneme that is mandatory in other contexts. Bird also addresses rule ordering, citing data from some dialects of English involving vowel lengthening and flapping. In this case, he explains away the problem posed for a monostratal theory by simply saying: "both rules are descriptively inadequate" (p. 99). Unfortunately, this does not address the more fundamental question: can all such cases be explained away or otherwise accounted for?

For this and other cases apparently requiring rule ordering, Bird could have presented an alternative account in which there was no ordering of processes by adding a rule (formulated as a constraint) to the grammar in order to compose the effects of the interacting rules and constraints. This general strategy has been highly developed in the well-known case of two-level phonology by Kaplan and Kay (1994). The move to compile out ordering into several rules that are unordered, although perhaps more

complex in their conditioning, is similar to the compilation of a logical theory with defaults into a nondefault theory.

Bird has answered the first important question, showing how destructive processes can be modeled in a monostratal theory; in some sense, this is extending the tradition begun by Gazdar (1981), who showed that certain syntactic processes could be modeled without transformations. But the issue of whether the result is a linguistically satisfying theory remains: can the relevant generalizations be captured? The lengthening rule and flapping rule are independently motivated, and it remains to be seen whether an explanation of their interaction in terms of monostratal rule composition will satisfy the theoretical demand for generalization. One alternative might be to follow Kaplan and Kay's lead in showing how a multistratal theory can be compiled into a bistratal one, thus satisfying the linguistic drive for generalization and the computational drive for monostratal constraints.

It might be interesting to consider a historical precedent in the matter of rule ordering and intermediate representations. Bird cites similarities between his constraint-based approach and Natural Generative Phonology (NGP) (Vennemann 1974; Hooper 1976; Hudson 1980). The primary tenets of NGP were the "true generalization condition," which disallowed abstract representations and constraints that did not hold of surface structure, and the "no ordering condition," which prohibited rule ordering. Together, these two principles indirectly constrained the abstractness of underlying representations by placing tight constraints on rules and derivations. Bird, we believe, incorrectly states that natural generative phonology did not survive "because it predicted that there is no neutralization, a claim which was anathema to phonologists at the time" (p. 39). Word-final devoicing in German is a classic case of a neutralization process, and it is perfectly acceptable as a process in NGP since it does indeed express a true generalization: all word-final obstruents are voiceless in German. The demise of NGP was more probably due to the fact that the constraints on rules and rule application were deemed to be too strong, preventing phonologists from expressing significant generalizations. The lack of rule ordering is one of the factors that undermined the success of NGP.

As in much of the phonological and constraint-based literature, Bird is rather vague as to the metaconditions on constraints. Although he is careful to segregate the underlying constraint system (a subset of first-order logic) from the linguistic theory (particular relations, individuals, and inference rules expressed in first-order logic), there is no clear picture of universal grammar presented from which we can determine which principles are universal and what kind of constraints might be available to particular languages. Although a fully fleshed-out theory is not necessary at this stage of research, it would be helpful to know the answers to some deep questions, such as what is the role of a lexeme and how is the theory meant to interface with morphology? Perhaps some of the power of rule ordering might be recaptured by allowing multistratal representations for morpheme combinatorics (following the HPSG approach to syntax, which allows phonological operations to generate the phonological representation of a mother from that of the daughters). Without clear indications of a metatheory or even a well-established grammar of a particular language, the newcomer is at a loss as to how to formulate new analyses in constraint-based phonology.

3.2 For the computational linguist

The key issue left unresolved for the computational linguist is the intended application of this system. There is some mention in the introduction about applications to speech systems, but the bulk of the work appears to be aimed at the theoretical linguist rather than the applied natural language processing system developer. Presumably,

Bird shares the goals of the theoretical linguist in laying out a cognitive model of human language knowledge and its use. Perhaps, as we suspect, this will be the shortest route to more robust speech-processing systems in the future.

Despite the fact that the theory is implemented, there is almost no computational motivation of data structures, search strategies, or of metatheoretical constraints. The linguist seeks a constrained universal grammar in order to describe the range of human languages and the human's ability to use them; the computer scientist might be interested in a constrained metatheory in order to develop efficient implementations. From a formal-languages point of view, it would be instructive to learn what kind of languages could be represented by the kind of system that Bird proposes. The primary constraint discussed, Mark Johnson's (1991) restriction to Schönfinkel-Bernays formulas in logic (a particular decidable subset of first-order logic), is only a constraint on the feature logic and not its application in a grammar.

Computational issues such as scaling up the task to handle a whole language and perhaps process corpora of natural data are not discussed. Very little is made of the complexity or representational power of the feature system, and the complexity of the entire system is not discussed. Thus we are not clear what the theoretical computer scientist is supposed to gain from the book other than an introduction to phonology and a partial example of an application of constraints to natural language processing.

One final computational point concerns default rules, which are introduced briefly (pp. 86–88) and later employed sporadically. Such rule systems are notoriously difficult to implement efficiently, even in a highly restricted propositional setting. Decisions must be made about key theoretical issues, such as what to do in the case of conflicting defeasible conclusions, and the default-logic literature presents a range of choices. In this regard, Bird is as guilty of lack of rigor as the generative phonologists he criticizes. If anything more is to be made of defaults in this system, such decisions must be made.

4. Relation to other work in computational phonology

For a book meant to introduce a new theoretical paradigm, there is surprisingly little attention paid to previous approaches to computational or theoretical phonology. One would have expected more than a cursory discussion of the two-level approach popularized in the Kimmo system, and the currently popular Optimality Theory (Prince and Smolensky 1993). The two-level approach is popular because of its programmability and its efficiency in practical applications (even though the system is NP-complete). Optimality Theory has grown in popularity because it solves several outstanding problems for which no other solutions are known. It is not clear what criteria could be used to convince a theoretical or computational linguist to adopt a constraint-based approach to phonology.

There is not even much discussion of the other work done in constraint-based computational phonology, such as Scobbie's (1991) Attribute-Logic Phonology, Russell's (1993) constraint-based version of government phonology, or even the author's own collaborative work on one-level phonology (Bird and Ellison 1994) and constraint-based phonology (Bird and Klein 1994). Integrating ideas from these sources would have greatly strengthened the presentation and the range of data covered. Discussion of the growing literature in constraint-based morphology would have also made a useful addition, due to the tight integration of phonology and morphology.

5. Conclusion

We would like to conclude by recommending this book to anyone interested in computational phonology. We would like to draw an analogy between this book and the

first HPSG book (Pollard and Sag 1987): they both provide great promise along with a number of useful and interesting partial analyses, raising as many questions as they answer. Whether or not this book is the harbinger of a new era of phonological modeling, it is worth reading to learn how fairly standard autosegmental surface representations can be produced by monostratal constraint resolution.

References

Bird, Steven and Ellison, T. Mark (1994). "One-level phonology: autosegmental representations and rules as finite automata." *Computational Linguistics*, **20**, 55–90.

Bird, Steven and Klein, Ewan (1994). "Phonological analysis in typed feature systems." *Computational Linguistics*, **20**, 455–491.

Bird, Steven and Ladd, D. Robert (1991). "Presenting autosegmental phonology." *Journal of Linguistics*, **27**, 193–210.

Browman, Catherine and Goldstein, Louis (1989). "Articulatory gestures as phonological units." *Phonology*, **6**, 201–251.

Clements, George N. (1985). "The geometry of phonological features." *Phonology Yearbook*, **2**, 225–252.

Gazdar, Gerald (1981). "Unbounded dependencies and coordinate structure." *Linguistic Inquiry*, **12**, 155–184.

Goldsmith, John (1990). *Autosegmental and metrical phonology*. Basil Blackwell.

Hooper, Joan B. (1976). *An introduction to natural generative phonology*. Academic Press.

Hudson, Grover (1980), "Automatic alternations in non-transformational phonology." *Language*, **56**, 94–125.

Johnson, Mark (1991). "Features and formulae." *Computational Linguistics*, **17**, 131–151.

Kaplan, Ronald M. and Kay, Martin (1994). "Regular models of phonological rule systems." *Computational Linguistics*, **20**, 331–378.

Pollard, Carl J. and Sag, Ivan A. (1987). *Information-based syntax and semantics: Volume I: Fundamentals*. Center for the Study of Language and Information.

Pollard, Carl J., and Sag, Ivan A. (1994). *Head-driven phrase structure grammar*. University of Chicago Press.

Prince, Alan and Smolensky, Paul (1993). "Optimality theory: constraint interaction in generative grammar." Technical Report 2, Center for Cognitive Science, Rutgers University.

Sagey, Elizabeth (1986). *The representation of features and relations in non-linear phonology*. Doctoral dissertation, Massachusetts Institute of Technology.

Russell, Kevin (1993). *A constraint-based approach to phonology and morphology*. Doctoral dissertation, Linguistics Department, University of Southern California.

Scobbie, James (1991). *Attribute value phonology*. Doctoral dissertation, Centre for Cognitive Science, University of Edinburgh.

Vennemann, Theo (1974). "Words and syllables in natural generative grammar." In *Papers from the Parasession on Natural Phonology*, Anthony Bruck, Robert A. Fox, and Michael W. La Galy (editors), 346–374, Chicago Linguistics Society.

Deirdre Wheeler is a senior research scientist in the Computational Linguistics Program at Carnegie Mellon University. She received her Ph.D. in Linguistics from the University of Massachusetts at Amherst in 1981. Her current research includes non-derivational approaches to phonology and computational models of phonological processing. *Bob Carpenter* is an associate professor in the Computational Linguistics Program at Carnegie Mellon University. He received his Ph.D. in Cognitive Science in 1989. His current research includes the theory and processing of attribute-value logics, and constraint-based approaches to phonology, syntax, and semantics. Wheeler's and Carpenter's address is: Computational Linguistics Program, Philosophy Department, Carnegie Mellon University, Pittsburgh, PA 15213; e-mail: wheeler+@cmu.edu and carp+@cmu.edu.