Optimality Theory and Human Sentence Processing[*]

Edward Gibson and Kevin Broihier, Massachusetts Institute of Technology

This paper applies the idea of a winner-take-all constraint satisfaction system like that found in Optimality Theory to human sentence parsing. Three sets of constraints, each of which is derived from a parsing theory from the current literature, are attempted. The empirical data considered here do not seem to be consistent with any of these theories. It is argued that a cumulative constraint weighting system accounts for the data considered here better than a winner-take-all approach as is assumed in Optimality Theory.

1. Introduction

To account for grammaticality facts under Optimality Theory (Prince and Smolensky 1993), it is assumed that there is a finite set of linguistic constraints which can be ordered in a given language such that for a set of competitor representations for the same underlying input, only the highest-ranked representation is grammatical, according to a winner-take-all ranking function (cf. Pesetsky, this volume, who assumes the possibility of constraint ties). For any two input representations X and Y, the function returns the representation with the fewest violations of the highest-ranked constraint at which the two representations differ. In particular, if representations X and Y have the same number of violations of constraints C_1–C_{n-1}, and differ on constraint C_n such that X has *m* violations of C_n while Y has $m+k$ violations of constraint C_n, then X is ranked higher than Y. This function is winner-take-all in the sense that even if representation Y has many fewer violations than representation X of some lower ranked constraints, it is still outranked by X according to Optimality Theory. For example, a representation which has no violations of constraints C_1 through C_n, but ten violations of constraint C_{n+1}, outranks a representation which has no violations of constraints C_1 through C_{n-1}, a single violation of constraint C_n and no violations of constraint C_{n+1}.

[*] We would like to thank the conference participants for their comments on the talk on which this paper was based. We also thank Colin Phillips, Carson Schütze and Ken Wexler for comments on an earlier draft of this paper. The second author was supported by the Research Training Grant "Language: Acquisition and Computation" awarded by the National Science Foundation to the Massachusetts Institute of Technology (DIR 9113607).

MIT Press/MIT Working Papers in Linguistics, 000-000

The point of this paper is to examine whether the winner-take-all strategy in Optimality Theory can be applied to the resolution of ambiguity in human sentence comprehension.[1] During sentence comprehension, local ambiguities confront the listener/reader with a set of structural possibilities that need to be ranked so that only a small number of structures are retained from one parse state to the next. Here we consider the consequences of the possibility that the parser uses an Optimality Theoretic system of constraints to execute this ranking. The proposal is a natural one because of the clear structural parallels between the process of selecting a grammatical output from a set of candidate structures and the process of heuristically selecting a structural analysis to pursue when ambiguity arises during parsing. Since OT grammars typically select a single grammatical output from a candidate set[2], the analogy works best if one adopts a serial view of sentence processing. According to this view, when ambiguity arises, exactly one structure is initially selected by the human sentence processing mechanism (HSPM).

We evaluate this view of sentence processing with respect to three constraint sets from the current sentence processing literature. The first system consists of the constraints of Minimal Attachment and Late Closure from the *garden path* theory of sentence processing due to Frazier (1978, 1987a). The second constraint set consists of two constraints based on thematic role assignments and a third constraint which prefers local attachments (Gibson 1991, in press, cf. Pritchett 1988). The final set of constraints which we analyze consists of two constraints: locality of attachment and proximity to a verbal predicate (Gibson et al., in press).

2. Preliminaries

In our attempts to apply OT to sentence processing in this paper, we are working with a serial view of parsing because this seems to be the most natural approach to using an OT sentence processor to select which candidate structure(s) to continue. Making this commitment to serialism, however, raises some difficult issues that need to be resolved, but will not be resolved here.

Several classic "garden path" effects receive a natural account on the serial view: since the processor is forced to choose an analysis when ambiguity arises, the wrong choice can eventually lead to processing difficulty. However,

---

[1] Of course, the question of whether or not OT works out as a theory of parsing is largely independent of whether or not it works out elsewhere as a grammatical theory. This paper will have nothing to say about whether OT provides a viable account of phonology, syntax or any other linguistic domain.

[2] If two candidates incurred exactly the same set of constraint violations, an OT system would identify both of them as optimal. Typically though, constraint systems are sufficiently "fine-grained" that this logical possibility does not come up.

because language is very ambiguous, and many ambiguities are very easy to comprehend, we are going to need a theory of reanalysis that explains when initially choosing the wrong parse will lead to processing breakdown, as in the examples in (1), and when it can be easily recovered from, as in (2). Note that in at least one of the examples in (2), a serial parser must initially have the wrong structure after processing *trains*.

(1) a. #The horse raced past the barn fell.

b. # The dog walked to the park chewed the bone.

(2) a. The desert trains young people to be tough.

b. The desert trains are tough on young people.

(Frazier and Rayner 1987, MacDonald 1993)

Pritchett (1988, 1992), Fodor and Inoue (1994), Frazier (1994) and Lewis (1993) all attempt to provide theories of reanalysis within serial systems.

How the apparent non-modularity of the HSPM should be handled in a serial model is also an open question. In what follows, we're abstracting away from the effects of context, pragmatics and semantics, which arguably interact with structural factors immediately on-line (see among others, Crain and Steedman 1985; Altmann and Steedman 1988; Trueswell, Tanenhaus and Garnsey 1994; MacDonald, Pearlmutter and Seidenberg 1994; Spivey-Knowlton and Tanenhaus 1994; Spivey-Knowlton and Sedivy 1995). It seems that there are two approaches that an OT parsing theory could take towards this non-modularity. First, an effort might be made to provide an OT theory of non-linguistic constraints as well as linguistic constraints and to interleave the two sets of constraints in OT fashion. This approach strikes us as implausible. Second, one might hope that it will turn out to be the case that syntax does in fact apply first, in modular fashion, but that standard techniques aren't yet fine-grained enough to pick this up. This claim is made explicitly in Frazier and Clifton (in press). For the purposes of this paper, we will assume the second approach will prove to be correct. Thus, the OT sentence processors that we consider will select among structural candidates on entirely syntactic grounds.

3. Constraint Set 1: Minimal Attachment and Late Closure

The first constraints that we will evaluate here with respect to a winner-take-all Optimality theoretic approach are the principles underlying the *garden path* theory of sentence processing: the principles of Minimal Attachment and Late Closure (Frazier 1978; Frazier and Fodor 1978; Frazier and Rayner 1982; Frazier 1987a). The statement of Minimal Attachment is given in (3):

(3) *Minimal Attachment*
Attach incoming material into the phrase-marker being constructed using the fewest nodes consistent with the well-formedness rules of the language (Frazier and Rayner 1982).

Given a structure for the current input and an ambiguous attachment of structures for the next word in the input string, Minimal Attachment favors attachments that cause fewer new phrase structure nodes to be created. Late Closure applies in cases where Minimal Attachment does not decide among representations:

(4) *Late Closure*
When possible, attach incoming lexical items into the clause or phrase currently being processed, i.e., the lowest possible nonterminal node dominating the last item analyzed (Frazier and Rayner 1982).

Late Closure prefers attachment to structures associated with the more recently occurring words in the input string over attachments to structures associated with words further back. Minimal Attachment applies in conflicts between Minimal Attachment and Late Closure. Hence, this system can easily be translated into a simple Optimality theoretic winner-take-all system, with two ranked constraints. This system is simple not only because it has just two constraints, but also because there is only one ordering of the constraints across all languages: Minimal Attachment is assumed to dominate Late Closure as a cross-linguistic universal.

Both constraints are currently defined in relative terms, defining the relative goodness of one structure as compared with another. To have an OT system, we need to be able to compute the number of violations in a computation local to each structure. Hence we will re-define these constraints so that they apply to a single structure such that the relative ranking of two structures is derived from the fact that one of the structures violates the absolute constraint more often. The absolute correlate of Minimal Attachment is Node Conservativity:

(5) *Node Conservativity*
Don't create a phrase structure node.

For each possible attachment, a violation of Node Conservativity is incurred for each phrase structure node that is created at the current parse step in order to allow the attachment. Thus an attachment that requires three new nodes will incur three violations of Node Conservativity. The absolute correlate of Late Closure is Node Locality:

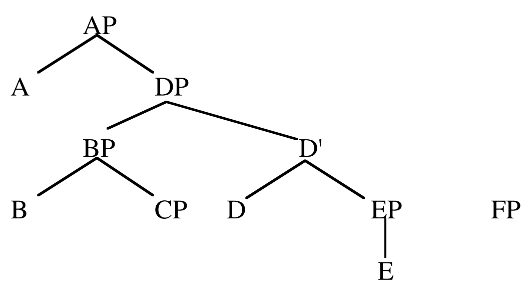
(6) *Node Locality*
Attach inside the most local maximal projection.

Non-local attachments are associated with violations of this constraint. One plausible definition of a violation of Node Locality is one which associates a Locality violation for each maximal projection on the right perimeter of the current structure (the prospective attachment sites) that is passed over in making an attachment:

(6') *Node Locality*
An attachment to structure XP at the node Y in XP is associated with one locality violation for each maximal projection on the right perimeter of XP that is dominated by Y.

Thus, attaching the current structure inside the most local maximal projection will never incur a Locality violation. Attaching to the next most local head on the right perimeter of the current structure will incur one violation, attaching to the second most recent head will incur two violations, and so on. For example, in (7), the attachment of FP below EP incurs no Locality violations:

(7)



Attaching immediately below D' adds one Locality violation, because the maximal projection EP is passed over. Attaching inside the least recent maximal projection AP incurs two violations because two maximal projections on the right perimeter of the structure are passed over to make this attachment.[3]

[3] There are other possible variations on what might count as a locality violation. For example, it would also be possible to count all maximal projections that are passed over in making an attachment, whether or not they are on the right perimeter of the structure. This would add two additional locality violations to the attachment of FP to AP, because two maximal projections down a left branch are also passed over in making this attachment. A second possibility would be to incur costs only for the maximal projections that grammatically allow attachment. This proposal would reduce the number of locality vio-

Here and throughout we are assuming that there is an inviolable constraint on parsing: words in the input can not be skipped. If this were not the case, a null structure with no nodes and no branches would always be preferred. Moreover, we are assuming, that the candidates that are evaluated at any point in the parse are grammatical structures.

The details of the phrase structure grammar are clearly crucial in determining the combined effects of Minimal Attachment and Late Closure. Frazier and Clifton (in press) assume a grammar in which there are no vacuous projections. For example, (8b) is generated, not (8a):

(8) a. [NP [N' [N John]]]

b. [NP John]

Ternary branching is also assumed to be permitted, as shown in (9):[4]

(9) [IP [NP Mary] [VP [V gave] [NP John] [NP a book]]]

The well-known "garden path" effects induced by sentences like (10a) and (10b) constitute a core case for the Minimal Attachment/Late Closure analysis. At the point that the first verb is encountered in these sentences—*raced* in (10a) and *walked* in (10b)—it could be analyzed as either the main verb of the matrix clause or, as turns out to be correct in this case, as part of a reduced relative clause.

(10) a. #The horse raced past the barn fell.

b. # The dog walked to the park chewed the bone.

---

lations incurred in attaching FP as an immediate daughter of AP in (7) if one of EP or DP do not allow grammatical attachment of FP.

Although the particular definition of locality will affect the number of violations incurred, the differences among definitions do not affect the evaluation of locality as a principle in a winner-take-all Optimality theoretic framework for the examples to be considered here. This is because the variations in the definition don't change the relative rankings of candidates in terms of the number of Locality violations they incur.

[4] In some sense, these structural assumptions are in keeping with the notion of Minimal Attachment. Not allowing vacuous projection results in fewer nodes (e.g. compare (8a) and (8b)). Similarly, the ternary branching structure for double object constructions shown in (9) contains fewer nodes than a possible alternative analysis which requires binary branching and the introduction of non-overt structural positions (e.g. Larson 1988).

At the point of processing *the horse*, the structure constructed for sentence (10a) is as shown in (11):[5]

(11) [IP [NP [Det the] [N horse]]]

The crucial ambiguity arises at the word *raced*. (12) contains the main clause analysis of *raced* and indicates the Conservativity and Locality violations that this analysis incurs:

(12) [IP [NP [Det the] [N horse]] [VP raced]]
1 new node: VP
1 Locality violation

The alternative reduced relative analysis and its performance with respect to the Conservativity and Locality constraints are shown in (13)[6]:

(13) [IP [NP [Det the] [N' [N horse] [CP [IP [VP raced]]]]]]
At least 4 new nodes: N', VP, CP, IP etc.
0 Locality violations

The reduced relative analysis in (13) incurs three more violations of Conservativity than the matrix clause analysis in (12). Since Conservativity is the highest-ranked constraint in the system, this predicts, correctly, that subjects will initially adopt the matrix clause analysis. Further downstream when the matrix clause analysis can't accommodate the "extra" verb, processing difficulty occurs. Crucially, Conservativity must outrank Locality because the reduced relative in (13) is actually the preferred structure with respect to the Locality constraint.

Experimental evidence about parsing preferences in cases where verbs subcategorize for either NP or CP complements is also successfully analyzed by the Minimal Attachment/Late Closure system, which we have translated into our Conservativity/Locality system. For example, the verb *know* can take either an NP argument, as in (14a), or a CP argument, as in (14b). The same is true for the verb *argued*, as shown in (14c) and (14d). When the first NP after the verb is

---

[5] It is assumed that the parser is top-down to some degree, predicting an IP from the presence of NP. This top-down component of the parsing algorithm is necessary to explain memory complexity effects (see e.g. Gibson 1991, in press; Babyonyshev and Gibson 1995) as well as certain ambiguity effects (see e.g., Frazier, 1987b; Bader and Lasser, 1994).

[6] As mentioned earlier, proponents of Minimal Attachment assume a phrase structure component in which there are no vacuous projections (Frazier 1990; Frazier and Clifton 1995). As a result, a modifier following a head which has no arguments is assumed to be attached as a sister to the head, even though this is not a theta-assignment position (see, e.g., (13). This assumption is made in order to account for English PP attachment facts, among other effects.

encountered, it is unclear whether it should be analyzed as the direct object of the main verb or as the subject of a CP argument of the verb. Although intuition is not overwhelming in this case, Frazier and Rayner (1982), Ferreira and Henderson (1991) provide experimental evidence that the CP continuations are slightly more complex (however, cf. Trueswell, Tanenhaus and Kello 1993). The additional complexity of the CP continuations makes sense in this system because the CP analysis requires the addition of several more nodes than the NP analysis.

(14) a. Bill knew John.

b. Bill knew John liked Mary.

c. The city council argued the mayor’s position forcefully.

d. The city council argued the mayor’s position was incorrect.

For example, consider the on-line parsing of either (14a) or (14b). After processing *Bill knew*, the structure is as shown in (15):

(15) [IP [NP Bill] [VP knew]]

The crucial ambiguity arises at *John*. The alternative structures, and their associated constraint violations, are shown in (16) and (17). Neither structure incurs any Locality violations, but the CP analysis in (17) is more costly in terms of Conservativity violations. For a serial OT parser which ranked Conservativity above Locality, the prediction is that the NP analysis in (16) would be adopted over the CP analysis in (17).

(16) [IP [NP Bill] [VP [V knew] [NP John]]]
2 new nodes: V, NP
0 Locality violations

(17) [IP [NP Bill] [VP [V knew] [CP [C e] [IP [NP John]]]]]
5 new nodes: V, CP, C, IP, NP
0 Locality violations

So far, Locality—our OT equivalent of Late Closure—has not played an active role in deciding between candidate parses. The reduced relative/main clause ambiguity in (10)-(13) provided an argument for ranking Conservativity above Locality. In the NP/CP complement ambiguity in (14)-(17), the alternative structures did not differ at all with respect to Locality. According to a theory in which Conservativity strictly dominates Locality, Locality’s effects should only be seen when two competing structures require the addition of exactly the same number of nodes. There are a number of such cases.

The first case that we'll consider involves the placement of temporal adverbs. The sentence in (18a) is globally ambiguous. On the interpretation where *yesterday* attaches more locally within the embedded clause, John left yesterday and Bill commented on John's leaving at some unspecified moment in the past. On the other hand, if *yesterday* modifies the main clause, the interpretation is that John left at some unspecified moment in the past and that Bill commented on John's leaving yesterday. Sentence (18b) provides another example of this global ambiguity. Although both readings are grammatically possible, there is a definite preference for modification of the embedded clause. The naturalness of this "downstairs" reading is highlighted by the examples in (18c) and (18d). In these examples the tense of the verb in the embedded clause is incompatible with the given temporal adverbs. For example, in (18c) since the event of John leaving has already happened, it makes no sense to modify the embedded clause with the temporal adverb *tomorrow*. This incompatibility between the temporal adverb and the embedded clause forces the temporal adverb to modify the main clause. As indicated in (18c) and (18d), when this "upstairs" reading is forced, the result is somewhat unacceptable.

(18) a. Bill said John left yesterday.

b. The Globe will predict that the Red Sox will win the World Series this year.

c. #Bill will say John left tomorrow.

d. # The Globe predicted that the Red Sox will win the series last year.

To see how Locality pushes parsing towards the "downstairs" reading, consider the processing of the example in (18a). After processing *Bill said John left*, the structure is as in (19):

(19) [IP [NP Bill] [VP [V said] [CP [C e] [IP [NP John] [VP left]]]]]

Both the "upstairs" and "downstairs" readings require an AdvP to host the temporal adverb. In both cases, adjunction of the temporal adverb also requires the creation of a new VP node. In terms of Conservativity, the structures are identical. In this case, then, Locality is the deciding factor. As indicated in (20), attachment within the embedded clause is the most local option available—no Locality violations are incurred. In contrast, in (21) attachment of the temporal adverb incurs three Locality violations because when the adverb adjoins to the main clause VP, the more local VP, IP and CP nodes of the embedded clause are all passed over.

(20) [IP [NP Bill] [VP [V said]
[CP [C e] [IP [NP John]
[VP [VP left] [AdvP yesterday]]]]]]
2 new nodes: AdvP, VP
0 Locality violations: attachment to the most local VP node
(VP "left")

(21) [IP [NP Bill] [VP [VP [V said] [CP [C e] [IP [NP John] [VP left]]]]
[AdvP yesterday]]]
2 new nodes: AdvP, VP
3 Locality violations: skipped a VP ("left"), IP, CP

The garden path sentences in (22) can also be explained purely in terms of Locality. In these sentences, the crucial ambiguity arises immediately after the verb in the modifying clause—*jogs* in (22a) and *ate* in (22b). These verbs are both optionally intransitive. The NPs that follow them could be interpreted as direct objects, but they could also be the subjects of the main clause. The preference for the direct object reading in (22) leads to later processing difficulty because it does not lead to a possible analysis for the structure as a whole.

(22) a. #Since she jogs a mile seems light work.

b. # While the cannibals ate missionaries drank.

Again, Locality is the key constraint. Consider the on-line analysis of the example in (22b). After processing *ate*, the structure is as in (23):

(23) [IP [CP [C While] [IP [NP the cannibals] [VP ate]]]]

At *missionaries*, two possibilities arise. On the low attachment reading shown in (24), *missionaries* is added as the direct object of the verb in the modifying clause. This, of course, requires adding an NP node. It also requires adding a V node since, because of the ban on vacuous projection, the VP *ate* did not have any internal structure up until this point. On the high attachment reading in (25), a new IP node is required in addition to the "mandatory" NP node.

(24) [IP [CP [C While] [IP [NP the cannibals]
[VP [V ate] [NP missionaries]]]]]
2 new nodes: V, NP
0 Locality violations

(25) [IP [CP [C While] [IP [NP the cannibals] [VP ate]]]
[IP [NP missionaries]]]
2 new nodes: IP, NP
3 Locality violations: VP, IP, CP

As in the case analyzed in (18)-(21), the competing analyses are identical with respect to Conservativity. With respect to Locality, however, the low attachment reading in (24) is a much better candidate. On this analysis, the new NP attaches to the most local available head, so there are no Locality violations. The high attachment reading, in contrast, requires the parser to attach the new IP node to the initial IP node, bypassing the intervening VP, IP and CP.

Proponents of the Minimal Attachment/Late Closure analysis, which we have straightforwardly recast in OT terms, provide a number of cases that argue for ranking Minimal Attachment above Late Closure. We have already seen one case above in the reduced relative/main clause ambiguity (10)-(13). There are a number of other such cases where the candidate that is better with respect to Minimal Attachment—our Conservativity—wins out, even though the other candidate is better with respect to Locality—our Late Closure.

Consider, for example, the garden path sentences in (26). The main clause verbs in these examples can select two arguments—an NP and a CP. In these examples, the key ambiguity arises when *that* is processed. At this point, it's possible that the NP argument is complete and that *that* is the beginning of the CP complement. However, it's also possible that *that* is the beginning of a relative clause modifying the NP argument.

(26) a. #The patient persuaded the doctor that he was having trouble with to leave.

b. # Dan convinced the child that the dog bit that cats are good pets.

For the sentences in (26), the relative clause analysis ultimately turns out to be the right one. The garden path effect can be explained within the current system, if the parser initially prefers the incorrect CP complement reading. Evaluation of the two candidate structures that arise when *that* is processed indicates that Conservativity will have to dominate Locality in order for this to happen. Immediately before processing the *that* in (26a), the parser's structure is as in (27):

(27) [IP [NP the patient] [VP [V persuaded] [NP [Det the] [N doctor]]]]

The CP complement reading of *that* is indicated in (28). The relative clause reading is indicated in (29).

(28) [IP [NP the patient] [VP [V persuaded] [NP [Det the] [N doctor]]
[CP that]]]
1 new node: CP
1 Locality violation: NP

(29) [IP [NP the patient] [VP [V persuaded]
[NP [Det the] [N' [N doctor] [CP that]]]]]
2 new nodes: CP, N'
0 Locality violations

The CP complement reading in (28) incurs a single Conservativity violation. It also incurs a single Locality violation because the CP attaches to VP rather than to the most local NP projection. Under the relative clause analysis in (29), on the other hand, the CP *does* attach to this closest NP projection. The structure in (29), therefore, incurs no Locality violations. If Locality dominated Conservativity, the parser would pursue this more optimal relative clause analysis. Since this relative clause analysis turns out to be the right analysis of (27), the prediction would be that the ambiguity at *that* does not result in processing difficulty. This prediction, however, is incorrect. The right prediction follows if we turn our attention to Conservativity. The structure in (29) incurs two Conservativity violations—one violation for the CP node and one violation for the N' node. The N' node is not added prior to processing *that* because of the ban on vacuous projection. If Conservativity dominates Locality, then the parser will initially go with the CP complement reading in (28). This will result, correctly, in eventual processing difficulty.

A similar argument/adjunct ambiguity provides an additional argument for ranking Conservativity above Locality. The main clause verbs in (30) select—obligatorily in the case of *put*, as one of several options for *gave*—for an NP theme followed by a PP goal.

(30) a. #I put the candy on the table into my mouth.

b. # Ron gave the letter to Nancy to the postman.

Ambiguity arises when the first preposition in these examples is processed because this preposition could be either the beginning of the PP argument or the beginning of a PP adjunct modifying the NP argument. In the examples in (30), the PP adjunct reading turns out to be the right choice. To get the garden path effect, though, we again need the parser to initially make the wrong choice. As in the example in (26)-(29), this will only happen if Conservativity dominates Locality.

Immediately before the first preposition, *on*, is processed in example (30a), the structure is as in (31):

(31) [IP [NP I] [VP [V put] [NP [Det the] [N candy]]]]

The PP complement reading, in (32), incurs one Conservativity violation because a PP node needs to be introduced and one Locality violation because the closest NP head is skipped in favor of attachment to the VP.

(32) [IP [NP I] [VP [V put] [NP [Det the] [N candy]] [PP on]]]
1 new node: PP
1 Locality violation: NP

The PP adjunct reading, in (33), does not incur any Locality violations. However, the adjunction to NP requires the introduction of an additional N' node.

(33) [IP [NP I] [VP [V put] [NP [Det the] [N' [N candy] [PP on]]]]]
2 new nodes: N', PP
0 Locality violations

With respect to Conservativity, then, (32) is the better structure. With respect to Locality, (33) is better. To capture the garden path effect, we need the parser to initially adopt the structure in (32). Therefore, Conservativity should outrank Locality.

Yet another argument for Conservativity outranking Locality can be constructed on the basis of the garden paths that arise from the second object/relative clause ambiguities in (34):

(34) a. #Arthur brought the dog the ball hit a bone.

b. # John gave the boy the cat scratched a dollar.

We have considered several examples for which it must be the case that Conservativity, our version of Minimal Attachment, dominates Locality, our version of Late Closure. The Minimal Attachment/Late Closure view and our OT restatement of this view require that this *always* be the case. However, there exist a number of cases for which it seems that Locality must dominate Conservativity. (Alternatively, one might consider introducing a high-ranking third constraint that was positively correlated with Locality in these cases, but, crucially, was negatively correlated with Locality in the cases discussed above.)

As a first problematic example, consider (35):

(35) a. John sent his company boxes for storage.

b. John sent his company boxes to the recycling plant.

c. I gave her earrings on her birthday.

d. I gave her earrings to Sally.

The strings *his company boxes* and *her earrings* allow two different analyses. On the “one-NP” reading of *his company boxes*, the NP *his company* is not an argument of the main clause verb. Instead it modifies *boxes*. Similarly, on the “one-NP” reading of *her earrings*, *her* is interpreted as possessive. (35b) and (35d) are examples of the “one-NP” reading. On the “two-NP” analysis, on the other hand, these strings are broken into two arguments of the main clause verb. In (35a), *his company* is the goal NP, while *boxes* is the head of the theme NP. In (35c), *her* is the goal NP, while *earrings* is the head of the theme NP.

Intuition suggests that the “one-NP” reading (e.g. (35a)) is easier to get than the corresponding “two-NP” reading (e.g. (35b)). Given the now-familiar logic of the serial view we’re adopting here, we want the parser to initially adopt the “one-NP” reading. If this can be made to work out, then the processing difficulty that arises in “two-NP” cases can be understood as a reflex of the need to perform some reanalysis. However, this can’t be made to work out with the current set of constraints.

The structure in (36) reflects the parse immediately after processing *company* in either (35a) or (35b):

(36) [IP [NP John] [VP [V sent] [NP [Det his] [N company]]]]

(37) [IP [NP John] [VP [V sent] [NP [Det his] [N company]] [NP boxes]]]
1 new node: NP
1 Locality violation: NP

(38) [IP [NP John] [VP [V sent] [NP [Det his] [N' [N company]
[N boxes]]]]]
2 new nodes: N', N
0 Locality violations

The key ambiguity arises at *boxes*. The “two-NP” candidate structure is shown in (37). The “one-NP” candidate structure is shown in (38). In the “two-NP” structure, *boxes* has attached to the main clause VP. This incurs a single Locality violation because the intervening NP node has been passed over. Only the minimal Conservativity violation is incurred; a single node needs to be created to accommodate the new noun. In the “one-NP” structure, there is no Locality violation because *boxes* adjoins within the local NP. This adjunction, however, requires the creation of structure that goes above and beyond the mandatory single Conservativity violation. If Conservativity dominates Locality the “two-NP” candidate should initially beat out the “one-NP” candidate. This, however, is contrary to fact. In this case, it seems we need Locality to dominate Conservativity.[7]

The contrast between (39a) and (39b) provides another case where we need the opposite ranking of Conservativity and Locality.

(39) a. While I talked with the woman John was ignoring at the party I came to like her.

b. While I talked with the woman John was ignoring me at the party as well as he could.

Immediately before *John* is processed, the structure is as shown in (40):

(40) [IP [CP [C While]
[IP [NP I]
[VP [V talked]
[PP [P with]
[NP [Det the] [N woman]]]]]]]

The question that arises when *John* is processed is whether *John* should attach "low" as the beginning of a relative clause modifying *the woman*, as in (41), or "high" as the subject of the main clause, as in (42). The "low attachment" structure, shown in (41), incurs no Locality violations, but creating and attaching the new relative clause requires quite a few new nodes. On the other hand, the only new nodes introduced in the "high attachment" structure are the obligatory NP node and an IP node to host the new sentence. Attaching to the top IP in this "high attachment" structure, however, requires five more local nodes to be skipped.

[7] The problems that the examples in (35) raise for the current system are actually even somewhat worse than we've been presenting so far. It's not at all clear on a serial view that there should even be an ambiguity at the point when *boxes* is processed. Notice that the "two-NP" structure in (37) is a straightforward extension of the structure in (36); the new NP *boxes* is simply attached to the VP. In the "one-NP" structure in (38), however, the noun *company*, which had previously headed an NP argument of the verb, has been shifted to a modifier position. Unless the reanalysis involved in the transition from (36) to (38), in some sense, comes for free, then the "two-NP" parse in (37) should be the only candidate parse at the point when *boxes* is processed. Again, this would appear to be contrary to fact, since the "one-NP" reading seems to be preferred.

(41) [IP [CP [C While]
[IP [NP I]
[VP [V talked]
[PP [P with]
[NP [Det the]
[N' [N woman] [CP [C e] [IP [NP John]]]]]]]]]
5 new nodes: N', CP, C, IP, NP
0 Locality violations

(42) [IP [CP [C While]
[IP [NP I]
[VP [V talked]
[PP [P with]
[NP [Det the] [N woman]]]]]]
[IP [NP John]]]
2 new nodes: IP, NP
5 Locality violations: NP, PP, VP, IP, CP

Since intuition suggests that the "low attachment" structure in (39a) is easier to process, we again need Locality to dominate Conservativity.

As a final example favoring Locality over Conservativity consider the examples from Phillips 1995 shown in (43):

(43) a. When Bill complained classes were boring they usually were.

b. # When Bill complained classes were annoying as a result.

Here, at the point *classes* is processed, it could attach as either a complement of *complained* or as the subject of the main clause. Intuition suggests that (43b), where *classes* attaches as the subject of the main clause, is more difficult. After processing *complained*, the structure is as in (44):

(44) [IP [CP [C When] [IP [NP Bill] [VP complained]]]]

The alternative structures and their corresponding violations are shown in (45) and (46).

(45) [IP [CP [C When] [IP [NP Bill] [VP [V complained]
[CP [C e] [IP [NP classes]]]]]]]
5 new nodes: V, CP, C, IP, NP
0 Locality violations

(46) [IP [CP [C When] [IP [NP Bill] [VP complained]]]
[IP [NP classes]]]
2 new nodes: IP, NP
3 Locality violations: VP, IP, CP

Once again, Locality needs to dominate Conservativity.

To sum up, then, it seems that our OT versions of Minimal Attachment and Late Closure can't be strictly ranked in the sense required by Optimality Theory. There are a number of cases that require the standardly proposed ranking of Minimal Attachment over Late Closure. However, there are also a number of cases that require the opposite ranking.

At this point, it is natural to ask whether the problems we have identified with the Minimal Attachment/Late Closure OT system can be solved by changing the way in which constraint violations are combined—say, for example, by moving to a system which assigned costs for constraint violations and additively combined them—or whether there is a more fundamental problem with the constraints. The contrast between the complement clause/relative clause CP attachment cases discussed in (26)–(28) and the "one-NP"/"two-NP" cases discussed in (35)–(38) indicates that there will be no coherent way of combining the constraints without some overhaul of assumptions about phrase structure. In (26)–(28), the standard argument is that the parser selects a candidate with one Conservativity violation and one Locality violation over a candidate with two violations of Conservativity. We've argued that in the case discussed in (35)–(38), the parser makes exactly the opposite choice. The "one-NP" reading with two violations of Conservativity is chosen over the "two-NP" reading with one violation of Conservativity and one violation of Locality. If this argument is correct, then the parser makes different decisions when faced with situations that, from the point of view of these constraints, are identical.

4. Constraint Set 2: Theta-violations and Recency Preference

One alternative to the phrase-structure-based theory of syntactic ambiguity resolution is the thematic-role-based theory of Gibson (1991, in press) (cf. Pritchett, 1988, 1992). Under the thematic-role-based theory, it is assumed that each partial structure that the human parser builds is associated with a short-term memory cost in terms of an abstract memory unit, a processing load unit (PLU). Unlike the Minimal Attachment/Late Closure system discussed above, constraints are not ranked under in winner-take-all fashion under the thematic-based parsing framework. Rather, constraints are associated with cost in terms of PLUs, so that the total cost associated with a structure resulting from an attachment is calculated across all constraints for a particular attachment. The costs of the competing structures are then compared to decide parsing preferences. The costs associated with structures are determined by a number of factors, including lexical, syntactic, semantic, pragmatic and contextual properties. The factors that are currently best understood in this framework are the thematic-role-based properties, which are derived from the θ-Criterion:

(47) *The θ-Criterion*
Each argument bears one and only one θ-role (thematic role) and each θ-role is assigned to one and only one argument (Chomsky 1981: 36).

The θ-Criterion is a constraint of grammar that rules out structures as ungrammatical. The idea behind the thematic-based processing theory is that there is a processing cost associated with *local* violations of the θ-Criterion. In particular, each argument that needs a θ-role needs to be kept in memory until a role-assigner is found, and each role that needs to be assigned has to be kept in memory until an argument for that role is encountered. Role assignment and reception therefore result in two processing properties of structure, each associated with a memory cost (definitions from Babyonyshev and Gibson 1995, Gibson in press; cf. Gibson 1991, Gibson, Hickok and Schütze 1994):

(48) *The Property of Thematic Reception*
Associate a load of x_{TR} PLUs of short term memory to the head of each chain that is in a position that can be associated with a thematic role, but which does not yet receive a thematic role.

(49) *The Property of Lexical Requirement*
Associate a load of x_{LR} PLUs of short term memory to each lexical requirement that is obligatory in the current structure, but is unsatisfied.

The underlying parser is assumed to have a predictive component, which hypothesizes nodes to the right of the lexical heads that are encountered in the input (see Gibson 1991, in press and Babyonyshev and Gibson 1995, for independent evidence for this assumption from arguments about computational load). In this framework, a lexical requirement is thematically unsatisfied when it consists of an hypothesized node which contains no thematic nodes (e.g., nodes having category N, V, or A) that are confirmed in the input. Furthermore, it is assumed that the local violations of each side of the θ-Criterion—assignment and reception—are equally costly, both requiring x_θ PLUs of memory. However, for the purposes of evaluating a winner-take-all constraint-ranking theory of sentence processing, we will consider the two constraints independently. The two constraints function straightforwardly as Optimality-style constraints whose violations can be calculated locally on a structure.

A third constraint is also necessary within this system: a *Recency Preference* property, which prefers attachments to structures associated with more recent words in the input over attachments to structures associated with words further back in the input. Recency Preference serves the same function as Late

Closure within the constraint set discussed in Section 3, preferring local attachments over attachments that pass over more recently built structures. The definition of Recency Preference from Gibson (1991) is given in (50):[8]

(50) *The Property of Recency Preference*
The load associated with the structure resulting from attachment to a (thematic, arbitrary) position = (number of more recent words that are also associated with a (thematic, arbitrary) attachment position) * x_{RP} PLUs.

Note that this definition of recency allows recency preferences to interact with thematic role preferences only as long as the most local (recent) attachment is a thematic (argument) attachment. However, if there is a thematic attachment and the most local attachment is not a thematic attachment, then recency does not count against the most local thematic attachment. Hence the statement of the Property of Recency Preference includes an implicit constraint-ranking: thematic attachments are preferred to non-thematic attachments. This constraint-ranking is difficult to implement within an OT system without creating spurious constraints which significantly overlap with the already existing thematic constraints. Furthermore, such a constraint-ranking is not in keeping with the general principles underlying the memory-cost framework, because cost-differentials between constraints already account for effects that constraint-rankings can explain. To the extent that two different kinds of mechanisms—constraint-ranking and constraint weights—are utilized within a parsing framework, that framework is dispreferred to a framework that only has one such mechanism.

We will therefore consider a different statement of Locality, one which does not implicitly interact with other required constraints. The definition that we will consider is the same as the one discussed in the previous section, (6'), with the added constraint that only those categories that participate in thematic role assignment (i.e., either assigning or receiving a thematic role) are counted as Locality violations:

(51) *Node Locality* (thematic version):
An attachment to structure XP at the node Y in XP is associated with one locality violation for each thematic maximal projection (either receiving or assigning a thematic role) on the right perimeter of XP that is dominated by Y.

---

[8] In fact, the definition in (50) is not verbatim from Gibson (1991): It has the same content, but it has been reworded slightly so that it makes more sense out of the context of Gibson (1991).

The motivation for this change in the definition of Locality violation comes from the fact that thematic role assignment is assumed to be so crucial to sentence processing under this theory. The advantage of the proposed definition appears when the parsing system is treated as a weighted-constraint system as opposed to an OT system.[9] (See Section 5 below for a discussion of these effects.) The definition change has no empirical impact on the examples to be considered here when this constraint system is a evaluated as a winner-take-all Optimality theoretic system. Similarly, there is no empirical difference between the proposed OT system and one which includes recency constraints more directly reflecting the definition in (50).

We will now evaluate the constraints set consisting of Thematic Reception, Lexical requirement and Node Locality as a winner-take-all Optimality system. First, let us consider examples that are taken as evidence for the existence of thematic-role-based properties. Similar to the Minimal Attachment effects discussed in the previous section, the thematic constraints must dominate the Locality constraint in order to account for a number of preferences. For example, consider once again the complement clause/relative clause ambiguity which occurs in (26a), repeated here as (52), when the complementizer *that* is initially encountered:

(52) #The patient persuaded the doctor that he was having trouble with to leave.

After processing the word *doctor* in (52), the human parser has arrived at the structure in (27), repeated here as (53):

(53) [IP [NP the patient] [VP [V persuaded] [NP [Det the] [N doctor]]]]

There are now two possible attachments of the complementizer *that*:

(54) *VP Argument attachment*

[IP [NP the patient] [VP [V persuaded] [NP [Det the] [N doctor]]
[CP that]]]
1 Lexical Requirement violation:
the IP complement of the complementizer *that*
0 Thematic Reception violations
1 Locality violation: NP

---

[9] There is also evidence within the weighted-constraint framework that there is a decay function associated with locality violations. That is, although cost increases with distance from the most local site, the function assigning cost is not linearly increasing with distance (Gibson, Pearlmutter, Canseco-Gonzalez and Hickok, in press). The existence of such a function is not relevant to a winner-take-all OT constraint system, so we will ignore it at this point. See Sections 5 and 6 for more about this function.

(55) *NP modifier attachment*

[IP [NP the patient] [VP [V persuaded]
[NP [Det the] [N' [N doctor$_i$]
[CP [NP Op$_i$] that]]]]]
2 Lexical Requirement violations:
the IP complement of the complementizer *that* and
the CP complement of the verb *persuaded*
1 Thematic Reception violation:
The non-lexical operator Op$_i$ requires a thematic role
0 Locality violations

Although Locality favors structure (55), both of the thematic preferences favor structure (54). Since the structure that people tend to follow in such an ambiguity is structure (54), at least one of the thematic properties must outrank Locality.[10] In fact, it is easy to show that both must outrank Locality. The processing of the PP attachment ambiguities from (30), repeated here as (56), provides evidence that Lexical Requirement must outrank Locality:

(56) a. #I put the candy on the table into my mouth.

b. # Ron gave the letter to Nancy to the postman.

At the point of processing the preposition *on* in (56a) there are two possible attachments, represented in (57) and (58):

(57) [IP [NP I] [VP [V put] [NP [Det the] [N candy]] [PP on]]]
1 Lexical Requirement violation: the NP complement of the P *on*
0 Thematic Reception violations
1 Locality violation: NP

(58) [IP [NP I] [VP [V put] [NP [Det the] [N' [N candy] [PP on]]]]]
2 Lexical Requirement violations:
the NP complement of the P *on* and
the PP complement of the verb *put*
0 Thematic Reception violations
0 Locality violations

---

[10] Under the weighted-constraint treatment of these constraints, a fourth property applies here as well: Predicate Proximity, which favors attachments to verbal predicates (see Section 5 for discussion of the motivation and effects of this preference property). This will favor high attachment in this and some of the following examples also. Some of the evidence discussed in Section 5 demonstrates that Locality must outrank Predicate Proximity in English, so we can ignore this constraint when evaluating a winner-take-all OT system on examples in which higher ranked constraints conflict.

Node Locality and Lexical Requirement conflict on these two representations, but Thematic Reception does not. Structure (57) is preferred if Lexical Requirement is the higher ranked constraint, while (58) is preferred if Locality is the higher ranked constraint. Because people prefer the VP attachment, structure (57), Lexical Requirement must outrank Locality in an OT treatment of these constraints. Similarly, it is easy to show that Thematic Reception also outranks Locality. Consider once again the second object, relative clause ambiguity in (59) (repeated from (34)):

(59) #Arthur brought the dog the ball hit a bone.

There are two possible attachments of the NP *the ball*, represented in (60) and (61):

(60) [IP [NP Arthur] [VP [V brought] [NP the dog] [NP the ball]]]
0 Lexical Requirement violations
0 Thematic Reception violations
1 Locality violation: NP

(61) [IP [NP Arthur] [VP [V brought]
[NP [Det the]
[N' [N' dog_i]
[CP [NP Op_i] [IP [NP the ball]]]]]]]
0 Lexical Requirement violations
2 Thematic Reception violations:
The NPs *ball* and the non-lexical operator Op_i require thematic roles.
0 Locality violations: NP

In this ambiguity, Thematic Reception and Node Locality conflict, while Lexical Requirement remains neutral. If Thematic Reception is ranked higher than Locality, then structure (60) is preferred, which is the desired result. Thus Thematic Reception must outrank Locality.

Locality is of course still necessary in this OT system, because there are numerous ambiguities for which Thematic Reception and Lexical Requirement are matched, and Locality gives the right preference. For example, consider the adverbial attachment example in (18a) once more:

(62) Bill said John left yesterday.

At the point of attaching the adverb *yesterday*, there are two possible attachments:

(63) [IP [NP Bill] [VP [V said]
[CP [C e] [IP [NP John]
[VP [VP left] [AdvP yesterday]]]]]]
0 Lexical Requirement violations
0 Thematic Reception violations
0 Locality violations: attaching to most local node (VP *left*)

(64) [IP [NP Bill] [VP [VP [V said] [CP [C e] [IP [NP John] [VP left]]]]
[AdvP yesterday]]]
0 Lexical Requirement violations
0 Thematic Reception violations
2 Locality violations:
The thematic assigning category VP headed by *left*; and the CP receiving a thematic role from *said*

People strongly prefer the local attachment of *yesterday*, structure (63). This is accounted for with the existence of the Locality constraint.

This constraint system makes the same predictions as the Minimal Attachment/Late Closure constraint system on the examples discussed thus far. The thematic-based constraint system also explains the preference in (65), which was a problem for the Minimal Attachment / Late Closure constraint system:

(65) a. When Bill complained classes were boring they usually were.

b. # When Bill complained classes were annoying as a result.

There are two possible attachments of the word *classes* in these examples, resulting in two possible structures:

(66) [IP [CP [C When] [IP [NP Bill] [VP [V complained]
[CP [C e] [IP [NP classes]]]]]]]
0 Lexical Requirement violations[11]
1 Thematic Reception violation: the NP *classes* needs a role
0 Locality violations

[11] The lexical requirements of the verb *complained* are satisfied by the CP headed by the non-lexical complementizer, which is confirmed in the input because there is lexical material to its right. Furthermore, the requirements of the non-lexical complementizer are satisfied by the presence of the NP headed by *classes* (a thematic category) which is immediately dominated by an X-bar projection of the predicted category. Thus this structure has no lexical requirement violations.

(67) [IP [CP [C When] [IP [NP Bill] [VP complained]]] [IP [NP classes]]]
0 Lexical Requirement violations
1 Thematic Reception violation: the NP *classes* needs a role
1 Locality violation: VP

The two representations each have the same number of Lexical Requirement and Thematic Reception violations, so Node Locality decides in favor of structure (66), which is the desired prediction. This prediction contrasts with the Minimal Attachment constraint set because although the two representations do not differ in the number of thematic violations each has, (66) has more new nodes than (67). The thematic-role-based constraint set is therefore the preferred constraint system in this case.

Although the thematic-role-based constraint set accounts for the preference in (65), it does not account for the other examples that were problematic for the Minimal Attachment/Late Closure constraint system. First, consider once again the sentence pair in (35), repeated as (68):

(68) a. John sent his company boxes for storage.

b. John sent his company boxes to the recycling plant.

The noun *boxes* can attach in two different locations, resulting in the following two structures:

(69) [IP [NP John] [VP [V sent] [NP [Det his] [N company]] [NP boxes]]]
0 Lexical Requirement violations
0 Thematic Reception violations
1 Locality violation: NP

(70) [IP [NP John] [VP [V sent] [NP [Det his]
[N' [N company] [N boxes]]]]]
1 Lexical Requirement violations: the lexical requirement of *sent*
0 Thematic Reception violations
0 Locality violations

As discussed earlier, structure (69) is intuitively the preferred structure for this and other similar examples. But we have already seen that Lexical Requirement violations must outrank Locality violations within this constraint set, so we have arrived at a contradiction.

A similar problem arises for the processing of the examples in (39), repeated here as (71):

(71) a. While I talked with the woman John was ignoring at the party I came to like her.

b. While I talked with the woman John was ignoring me at the party as well as he could.

After processing the noun *woman* in these examples, there are two possible structures:

(72) [IP [CP [C While]
[IP [NP I]
[VP [V talked]
[PP [P with]
[NP [Det the]
[N' [N woman$_i$]
[CP [NP Op$_i$] [C e] [IP [NP John]]]]]]]]]
0 Lexical Requirement violations
2 Thematic Reception violations:
The NP headed by *John*, and the non-lexical operator Op$_i$
0 Locality violations.

(73) [IP [CP [C While] [IP [NP I] [VP [V talked]
[PP [P with]
[NP [Det the] [N woman]]]]]]
[IP [NP John]]]
0 Lexical Requirement violations
1 Thematic Reception violation: The NP headed by *John*
2 Locality violations: NP, VP

Intuitions suggest that the local attachment, the relative clause structure (72), is preferred to the non-local attachment, in spite of the fact that the non-local attachment has fewer Thematic Reception violations. For this comparison to work, Locality will have to outrank Thematic Reception, which is inconsistent with earlier requirements. We therefore run into similar difficulties in analyzing this set of constraints as a constraint ordering system as were encountered in the analysis of the Minimal Attachment/Late Closure system.

5. Constraint Set 3: Cross-linguistic differences in recency preference

The discussion of the constraints in the earlier sections of this paper has been restricted to a single language, so that only one (possibly universal) ranking of each of the constraint sets has been investigated. Although it is possible that the human sentence processing mechanism is universal cross-linguistically with no parameterizations, it is also possible that there are differences between languages in how the parser operates. To the extent that such differences exist, a winner-take-all Optimality Theoretic parsing system will implement such differences in terms of different parsing constraint rankings across languages.

The best-documented example of a cross-linguistic difference in parsing preferences involves relative clause attachment to two prospective NP heads, as in the English example (74a) and the Spanish example (74b) (Cuetos and Mitchell 1988; Mitchell and Cuetos 1991):

(74) a. The journalist interviewed the daughter of the colonel
[CP who had had the accident]

b. El periodista entrevisto a la hija del coronel
[CP que tuvo el accidente]

In English, there is a preference to attach the relative clause *who had had the accident* to the most local NP site *the colonel*, whereas the preference in Spanish is for attachment to the less local site *la hija* 'the daughter' (Cuetos and Mitchell, 1988; Mitchell and Cuetos, 1991; Clifton, 1988; cf. Gilboy et al., 1995). In order to account for this cross-linguistic preference difference within an Optimality theoretic framework, it might be proposed that there are two constraints which are ranked differently across the two languages: one favoring local attachment (e.g., Locality) and another favoring non-local attachment. This is very much like a proposal of Gibson, Pearlmutter, Canseco-Gonzalez and Hickok (in press), who suggest that the local preference factor is Recency (or Locality) preference and the non-local preference constraint is Predicate Proximity, which favors attachments to heads which are structurally closer to the head of a predicate phrase (e.g., verb phrase):

(75) *Predicate Proximity*
Attach as close as possible to the head of the predicate.

Gibson et al.'s proposal is to treat Recency Preference and Predicate Proximity as constraints associated with weights, so that preferences are determined by summing the combinations of these weights for a particular structure. When considering only preference differences like those in (74), this approach is also consistent with an Optimality theoretic winner-take-all constraint ranking system. In English, Recency Preference dominates Predicate Proximity to result in the low (local) attachment preference in (74a). In Spanish, Predicate Proximity dominates Recency preference to result in the high (non-local) attachment preference in (74b).

However, unlike a winner-take-all constraint ranking, the constraint-cost framework allows the possibility that multiple violations of low-ranked constraints might outweigh the effects of a single higher-ranked constraint. This is not possible under a winner-take-all constraint ranking approach. A winner-take-all approach predicts that, if two structures for the same input differ on the highest ranked constraint, then this difference should be sufficient to determine the preferred structure of the two choices, independent of other properties of these structures. It should not matter if there are large differences between the two structures with respect to a lower ranked constraint. Hence a winner-take-all approach to the Spanish preferences predicts that large differences in the number of Locality violations should not be enough to override differences in the number of predicate proximity violations. In order to investigate this prediction, Gibson et al. examined relative clause attachment ambiguities involving three NP attachment sites in Spanish like the ambiguity in (76):

(76) Pedro encendió
[NP_1 la lámpara cerca de [NP_2 la pintura de [NP_3 la casa]]]
[CP que fue dañada en la inundación]
'Pedro turned on the lamp near the painting of the house that was damaged in the flood.'

The relative clause *que fue dañada en la inundación* 'that was damaged in the flood' can grammatically attach to any of the three preceding NP sites. Predicate Proximity favors attachment to NP_1 because this site is structurally closest to a verbal predicate. Attachment to NP_3 is associated with the most Predicate Proximity violations, because this site is furthest from the predicate. Depending on how Predicate Proximity violations are counted, attachment to NP_3 could be a single violation or multiple violations. On the other hand, Locality favors attachment to the lowest, most recent site, NP_3, and least favors the highest site NP_1. Depending on how Locality is defined, there are at least two Locality violations in attachment to NP_1, one for each of the prospective attachment sites passed over (NP_3 and NP_2), and possibly many more violations, if all maximal projections that are passed over also count. Because 1) the preference in two-site ambiguities favors the less recent site in Spanish and 2) the same set of constraints are involved in processing three-site ambiguities, a winner-take-all approach therefore predicts that the less recent site preference will remain in a three-site ambiguity.

To determine the preferences among the three sites, Gibson et al. measured reading times and gathered on-line grammaticality judgments on disambiguated versions of attachments to all three prospective sites in examples closely related to (76), using number agreement to disambiguate[12]:

[12] Unlike (76), the items that Gibson et al. tested were NPs without an initial subject and verb (e.g., items like (76) without the initial words *Pedro encendió* 'Pedro turned on'. Intuitions suggest that the preferences for two- and three-site ambiguities are the same whether or not an initial subject and verb are present, so this difference should not affect the experimental results.

The fact that the preferences do not seem to be affected by the location of the

(77) a. *Low*
[NP_1 las lámparas cerca de [NP_2 las pinturas de
[NP_3 la casa] [$_{CP}$ que fue dañada en la inundación]]]

b. *Middle*
[NP_1 las lámparas cerca de [NP_2 la pintura de
[NP_3 las casas]] [CP que fue dañada en la inundación]]

c. *High*
[NP_1 la lámpara cerca de [NP_2 las pinturas de
[NP_3 las casas]]] [CP que fue dañada en la inundación]

'the lamp(s) near the painting(s) of the house(s) that was damaged in the flood'

The results of this experiment were that low attachments (as in (77a)) were read fastest, followed by high attachments (as in (77c)), with middle attachments the slowest. The same preference ordering (Low, High, Middle) was also observed in the on-line grammaticality judgment data. Hence, contrary to the prediction of the winner-take-all constraint-ranking, low attachments are preferred to high attachments in the three-site attachments. This indicates that a winner-take-all system is probably not being applied in the resolutions of these ambiguities.

On the basis of these and other data, Gibson et al. argue that the parser considers multiple weighted constraints at the same time in the process of ambiguity resolution. The two constraints that are necessary to account for the relative clause attachment preferences (and modifier attachments more generally) are Recency Preference and Predicate Proximity. To explain the difference in preferences across English and Spanish with respect to two-site ambiguities, it is assumed that the cost associated with violating Predicate Proximity is variable across languages. (See Gibson et al. for the motivation for this claim.) Furthermore, it is assumed that the cost associated with violating Predicate Proximity is uniform for all attachments to sites that are not maximally proximate to the predicate, while the cost associated with attaching non-locally increases with distance, according to an increasing decay function motivated from the short

---

initial NP—either before or following a θ-role assigning predicate—means that Predicate Proximity must be defined in such a way that it applies before the predicate has been identified in the input. In order to implement this, Gibson et al. assume that the human parser is partially top-down, predicting categories to the right before they have been identified. (For a summary of independent evidence for this assumption from arguments about computational load, see Gibson (1991, in press) and Babyonyshev and Gibson 1995.) In particular, it is assumed that an NP always causes the prediction of an IP, which includes a predicate phrase. Thus attachment to the initial NP is preferred by Predicate Proximity even before the content of the predicate has been identified.

term memory literature. These assumptions allow 1) an explanation of the switch in preferences between high and low sites in Spanish two- and three-site ambiguities and 2) an explanation for the fact that the preference ordering among two-NP-site ambiguities differs between English and Spanish, but is the same for three-NP-site ambiguities. See Gibson et al. (in press) for more details.

6. A weighted constraint theory of processing

Unlike a winner-take-all constraint ranking approach, the constraint-cost framework allows the possibility that multiple violations of low-weighted constraints might outweigh the effects of a single higher-ranked constraint. It turns out that if costs are associated with the constraints discussed here along the lines in Gibson (1991), Gibson et al. (1994), then a consistent theory of the highest ranked structure in syntactic ambiguities of the sort discussed in Sections 2 and 3 also results. In particular, if the Thematic Reception and Lexical Requirement properties are associated with one processing load unit (PLU) each, and Locality is associated with slightly more cost than this, say 1.5 PLUs, then many of the desired patterns of preferences are obtained. As mentioned in footnote 9, one additional property is also required: Predicate Proximity, which prefers attachments to be as close to verbal predicates as possible. As will be discussed in more depth in the following section, a Predicate Proximity violation is assumed to be slightly less costly than a Locality violation in English. For our purposes here, let us assume that a Predicate Proximity violation is associated with one PLU.

Given these four constraints, the preference for the argument attachment of CP headed by *that* in (26a) (repeated here as (78), and reanalyzed in (79) and (80)) is accounted for by comparing the relative costs at the ambiguity:

(78) #The patient persuaded the doctor that he was having trouble with to leave.

(79) VP Argument attachment:
[IP [NP the patient]
[VP [V persuaded] [NP [Det the] [N doctor]] [CP that]]]
1 Lexical Requirement violation:
the IP complement of the complementizer *that*
0 Thematic Reception violations
1 Locality violation: NP
0 Predicate Proximity violations:
Attaching to as close a predicate as possible

(80) NP modifier attachment:
[IP [NP the patient] [VP [V persuaded]
[NP [Det the]
[N' [N doctor$_i$]
[CP [NP Op$_i$] that]]]]]
2 Lexical Requirement violations:
the IP complement of the Comp *that* and the CP complement of the verb *persuaded*
1 Thematic Reception violation:
The non-lexical operator Op$_i$ requires a thematic role
0 Locality violations
1 Predicate Proximity violation:
Attaching to one XP removed from the predicate.

The total cost associated with structure (79) is 2.5 PLUs, while the cost associated with structure (80) is 4 PLUs. Structure (79) is therefore preferred.

Similar explanations of the other examples discussed here so far also apply. Furthermore, this theory also allows explanations of the problematic examples. Consider first the ambiguity (43), repeated here as (81), which was a problem for the Minimal Attachment / Late Closure constraint system:

(81) a. When Bill complained classes were boring they usually were.

b. # When Bill complained classes were annoying as a result.

The two possible structures at the point of processing *classes* are repeated, with a new cost analysis, as follows:

(82) [IP [CP [C When] [IP [NP Bill] [VP [V complained]
[CP [C e] [IP [NP classes]]]]]]]
0 Lexical Requirement violations
1 Thematic Reception violation: the NP *classes* needs a role
0 Locality violations
0 Predicate Proximity violations

(83) [IP [CP [C When] [IP [NP Bill] [VP complained]]] [IP [NP classes]]]
0 Lexical Requirement violations
1 Thematic Reception violation: the NP *classes* needs a role
1 Locality violation: VP
0 Predicate Proximity violations

The total cost associated with structure (82) is 1 PLU, while the cost associated with structure (83) is 2.5 PLUs. Structure (82) is therefore preferred, as desired.

Consider now the problematic sentence pair in (35), repeated here as (84) once again, along with the two structures at the ambiguous attachment point:

(84) a. John sent his company boxes for storage.

b. John sent his company boxes to the recycling plant.

(85) [IP [NP John] [VP [V sent] [NP [Det his] [N company]] [NP boxes]]]
0 Lexical Requirement violations
0 Thematic Reception violations
1 Locality violation: NP
0 Predicate Proximity violations

(86) [IP [NP John] [VP [V sent]
[NP [Det his] [N' [N company] [N boxes]]]]]
1 Lexical Requirement violations: the lexical requirement of *sent*
0 Thematic Reception violations
0 Locality violations
0 Predicate Proximity violations

The total cost associated with structure (85) is 1.5 PLUs, while the cost associated with structure (86) is 1 PLU. Structure (85) is therefore slightly preferred, as desired.

Consider now the final problematic sentence pair in (39), repeated here as (87), once again, along with the two structures at the ambiguous attachment.

(87) a. While I talked with the woman John was ignoring at the party I came to like her.

b. While I talked with the woman John was ignoring me at the party as well as he could.

(88) [IP [CP [C While]
[IP [NP I]
[VP [V talked]
[PP [P with]
[NP [Det the]
[N' [N woman$_i$]
[CP [NP Op$_i$] [C e] [IP [NP John]]]]]]]]]]
0 Lexical Requirement violations
2 Thematic Reception violations:
The NP headed by *John* and the non-lexical operator Op$_i$
0 Locality violations
0 Predicate Proximity violations

(89) [IP [CP [C While] [IP [NP I] [VP [V talked]
[PP [P with]
[NP [Det the] [N woman]]]]]]
[IP [NP John]]]
0 Lexical Requirement violations
1 Thematic Reception violation: The NP headed by *John*
2 Locality violations: NP, VP
0 Predicate Proximity violations

The total cost associated with structure (88) is 2 PLUs, while the cost associated with structure (89) is 4 PLUs. Structure (88) is therefore preferred, as desired.[13]

Thus the weighted-constraint theory offers plausible explanations for parse preference differences that are difficult to handle in a constraint-ranking approach. The weighted-constraint approach also has the advantage of being easily compatible with the non-modularity of syntactic processing. Under this approach it is plausible to assume that constraints from multiple information sources are associated with processing cost in a central processing system. Furthermore, although a constraint-ranking approach is more easily reconcilable with a serial approach, the weighted-constraint approach is easily compatible with either parallel or serial processing. So to the extent that there is evidence for parallel processing of certain structures, this can be readily accommodated with a weighted-constraint system. In fact, in the weighted-constraint frame-

[13] In fact, there is also evidence that all locality violations are not equally costly. In particular, there is evidence that preference for locality of attachment is associated with a decay function according to which less recent sites incur less additional cost than more recent sites. (See the previous section for evidence for this decay function.) For example, one candidate for such a decay function would be one that incurs half as much additional cost for each additional Locality violation. This would result in a total cost of 3.25 PLUs for structure (89) rather than 4 PLUs.

Although not crucial to explain the initial preference facts discussed here, this function is necessary to explain the locality effects in the previous section, among others.

work of Gibson (1991, in press), it is assumed that the underlying parser is parallel.

Two additional advantages of the weighted-constraint approach over the constraint-ranking approach are as follows. First, the weighted-constraint framework extends naturally to explaining the *strength* of preferences: easy ambiguities or hard garden-path effects. Under such a system, it is natural to assume that the larger the difference, the harder the reanalysis (Gibson 1991). The ranked constraint system makes no such predictions: an independent theory of reanalysis is needed in addition. The weighted-constraint approach also extends naturally to explain processing overload effects, such as the difficulty experienced in processing center-embedded structures. There is no such explanation within the ranked constraint system. See Gibson (1991, in press) for more complete explanations of these kinds of phenomena within a weighted-constraint framework.

7. Summary and Conclusions

In this paper, we have given a few examples of possible ranked constraint systems for the HSPM, each of which has some difficulties when interpreted as an OT system. Although this is not a proof that there is no constraint ranking theory of the HSPM, it casts doubt on the possibility that such a system exists. For our OT recast of the Minimal Attachment/Late Closure system, there were a number of conflicting arguments about which way the constraints should be ordered. Moreover, we raised a more general concern about whether the constraints could work out under any theory of constraint combination. Although the situation improved somewhat, the problem of conflicting constraint ranking arguments remained in the second Theta-violations/Recency system we considered, which aimed at capturing the same range of phenomena as the Minimal Attachment/Late Closure system. Here, we argued that summing constraint costs led to a more successful application of the constraints than the "winner-take-all" approach central to OT. Our analysis of the Recency/Predicate Proximity system also provided an argument for an additive model of constraint combination.

References

Altmann, Gerry & Mark Steedman. (1988). Interaction with context during human sentence processing. *Cognition* 30, 191–238.

Babyonyshev, Maria & Edward Gibson (1995). Processing overload in Japanese. In Carson Schütze, Jenny Ganger & Kevin Broihier (eds.), *Papers in language processing and acquisition*, MITWPL 26, Cambridge, MA: MIT Working Papers in Linguistics, 1-35.

Bader, Marcus & I. Lasser. (1994). German verb-final clauses and sentence processing: Evidence for immediate attachment. In Charles Clifton, Jr., Lyn Frazier & Keith Rayner (eds.), *Perspectives on sentence processing*. Hillsdale, NJ: Lawrence Erlbaum, 225–242.

Chomsky, Noam (1981) *Lectures on government and binding*. Dordrecht: Foris.

Clifton, Charles Jr. (1988). Restrictions on late closure: Appearance and reality. Paper presented at the 6th Australian Language and Speech Conference, University of New South Wales, August 19–21.

Crain, Stephen & Mark Steedman (1985). On not being led up the garden path: The use of context by the psychological parser. In D. R. Dowty, L. Karttunen, & A. M. Zwicky (eds.), *Natural language processing: Psychological, computational and theoretical perspectives*. Cambridge: Cambridge University Press.

Cuetos, F. & Donald C. Mitchell (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition* 30, 73–105.

Ferreira, Fernanda & J. M. Henderson (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language* 30, 461–476.

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

Frazier, Lyn (1978). *On comprehending sentences: Syntactic parsing strategies*. Ph.D. Dissertation, University of Connecticut.

——— (1987a). Sentence processing: A tutorial review. In M. Coltheart (ed.), *Attention and performance XII*. Hillsdale, NJ: Lawrence Erlbaum.

——— (1987b). Syntactic processing evidence from Dutch. *Natural Language and Linguistic Theory* 5, 519–559.

——— (1990). Parsing modifiers: Special purpose routines in the human sentence processing mechanism? In D. A. Balota, G. B. Flores d'Arcais & Keith Rayner (eds.), *Comprehension processes in reading*, Hillsdale, NJ: Lawrence Erlbaum.

——— (1994). Sentence reanalysis. Paper presented at the 1994 CUNY Sentence Processing Conference.

Frazier, Lyn & Clifton, Charles Jr. (in press, 1995). *Construal*. Cambridge, MA: MIT Press.

Frazier, Lyn & Janet D. Fodor (1978). The sausage machine: A new two-stage parsing model. *Cognition* 6, 291–325.

Frazier, Lyn & Keith Rayner (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14, 178–210.

——— (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language* 26, 505–526.

Gibson, Edward (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Ph.D. dissertation, Carnegie Mellon University.

——— (in press). *Memory limitations and linguistic processing breakdown*. Cambridge, MA: MIT Press.

Gibson, Edward, Gregory Hickok & Carson Schütze (1994). Processing empty categories: A parallel approach. *Journal of Psycholinguistic Research* 23, 381–406.

Gibson, Edward, Neal J. Pearlmutter, Enriqueta Canseco-Gonzales & Gregory Hickok (in press, 1996). Recency preference in the human sentence processing mechanism. *Cognition* 59, 23-59.

Gibson, Edward & James Thomas (1995). The processing complexity of center-embedded and self-embedded English structures. Ms., MIT.

Gilboy, E., J. M.. Sopena, Charles Clifton, Jr. & Lyn Frazier (1995). Argument structure and association preferences in Spanish and English complex NPs. *Cognition* 54, 131–167.

Larson, Richard K. (1988). On the double object construction. *Linguistic Inquiry* 19, 355–391.

Lewis, Richard L. (1993). *An architecturally-based theory of human sentence comprehension*. Ph.D. dissertations, Carnegie Mellon University.

MacDonald, Maryellen C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language* 32, 692–715.

MacDonald, Maryellen C., Neal J. Pearlmutter & Mark S. Seidenberg (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101, 676–703.

Mitchell, Donald C. & F. Cuetos (1991). The origins of parsing strategies. In *Conference proceedings: Current issues in natural language processing*. University of Texas at Austin.

Pesetsky, David (1995). Some principles of sentence pronunciation. In this volume.

Phillips, Colin (1995). Generalized right association. Ms., MIT.

Prince, Alan & Paul Smolensky (1995). *Optimality theory*. Cambridge, MA: MIT Press.

Pritchett, Brad (1988). Garden path phenomena and the grammatical basis of language processing. *Language* 64, 539–576.

Pritchett, Brad (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.

Spivey-Knowlton, Michael & Julie Sedivy (1995). Resolving attachment ambiguities with multiple constraints. *Cognition* 55, 227–267.

Spivey-Knowlton, Michael & Michael K. Tanenhaus (1994). Referential context and syntactic ambiguity resolution. In Clifton, Charles, Jr., Lyn Frazier & Keith Rayner (eds.), *Perspectives in sentence processing*, Hillsdale, NJ: Lawrence Erlbaum, 415–439.

Trueswell, J. C., Michael K. Tanenhaus & S. M. Garnsey (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*.

Trueswell, John C., Michael K. Tanenhaus & C. Kello (1994). 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.