

## Principle-Based Parsing: Computation and Psycholinguistics

Robert C. Berwick, Steven P. Abney, and Carol Tenny, editors  
(MIT, Bell Communications Research, and University of Pittsburgh)

Dordrecht: Kluwer Academic  
Publishers (Studies in Linguistics and  
Philosophy 44), 1991, vii + 408 pp.  
Hardbound, ISBN 0-7923-1173-6, \$89.00,  
£58.00, Dfl 165.00

*Reviewed by*  
Geoffrey K. Pullum  
University of California, Santa Cruz

This book is a collection of 13 essays on various aspects of parsing:

"Principles of principle-based parsing" (an introduction to the volume)  
by Robert Berwick (pp. 1–37)

"Deductive parsing: The use of knowledge of language" by Mark  
Johnson (39–64)

"The computational implementation of principle-based parsers" by  
Sandiway Fong (65–82)

"Empty categories, chain binding, and parsing" by Nelson Correa  
(83–121)

"Parsing Warlpiri—A free word order language" by Michael B. Kashket  
(123–151)

"Principle-based parsing for machine translation" by Bonnie Jean Dorr  
(153–183)

"Principle-based interpretation of natural language quantifiers" by  
Samuel S. Epstein (185–198)

"Avoid the pedestrian's paradox" by Edward P. Stabler, Jr. (199–237)

"Parsing with changing grammars: Evaluating a language acquisition  
model" by Rick Kazman (239–256)

"Parsing by chunks" by Steven P. Abney (257–278)

"Subcategorization and sentence processing" by Paul Gorrell (279–300)

"Subjacency in a principle-based parser" by Bradley L. Pritchett  
(301–345)

"Locating wh-traces" by Howard S. Kurtzman, Loren F. Crawford, and  
Caylee Nychis-Florence (347–382)

The essays are diverse; they run the gamut from overview-of-my-system term papers in computational linguistics (Fong's paper even includes screen dumps of Lisp machine displays) to experimental papers in psycholinguistics (like the last three papers and Kazman's experimental acquisition study).

The “principle-based parsing” that all the essays are supposed to have in common seems to be entirely a rhetorical construct. Insofar as there is clear and explicit work on parsing in this book, it all looks suspiciously like plain old context-free parsing (sometimes, in fact, clearly deterministic context-free). Where the essays are too vague for this to be said of them, the only connecting theme seems to be a link to MIT-originated linguistic concepts.

Robert Berwick’s introduction opens in breathlessly self-congratulatory mode: “This book chronicles the first stirrings of a revolution in the study of natural language processing, language variation, and psycholinguistics—what some have called *principle-based parsing*” (p. 1). But of course “principle-based parsing” (henceforth PBP) is the slogan of Berwick’s own group at MIT; it is highly disingenuous for him to imply (with his “what some have called”) that it came from out there in the community and he is going along with it. PBP is not only a brand name exclusive to Berwick and his students, but also seems to have an exclusively sociological definition: the preface of this book suggests that all those who were invited to give papers at the MIT Parsing Project Lecture Series between 1987 and 1989 get the PBP rosette pinned on their work no matter what its content, and nobody else does, no matter how much they may build “principles” into their systems.

I cannot find any sign of an incipient revolution chronicled in this book; not even the outline of a new position is visible. Berwick’s claim is that once upon a time people used to use “thousands of individual, language-particular, and construction-specific rules” when writing grammars or parsers, and PBP “replaces this standard paradigm with another world view” under which a small number of general “principles” do the work. These principles turn out to be vaguely stated generalizations from Government-Binding (GB) syntax, of course (to a large extent, PBP practitioners seem to be just GB linguists with Lisp machines). Some of the “principles” are close to vacuous, e.g., “verb phrases in sentences must either *begin* with a verb . . . or *end* with a verb” (trivially true given the Bloomfieldian doctrine that all branching is binary, which GB syntacticians appear to maintain). Some are quite traditional, e.g., “all pronounced or *lexical* noun phrases . . . must receive Case” (i.e., Case is an obligatory grammatical category for nouns—provided you accept that in Chinese or in the English noun it shows no morphological effects). At least one is false as stated, namely “every verb must discharge its *Thematic arguments* and every noun phrase must receive a thematic role” (i.e., a verb denoting an *n*-place relation needs exactly *n* argument NPs; Berwick seems to have forgotten about nonthematic NPs like idiom chunks and ‘dummy’ or ‘expletive’ NPs).

It is hard to see any parser design emerging out of such “principles,” and sure enough, Berwick rapidly admits (p. 8) that “two related computational difficulties lie at the heart of principle-based parsing: *overgeneration* and *slow parsing*.” Producing incorrect results, and doing it slowly. Not very promising. But Berwick faces this embarrassment with more feats of rhetoric, drawing inspiration from the language of product advertising (“Astonishingly . . . Fong’s parser can actually parse basically all of the several hundred example sentences in Lasnik and Uriagereka’s textbook . . .”—tests have shown!), political history (“the central achievement of principle-based parsing stands . . .”), revolutionary philosophy (“Thinking about principles liberates us . . .”) and military history (“results described in this volume form the beachhead of a much broader wave of principle-based research to come”). Such stuff will make some grin and others wince. Maybe it will convince a few to join the movement — but it is not clear to me what they will be joining.

The reference to the days when unnamed people proposed “thousands of individual, language-particular, and construction-specific rules” is, of course, an oblique attack

on generalized phrase structure grammar (GPSG) and all the computational work that has evolved from it at sites like Hewlett-Packard Laboratories, Stanford, SRI International, Carnegie-Mellon, Ohio State, Edinburgh, Cambridge, Sussex, etc. But anyone who knows that line of work will be aware that it has *always* sought general principles rather than lists of rules (though where rules were called for, they were at least explicitly stated rather than arm-wavingly assumed). The very earliest publications on GPSG were interested in general principles such as the Head Feature Convention. And by about 1984, head-driven phrase structure grammar (HPSG) was emerging from research at Hewlett-Packard Laboratories and had reduced the number of separately stated syntactic rules in a substantial working grammar for English down to less than half a dozen (most of the work being done by just two, one for NP–VP structures and the other for subcategorization structures). The PBP slogan had appeared nowhere previously when Edward Barton issued an entirely programmatic MIT AI Lab memo with that title in 1984. But with 1984-style goodthink, Berwick portrays 1984-style GPSG/HPSG as *falling in step with PBP*, reporting that they “gradually shifted” (p. 35, n. 1) toward transconstructional declarative constraints. The alleged “shift” toward PBP thinking had already taken place before PBP work had even started.

Given Berwick’s anti-rule rhetoric, the innocent reader may be amazed to find that in the clearer papers in this volume one does encounter rules. We find context-free expansion rules on pages 87ff of Correa’s paper (a workmanlike though hardly revolutionary piece of work on formalizing GB-style syntax with attribute grammars); on page 132 and implicitly elsewhere in Kashket’s paper (an extremely limited effort to show how one might parse a language with complex morphology but no word order rules, showing no acquaintance with the relevant literature, e.g., the work on ID/LP format and liberation schemata); in the schema on page 165 and elsewhere in Dorr’s paper (153–183); coded in Prolog throughout Stabler’s paper (199–237); and in pages 262ff of Abney’s paper (a straightforward piece of deterministic context-free LR parsing). So much for principles replacing rules in actual work.

But of course, parsers are always based on rules (however general and schematic or detailed and specific they may be), so the enterprise of appearing to recast parsing without them can only be one of smoke and mirrors. The technique used for keeping rules from making more of an appearance in this book is a combination of wishful thinking and extreme vagueness. I will give one example. Dorr (p. 158) explains approvingly how the grammarian describing Spanish will not need these rules:

- (1)  $S \rightarrow NP VP$
- (2)  $S \rightarrow VP$

but “need only set the *null subject* parameter.” Later (p. 167) she explains what this is: “a minimal binary difference that does or does not allow empty noun phrases to occupy subject position.” Now, nothing much hangs on the switch from missing noun phrases to “empty noun phrases”; it means replacing rule (2) by a rule saying

- (2′)  $S \rightarrow NP[+NULL] VP$

(where NP[+NULL] expands to the empty string). Rule (1) is assumed to be available for all languages (or all SVO languages). The question is whether (2′) is also assumed. Clearly, what Dorr means by setting the “null subject parameter” to *Yes* is assuming that rule (2′) is available to admit local trees rooted in S, and what she means by setting it to *No* means assuming that rule (2′) is not available.

Dorr adds a footnote of boilerplate about how a “rule-based approach” like GPSG would need lots of different rules (p. 180, n. 18), but there is no attempt to show that this is true (the rules she is talking about are actually reduced to one in GPSG work of 1985 and later), and there is (crucially) no attempt to show how the notion “subject position” is defined *without rules* so that the “null subject parameter” can be stated. Dorr states nothing relevant with any precision (there is just a casual remark in her text about whether “empty noun phrases” can “occupy subject position” on page 167). Since Dorr claims to have a machine translation system up and running, she presumably has something tantamount to a statement of the null subject parameter buried in the code of her system. But she does not attempt to compare the information content of this buried statement with the information content of an explicit GPSG grammar fragment for the same fragment of Spanish so that her vague aspersions can be given some meaning. As things stand, all one can say with reasonable certainty is that Dorr wants to be seen as loyal to the approach associated with Berwick’s slogans. Her assumptions are just those of GB linguistics, and the faults in GB tend to carry over to her computational work.

There are some papers in this book that are worth reading, and they are generally the ones that ignore Berwick’s sloganeering completely. Stabler’s paper, for example, is a small but serious contribution to Prolog-based computational linguistics — though I find absolutely nothing in it that refers to GB concepts or PBP. Johnson’s is a clear statement of a simple if not particularly original idea (that using Prolog, parsing can be simply a process of deduction in the logical sense — an idea that seemed to be apparent to Prolog enthusiasts in Europe over a decade ago) and an exploration of a few (five) different control structures that a GB parser might use; and Correa’s work on attribute grammars, as noted above, might repay the reader’s attention. But the level of many of the papers is depressingly low. Fong seems to have built a system for the purpose of testing alternative orders of application for 16 different principles and tests a few without noting that there are  $16! = 20,922,789,888,000$  alternatives to be tested; Kashket has a one-page section called “The problem with context-free parsers,” (pp. 126–7) which does nothing but assert the entirely false claim that a context-free grammar cannot generate structures for VSO sentence types in NP–VP languages; such chapters represent feeble contributions at best, managing to sound both amateurish and insular.

In a way, the most favorable thing I can say about the various kinds of work on parsing in this book is that the contributors pay so little regard to the promises issued on their behalf by the blatant marketing hype of Berwick’s introduction. After all, one of the ways in which computational linguistics holds out promise for linguistics in general is by providing practical tests that will inject some honesty into the testing and comparison of linguistic theories, and that promise is betrayed when routine work on parsing problems is decked out with the kind of vagueness and self-deception that are so evident in the introduction to this book.

*Geoffrey K. Pullum* is on the faculty of the University of California, Santa Cruz, where he has been Professor of Linguistics since 1981 and Dean of Graduate Studies and Research since 1987. Pullum’s address is: Cowell College, UCSC, Santa Cruz, CA 95064; e-mail: pullum@cats.ucsc.edu.