The Language Complexity Game

Eric Sven Ristad
(Princeton University)

Cambridge, MA: The MIT Press, 1993,
xvii + 148 pp.
Hardbound, ISBN 0-262-18147-9, $30.00

Reviewed by
Alexis Manaster Ramer
Wayne State University

The book under review makes certain claims about natural languages couched in mathematical terms. And so I begin with a caveat (see also Manaster Ramer 1992): as in far too many other books and articles in our field, Ristad consistently and mistakenly uses the words *prove* and *proof* when speaking not only about the results concerning well-defined mathematical objects (i.e., formal languages), but also about the claims that (some) NLs are properly modeled by these mathematical objects and hence share their properties. In this review, I will consistently distinguish between *proofs* of properties of formal languages and *arguments* about the properties of NLs.

With this caveat firmly in mind, we may now specify that Ristad's book presents the argument that the understanding of anaphora in NLs, specifically in English, is NP-complete, i.e., it can be computed in nondeterministic polynomial time and is also NP-hard, that is, as hard as the "hardest" problems in the class of problems solvable in nondeterministic polynomial time (see, e.g., [Hopcroft and Ullman 1979 pp. 320–341] for a more formal introduction to these concepts). We are thus dealing with an attempt to characterize the properties of NLs in terms of complexity theory rather than the more familiar formal language theory. Moreover, the problem studied is not that of distinguishing grammatical from ungrammatical sentences, as in most work on mathematical linguistics, but rather that of determining what meanings are possible for a given set of sentences. In particular, Ristad's argumentation revolves around various aspects of the problem of determining which anaphoric elements in a given sentence can refer to which potential antecedents. Ristad takes the reader through five rounds of what he calls a "complexity game," which is a contest between a maximizer, who tries to make natural languages as complex as possible, and a minimizer, who seeks to reduce the complexity to a bare minimum.

In the first round, we read an argument purporting to demonstrate the NP-hardness of any language whose anaphora are required to agree in features such as number, gender, and so on with their antecedents. In the second round, this argument is refuted by the minimizer, who claims that the standard theory of how agreement works is wrong and proposes a new theory, under which anaphoric agreement is now recognizable in deterministic polynomial time. In the third turn, a new set of data leads to the central argument in the book, according to which the anaphora problem is NP-hard after all. The facts crucial to this argument have to do with obviation—that is, conditions under which coreference is impossible, as in *He saw him*. Not content with this result, the maximizer goes on to the fourth round, in which he presents yet another body of data (dealing with anaphora in elliptical structures) on the basis of which he argues that the anaphora problem is PSPACE-hard (i.e., as hard as the "hardest" problems in the class of problems solvable in deterministic polynomial space, a class which contains $\mathcal{NP}$). But the minimizer replies, in the fifth and final round, by arguing against the theory of ellipsis presupposed in this argument and proposes another theory, under which the part of the anaphora problem involving ellipsis is no more difficult than the part dealing with obviation that was argued earlier to be merely NP-hard.

The game thus ends with the result that the English anaphora problem is NP-hard, but within the class $\mathcal{NP}$, that is to say, NP-complete. This proposal is endorsed by Robert C. Berwick in a glowing seven-page foreword, in which he describes these results as no less revolutionary than those presented by Chomsky in the 1950s (while at the same time making fun of other current work on the mathematical properties of NLs). It is perhaps unusual in a review to discuss a foreword, but then this foreword is itself unusual, as it is my painful duty as a reviewer to mention.

Specifically, Berwick (pp. xiii–xiv) tells us that Ristad's NP-hardness result has to do with the problem of assigning reference to pronouns in sentences such as (1), which is claimed to be NP-hard because supposedly one can reduce to it the problem of graph *k*–coloring, which is an NP-complete problem.

(1) Before Bill, Tom, and Jack were friends, he wanted him to introduce him to him.

Not only is there a mathematical proof of this fact, says Berwick, but "If you don't believe that, just try to figure out the links between the different *him*s in Ristad's example sentences such as [(1)]."

However, (1) is actually example (17) from Ristad's dissertation (1990, p. 69), appears in a different form in the book (p. 54), and is completely irrelevant to the NP-hardness argument. As Ristad (1990, p. 69) points out (see also p. 55 of the book), "The configurations used to construct the sentence (17) can only give rise to very simple obviation graphs on their own, and therefore the proof of" NP-hardness "must build obviation graphs using the agreement condition."

The agreement condition referred to by Ristad is simply the statement that pronouns have to agree with their antecedents in various inflectional features. Of course, in (1), this condition does not come into play because all the pronouns are masculine singular (*him*). To get NP-hardness one would need to consider examples in which pronouns with different features are available. This is why (1) is too "easy" to be an example of the kind of complexity Ristad is after. Examples such as this, even if they are difficult for people to figure out, have nothing whatever to do with NP-hardness.

In fact, the particular NP-hardness result we are discussing assumes that there is an infinite set of inflectional features. If there is only a finite set of features, as some might argue, then this particular argument for NP-hardness collapses. In Ristad (1990, pp. 101–102) there is an extended attempt to justify the assumption of an infinite number of features as a reasonable idealization about NLs, but this has been left out of the book, which only cites the discussions by Ristad (1990) as well as by Pullum (1983). This is a pity, because the reader has to go back to Ristad (1990, p. 101) to find the *Guinness Book of World Records* cited as the source for the impressively large (though still finite) number of cases in the North Caucasian language Tabassaran.

Moreover, the whole argument about the NP-hardness of the anaphora problem in agreement situations is no longer even accepted in Ristad's book. To be sure, this argument was the centerpiece of his dissertation, but it is now presented merely as the maximizer's first (flawed) attempt at showing NP-hardness (without any mention of the claims that were made in 1990). Specifically, the second round of the game is devoted to refuting this argument on grounds that I will discuss momentarily. Thus, not only did Berwick misinterpret what the result is all about, but he also is lavishing praise on an argument that, although advanced by Ristad in 1990, is now presented as a straw man.

The grounds given by Ristad for rejecting this argument have to do with the way that agreement is defined. The argument assumed that agreement means that all anaphors referring to a given antecedent are nondistinct in their feature specifications from it and from each other. In other words, agreement was assumed to mean nondistinctness rather than identity of feature specifications.

But the minimizer now claims that this conception of agreement is incorrect and that agreement requires more than mere lack of conflict in feature specifications. Ristad bases his new theory of anaphoric agreement on certain facts about English. For example, the normally plural *they* can sometimes be used with a singular antecedent:

(2) About your doctor, do they seem competent?

Ristad takes this to mean that *they* subsumes *he* (as well as *she*). If agreement were a matter of nondistinctness, then it should be normal to use both *they* and *he* to refer to the same antecedent in the same sentence. Yet, this is not the case.

(3) About your doctor, do they seem competent when he examines you?

In (3), *they* cannot be read as coreferent with *he*. Because of facts such as this, Ristad claims that, universally, a linguistic element β with an inflectional feature c may antecede an anaphoric element α only if α represents the most specific category subsuming c that exists in the paradigm of the anaphoric elements. (Thus, agreement would require identity of feature specifications, unless certain feature combinations are simply not realized in the anaphoric elements available in the language.)

If this is granted, the maximizer's NP-hardness argument collapses. However, it is not, apparently, possible to maintain this proposed universal. Bloomfield (1962, pp. 38–39) cites Menomini examples of coreference that seem to involve nondistinctness, and Sapir (1930, pp. 187, 194, 201, 202) gives similar examples from Southern Paiute. I have also been able to find an apparent example in Spanish. In this language, there are several ways of expressing third-person possessive pronouns. The form *su*, which is not specified as to gender or number, would seem to subsume forms such as *de él* 'his', *de ella* 'her', etc. According to Ristad's theory, therefore, it should not be possible to use *su* and *de él* coreferentially. Nonetheless, my informants accept sentences such as

(4) Juan le dió a su esposa el reloj de él.
Juan her gave to his wife the watch of him
'Juan gave his wife his watch.'

Yet, although (and indeed precisely because) the linguistic facts remain to be clarified, I believe that Ristad has identified a genuine problem for linguistic theory. Although his account of coreference is not correct in general, there are clearly many cases in which mere nondistinctness is insufficient to allow coreference. Thus for theoretical and computational linguistics, there is considerable interest in Ristad's claims about the conditions under which coreference obtains.

Yet, no matter what we decide about the way that coreference works, it seems to me that the maximizer's NP-hardness argument based on agreement must be rejected because it assumes an infinite set of features. We do not have to accept the minimizer's counterargument to agree that the maximizer's argument is not compelling.

Thus, the first two rounds of the complexity game have yielded nothing more than a refutation of Ristad's original 1990 NP-hardness argument and a number of mostly open questions about how agreement really works.

In the third turn comes the NP-hardness "proof" that Ristad now holds to be valid and that is the centerpiece of the book. This argument involves a reduction of the 3-SAT problem (Hopcroft and Ullman 1979 pp. 330–331), which, like Graph *k*–Coloring, is an example of NP-completeness, to a new set of examples involving anaphoric elements in English. In the new examples, the troubling issues of how many features there are and of how to define agreement no longer rear their ugly heads. This time all the pronouns are *him*. The required complexity comes not from agreement (as might appear from a rather misleading footnote in Manaster Ramer 1993a, p. 10, fn. 6), but from the interaction of various conditions of coreference and obviation in English.

It would be nice at this point in the review to be able to give an illustration, but Ristad never supplies even a single complete example. All we are offered are several hints about what sentences are intended, in the form of examples of certain parts of such sentences, together with partial narrations and some diagrams showing to some extent how these parts are to be put together. However, the fact remains that his proof does not state what the reduction from 3-SAT is to, which is a serious matter for a reduction proof.

Of course, we can try to fill in the gaps, but in any event, it is best to begin by noting that the central factual claim appears to be what Ristad calls "invisible obviation," which is supposed to show up in examples such as these:

(5a) Sue wanted Bill to kiss Mary and he did too.

(5b) Romeo wanted Rosaline to love him before wanting himself to.

According to Ristad, in (5a), *he* cannot be understood as coreferential with *Bill*, and likewise in (5b), *him* cannot be understood as coreferential with *Romeo*. In Ristad's terms, in each of these cases the "invisible" verb phrase contains an "invisible" direct object pronoun that obviates the same noun phrases as an overt direct object pronoun in an overt verb phrase would. Thus, in (5a), the second conjunct can be understood as (6), but since in (6), *he* obviates *him*, the same is claimed to be true in (5a). This would mean that the "invisible" *him* cannot be understood as coreferent with *he*. But since the "invisible" *him* must be coreferent with *Bill*, it then follows that *he* and *Bill* cannot corefer.

(6) He wanted him to kiss Mary, too.

I personally am not sure that this observation is correct, and I have been present at meetings where several other linguists, who are native speakers of English, have flatly rejected it. Yet these subtle judgments are the linchpin of the whole argument. Moreover, Ristad himself seems to admit that the forbidden interpretations are possible under contrastive stress (p. 58). It has been my experience, in many cases in which it is claimed that some interpretation can occur only with contrastive stress, that such stress is merely helpful in getting the interpretation in question and not absolutely required. This makes the whole thing even more subtle. In addition, since contrastive stress is not usually marked in written English in any way, the NP-hardness argument at issue can at best be made only for spoken English (and hence would be of little relevance to that body of computational work that, as noted by Ristad on page 118, is concerned with "orthographic forms").

Also, as already mentioned briefly, Ristad never states clearly how to bridge the gap between such English examples and the mathematical result about NP-hardness. We are supposed to accept that there is a reduction from 3-SAT to the English anaphora problem under discussion, yet the translation from Boolean formulas to English sentences is not fully specified. Ristad gives only a number of partial indications of how the reduction might be performed. It thus appears that given a Boolean formula such as (*A* or *B* or *C*), we must construct a quite complex English sentence made up of the following parts. First, there have to be three sentences (one for each Boolean variable) such as (7).

(7) He persuaded him to introduce him to Hector.

Next, there has to be a sentence containing three elliptical verb phrases (one for each occurrence of a Boolean variable) of the form in (8), so that we can get the effects of invisible obviation ([e] is used to mark the spots where invisible verb phrases are supposed to occur).

(8) True met him, who he expected him to want him framed, who he
believed he did [e] with, after exposing himself to [e] for, before
telling himself to [e].

Finally, we are somehow supposed to fit the three sentences of the form in (7) together and combine the result with (8). The diagram on page 67 would seem to suggest that these combinations are perhaps all coordinations, but the details are not clear.

In any event, in the problem of 3-SAT we are concerned with whether a Boolean formula of a particular form is satisfiable (i.e., whether any assignment of the truth values **true** and **false** to the variables yields **true** as the value of the whole formula), and of course, each occurrence of a given variable has to be assigned the same value. In the English examples intended by Ristad, whatever their exact form is supposed to be, coreference with the noun phrase *True,* which is taken to be a proper name, corresponds to the assignment of the truth value **true** to a variable in a Boolean formula, and the occurrence of multiple coreferent expressions corresponds to the multiple occurrences of the same variable in various places in a Boolean formula. Now, the crucial thing is supposed to be that, because of various conditions on coreference and obviation (including "invisible" obviation), at least one of the "invisible pronouns" in sentences such as (8) must be understood to refer to the name *True* (and so must any overt pronouns that corefer with the "invisible" one), thus corresponding to the requirement that at least one of the variables in the Boolean formula (*A* or *B* or *C*) must have the truth value **true** in order for the whole formula to have this value.

Given the complexity of such examples as (8)—and (8) corresponds to the very simplest Boolean formula under consideration—it is obviously going to be difficult to determine whether the coreference possibilities recognized by a speaker of English really have this property. To be sure, Ristad gives a diagram intended to show which expressions corefer and which obviate each other in (8). However, this diagram and the accompanying explanatory text contain a number of omissions and typos, so that it is in part up to the reader to figure out what should corefer with (or obviate) what, before we can sit down to decide whether we are willing to grant the judgments demanded by Ristad.

There are thus three reasons to doubt the validity of the whole argument. First, it seems to rest on judgments that some speakers other than Ristad do not share. Second, even Ristad seems to admit that the crucial judgments depend on stress, that last refuge of the formal grammarian. Third, he never provides a full description, or a single complete example, of the sentences he has in mind and of the coreference and obviation relations that are supposed to hold in these sentences. All three points would need to be addressed before the NP-hardness argument could be entertained, much less accepted.

Yet Ristad concludes that the third round has yielded a "proof" that the anaphora problem for English is NP-hard, and he then goes on to two more turns of the complexity game, in which the question is whether English might be even harder. Here the maximizer tries to show that the English anaphora problem is PSPACE-hard, but the minimizer refutes this argument. The issues here are quite complicated, but the central point is that they deal with ellipsis, as in (9).

(9) Juliet thought that the Friar poisoned her without realizing that she did.

As Ristad argues, the anaphora problem in elliptical structures would be PSPACE-hard if we assumed a copy (i.e., basically, transformational) model of ellipsis, but he then develops a theory of ellipsis as a kind of higher-order predicate sharing and offers a nondeterministic polynomial time algorithm for processing elliptical structures according to this theory. The game thus ends with the conclusion that the understanding of English anaphora is NP-hard (as argued in round 3) and within the class $\mathcal{NP}$. Although I have made it clear that I find Ristad's arguments quite unconvincing, there are three issues that come to mind even if we were to accept the NP-hardness thesis.

First, it should be noted that there are two parts to the claim that the understanding of anaphora is NP-hard but is within the class $\mathcal{NP}$, and these two are not of the same type (pace Ristad's claims on pp. 7–8). The former, the lower-bound claim, could arguably be established by a single convincing set of examples (although I have yet to see this done), but the latter, the upper-bound claim, cannot, since the existence of an $\mathcal{NP}$ algorithm for one set of examples does not ensure that all sets of examples (especially if we consider all possible natural languages) will have such algorithms.

Second, we must ask whether the NP-hardness result would explain why some parts of natural languages are difficult for people to handle. One might have expected such a claim, especially since this is what Berwick tells us in his introduction (p. xiii), even offering an example of this (even if the example in question does not in fact involve NP-hardness, as noted). However, Ristad (pp. 121–124) explicitly rejects this sort of connection, arguing that there are many natural-language examples that are hard for people but easy for machines, and vice versa, and excoriating those linguists who have sought to explain performance difficulties in terms of limitations on computational resources. Yet, is it not inevitable, if natural languages really are NP-hard, that there would have to be classes of examples that are difficult for both people and machines (and this precisely because of the NP-hardness)?

Third, Ristad emphatically and repeatedly asserts that his NP-hardness argument "relies only on the empirical facts ..., and on the uncontroverted assumption that these facts generalize in a reasonable manner" (p. 66 and passim). Yet the same was claimed by Ristad (1990, p. 70) with regard to the argument (involving agreement) that is now presented as a straw man on the grounds that the theory of agreement assumed in 1990 is incorrect. Perhaps part of the problem is that it is rarely uncontroversial just how a finite array of examples generalizes in a reasonable manner. It is thus difficult for me to accept that nothing beyond theory-independent observations about English is involved in the argument for NP-hardness, and a spelling out of the theoretical assumptions would thus be one more thing to require before one signs on the dotted line.

There is, of course, much more to the book under review, including a compact introduction to the relevant concepts of complexity theory (pp. 133–138) and discussions of two theoretical points that could have been, had they been developed properly, of potentially great interest.

One is a sermonette (pp. 13, 109–110, and passim) about a new version of the I-language view of natural languages that purports to be more Chomskyan (i.e., more mentalist) than Chomsky. The I-language perspective (see, especially, Chomsky 1986) holds that natural languages should not be conceived of as infinite sets of grammatical sentences, partly because well-formedness comes in different degrees (and different kinds) and partly because an infinite language is an abstraction of doubtful ontological status. According to Chomsky, then, the real object of linguistic study should be the finite set of sentences that are available to language learners as well as the grammar that the learners are assumed to develop in their minds upon exposure to this set.

It is this mental grammar that is now (re)named "I(nternalized)-language," whereas the term "E(xternalized)-language" is applied to the now-rejected concept of the infinite set of grammatical sentences. With the help of these terms, Chomsky contrasts the I-language view (or perspective), which he claims to have consistently been his own approach, with the E-language view (or perspective), which is said to be the approach of most nongenerative linguists and of mathematical linguists (for a rebuttal, see Manaster Ramer 1993a, 1993b, to appear).

As for Ristad, he proposes that natural language is "a completely internalized cognitive system, more internalized than the I-language view of Chomsky," and, more specifically, that "each particular human language is a finite computing machine that executes a computation from some extralinguistic information to a linguistic representation of that information" (p. 13). By "extralinguistic information" Ristad (p. 12) means things such as the "sensation arising from ... an acoustic or visual signal" (i.e., from the hearing or reading of a sentence) and "intentions, sensations, beliefs, models of agents (other agents as well as oneself), mental lexicon, conceptual system, and so forth" (p. 9) (i.e., what we would normally describe as the semantic and pragmatic content of a sentence). Thus, the crucial feature of Ristad's proposal would appear to be that instead of phonetics and semantics/pragmatics as such, he wants us to talk about the mental representations of phonetics and semantics/pragmatics.

Now, it is by no means clear to me or to Chomsky (personal communication) that this is original, that is, any different from, for example, Chomsky's approach. Moreover, Ristad gives no basis for preferring this view of language to what he takes to be Chomsky's version of the I-language perspective. Indeed, whereas Chomsky has sought to show that there are factual reasons for adopting his I-language view (in preference to the E-language view) and significant consequences if we adopt it, Ristad does not attempt anything of the sort and appears to admit that the substantive part of his work, the NP-hardness argument, not only is independent of the proposed change from one kind of I-language perspective to the next (assuming these *are* different), but is even consistent with the E-language approach (pp. 117–120 and passim).

Ristad's other theoretical point is a strident attack on the distinction between competence and performance (pp. 116–117, 122–124), summarized in the words "There is simply human language itself, exactly as it is and how it operates" (p. 117). What is troubling here is that, unlike the many linguists (none of them mentioned by Ristad) who have argued over the years against the competence/performance distinction (or at least some versions of it), Ristad does not attempt to deal with any of Chomsky's well-known arguments for it.

Worse, Ristad's own NP-hardness argument appears to presuppose crucially the difference between competence and performance, that is to say, the difference between (un)grammaticality and (un)acceptability. For, if there were no such difference, then sentences such as (8) could not be both grammatical and unacceptable. Their status in the "language itself, exactly as it is and how it operates" would then have to be either unequivocally good (which surely no speaker of English would assent to) or unequivocally bad (in which case they do not belong to English, and hence the whole argument collapses).

Overall, then, the book under review pinpoints a number of interesting linguistic facts and asks some very good questions about the theory of anaphoric elements, about the mathematical properties of natural languages, and about the foundations of linguistics, but then throws out answers that are supported either inadequately or not at all.

Acknowledgments

Thanks to Geoffrey K. Pullum, Walter J. Savitch, and Daniel Radzinski for reading and commenting on the draft of this review, and suggesting some corrections. The opinions expressed, and any remaining errors, are, of course, my own responsibility.

References

Bloomfield, Leonard (1962). *The Menomini Language.* New Haven, Connecticut and London: Yale University Press.

Chomsky, Noam (1986). *Knowledge of Language: Its Nature, Origins, and Use.* New York: Praeger.

Hopcroft, John E., and Ullman, Jeffrey D. (1979). *Introduction to Automata Theory, Languages, and Computation.* Reading, MA: Addison-Wesley.

Manaster Ramer, Alexis (1992). Review of Barbara H. Partee, Alice ter Meulen, and Robert E. Wall. *Mathematical Methods in Linguistics.* Dordrecht: Kluwer. 1990. In *Computational Linguistics* 18(1):104–107.

Manaster Ramer, Alexis (1993a). "Capacity, complexity, and beyond." (Presidential address to MOL 2). *Annals of Mathematics and Artificial Intelligence* 8(1/2). (*Mathematics of Language,* edited by Wlodek Zadrozny, Alexis Manaster Ramer, and M. Andrew Moshier), 1–16.

Manaster Ramer, Alexis (1993b). "Towards transductive linguistics." *Natural Language Processing: The PLNLP Approach,* edited by Karen Jensen, George E. Heidorn, and Stephen D. Richardson, 13–27. Boston/Dordrecht/London: Kluwer Academic Publishers.

Manaster Ramer, Alexis (to appear). "The uses and abuses of mathematics in linguistics." *Lenguajes naturales y lenguajes formales XI,* edited by Carlos Martín Vide. Barcelona: PPU Promociones y Publicaciones Universitarias, S.A.

Pullum, Geoffrey (1983). "How many possible human languages are there?" *Linguistic Inquiry* 14:447–467.

Ristad, Eric Sven (1990). *Computational structure of human language.* Doctoral dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Sapir, Edward (1930). "Southern Paiute language." In *Proceedings, American Academy of Arts and Sciences* 65:1–296.

Alexis Manaster Ramer served as the first President of the Association for Mathematics of Language (SIGMOL), edited the proceedings of *MOL 1* (published under the title *Mathematics of Language* by John Benjamins), and coedited those of *MOL 2* (published as *Annals of Mathematics and Artificial Intelligence* 8(1/2)). He has a doctorate in linguistics from the University of Chicago and used to teach that subject at the University of Michigan. He now teaches theoretical computer science at Wayne State University. Manaster Ramer's address is Department of Computer Science, Wayne State University, Detroit, MI 48202. E-mail: amr@cs.wayne.edu