Distinguishing Serial and Parallel Parsing

Edward Gibson[1,3] and Neal J. Pearlmutter[2,3]

This paper discusses ways of determining whether the human parser is serial maintaining at most, one structural interpretation at each parse state, or whether it is parallel, maintaining more than one structural interpretation in at least some circumstances. We make four points. The first two counterclaims made by Lewis (2000): (1) that the availability of alternative structures should not vary as a function of the disambiguating material in some ranked parallel models; and (2) that parallel models predict a slow down during the ambiguous region for more syntactically ambiguous structures. Our other points concern potential methods for seeking experimental evidence relevant to the serial/parallel question. We discuss effects of the plausibility of a secondary structure in the ambiguous region (Pearlmutter & Mendelsohn, 1999) and suggest examining the distribution of reaction times in the disambiguating region.

This paper addresses the question of determining how many structural interpretations the human parser retains during its normal first-pass operation: whether the parser is serial, maintaining at most one structural interpretation at each parse state, or whether it is parallel, maintaining more than one structural interpretation in some circumstances. It has long been known that the human parser does not retain all possible structural interpretations for an ambiguous input string in parallel, because of the existence of garden-path effects. Thus the question of serial vs. parallel processing reduces to whether or not there are some circumstances in which multiple structural interpretations are retained.

The serial/parallel processing dimension is orthogonal to a number of other dimensions of interest in the sentence-processing mechanism. Some

[1] Department of Brain and Cognitive Sciences, NE20-459, Massachussetts Institute of Technology Cambridge, Massachusetts 02139.

[2] Psychology Department, 125 NI, Northeastern University, Boston, Massachusetts 02115.

[3] Send correspondence to either author: Email: gibson@psyche.mit.edu, pearlmutter@neu.edu.

of these include (for relevant reviews, see Gibson & Pearlmutter, 1998; Tanenhaus & Trueswell, 1995): (1) Whether syntactic information is available in advance of, or simultaneously with, other information; (2) Which sources of information are used in determining structural preferences (e.g., syntactic information, lexical frequencies, semantic information, plausibility, and discourse); (3) Whether the parser uses the information available to it probabilistically or deterministically; (4) Whether reanalysis is repair-based or involves shifting to an alternative representation. In this paper, we first respond to two of Lewis' (2000) claims, and then we suggest two kinds of evidence relevant to the serial/parallel question.

RESPONSES TO LEWIS'S CLAIMS

Lewis raises two potential ways of distinguishing serial and parallel processing that we will discuss here. First, Lewis observes that differences in the type of disambiguating cue can affect the ease or difficulty of reanalysis (Fodor & Inoue, 1994). In (1), for example, the prepositional phrase "in the bowl" can attach as either the destination argument of the verb "put" or as a locative modifier of the NP "the strawberries":

(1) a. Mary put the strawberries in the bowl in the ice cream before dinner.
    b. Mary put the strawberries in the bowl into the ice cream before dinner.

The destination argument attachment is strongly preferred, but this turns out to be incorrect in both (1a) and (1b). In (1a), the PP "in the ice cream" cannot plausibly attach to the preceding syntactically available NP or VP and reanalysis is initiated. The PP "in the bowl" is reanalyzed as a modifier of the NP "the strawberries", so that the second PP can be successfully analyzed as a destination argument for "put." The same reanalysis is necessitated in (1b), but an additional cue to this reanalysis is available: the destination preposition "into," which cannot be a locative modifier. This additional cue to reanalysis makes the reanalysis intuitively easier, as verified in a self-paced reading experiment on items like (1a) and (1b) (Babyonyshev, Gibson & Kaan, in preparation).

Lewis observes that, in the parallel-processing models developed so far (e.g., Gibson, 1991; Gorrell, 1987; Kurtzman, 1985; MacDonald, Pearlmutter, & Seidenberg, 1994; Spivey & Tanenhaus, 1998), the type of disambiguating cue has not been hypothesized to affect the ease or difficulty of reanalysis. Then he claims that if it is found that the kind of disambiguating cue affects the difficulty of reanalysis (structure reranking), this would constitute evi-

dence against these parallel models. While it is true that previous parallel models have not proposed that the kind of disambiguating cue affects the difficulty of reanalysis, this is only because parallel models have not addressed this question to date, not because parallel models cannot account for such reanalysis effects. Like a complete serial model of sentence comprehension, a parallel model of sentence comprehension needs to include a theory of reanalysis. However, evidence gathered from varying the kinds of cues in the disambiguating region tells us nothing about the serial/parallel distinction. This type of manipulation will tell us how reanalysis/structure reranking occurs in either a serial or a parallel model, whichever turns out to be correct.

The second of Lewis's claims that we will address is that parallel models predict a slowdown during the ambiguous region for more ambiguous material. He provides the following locally ambiguous sentence initial fragment, which he hypothesizes could be used to test this claim:

(2) Mary suspected the students who saw her . . .

The verb "suspected" has two possible subcategorizations: an NP complement or an S complement. The verb "saw" also has two possible subcategorizations: an NP complement or VP small-clause complement. Finally, the pronoun "her" is ambiguous between genitive and accusative readings. Embedding the ambiguities as in (2) leads to eight distinct possible syntactic interpretations at the point of processing "her" in (2).

Lewis claims that serial and parallel approaches make distinguishable predictions with respect to the processing of highly ambiguous materials like (2). In particular, Lewis claims that a parallel model predicts a slowdown during the ambiguous region relative to an unambiguous control, whereas a serial model predicts no such slowdown. One way that a parallel model might make this prediction is if its processing speed is resource-based, such that the greater the number of structures that are retained, the smaller the quantity of working memory resources that is available for integrating new words into each, and, hence, the slower these integrations proceed. As the system reaches its resource limit, it slows down (cf. Just & Carpenter, 1992).

Although this is the prediction of one plausible parallel model, extending it to parallel models in general runs into at least three problems. First, even if processing speed is resource-based, Lewis's example may not require a lot of resources. A lot of structure may be shared in the parser's representation of this ambiguity (e.g., Earley, 1970; Pearlmutter & Mendelsohn (1999). Second, even if processing speed is resource-based, the control of alternatives in a parallel model may not be directly resource-based, but rather indirectly resource-based, such that perhaps only the best two or three structures are retained, or such that structures that are heuristically evaluated to be much worse than other co-extant structures are pruned from consideration (Gibson, 1991, 1998;

Jurafsky, 1996). In these cases, the parser may never be close to its resource limits, so it will process the ambiguous region quickly. Of course, if some structures are pruned from consideration along the way, then there will be ambiguity effects in the disambiguating regions for these continuations.

Another possibility is that the control of alternatives may be competition-based (Spivey & Tanenhaus, 1998; Stevenson, 1994) and there might not be much competition among the structures in Lewis's case. For example, competition might occur primarily between interpretations rather than syntactic structures, and in this case there is little or no interpretative conflict. In particular, the lexical NPs can be interpreted as the objects of the verbs ("the students" as object of "suspect"; "her" as object of "saw"), but these NPs cannot be interpreted as subjects of upcoming verbs until these verbs are processed.

In each of these cases, no slowdown is necessarily expected in the ambiguous region. Thus, a ranked parallel model does not necessarily predict increasingly slow reading times as the number of alternatives increases. To interpret the results in terms of the serial/parallel issue, additional assumptions are required about how processing speed and alternative interpretations are controlled.

EFFECTS OF AN UNPREFERRED INTERPRETATION IN THE AMBIGUOUS REGION

Although parallel models do not necessarily predict a slowdown during an ambiguity, even if multiple structures are being retained, there should be some way to measure effects of secondary structures in the ambiguous region. One way to look for evidence of the presence of a secondary interpretation is to manipulate properties of the secondary interpretation and look to see whether the manipulation affects comprehension performance (e.g., Gorrell, 1987; Kurtzman, 1985). Pearlmutter and Mendelsohn (1999) made use of such a design, manipulating the plausibility of a secondary interpretation. Relative implausibility increases processing difficulty, and thus if readers are considering a secondary interpretation, they should experience more difficulty when that interpretation is implausible than when it is plausible.

Pearlmutter and Mendelsohn constructed stimuli like (3), where the string "that the dictator described" is temporarily ambiguous between a full sentence complement (SC) interpretation (as in 3a) and a relative clause (RC) interpretation (3b). The presence or absence of the direct object ("the country") after the embedded verb "described" disambiguates. When both interpretations were plausible, as in (3), the SC interpretation created no difficulty compared to an unambiguous control, whereas the RC interpretation did. This difference indicates that the SC interpretation was preferred. Therefore,

sensitivity to the plausibility of the RC interpretation prior to disambiguation would indicate that readers were considering both interpretations.

(3) a. The report that the dictator described the country was not evaluated until later.
    b. The report that the dictator described was not evaluated until later.

To manipulate the plausibility of the secondary (RC) interpretation, Pearlmutter and Mendelsohn constructed stimuli like those in (3a), except that they varied the embedded verb as in (4), where "bombed" renders the RC interpretation implausible without affecting the plausibility of the SC interpretation (measured in separate ratings).

(4) The report that the dictator bombed the country was not evaluated until later.

The SC interpretation in (3a) and (4) is initially preferred, is plausible throughout, and is never incorrect. Therefore, a deterministic serial model, in which the parser always selects the same alternative for a given structural ambiguity, should never consider the RC interpretation and thus should show no sensitivity to its plausibility. However, Pearlmutter and Mendelsohn found that readers slowed down at "bombed" in (4) relative to its unambiguous control, indicating that the implausibility of the RC had affected processing. At "described" in (3a), on the other hand, readers showed no additional difficulty relative to the unambiguous control. Thus, in at least some cases, the parser must have computed the RC interpretation prior to disambiguation.

However, nondeterministic (probabilistic) serial models can still account for these results, because such models might compute the preferred (SC) interpretation most of the time, while instead computing the secondary (RC) interpretation on a minority of trials. In such models, the difficulty at the embedded verb in the implausible RC cases (4) would arise from this minority of trials. Ranked parallel models also account for these results, of course, because although they compute the preferred SC interpretation, they also compute the secondary RC interpretation and thus its plausibility can have an effect on processing.

To differentiate probabilistic serial models and ranked parallel models, Pearlmutter and Mendelsohn considered a further property of such models, which is how the relative preference for different interpretations varies across particular items. In probabilistic serial models, the alternatives are necessarily in complementary distribution because only a single interpretation is computed. Thus if, for example, the SC interpretation is very strongly preferred (very often computed) for a particular item, the RC interpretation must be very weak (very rarely computed) for that item.

For ranked parallel models, this prediction does not necessarily hold, depending on how competition between alternatives works. To the extent that the two alternatives compete, ranked parallel models will behave like probabilistic serial ones, because for items in which one interpretation is strongly supported, the other interpretation will be weakly supported. Thus Pearlmutter and Mendelsohn differentiated competitive and noncompetitive ranked parallel models. In the former case, a ranked parallel system will behave like a probabilistic serial one. In a noncompetitive ranked parallel model, however, where the strength of the two interpretations can vary independently, items with a strong SC interpretation (for example) may or may not have a strong RC interpretation.

To test these predictions concerning variation across items, Pearlmutter and Mendelsohn examined correlations between reading difficulty and a lexical property hypothesized to influence the relative preference for the ambiguity: argument structure frequency bias (e.g., MacDonald *et al.*, 1994; Trueswell, Tanenhaus, & Kello, 1993). The relevant argument structure frequency bias was SC preference, which was measured as each ambiguity initiating noun's relative preference for an SC versus an RC. Pearlmutter and Mendelsohn first showed that the noun's SC preference was negatively correlated across items with ambiguity effect size at the disambiguation, in the SC-disambiguated conditions where both interpretations were plausible (3a vs. its unambiguous control). This confirmed that SC preference predicts the strength of the preference for the (correct) SC interpretation, and it fits with most theories that allow for an influence of lexical properties on ambiguity resolution (e.g., Ferreira & Henderson, 1990; MacDonald *et al.*, 1994; Trueswell *et al.*, 1993).

Given that SC preference partially controls the strength of the SC interpretation at disambiguation, the critical question is whether it similarly controls the strength of the RC interpretation during the ambiguity. In both probabilistic serial and competitive ranked parallel models, SC preference should have a similar influence on the RC interpretation, because the two interpretations will be in complementary distribution. In a probabilistic serial model, a strong SC preference will lead to very frequent selection of the SC alternative and thus very rare selection of the RC alternative, so that the RC's implausibility will rarely be detected. Similarly, in a competitive-ranked parallel model, a strong SC preference will lead to strong support for the SC alternative and weak support for the RC alternative, so that the implausibility of the latter will again have little impact on processing.

In noncompetitive-ranked parallel models, however, a strongly supported SC alternative will not tend to weaken support for the RC alternative, and thus SC preference will not correlate with RC strength. To examine these predictions, Pearlmutter and Mendelsohn examined the correlation

between SC preference and ambiguity effect size at the embedded verb, in the conditions where the RC interpretation was implausible (4 and its unambiguous control). The embedded verb ambiguity effect should increase in size as the RC interpretation increases in strength across items, because a stronger RC interpretation will result in an increased effect of its implausibility. However, this correlation, unlike that between SC preference and ambiguity effect size at disambiguation in the plausible-RC conditions, was not statistically reliable.

This combination of reading time results and correlations thus argues against both deterministic and probabilistic serial models. The former cannot account for sensitivity to properties of a secondary interpretation for an ambiguity at all. The latter cannot account for the difference between (1) correlations at disambiguation, which illustrate the reliable effect of a lexical factor on the ease of handling the preferred interpretation when necessary, and (2) the lack of correlation during the ambiguity, which (in view of the proven sensitivity to the lexical factor) indicates that the strengths of the two interpretations must be independent, an impossibility in a serial model.

The correlational results, in particular, also argue against some ranked parallel models, namely, those in which the alternatives necessarily compete directly for resources, because in competitive parallel models as in serial models, the relative support for the two interpretations cannot be independent. Thus Pearlmutter and Mendelsohn argue that a critical factor in understanding ambiguity resolution is the *compatibility* of alternatives: To the extent that alternatives are compatible, they will not compete for resources, and the parser will be able to maintain multiple compatible interpretations for an ambiguity. They hypothesize that compatibility is a ratio of the amount of overlap between the representations for two interpretations, to the total content of the representations. Whether it is more appropriately measured in terms of syntactic representations or semantic/discourse representations is unclear. In the case of the SC versus RC ambiguity, the two interpretations overlap substantially in terms of both syntax (embedded clause subject, verb, most clausal head information, and a direct object position) and interpretation (word senses, subject–verb argument relationship). Prior to disambiguation, they differ only in the relationship between the matrix subject head ("report") and the embedded clause, which may involve no more than the addition of an operator position at the beginning of the clause.

Many other ambiguities in the literature will tend not to have compatible alternatives and thus are predicted to show stronger competition effects. For example, in the main verb versus reduced relative ambiguity, a variety of the verb's lexical properties (voice, morphological tense, and possibly argument structure; MacDonald *et al*., 1994) and argument relations vary between the two interpretations. There are also substantially larger differ-

ences between the syntactic structures involved in the two interpretations and between their discourse representations. In lexical–semantic ambiguities, too, although the alternative representations are less complex than those involved in syntactic ambiguities, the ratio of overlap is likely to be much smaller than in the SC versus RC case: The "financial institution" and "river's edge" meanings of "bank," for example, share very little in terms of semantics, and thus they will compete with each other.

THE DISTRIBUTION OF REACTION TIMES AT DISAMBIGUATION: BIMODAL OR UNIMODAL

Another potential method for distinguishing serial and parallel models of sentence comprehension is to examine the distribution of reading times at the disambiguation of an ambiguity, relative to the same location in an unambiguous control. Consider a temporary ambiguity with two possible structural interpretations, such as the SC/RC ambiguity. A serial model predicts a bimodal distribution of ambiguous condition reaction times: one mode corresponding to the processor getting the analysis right the first time and a second mode corresponding to reanalysis taking place. In contrast, if both interpretations of the ambiguity are carried in parallel, then a unimodal pattern of data is predicted. In particular, if the ambiguity is resolved toward the preferred reading, then a unimodal pattern of data is expected, with similar reaction times to the unambiguous control. On the other hand, if the ambiguity is resolved toward the less preferred reading, then another unimodal pattern of data is expected, one which is centered on a time greater than that of the unambiguous control.

Consider these predictions with respect to the SC/RC ambiguity. Pearlmutter and Mendelsohn observed a 30 ms ambiguity effect in the disambiguating region of the RC continuation (3b relative to its unambiguous control). Consider first the predictions of a serial model. Suppose that readers initially get the RC interpretation half the time, so that no reanalysis is necessary in these trials. In the other half of trials, readers initially follow the SC interpretation and reanalysis is necessary. Thus we would expect one mode corresponding to the initial correct analysis and a second mode 60 ms slower, corresponding to the reanalyzed cases. These average to give a 30 ms reanalysis effect. If the RC is pursued a smaller fraction of the time, then the second mode will be larger and closer. In contrast, a parallel model predicts a single mode, centered on a time 30 ms slower than the unambiguous control.

Although this type of analysis offers the potential to distinguish serial and parallel models, it presents several methodological difficulties. First, the larger the second mode is, the closer it is to the first mode, making the

two modes hard to differentiate. Alternatively, if the second mode is farther from the first mode, then it must be smaller and, hence, it is difficult to distinguish from the tail of the distribution. Second, a great deal of data is needed to accurately estimate modes: thousands of data points, not tens or hundreds as in typical sentence processing experiments. Furthermore, it may not be possible to collapse data across subjects and/or items. Third, the strength of the conclusions from this analysis may be somewhat limited. If a second mode is found, this only shows that processing is serial for the ambiguity being examined. Processing still might be parallel for other ambiguities. On the other hand, evidence supporting parallel processing would be the lack of a second mode, which is a null result. These concerns make it unclear whether a bimodality analysis can provide convincing evidence about the serial/parallel question, but a sufficiently large data set and a detailed theory about which ambiguities are more or less likely to be handled in parallel could make such an analysis feasible.

REFERENCES

Babyonyshev, M., Gibson, E., & Kaan, E. (in preparation). *Syntactic and non-syntactic cues to sentence reanalysis*. Manuscript, Massachusetts Institute of Technology.

Earley, J. (1970). An efficient context-free parsing algorithm. *Communications of the Association of Computing Machinery, 13,* 94–102.

Ferreira, F., & Henderson, J. M. (1990). The use of verb information in syntactic parsing: A comparison of evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16,* 555–568.

Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden-paths. *Journal of Psycholinguistic Research, 23,* 407–434.

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68,* 1–76.

Gibson, E., & Pearlmutter, N. (1998). Constraints on sentence comprehension. *Trends in Cognitive Science, 2,* 262–268.

Gorrell, P. G. (1987). *Studies of Human Syntactic Processing: Ranked-Parallel versus Serial Models*. Unpublished doctoral dissertation, University of Connecticut, Storrs, CT.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation: *Cognitive Science, 20,* 137–194.

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, MIT, Cambridge, MA.

Lewis, R. L. (2000). Serial and parallel parsing. *Journal of Psycholinguistic Research, 29,* 241-248.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703.

Pearlmutter, N. J., & Mendelsohn, A. (1999). Serial versus parallel sentence comprehension. Submitted manuscript.

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.

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research, 23,* 295–322.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Speech, language, and communication* (pp. 217–262). San Diego, CA: Academic Press.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 528–553.