

Falsifying Serial and Parallel Parsing Models: Empirical Conundrums and An Overlooked Paradigm

Richard L. Lewis¹

When the human parser encounters a local structural ambiguity, are multiple structures pursued (parallel or breadth-first parsing), or just a single preferred structure (serial or depth-first parsing)? This note discusses four important classes of serial and parallel models: simple limited parallel, ranked limited parallel, deterministic serial with reanalysis, and probabilistic serial with reanalysis. It is argued that existing evidence is compatible only with probabilistic serial-reanalysis models, or ranked parallel models augmented with a reanalysis component. A new class of linguistic structures is introduced on which the behavior of serial and parallel parsers diverge the most radically: multiple local ambiguities are stacked to increase the number of viable alternatives in the ambiguous region from two to eight structures. This paradigm may provide the strongest test yet for parallel models.

When the human parser encounters a local structural ambiguity, are multiple structures pursued (parallel or breadth-first parsing), or just a single preferred structure (serial or depth-first parsing)?² The purpose of this note is twofold. First, I will examine four important classes of serial and parallel models and show how each can or cannot be falsified. Second, I will introduce a class of linguistic structures, which may provide the strongest test yet for parallel models.

The author acknowledges very helpful comments from Chuck Clifton, Janet Fodor, Ted Gibson, Neal Pearlmuter, and audiences at the 1999 CUNY Sentence Processing Conference and at Potsdam University. This work was conducted while the author was visiting the Departments of Psychology and Linguistics at the University of Potsdam, Potsdam, Germany.

¹ Department of Computer and Information Science, Department of Linguistics and Center for Cognitive Science, The Ohio State University, Columbus, Ohio 43210. E-mail: rick@cis.ohio-state.edu.

² Of course, the answer may be “neither”, if it is possible for parse trees to be underspecified. No discussion of the serial vs. parallel debate would be complete without also considering underspecification, but there is not space here to do so.

SERIAL AND PARALLEL MIMICRY: FOUR EMPIRICAL CONUNDRUMS

To illustrate the issues in empirically distinguishing serial and parallel models, I will focus on falsifying four specific classes of models: limited parallel, ranked limited parallel, deterministic serial with reanalysis, and probabilistic serial with reanalysis. These four models were chosen, in part, because they are notoriously good at “mimicking” important signature results of the opposing theories.

Falsifying Limited Parallel Models

As Gibson and Pearlmutter (this issue) point out, the existence of strong garden-path effects clearly rules out a parallel parser that always maintains all structural interpretations of a sentence. All parallel models proposed to date posit a kind of limited parallelism: in some cases (intended to be those cases corresponding to strong garden-path effects), the alternative structures are not pursued. For example, Gibson (1991) proposed a parallel model that prunes a structure from consideration when the relative cost of that structure (computed by a specific memory metric) exceeds a certain threshold. Thus, like a serial model with limited reanalysis, limited parallel models can account for the existence of both strong garden-path effects (when the required structure has been pruned) and easy ambiguities (when both structures are still available).

Limited parallel models and serial parsers still diverge in their predictions in one crucial area: the disambiguating region. A serial parser augmented with a capability for reanalysis will always predict some cost (even if very small) for recovering the dispreferred interpretation in an easy ambiguity and this asymmetry should be reflected in reading times in the disambiguating region. A simple limited parallel model predicts no asymmetry: if both structures can be recovered, it is because both structures have been maintained and, if both are maintained, they should be equally easy to recover.

There is, of course, ample empirical evidence for such asymmetries: relative to unambiguous baselines, dispreferred structures take longer in the disambiguating region and these asymmetries constitute the core evidence for garden-path models (Frazier, 1987). To account for this evidence, limited parallel models must be modified to include some way of *ranking* the alternatives.

Falsifying Ranked Parallel Models

In a *ranked parallel model* (e.g., Gibson, 1991; Gorrell, 1987; Kurtzman, 1985; Spivey & Tanenhaus, 1998), not all interpretations are created equal: both (all) structures may be pursued, but one structure may be more highly

preferred. If a ranked parallel model, furthermore, predicts a processing cost for switching from the preferred to the dispreferred interpretation, it can account for *both* of the signature phenomena of a serial-reanalysis model: it will fail on severe garden paths and will predict the asymmetries at the disambiguating region for the weak garden paths.

However, it is still possible to distinguish a simple ranked parallel model from certain serial-reanalysis models. In a ranked parallel model, the availability of alternative structures is a function of their ranking: the activation or preference assigned to each structure. The availability of the alternatives should *not* vary as a function of the disambiguating material. That is, if some alternative can be recovered via some disambiguating material, then it should be recovered for *any* disambiguating material that requires it (Frazier, 1998).

However, work on diagnosis and repair models has yielded intuitive evidence that strongly suggests that the ability to recover alternative structures is a function of the disambiguating material (e.g., Fodor & Inoue, 1994). Consider the following example from Fodor and Inoue:

- (1) a. Have the soldiers marched to the barracks tomorrow.
b. Have the soldiers marched to the barracks already?
- (2) Have the soldiers marched to the barracks, would you.

The strong garden-path effect in (1a) (an imperative temporarily misanalyzed as a question) can be significantly reduced or eliminated by changing the cue for disambiguation from the adverbial *tomorrow* to the tag question, *would you*.

A ranked parallel model would not necessarily predict such a difference. For example, the Gibson (1991) model would assign the same ranking to the imperative and question structures independently of what the disambiguating cue is. If the imperative structure is available in (2), it should be equally available in (1a), but it is not. Any ranked parallel model that computes the ranking or preferences for alternative structures based only on the ambiguous material and the prior context will suffer the same empirical fate. A serial-reanalysis model can easily account for the effect of the disambiguating cue because computing the required analysis can take account of both the cue and the preferred structure. A ranked parallel model could also accommodate such effects, but at some loss of perspicuity, because it would have to be augmented with a reanalysis component that is sensitive to the cue (Gibson & Pearlmutter, *this issue*).

Falsifying Serial Reanalysis Models

Now let us consider how serial models might be falsified. In this case it is most useful to consider effects in the ambiguous region. If we can find

signs that the dispreferred structure is computed in the ambiguous region *prior* to disambiguation, that should count as striking evidence against a serial model (Hickok, 1993; MacDonald, Just, & Carpenter, 1992; Nicol & Pickering, 1993).

Nicol & Pickering (1993), Hickok (1993), and Pearlmutter & Mendelsohn (1998) have all pursued empirical studies seeking to falsify serial models in just this way. I will focus here on Pearlmutter and Mendelsohn's work, since it is discussed in the companion note by Gibson and Pearlmutter, and provides the strongest evidence to date against serial models. The key finding is an effect of the plausibility of the dispreferred relative clause interpretation in complement/nominal structures, such as (3) below:

- (3) The report that the dictator described/bombed the country seemed to be false.

The verb *bombed* shows more difficulty than *described*, compared to unambiguous baselines (*The report showing that the dictator described/bombed the country seemed to be false*). This is most naturally ascribed to the implausibility of the RC interpretation at that point: it is less plausible that a dictator would bomb a report than describe it. A serial parser pursuing only the preferred complement structure should have been insensitive to properties of the relative clause analysis.

This result has all the right properties to make it a serious challenge to a serial model: it is probing in the ambiguous region, before the disambiguating material; the probe is not immediately at the point of ambiguity, but a few words later; it is an on-line task; and there is no secondary task to complicate the interpretation of the results (cf. Hickok, 1993; Nicol & Pickering, 1993).

Nevertheless, it is possible to accommodate the plausibility effect in a serial model, as Pearlmutter and Mendelsohn and Gibson and Pearlmutter point out. The key is to adopt a *probabilistic* serial model, which would pursue one interpretation some of the time and the other interpretation another part of the time. The asymmetric effects at disambiguation will arise because reanalysis will be required when the incorrect structure is pursued. There will also be an effect of the dispreferred interpretation during the ambiguous region, because it is pursued some of the time.

Falsifying Probabilistic Serial Models

Pearlmutter and Mendelsohn present further evidence that they argue should rule out a probabilistic serial model. It is useful to carefully consider the argument because it illustrates how difficult it is to falsify such models. The argument can be summarized as follows. They also note that a noun's preference for a sentential complement (SC) versus a relative clause corre-

lated negatively with the size of the ambiguity effect at disambiguation (which always required the SC reading). This could be taken to suggest that the SC preference of the noun affects how the SC/RC ambiguity is resolved by a serial parser: a clause following a strong SC noun would be more often interpreted as an SC, hence reducing the ambiguity effect at disambiguation. Furthermore, the effect of the plausibility of the RC interpretation should increase as the SC bias decreases, since the SC and RC interpretations are in complementary distribution in a serial model. However, Pearlmutter and Mendelsohn report that this latter correlation (a negative correlation between SC preference and the RC implausibility effect) was not statistically reliable. Pearlmutter and Mendelsohn and Gibson and Pearlmutter therefore conclude that a probabilistic serial model cannot account for the results.

This conclusion rests on a null finding and is, to that extent, insecure. However, it is possible to refute it on more interesting grounds. A key assumption in the argument is that the ambiguity resolution process in the serial model must be sensitive to the SC preference of the noun, in order to account for the negative correlation between SC preference and effect of ambiguity at disambiguation. However, this might not be so. There is another alternative: this negative correlation may simply be due to the increased ease of interpreting the SC structure for those nouns with higher SC preference. Suppose, for example, that the SC interpretation was chosen 60% of the time, and the RC interpretation 40%. Then for 40% of the items, reanalysis is required to get the SC structure. This will produce an ambiguity effect at the disambiguating region. It seems quite natural to assume that the computation and interpretation of that SC structure will be easier for those nouns that occur more frequently with SC structures. In the unambiguous conditions, the cost of computing the SC structure has already been paid before encountering the embedded verb. Thus, we should expect a negative correlation of SC bias with ambiguity effect size at the disambiguating region. Crucially, there will be no correlation between SC bias and the plausibility effect size, since the ambiguity resolution did not depend on the SC bias.

WHERE TO LOOK NEXT: A CRITICAL DIFFERENCE BETWEEN SERIAL AND PARALLEL PARSING

In all previous explorations of serial and parallel parsing, a single ambiguity has been used, usually with just two structural alternatives. However, these are hardly the structures that will press parallel parsing to its limits. What is needed is a series of local ambiguities that can be stacked together to yield an ambiguous region with a large set of structural interpretations, *all* of which are easily recovered.

Multiple Unresolved Ambiguities

Let A1, A2, and A3 be three local ambiguities, each with two structural interpretations. Let D1, D2, and D3 be the disambiguating region for each. A string of the form A1 . . . D1 . . . A2 . . . D2 . . . A3 . . . D3 will have at most two structures possible at any point. However, a string of the form A1 . . . A2 . . . A3 . . . D3 . . . D2 . . . D1 will have *eight* possible structures, beginning at A3. This is not a concern for either serial or parallel models, if it turns out that only one or two of these structures is recoverable by the human parser. However, in the following example (4), it does seem that all eight structures are easily recovered. All of the continuations a–h seem to be readily accessible to the parser. This example strings together three familiar ambiguities: the subject/object ambiguity following NP/SC verbs (like *suspected*), the subject-of-small-clause/object ambiguity following *saw*, and the genitive/accusative ambiguity of *her*:

- (4) Mary suspected the students who saw her . . .
- a. . . yesterday.
 - b. . . jogging yesterday.
 - c. . . dogs yesterday.
 - d. . . dogs fighting yesterday.
 - e. . . yesterday were cheating on the exam.
 - f. . . jogging yesterday were cheating on the exam.
 - g. . . dogs yesterday were cheating on the exam.
 - h. . . dogs fighting yesterday were cheating on the exam.

No experiments on such structures have yet been conducted, but to intuition it is clear that none of these continuations give rise to strong garden-path effects. This suggests that (1) for a serial parser, they must all be either preferred or within the scope of the reanalysis mechanism and (2) for a parallel parser, all eight of the structures must be computed and carried forward simultaneously in the parse.

Note that the standard manipulation of removing the overt complementizer *that* after *suspected* no longer changes the number of possibilities from 1 to 2, but rather from 4 to 8. Crucially, some serial models, such as the standard garden-path model (Frazier, 1987), would actually predict a *decrease* in reading time in the ambiguous region as the number of structural alternatives increases dramatically. The reason is simple: the added alternatives in this case are simpler structures (attaching *the students* as the direct object of *suspected*, rather than subject of an incoming clause) and simpler structures are favored in the serial model and computed more rapidly. No existing parallel model makes such a prediction and indeed some make the opposite prediction: as the number of alternatives increases, reading times in the ambiguous region should increase (Just & Carpenter, 1992; MacDonald, Pearlmutter, & Seidenberg, 1994).

Could a parallel model be modified so that it could also predict a decrease in reading times in highly ambiguous regions? One possibility might be to posit that reading times are a function only of the most rapidly computed alternative. As soon as one alternative is computed, the eyes advance. This would be a kind of *race-based parallel model* that could mimic the predictions of a standard serial model. It is not yet clear how viable such a model is. If the parser is moving along at the pace of the most rapidly computed structure, then it must be moving along at a pace that is too fast for the alternative structures, so how could they be computed? This might not be a serious problem in the case of a single ambiguity in which all of the alternatives are relatively easy to compute—the time differential may be small enough that the alternatives could “catch up” before the disambiguation arises. However, in the case of more strongly dispreferred (i.e., slower) structures, such a parser seems at dire risk of falling further and further behind, particularly in the face of additional ambiguities.

SUMMARY AND CONCLUSIONS

1. It seems quite difficult, if not impossible, to distinguish empirically between serial and parallel models as broad classes—it is only possible to test very specific instantiations of these models. Probes in the ambiguous region offer the most direct tests of serial and parallel parsing.

2. The current evidence seems to rule out simple limited parallel, ranked parallel, and deterministic serial models. Ranked parallel models with a reanalysis component, and probabilistic serial models with a reanalysis component are viable at present. While “serial” and “parallel” have not been ruled out as broad classes, this narrowing down to quite specific subclasses represents significant progress.

3. A critical difference between serial and parallel parsing lies in how they handle multiple, unresolved local ambiguities. Intuitive evidence suggests that it is possible to create locally ambiguous regions with many (at least eight) distinct structural interpretations, all of which are easily recoverable. In some cases, serial models may predict decreases in reading time in the ambiguous region as the number of alternatives increases. No current parallel model makes this prediction.

REFERENCES

- Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research*, 23, 407–434.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading*. East Sussex, U.K.: Erlbaum.

- Frazier, L. (1998). Getting there . . . slowly. *Journal of Psycholinguistic Research*, 27(2), 123–146.
- Gibson, E. A. (1991). *A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown*. Unpublished Ph.D. dissertation, Carnegie Mellon.
- Gorrell, P. (1987). *Studies of Human Syntactic Processing: Ranked-Parallel Versus Serial Models*. Unpublished Ph.D. dissertation, The University of Connecticut, Storrs, Connecticut.
- Hickok, G. (1993). Parallel parsing: Evidence from reactivation in garden-path sentences. *Journal of Psycholinguistic Research*, 22, 239–250.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.
- Kurtzman, H. S. (1985). *Studies in Syntactic Ambiguity Resolution*. Unpublished Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts.
- MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, 24, 59–98.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- Nicol, J. L., & Pickering, M. J. (1993). Processing syntactically ambiguous sentences: Evidence from semantic priming. *Journal of Psycholinguistic Research*, 22, 207–237.
- Pearlmutter, N. J., & Mendelsohn, A. (1998, March). Serial versus parallel sentence processing. Paper presented at the The Eleventh Annual CUNY Sentence Processing Conference, Rutgers University, New Brunswick, NJ.
- Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24, 1521–1543.