A Minimalist Theory of Human Sentence Processing

By Amy Weinberg

Linguistics Department/UMIACS

University of Maryland

weinberg@umiacs.umd.edu[1]

I. Introduction

Research in the theory of human sentence processing can be characterized by 3 styles of explanation. Researchers taking the first track have tried to motivate principles of structural preference from extralinguistic considerations like storage capacity in working memory, or bounds on complexity of incremental analysis. Frazier and Rayner’s (1982) Minimal Attachment and Right Association principles, and Gorrell’s simplicity metric, are examples of this type of theory.

The second track eschews "parsing strategies", replacing them with a fairly complex tuning by speaker/hearers to frequency in the hearer's linguistic environment. The difficulty of recovering an analysis of a construction in a particular case is a function of how often similar structures or thematic role arrays appear in the language as a whole. The work of Trueswell et al (19 4), Jurafsky (1996) and MacDonald et al (1994) are examples of frequency or probability based constraint satisfaction theories.

The third track takes a more representational view and ties processing principles to independently needed restrictions derived from competence and language learning. This approach claims that the natural language faculty is extremely well designed in the sense that the same set of principles that govern language learning also contribute to a theory of sentence processing. This track is represented by the work of Gibson (1981), Gorrell (1995) Pritchett (1992), Philips (1995, 1996) and Weinberg (1992), who argue that processing can be seen as the rapid incremental satisfaction of grammatical constraints such as the Theta Criterion, which are needed independently to explain language learning or language variation. A variant of this approach, represented by Crain and Steedman (1985) among other restraints the grammatical source for parsing principles but locates these principles within a discourse or semantic, rather than a syntactic component.

This paper proposes a model of the last type. We argue that a particular version of the Minimalist Program (Chomsky (1993), Uriagereka (this volume)) provides principles needed to explain both initial human preferences for ambiguous structures and provides a theory of reanalysis, explaining when initial preferences can be revised given subsequent disconfirming data, and when they lead to unrevisable garden paths. We will then argue that this type of theory is to be preferred to theories motivated on extralinguistic principles.

In the first section of this paper we discuss the Minimalist Theory of syntax upon which we will base our parsing proposals. Features that distinguish this theory from precursors are:

(1) The theory is derivational, providing principles for how an analysis is constructed rather than filtering conditions that constrain output representations. The main derivational constraints are the so-called Economy Conditions (Chomsky 1993).

(2) The theory applies constraints strictly locally. Derivations are evaluated at each point in the analysis. They are optimized with respect to how well they satisfy constraints of a given item that is a candidate for integration into the structure at each point. How a proposed structure satisfies constraints imposed by the derivation as a whole is irrelevant.

(3) The theory incorporates a claim about a one to one mapping between precedence order and structural hierarchy or dominance that is embodied in the <u>Linear Correspondence Axiom</u> (Kayne (1994), Uriagereka (this volume)**)).**

Next, we show how to interpret Minimalist principles as a parsing algorithm. We will show that the Economy conditions below define a crosslinguistically attested theory of preference judgments. (2) and (3) combined distinguish cases where an initial preference can be reanalyzed from those cases where reanalysis into the appropriate structure is impossible with a resulting garden path.

The next section compares our models with Colin Phillips' model of sentence processing. Phillips shares our view that principles of grammatical theory should form the basis of the theory of sentence processing. The processing principles that he invokes are based on a slightly different grammatical theory one that he claims is identical to the theory of linguistic competence. We will first discuss what we see as strengths of his theory and then discuss three types of problems with his approach.

The final section argues that this type of theory has advantages over theories relying on extralinguistic frequency or parsing strategy principles.

II. Some Minimalist Assumptions:

Readers of this volume are already familiar with many of the features of the minimalist system. We provide a brief review here of the features that are important for the construction of our parsing algorithm.

The two most salient features of this system are its derivational character and the role that Economy conditions play in regulating possible derived structures. At least at the level of competence, the model has moved away from the overgeneration and filtering character of it's Government and Binding precursor. Structures that do not pass the Economy conditions are simply not generated. The two major grammatical operations (Merger and Movement), used to enerate structure are seen as feature checking. Categories are input from the lexicon with features such as Case and theta role that have to be checked. Checking is satisfied when a category needing a feature is in construction with some other element that can supply that feature in the sentence. Movement or merger operations are only licensed if they allow feature checking to occur. Movement or merger serve to allow an element to transfer a feature necessary to satisfy some constraint. The relevant conditions that rule out overgeneration are the following:

(4) Last Resort: Operations do not apply unless required to satisfy a constraint. A minimal number of operations is applied to satisfy the constraint.

(5) Greed: "The operation cannot apply to α to enable some different element β to satisfy its properties...Benefiting other elements is not allowed**.”**

III. Multiple Spell -out:

A corollary assumption that has been incorporated into the Minimalist program has been the derivation of a correlation originally due to Kayne (1994). Previous grammatical formalisms had argued that restrictions on linear precedence and immediate dominance were the product of two separate subsystems. Kayne (994) suggested hat these two systems were linked, and that one could derive precedence order from information about dominance. This conjecture is known as The Linear Correspondence Axiom (LCA) given in (6).

(6) LCA:

Base Step:

If (α) precedes (β), then (α) c-commands (β).

Induction Step:

If (γ) precedes (β), and (γ) dominates (α), then (α) precedes (β).

C-command is defined as in Epstein (this volume, repeated below)

(7) (α) c-commands al and only terms with which (α) was paired b Merge or Move in the course of the derivation.

(8) illustrates the relationships licensed by these definitions.

(8) IP

DP I’

D NP I VP

the *man* *tense* *slept*

The precedence relations among elements in the subject are licensed because the determiner c-commands and precedes the NP (man). The second part of the definition is needed since the terminal elements in the subject position did not directly combine with the elements in the VP by either Merge or Move. Therefore they do not c-command these VP elements even though the terminals in the subject precede those in the VP as required by the base step of the LCA. Their presence is allowed however, by the second clause in the definition because the DP dominating both these terminals precedes the VP and dominates both the determiner and the NP, which inherit precedence by a kind of transitivity. Uriagereka, (this volume) argues that the base step of the definition follows from the kind of “virtual conceptual necessity” inherent in the Minimalist program. The simplest kind of mapping between precedence and dominance is one to one, and therefore we might expect a grammar that specifies linear and dominance order to have this simplifying restriction (see Uriagereka (this volume) for details). We cannot so derive the induction step, which appears only to allow terminals in a c-command relation to co-exist in a structure. General goals of the Minimalist program, which try to derive features of the grammatical system from "virtual conceptual necessity" force us to either derive the induction step from other considerations, or eliminate it from the system. Uriagereka adopts the latter course.

Uriagereka (this volume) claims that we can maintain the simple relationship between command and precedence given by the base step in (6) if we allow the operation of Spell-Out to apply many times during the course of the derivation. Spell-out is the operation that removes material from the syntactic component and feeds it to the interpretive components of Logical Form (LF) and Phonetic Form (PF) when that material is ready for interpretation. Uriagereka points out that since the minimalist system dispenses with a global level of s-structure as the conduit to the interpretive components, there is nothing to stop material from being passed for interpretation multiple times.

We assume that spell-out applies whenever two categories cannot be joined together by the Merge operation. If Merge doesn't apply then the category currently being built is spelled out or reduced. We retain the notion from earlier theories of grammar [2]that Spell-out is a conduit between the syntax and the phonology. It is well known that the constituency established by the syntax is not relevant for phonological processes.[3] Spell-out, turns a syntactic structure with relevant constituent relationships into a string ready for phonological interpretation. Uriagereka uses "Spell-Out" as a repair mechanism to retain one to one correspondence between domination and precedence. He assumes that both precedence and dominance must be established between terminal elements at all points of the derivation. Precedence implies merger and merger is only possible when a chain of domination can be established. When merger is not possible, the string is linearized (turned into a unstructured string where only previously established precedence relations are preserved) Since the elements that have been linearized are invisible in the syntax, precedence does not have to be established between them and other items in the structure. Thus, when two categories cannot be combined through merger or movement (the only syntactic operations) to form a dominating category, the material that has been given structure so far is "spelled out" or linearized.

(9) "L is an operation L(c)=p mapping command units"(units that can be formed through merger(ASW) c to intermediate PF sequences p and removing phrasal boundaries from c representations"

Uriagereka (this volume)

This idea preserves the one to one mapping between precedence and dominance but at the cost of ever building a single phrase marker. Instead one builds blocks (Uriagereka calls them "command blocks”) where all elements stand in a c-command relation to each other. When this c-command relation is interrupted, the unit is spelled out, with an unstructured unit shipped to the phonology for phonological interpretation and a structured unit shipped to LF (logical form) for semantic interpretation. The result of "Spell-Out” is an unstructured string (a syntactic word) with no further internal phrase structure. Within the context of the Minimalist system, "Spell-Out" is a grammatical operation, on a par with movement transformations. As such it is governed by conditions on transformations, in particular by the Economy Conditions discussed above. This economy condition establishes a preference for derivations which utilize the fewest number of operations possible. An operation is applied only to satisfy some independent grammatical condition. In this case, this means that we will Spell-Out or linearize only when we could not otherwise establish a chain of precedence.

III. Minimalist Principles as a Parsing Algorithm

We will now apply a theory incorporating economy conditions and multiple spell-out to parsing. We assume that the algorithm applies left to right and evaluates ambiguities with respect to the economy conditions. As in minimalist theory, items are inserted into the derivation (or moved) with the goal of checking features. The feature checking aspect of the theory will impose an argument over adjunct attachment preference along the lines of Pritchett (1992) and Gibson (1991) on the assumption that theta roles are relevant features for checking. Attachment as an adjunct will never lead to receipt or transfer of theta, case or other features, whereas insertion into an argument position will allow this transfer to occur. We will see that this preference is well attested. Unlike Pritchett (1992) and Gibson (1991), feature transfer is optimized locally. Pritchett and Gibson allowed the parser to scan the entire derivation at of an item's attachment and to compare whether the attachment of a category optimized the assignment of features over all elements of the tree built so far. By contrast, since feature checking is subject to Greed in the Minimalist system, this theory only allows optimal feature checking on the particular category that is being attached irrespective of whether this optimizes feature checking across the derivation as a whole. We will see that this is crucial for some of our examples below.

Insertion or movement is governed by the Economy Conditions discussed above. The preference to attach a category using minimal structure follows immediately from this notion of Economy. At each point a category is inserted using the least number of operations necessary for feature transference or merger. This ban on unnecessary operations subsumes Frazier and Rayner (1982)'s minimal attachment and Gorrell's (1995) simplicity condition with the advantage of following from independently motivated grammatical principles.

Following Uriagereka, we assume that Spell-out occurs whenever a derivation would otherwise violate the LCA (now containing only the base step). The spell out conditions thus also provide us with an independently motivated theory of reanalysis. If a preferred reading induces a precedence/dominance mismatch, the category that precedes but does not dominate will be spelled out. Again following Uriagereka, this means that the material inside the spelled out category is linearized and all internal syntactic structure is removed, creating a nondecomposable syntactic word. Given this, reanalysis from the preferred to dispreferred reading that requires either extraction of material from, or insertion of material into this syntactic word, will be impossible. As a lexical item, the spelled out material is an atomic unit, which can no longer be decomposed into its component pieces. If however, reanalysis occurs within a domain where Spell-Out has not applied, then material can be accessed and the preferred reading can be transformed into the dispreferred structure. Incorporating Spell-out and Economy conditions into the grammar also explains the preference for right branching derivations without the need for extra explicit principles which favor this type of derivation.[4]

As a grammatical operation, Spell-out is governed by Economy. Since it does not allow the checking of any features it is an operation of the last resort. As such, it will only be invoked when no other feature checking operation can apply and the minimal number of spellouts to guarantee satisfaction of the LCA will operate at each time step in the derivation. A right branching structure insures that an element that proceeds will also dominate a category and thus minimize the need for Spell-Out. Therefore, right branching structures will be preferred because they economize on the need for spell-out.

The algorithm in **(10)** embodies these principles

(10) A derivation proceeds left to right. At each point in the derivation, Merge using the smallest number of operations needed to check a feature on the category about to be attached. If Merger is not possible, try to Move within the current Command path. If neither merger nor movement is licensed, Spell-Out the command path. Repeat until all terminals are incorporated into the derivation.

IV. Some Cases:

(i) Argument/Adjunct attachment ambiguities:

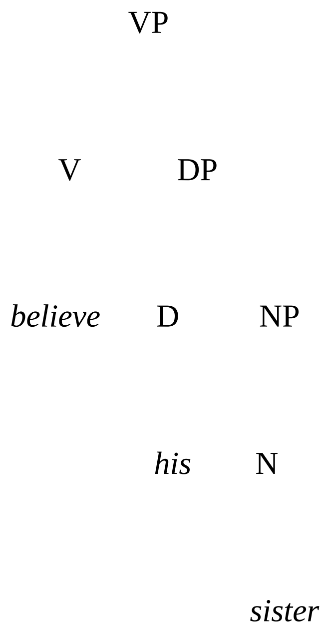
These cases illustrate the role of optimizing feature checking relative to Economy conditions. In all cases, attachment as an argument is preferred because it allows assignment of features.

(a) <u>Direct object/complement subject ambiguity:</u>

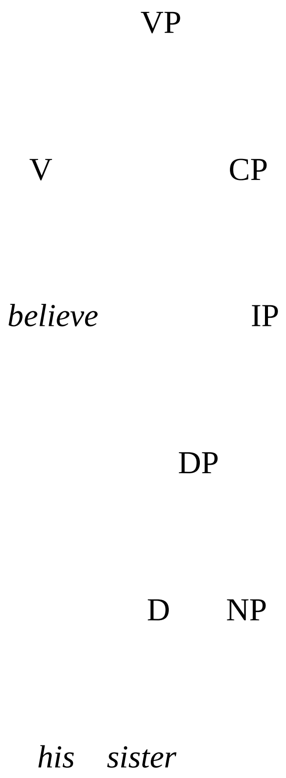
The sentences and the relevant structures are given in **(11).**

(11) a. The man believed his sister would win the Nobel Prize.
b. The man believed his sister.

c.



d



The DP 'his' 's ster' will be assigned both case and theta features by the preceding verb if it is attached as the direct object. Case and theta features can only be assigned by the Case and Theta assigner, the Head of the complement clause. Since this category has not yet been processed, no features will be assigned by an attachment of 'his sister' as the subject of the complement clause. Therefore (c) is the preferred structure. It is also the structure that is the most economical, involving fewer operations, although this is not a crucial determinant of attachment for this case. In neither case is Spell-Out necessary at the site of attachment of 'his, 'sister'.[5]

Notice that the attachment motivated by the desire to check features does not cause any spell-out within the VP. Both the verb and the object are available when the embedded verb is encountered in a case like (a). Therefore, the object NP is available for reinsertion as the embedded subject in (d) even though the initial structure chosen for this case is (c). All elements remain on the command path.

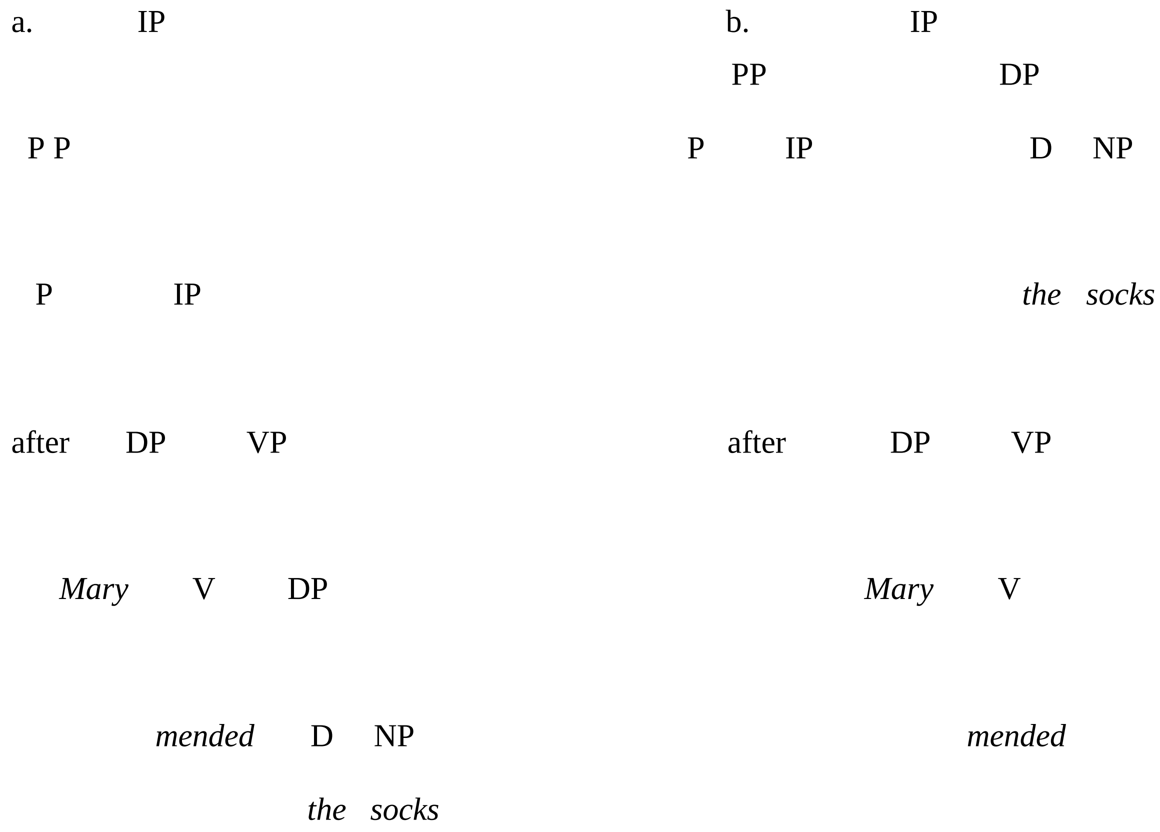
b. <u>Preposed object/matrix subject ambiguity</u>.

Next consider **(12)** where there is a preference to treat the word following the first verb as an object in the preposed adverbial, rather than as the subject of the matrix sentence**.**

(12) a. After Mary mended the socks fell off the table.
After Mary mended the socks they fell off the table.

Again, incorporation as an object allows case and theta features to be checked off from the phrase 'the socks'. Incorporation as the matrix subject does not allow any case or theta feature checking, againbe cause the case and theta assigning head of the IP has not yet been incorporated into the structure. The relevant structures are given in (13).

(13)



We do not expect reanalysis to be possible given the algorithm 10). After building the optimal structure in (13) , the phrase ‘fell’ cannot be incorporated into the preposed adverbial clause. A globally optimizing algorithm might look to see what series of transformations could be made to incorporate this category. However, our algorithm is a dumb one that acts only to incorporate local material. Since the second verb phrase phrase can’t incorporate into any node within the preposed adverbial, the adverbial is spelled out in a phrase by phrase manner, leaving the structure in

(14). This structure respects the LCA.

(14) **IP**

AP fell

After Mary sewed the socks

However, there is no way to incorporate the structure into this remnant either. The preceding material has been spelled out and so there is no way to retrieve anything from this phrase to be inserted as the necessary matrix subject. Since no further operations apply, and there is remaining unincorporated terminal material, the parse fails and a garden path is detected.

<u>c. Ditransitive/complex transitive object ambiguity</u>:

(15) a. John gave the man the dog for Christmas**.**

\# b. John gave the man the dog bit a bandage.

The preferred reading for (b) is to treat ‘the dog’ as a ditransitive object as in (16) as opposed to treating this category as the subject of a relative clause modifying ‘the man’ as in (17)

(16) VP

V VP

give $_i$ DP VÃ

the man V DP

e_i D NP

the dog

(17) VP

VP

V DP V’

gave$_i$

DP CP V

the man C IP

DP

the dog e$_i$

Clearly (17) is more complicated and requires more mergers than (16), violating Economy. This is again not crucial because the analysis as an indirect object allows features to be checked on the DP ‘the dog’ while attachment as material in the matrix subject does not allow feature transference.

Reanalysis is not possible in this structure. We crucially assume the Larsonian shell structure in (16) to explain wh . Reanalysis would involve incorporation of the category in the indirect object position originally part of the relative clause on the direct object. This however, cannot be accomplished while the trace of the moved V remains in the structure because a relative clause inside the direct object would not command the verb trace. Therefore, maintenance of the terminals of the preceding relative and the verb trace in the same tree would violate the LCA. Therefore, the V’ in (17) must be spelled out. If this category is spelled out however, there is no host site for subsequent attachment of the true indirect object because all structure under the V’ node is no longer accessible.

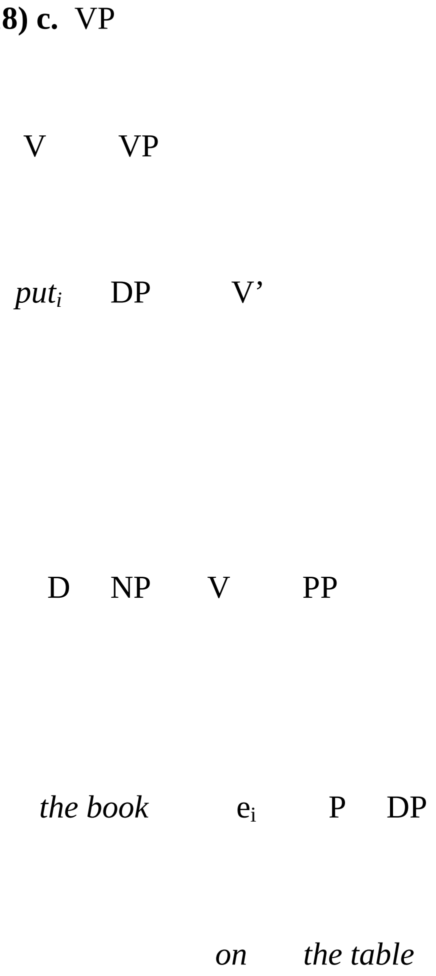
d. Subcategorized PP/NP modifier ambiguities:

There is a preference to treat the PP ‘on the table’ as an argument of the verb ‘put’ rather than as a modifier of the NP ‘the book’. We will assume (non crucially) the Larsonian analysis of PP complements as well. Whatever the structure is, the attachment as an argument allows the PP to receive and the V to discharge features. The structures are given in (18) c and d.

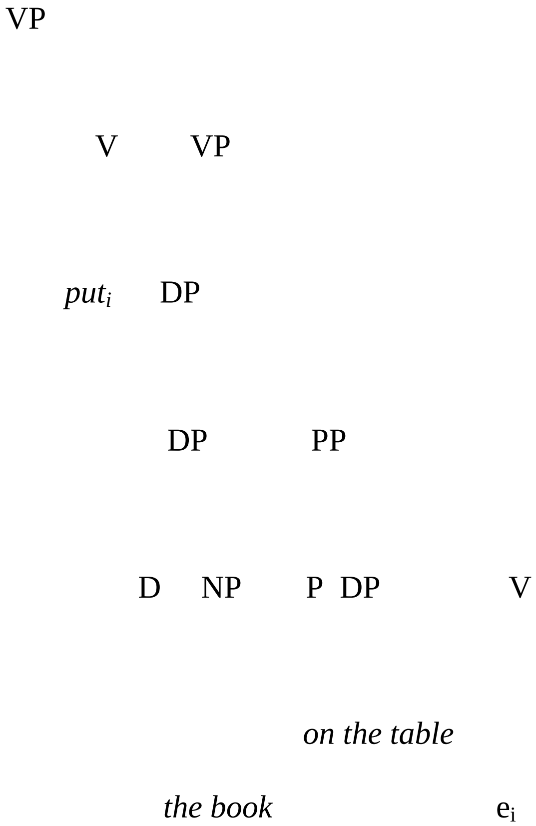
(18a) I put the book on the table

b) I put the book on the table into my bag.

(18) c.



d.



Reanalysis is not possible for the same reason as the ditransitive case above. To reanalyze the PP as part of the direct object as the adjunct to 'the book' requires Spell-Out of the V', since material inside the relative would not command this category. If this category is spelled out though, there is no site for the true locative PP ‘ into my bag’ to merge to.

The final case of an argument / adjunct ambiguity is the famous main clause/relative clause ambiguity exhibited in cases like (19)

(19) The horse raced past the barn fell.

These are strict garden paths, with native speakers preferring a main clause reading “ The horse raced past the barn’ for these cases and being unable to reanalyze these as reduced relative clauses.

Interestingly, Pritchett (19 2) and Stevenson and Merlo (19 7) have suggested that these types of ambiguities do not always yield garden paths. When transitive and unaccusative verbs replace the unergatives like those in (19), the sentences become quite easy to process as shown in (20).

(20) The student found in the classroom was asleep.
(b) The butter melted in the pan was burnt.[6], [7]

Within the context of the Minimalist account, these subtle facts are accounted for because both transitives and unaccusatives must have traces inserted in the postverbal position, whether or not these structures are analysed as main clauses or relative clauses. This is because the theta grid of both transitives and unaccusatives signals to the parser that these verbs both require NP objects. Since there is no overt object in the postverbal position, a trace must be inserted here. So, even if the preferred analysis for these cases is as main clauses, the material needed to appropriately interpret these structures as open sentences, with traces in the post verbal position, are built as part of the main clause analysis, BEFORE the spellout required by the disambiguating main verb for cases which are truly reduced relatives. The initial analyses are given in (21a), (21b). The reanalysis proceeds along the lines discussed above. The material preceding the main verb is initally analysed as a main clause. When the true matrix verb is encountered, spellout occurs of everything preceding the verb in accordance with the LCA. Now, however the spelled out material can be appropriately interpreted as a relativ clause, and so no garden path results.

(21a) $[_{IP}$ The student$_i$ $[_{VP}$ found e_i [in the classroom]

b.) $[_{IP}$ The butter$_i$ $[_{VP}$ melted e_i [in the pan]

In all of the above cases, Economy seemed to redundantly track feature checking in the sense that the most economical structure was also the one that allowed features to be checked. We now turn to cases where local economy is crucial to predicting both preference and reanalysis judgments. These cases deal primarily with instances where the ambiguity is between two different types of adjunct attachment. In neither case will a feature be checked so Economy is the only factor in play.

ii. Adjunct/Adjunct Attachment:

(a) Adverb or particle placement:

The grammar presents multiple attachment sites directly after the italicized words in all of the cases in
(21).

The parser always chooses the position after the most recently encountered word as the preferential site of attachment.

(21) I told Mary that I will come yesterday
I called to pick the box up

I yelled to take the cat out.

In the first case, the adverb ‘yesterday’ is construed with the embedded verb despite the fact that this reading is semantically anomalous and despite the fact that an alternative attachment to the matrix verb would result in an acceptable reading. The other two cases show that the particle prefers low attachment as well.

These preferences can be explained on the assumption that Spell-out, as one of the grammatically licensed operations is also subject to the Economy conditions. Therefore mergers involving fewer spell-outs will be preferred. Consider (22) at the point when the adverb ‘yesterday’ enters the derivation.

(22)

VP

V VP

told$_i$ DP V’

Mary

V CP

e$_i$ C IP

that DP IÃ

##I## I VP

will V VP

come V’

V AP

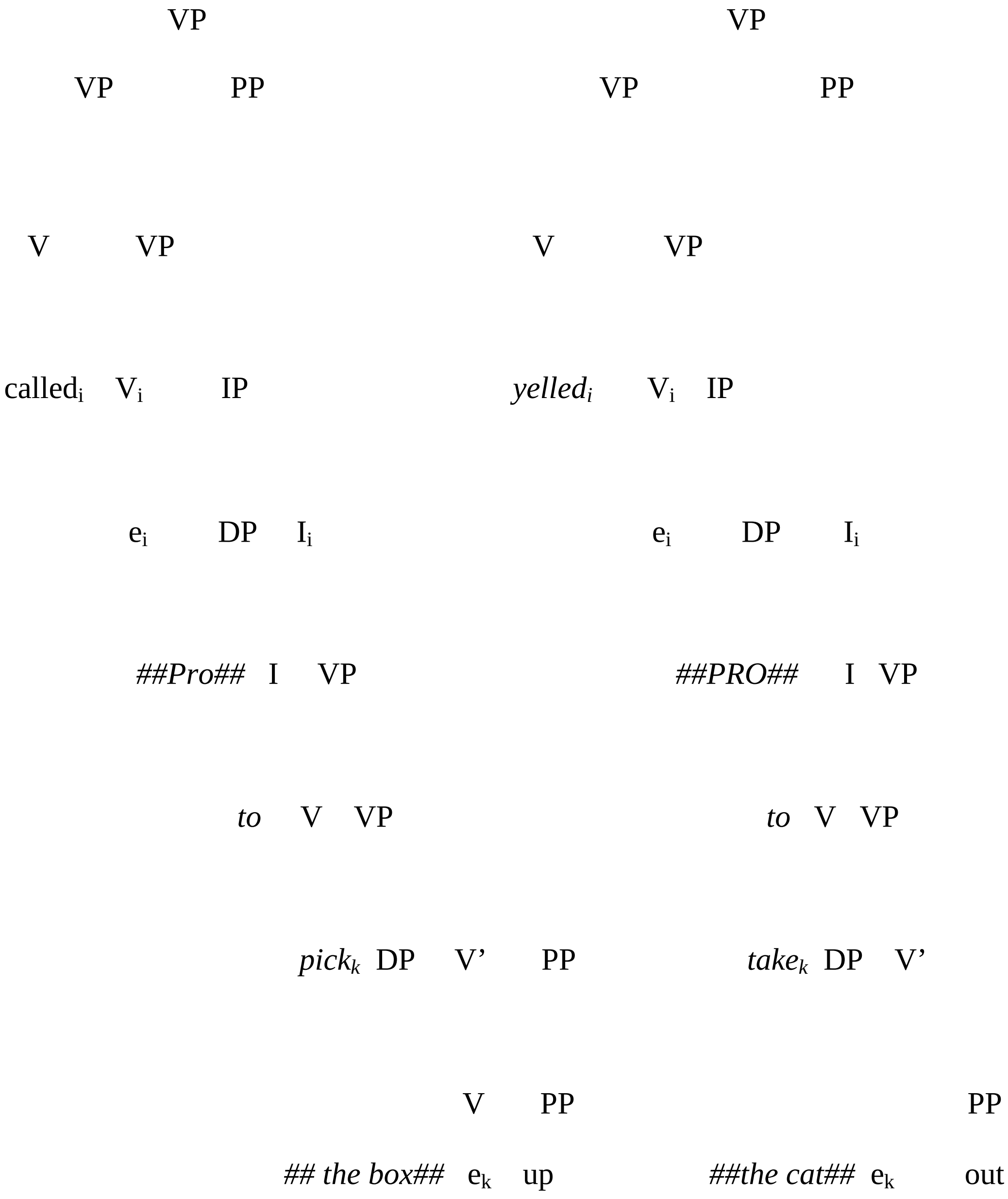
e$_i$

Assuming attachment into a Larsonian shell associated with the lowest verb, where adverbs assume the position of complements would require no Spell-Outs at this point.

The adverb would simply be merged under the italicized phrase. Assuming Uriagereka's version of the LCA though, attachment as an adjunct to the higher verb would require Spell-out of the lower VP, I' and IP respectively given the algorithm in (10). The algorithm in (10) requires spellout of only the material that would not c-command the site of a potential merger. Therefore, if the parser has processed everything up to the lowest clause in the preposed position, it wil require multiple spellouts to return to the highest level of the preposed adverbial. In the competence model, one could think of high or low attachment as requiring an equal number of spellouts, each with a different number of phrases in the spelled out component of the analysis. In a parser however, one does not keep the whole structure in memory at a given point and therefore, one must provide an explicit procedure for dealing with previously processed material. The parser cannot retrieve a site for attachment in this case with out successive iterations of spellout, given (10). Since lower attachment involves fewer iterations of the spellout procedure, economy conditions thus favor this attachment choice. This will be true for the rest of the cases in (

(21). Attachment of the particle to the higher verb will cause spell-out of the phrases remaining on the c-command path of the lower clause. These phrases are italicized in (23). Attachment as the particle of the lower verb requires no Spell-out and will again be preferred by Economy.

(23)



The next case was discussed in Phillips and Gibson (in press). Normally relative clause attachments are dispreferred, but in this case they are the favored reading.

(24) Although Erica hated the house she had owned it for years.
Although Erica hated the house she owned her family lived in it for years

Philips and Gibson presented sentences like these with either temporal or non-temporal adverbial modifiers in a word by word self-paced reading task with a moving window display. At the disambiguation point (either ‘it’ or ‘her family’), subjects showed a clear preference for the attachment of the preceding clause as a relative clause modifying the noun ‘the house’. There was a significant increase in reaction time at the disambiguation point if the ambiguous noun phrase was disambiguated as the matrix subject.

We can explain this preference again with reference to economy of spell-out. Again at the relevant point, neither attachment will allow the discharge of a feature. However attachment as a relative clause permits much more of the preceding material to remain in the derivation as it would still command the incoming merged material. Attachment as the matrix subject requires spell-out of the entire preposed adverbial. The relevant structure with the number of nodes needed to be spelled out in italics for the matrix subject reading, and underlined for the relative clause reading are given in (25).

(25) IP

AP DP

A IP *she*

although DP I’

Erica VP

V DP

hated DP CP

the house ## IP

DP

she

V. Right Branching Structure in the Grammar and in the Parser- A Comparison with Collin Phillips’ Approach

Phillips (1995), (1996) presents very interesting work that argues for an alternative grammatically based theory of processing. In fact, Phillips claims that there is no distinction between the parser and the grammar. Derivations in both the competence and performance systems are built up incrementally left to right.

Given this grammatical underpinning, Phillips tries to link performance preferences to the grammar in the following way. First, he defines a condition called Branch Right given in (26) below

(26) Branch Right:

“Select the most right branching available attachment of an incoming item
Reference Set: all attachments of a new item that are compatible with a given interpretation”[8]

The preference for right branching structure is in turn derived from a principle that insures that the base step of the Linear Correspondence Axiom (LCA) discussed above is incrementally satisfied to the greatest extent possible. As Phillips writes:

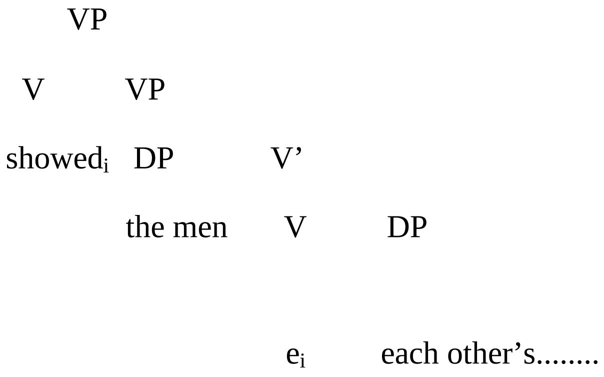
"I assume that a structure is right branching to the extent that there is a match between precedence relations among terminal elements and c-command relations among terminal elements"[9]

Phillips couples this with the idea that grammatical as well as parsing derivations proceed left to right to handle a variety of bracketing paradoxes. Consider (27)

(27) **a.** John showed the men each other's pictures.
b. John showed each other the men's pictures.

These examples suggest that double object constructions have right branching structures where the indirect object c-commands the direct object as in (28)

(28)



The fact that (29 a-c) are gram atical as VP fronting structures suggests that the structure for the PPS should be left branching, allowing the right subparts to be constituents. The structure is given in (30).

(29) I said I would show the men the pictures in libraries on weekends, and
(a) show the men the pictures in libraries on weekends, I will
(b) show the men the pictures in libraries, I will on weekends.

(c) show the men the pictures, I will in libraries , on weekends

(30) V’

V’ PP

V’ PP P DP

V’ P DP on weekends

V’ in libraries

V’ DP

V’ the pictures

V’ DP

show the men

Phillips shows that we can derive the effects of a structure like (30) without the need to assume it, by assuming that Branch Right applies from left to right with the seeming left branching structures actually being intermediate structures in the derivation. Branch Right for example, would first combine ‘show’ and ‘the men’ to form a constituent. This constituent would then be reconfigured when subsequent material was uncovered. Phillips (1996) presents a variety of advantages for his approach over other treatments of paradoxical constituency. The definition in (26) suffices to handle all of these paradoxes.

Philips claims that we can use Branch Right to resolve various parsing ambiguities. In order to do this, he redefines Branch Right as (31).

(31)

Metric: Select the attachment that uses the shortest path(s) from the last item in the input to the current input item.

Reference Set: all attachments of a new item that are compatible with a given interpretation.[10]

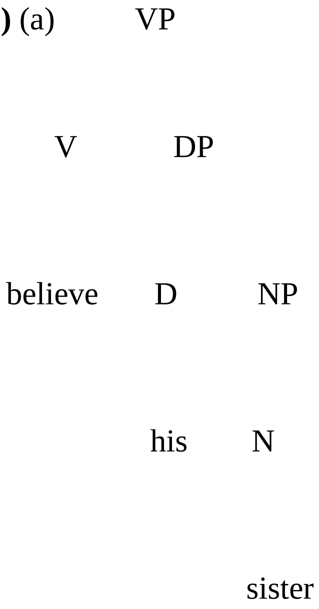
(32), repeated from above is a simple illustration of how the principle works.

(32)a. The man believed his sister would win the Nobel Prize.

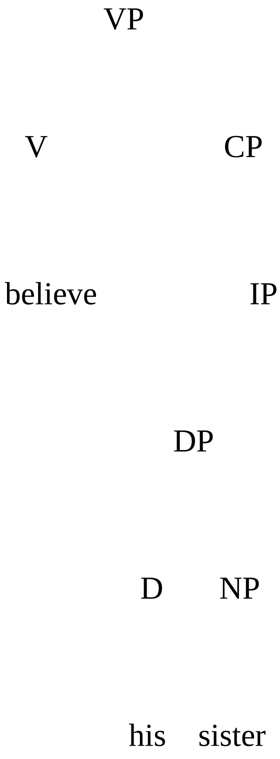
b. The man believed his sister.

Branch Right predicts this preference because there are fewer branches in the path between ‘believed’ and ‘his sister’ if one construes the post verbal noun phrase as a direct object, than if one construes this phrase as the subject of an embedded clause as shown in(33)

(33) (a)



(b)



Path = 1 step up from V to VP
1 step down from DP to D

Path = 1 step up from V to VP
4 steps down from VP to CP, IP,
DP and D

Since the embedded subject reading takes more steps on the downward path, it is dispreferred.

Phillips uses this simple principle to handle a wide range of data in English, and illustrative cases from German and Japanese. The empirical coverage of this simple principle is impressive. In addition, the use of Branch Right is argued to be independently justified by the LCA or at least from its ability to handle bracketing paradoxes, so it appears that we are getting a parsing principle for free from independently needed competence principles. For these reasons the proposal is quite interesting. Nonetheless I will argue against this approach on several grounds.

a. **Problems with "Branch Right":**

The range of problems which we now turn to do not focus on the empirical coverage of the theory per se. We note in passing however, that this theory is intended merely to be a theory of initial preference. It is well known that certain initial preferences, such as (32) above can be overridden given subsequent disambiguating material, while cases like (34) are not subject to reanalysis, and remain garden paths.

(34) The horse raced past the barn fell[11]

(34) is initially interpreted as a main clause 'The horse raced past the barn'. Reanalysis as a reduced relative ''The hor e that was raced past the barn' is impossible. The availability of this grammatically licensed interpretation has to be pointed out to naive speakers, as is well known. Phillips’ theory is silent on the issue of when reanalysis is possible. Phillips claims that reanalysis should not be part of the theory of sentence processing .

> " ...it is not clear that should want Branch Right to account for recovery from error. I assume that Branch Right is a property of the system that generates and parses sentences in a single left- to-right pass, and that reanalyzes require backtracking and are handled by other mechanisms."[12]

This claim depends on the unargued for presupposition that the sentence processor proceeds in a purely left-to-right manner. However, we know from eyetracking studies that backwards saccades even in processing unambiguous and perfectly understandable text is the norm. Secondly, given that both interpretations in (32) are easily processable, it is hard to see why these reanalyses are not the domain of the human sentence processor. We agree with Phillips that the actual mechanisms of reanalysis, particularly in cases where conscious breakdown occurs, may not be the domain of the processor. We see no reason however not to demand that a full theory of sentence processing distinguish cases where these mechanisms can apply; where the human sentence processor presents the appropriate representations for these mechanisms to operate on, from cases where the sentence processor does not present the appropriate representations for the operation of potentially external general purpose reanalysis mechanisms. Phillips' theory is mute on this domain of empirical prediction.

We turn now from the domain of prediction to that of independent motivation. Part of the main appeal of the <u>Branch Right</u> theory is its independent motivation in terms of the

LCA and the bracketing paradoxes. We get a processing principle for nothing. However, we will see that this motivation is partial at best.

(32) above illustrates the next two problems with this condition. (32) crucially relies on a comparison of the number of steps needed to derive both possible readings independently of whether either of these readings causes a precedence/c-command mismatch. Both of the structures in (32) respect the grammatically relevant version of <u>Branch Right</u> given in (26) above where ‘right branchingness’ is defined in terms of respect for the base step of the LCA. In both structures the verb both precedes and dominates the following NP whether or not it is construed as the direct object as in (b), or the complement subject as in (a). Nonetheless speakers have a clear preference for the interpretation of (b) over (a). This prediction thus rests on the notion of **shortest** path. This however is not independently motivated by any of Phillips’ grammatical considerations . In effect, Phillips has sneaked in the grammatically unmotivated Minimal Attachment principle of Frazier and Rayner (1982), yielding a combined Minimal Attachment/Branch Right principle which is only half motivated by the grammar. Without the ‘minimal attachment’ part of this principle, the theory is too weak to predict the preference for ( ) over (a).

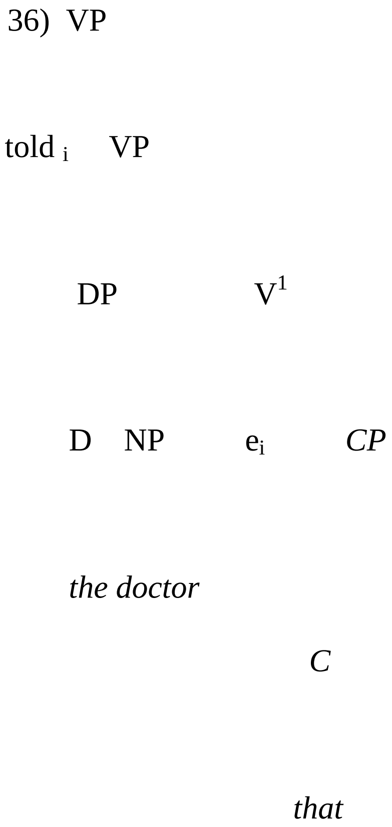
In (36) we present a case where the theory, without the minimal attachment addendum is too strong.

(35) a The man told the doctor that he was having trouble with his feet**.**

\# b The man told the doctor that he was having trouble with to leave.

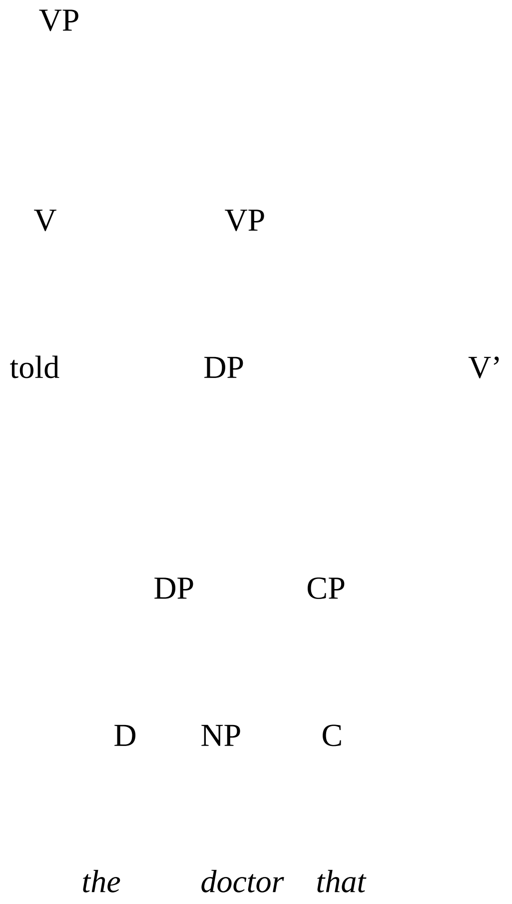
Building either structure at the ambiguous point involves creating a precedence/dominance mismatch. Nonetheless there is a strong preference for (36) over (36b). Phillips assumes that the preferred structure is analyzed as a VP shell. As such the structure would be as in (**36**) with the ambiguous material italicized.

(36)



In this structure the direct object ‘the doctor dominates neither the trace of the verb ‘told’, nor the complementiser of the complement clause. This structure induces a precedence/dominance mismatch. The same is true in the less highly valued relative reading.

(37)



The difference in these cases is then again, not attributable to the metric of precedence/ dominance correspondence or mismatch, but to the length of the path between ‘man’ and the next terminal node. Again, this reduces to the unprincipled ‘minimal attachment’ portion of Phillips’ Branch Right.

To sum up, we have identified two problems with Phillips’ Branch Right. First, it fails to provide a theory of reanalysis; or more precisely does not distinguish representations in such a way as to form a basis even for an independent theory of reanalysis. Second, it incorporates a ‘minimal path’ condition as well as a preference for right branching structure in such a way that the minimal path condition cannot be derived from the latter part of the condition. As such, there is a large portion of the constraint that is not grammatically motivated. Without this unmotivated portion, the theory is both empirically too strong and too weak.

VI. Constraint-Based Theories:

In this section, I would like to review some data presented above with the goal of contrasting a grammatically based view, such as the two previously discussed, with frequency based or probabalistic constraint based theories. MacDonald, Pearlmutter, and Seidenberg present a theory of this type. The main tenet of this theory is summarized as follows:

"Processing involves factors such as the frequencies of occurrence and co-occurence of different types of information and the weighing of probabalistic and grammatical constraints" Our approach has suggested ...that syntactic parsing, including ambiguity resolution, can be seen as a lexical process."[13]

Structural heuristics under this view are replaced with frequency about use of either a lexical item , or in some theories a construction type. [14] For example, the "minimal attachment preference"in (33) bove would not derive from a minimal atttachment preference, or from its grammtical derivation through economy. Rather, speakers can tune to the fact that 'believe' is either used much more frequently with a simple NP as its direct object, than with a sentential complement, or to the fact that simple sentences occur more frequently in the language than sentences with embeddings. Since this theory is "verb sensitive", it can easily account for the verb sensitivity of a variety of preference judgements. For example, verbs like ‘decide’ , which occur much more frequently with sentential complements are correctly predicted to be immune from the “minimal attachment effect. [15]

(38) John decided the contest was fair.

I would like to argue that, while speakers may very likely track frequency, this variable works in tandem with independent grammatical constraints. If a structure occurs very frequently in a given construction, it can influence the initial preferred analysis, but once an analysis is chosen , based on an amalgum of frequency and grammatical variables, the grammatically driven reanalysis principles decide what will or will not be a garden path.

In (20) above repeated as (39), we considere a case where lexical choice was also relevant to preference judgements. Stevenson and Merlo (1997) showed that unaccusative and transitive cases were much better as reduced relative clauses than were unergative verbs.

(39) The student found in the classroom was asleep.
(b) The butter melted in the pan was burnt.

Table I. gives grammaticality ratings for unaccusative versus unergative single argument verbs. They found that unaccusatives were indistinguishable from transitive cases with respect to grammaticality judgements, yielding a two way distinction, with unergatives being terrible, and transitives unaccusatives being fine as reduced relatives.

	Ambiguous		Unambiguous	
	VERB	SCORE	VERB	SCORE
Unaccusative	melt	2	begin	2

	mutate	1.66	break	1
	pour	1.66	freeze	1.5
	reach	1	grow	1
			sink	3.25
Unergative	advance	5	fly	4.25
	glide	5	ring	3.75
	march	5	run	5
	rotate	5	withdraw	3.40
	sail	5		
	walk	3.75		

Table I: Grammaticality Ratings (1=perfect- 5 terrible)

Merlo and Stevenson surveyed corpora with the goal of determining whether this two-way distinction could be derived from frequency of occurrence in a corpus. Using the Wall Street Journal corpus as the reference, they counted how many times a structure appeared as a reduced relative versus how many times it appeared as a main clause. The results are given in Table 2. The important thing to notice here is that both unaccusatives and unergatives occur extremely infrequently as reduced relatives. Nonetheless, they yield radically different judgements with respect to whether clauses containing them yield garden paths or not. Unergatives are unerringly garden paths, whereas unaccusatives are not. Thus , frequency along this dimension does not predict this distinction

	RR	MV	Totals
Unergatives	1	327	328
Unaccusatives	6	358	364
Ordinary	16	361	377

Table 2: Number of Reduced relatives vs Main clause in 1.5 million word Wall Street Journal

Next, they looked at the number of times a verb appeared as a transitive or an intransitive verb. Since reduced relative clauses are uniformly transitive, perhaps this variable is what is tracked and the frequency of occurrence as a transitive is what predicts ability to appear in a reduced relative. Interestingly, these data seem to show a three way distinction with unergatives normally used with one argument, unaccusatives showing a more even distribution, and a third class of (ordinary) verbs showing a distinct tendency to be transitive. ordinary verbs are distinguished from unergative and unaccusative verbs in that adding the second argument does not invoke a “causative interpretation on the predicate”. A

paradigm is given in (40).

(40) I raced the horse (cause the horse to race) vs. The horse raced
I broke the vase (cause the vase to break) vs. The vase broke.

I played baseball vs. I played

	trans	intrans	totals
Unergatives	86	242	328
Unaccusatives	176	228	404
Ordinary	268	114	382

Table 3: Number of Transitive vs. Intransitive frames from Penn treebanked subsection of Wall Street Journal[16]

The interesting point here is that this three way distinction is again not mirrored in native speaker judgements. Ordinary verbs pattern like unergatives unless there are extra pragmatic clues as shown in (41).

(41) The author studied in the English class was boring.

Results like these suggest a picture where frequency has a role to play, but is filtered through grammatically justified constraints. Given the Minimalist theory discussed above, ordinary verbs pattern like unergatives because when they are given their preferred interpretations as main clauses, they are pure intransitives with no trace in the object position. The main verb 'was' in a case like (2), triggers reanalysis as a relative clause, but by that time the material preceding it is already spelled out, and the trace neccessary for interpretation as a reduced relative cannot be inserted. The structure is given in (42). Frequency , coupled with the Economy driven conditions may drive the initial preference for a given verb to be either a main clause or reduced relative, but if this preference is incorrectly set to a main clause in the first analysis, reanalysis as a reduced relative will be impossible.

(42) ## IP

DP I

#The author studied# was

This contrasts with the unaccusative cases, as discussed above. These cases must insert a trace in the post verbal position whether or not the structure is interpreted as a main clause or as a reduced relative. Therefore, whether or not the main clause or relative clause reading is intially chosen (perhaps based on frequency) reanalysis is possible. If this analysis is correct, we are driven to a theory where frequency information interacts with grammatically based principles, but frequency does not replace these principles.

VII. Conclusions:

In this paper we have argued for a theory of processing preference and reanalysis that is heavily based on independently needed conditions within Chomsky's grammatical theory. There are no independent "parsing principles." In this case, the theory of preference is grounded in the Economy Conditions of Chomsky's (1993) minimalist theory. We contrasted our approach with one proposed by Collin Phillips. These theories are similar in that principles are all independently motivated by grammatical considerations. We argued however that

these economy conditions allow us to derive the unmotivated 'shortest path' portion of Phillips' <u>Branch Right</u>. The principle of Least Effort discussed above favors feature passing that involves the minimal number of steps. Next, we follow Uriagereka in eliminating the induction step of the LCA in favor of a theory involving multiple spellouts. We have shown show that multiple Spell-Out when combined with the independently motivated economy conditions also provides a motivation for the preference for right branching structures and an independently motivated theory of reanalysis. In the last section, we argued that these principles interact with frequency derived parsing constraints in interesting ways and can explain subtle differences between the garden path status of reduced relatives derived from unergatives, unaccusatives,and transitives that are otherwise mysterious. This argues in turn for a theory where grammatical principles are supplemented, but not replaced by considerations of frequency or probablility.

References:

Chomsky, N. (1993) “A minimalist program for linguistic theory in K. Hale and S.J. Keyser eds, The View from Building 20: Essays in Honor of Sylvain Bromberger. MIT Press.

Frazier, Lyn and K. Rayner (198) “Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences.” Cognitive Psychology 14. pp. 178-210.

Gibson, E. (1991) A Computational Theory of Human Language Processing: Memory Limitations and Processing Breakdown. unpublished Carnegie-Mellon PhD dissertation.

Gorrell, Paul (1995) Syntax and Parsing. Cambridge University Press.

Hale, K. and SJ Keyser (1993) On Argument Structure and lexical Expression of Syntactic relations.” in SJ Keyser, ed. The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger. MIT Press.

Jackendoff(1972) Semantic Interpretation in Generative Grammar. MIT Press.

Jurafsky, Daniel (1996) “ A Probablistic Model of Lexical and Syntacttic Access and Disambiguation.” in Cognitive Science pp. 137-194.

Kayne, R. (1994) The Antisymmetry of Syntax. MIT Press.

Larson, R.(1990) “Double objects revisited: reply to Jackendoff . Linguistic Inquiry 21. 589-632.

MacDonald, ME, D. Pearlmutter, and M. Seidenberg (1994) “The lexical Nature of Syntacttic Ambiguity resolution” Psychological Review 101 pp.678-703.

Merlo, P. (1994) “A Corpus based Analysis of verb Continuation Classes for Syntactic Processing.” Journal of Psycholinguistic Research 23:6 pp. 676-703.

Phillips, Colin, (1995) “Right Association in parsing and grammar.” in C. Schutze, J. Ganger, and K. Broiher, eds., Papers in Language Processing and Acquisition. MITWPL. 26, pp. 37-93.

Phillips, Colin (1996) Order and Structure. unpublished MIT PhD dissertation.

Phillips, C. and E. Gibson (in press) “ On the strength of the local attachment preference.”
Journal of Psycholinguistic Research .

Pritchett, B. (1992) Grammatical Competence and Parsing Performance. University of Chicago Press.

Steedman, Mark(1996) Surface Structure and Interpretation. MIT Press.

Stevenson, S. (1993) “A Competition-based explanation of syntacttic attachment preferences and garden path phenomena.” Proceedings of the 31st Annual Association for Computational Linguistics. pp. 266-273.

Stevenson, S. and P. Merlo(1997) “Lexical Structure and Processing Complexity.” in Language and Cognitive Processes vol 12.2/3, pp. 349-399.

Trueswell. JC & Tanenhaus, MK (994) “Towards a lexicalist Framework for Constraint Based Syntacttic Ambiguity Resolution. in C. Clifton, L Frazier, and K Rayner, eds. Perspectives on Sentence Processing. pp. 155-179.

Uriagereka (this volume) “Multiple Spell-out

Weinberg (1992) “Parameters in the theory of sentence Processing: Minimal Commitment Theory Goes East”. Journal of Psycholinguistic Research 22.3 pp. 339-364.

Notes

---

[1] The author gratefully acknowledges support from the following agendcies: The Army Research Laboratory under contract DAAL0197K0135, The National Security Agency under contract MDA90496C1250, and MicroAnalysis and Design under SBIR contract 7030. Thanks to the audiences at the CUNY Sentence Processing and Computational Psycholinguistics Conferences as well as audiences at Yale University and the Basque Country Summer Courses, as well as to Robert Berwick, Stephen Crain, Samuel Epstein, and Juan Uriagereka for helpful discussion. Errors remain my own.

[2] See Jackendoff((1972)

[3] But see Steedman(1996) for a different view with in the framework of Combinatory Categorial Grammar).

[4] This point will become important in our discussion of Phillips below.

[5] The complement cannot be identified as an ECM construction until the complement head is encountered,

so case can’t be assigned by the higher verb ‘beleive’ either.

[6] Stevenson and Merlo explain this contrast using a lexical theory devised by Hale and Keyser (1993). This theory is embedded in the processing model proposed in Stevenson (1993). The main claim is that the lexical representations needed to process unergative reduced relative clauses cannot be computed by a Stevenson-type model, given the Hale-Keyser lexical representations. The account is extremely interesting, however it is too strong in that even unergatives can be improved in the appropriate discourse context as shown in (a)

(a) Only horses raced at good tracks are eligible for the Triple Crown**.**

I saw a horse raced with a good jockey, and a horse raced with a bad one.

A horse raced with a good jockey will learn to go for the inside

[7] This type of improvement due to different types of quantification was first pointed out in Crain and Steedman (1985) and has been elaborated on in much subsequent work by these authors.

[8] Phillips(1996) pg 29.

[9] ibid

[10] Phillips(1996) pg. 111

[11] This is not to say that the initial preference as either a main or relative clause cannot be shifted. This is merely to say that the initial preference is difficult to sense once it is set.

[12] Phillips ibid pg. 161.

[13] MacDonald Pearlmutter and Seidenberg (1994) pg. 700

[14] Jurafsky(1996) explores use of construction type frequency.

[15] Space wil prevent us from fully contrasting a grammatically based account with the constraint based theory. Important topics that we will not have a chance to review or contrast involve the importance of pragmatic or semantic factors. Do pragmatic factors interact with frequency or syntactically derived constraints in a modular or non modular way? Our goal in this section wil simply be to discuss the role of grammatical principles in a theory where we assume that frequency plays some role.

---

[16] All tables from Merlo and Stevenson(1966).