

An Emergentist Approach to Syntax*

William O'Grady

1. Introduction

The preeminent explanatory challenge for linguistics involves answering one simple question—how does language work? The answer remains elusive in the face of the extraordinary complexity of the puzzles with which we are confronted. Indeed, if there has been a consensus in the last half century of work in formal linguistics, it is probably just that the properties of language should be explained by reference to principles of grammar. I believe that even this may be wrong, and that emergentism may provide a more promising framework for understanding the workings of language.

In its most fundamental form, emergentism holds that the complexity of the natural world results from the interaction of simpler and more basic forces. In this spirit, emergentist work in linguistics has been pursuing the idea that the core properties of language are shaped by non-linguistic propensities, consistent with Bates & MacWhinney's (1988:147) suggestion that language is a 'new machine built out of old parts.' O'Grady (2008a,d) presents an overview of some recent emergentist contributions to the study of language.

Syntax constitutes a particularly challenging area for emergentist research, as traditional grammar-based frameworks have reported significant success in their analysis of many important phenomena. This chapter reconsiders a number of those phenomena from an emergentist perspective with a view to showing how they can be understood in terms of the interaction of lexical properties with a simple efficiency-driven processor without reference to grammatical principles.

The ideas that I wish to put forward rest on two key claims, which can be summarized as follows.

- Syntactic theory can and should be unified with the theory of sentence processing.
- The mechanisms that are required to account for the traditional concerns of syntactic theory (e.g., the design of phrase structure, pronoun interpretation, control, agreement, contraction, scope, island constraints, and the like) are identical to the mechanisms that are independently required to account for how sentences are processed from ‘left-to-right’ in real time.

The proposed unification thus favors the theory of processing, which for all intents and purposes simply subsumes syntactic theory.

A metaphor may help convey what I have in mind. Traditional syntactic theory focuses its attention on the architecture of sentence structure, which is claimed to comply with a complex grammatical blueprint. In Principles and Parameters theory, for instance, well-formed sentences have a Deep Structure that satisfies the X-bar Schema and the Theta Criterion, a Surface Structure that complies with the Case Filter and the Binding Principles, a Logical Form that satisfies the Empty Category Principle, and so on. The question of how sentences with these properties are actually built in the course of language use is left to a theory of ‘carpentry’ that includes a different set of mechanisms and principles (parsing strategies, for instance).

I propose a different view. Put simply, there are no architects; there are only carpenters. They design as they build, limited only by the materials available to them and by the need to complete their work as quickly and efficiently as possible. Indeed, drawing on the much more detailed proposals put forward in O’Grady (2005), I suggest that

efficiency is THE driving force behind the design and operation of the computational system for language.

2. Representations

As a first approximation, I assume that the investigation of the human language faculty requires attention to at least two quite different cognitive systems—a lexicon that draws primarily on the resources of declarative memory, and a computational system whose operation is supported by working memory, sometimes called procedural memory (Ullman 2001).

I adopt a very conventional lexicon that serves as a repository of information about a language's words and morphemes, including information about their category membership (N, V, etc.¹) and their combinatorial propensities. Thus, the entry for *drink* indicates that it is a verb and that it takes two nominal arguments. ('N' here stands for 'nominal category,' not just 'noun.')

(1) *drink*: V, <N N>
 ↑ ↑
 category argument grid

The computational system operates on these words and morphemes, combining them in particular ways to construct phrases and sentences, including some that are extraordinarily complex. Its operation is subject to the following simple imperative.

(2) Minimize the burden on working memory.

I take working memory to be a pool of operational resources that not only holds representations but also supports computations on those representations (e.g., Carpenter, Miyake & Just 1994, Jackendoff 2002:200). An obvious consequence of seeking to minimize the burden on these resources is that the computational system should operate in the most efficient manner possible, carrying out its work at the first opportunity.

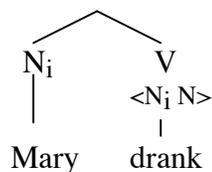
(3) The Efficiency Requirement

Dependencies (lexical requirements) must be resolved at the first opportunity.

As we will see as we proceed, many core properties of English (and, presumably, other languages) follow from this simple constraint, opening the door for a memory- and processing-based emergentist account of syntax.

In forming a sentence such as *Mary drank water*, the computational system begins by combining the verb *drink* with the nominal to its left, yielding the representation depicted below. (I assume that categories are ‘directional’—in English, a verb looks to the left for its first argument and to the right for subsequent arguments; a preposition looks rightward for its nominal argument; and so forth.)

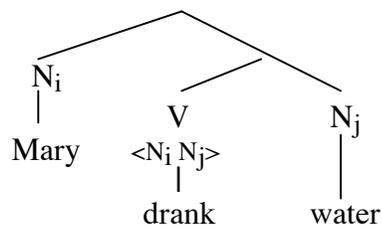
(4) Step 1: Combination of the verb with its first argument



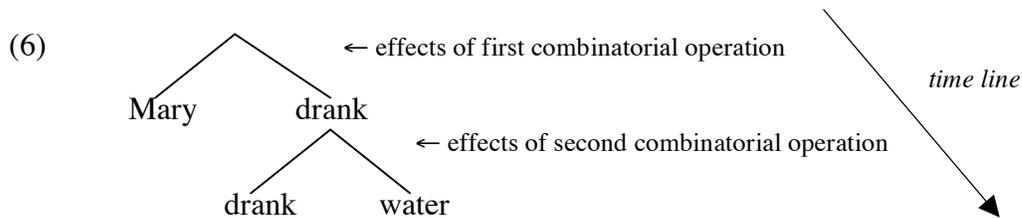
The resolution of an argument dependency is indicated by copying the nominal's index (representing its interpretation, as in Sag & Wasow 1999:106-08) into the verb's argument grid. Thus, the index of *Mary* in (4) is copied into the first position of the grid of *drink* at the point where the two are combined.

The computational system then proceeds to resolve the verb's second argument dependency by combining the verb directly with the nominal to its right, giving the result depicted below.

(5) Step 2: Combination of the verb with its second argument



Syntactic representations formed in this way manifest the familiar binary-branching design, with the subject higher than the direct object, but not because of a grammatical blueprint like the X-bar Schema. As I see it, syntactic structure is nothing but a fleeting residual record of how the computational system goes about combining words—one at a time, from left to right, in accordance with the demands of the Efficiency Requirement. Thus, the structure in (4) exists only as a reflex of the fact that the verb combined with the nominal to its left as soon as there was an opportunity to do so. And the structure in (5) exists only because the verb then went on to combine with the nominal to its right as soon as the opportunity arose. A more transparent way to represent these facts (category labels aside) might be as follows.



The time line here runs diagonally from left to right, with each ‘constituent’ consisting of the verb-argument pair acted on by the computational system at a particular point in the sentence’s formation.

3. Binding

Pronoun reference has long occupied an important place in theorizing about the computational system for language. The centerpiece of traditional UG-based theories is Principle A, which requires that reflexive pronouns be bound (i.e., have a c-commanding² antecedent), roughly in the same minimal clause. Thus, (7a) is acceptable, but not (7b) or (7c).

(7) a. The reflexive pronoun has a c-commanding antecedent in the same clause:

Harry_i described himself_i.

b. The reflexive pronoun has a non-c-commanding antecedent in the same clause:

*[Harry’s_i sister] described himself_i.

c. The reflexive pronoun has a c-commanding antecedent, but not in the same clause:

*Harry_i thinks [_S Helen described himself_i].

In the computational system that I propose, Principle A effects follow from the Efficiency Requirement. The key assumption is simply that reflexive pronouns introduce a referential dependency—that is, they require that their reference be determined by another element. In order to see how this works, let us assume that referential dependencies are represented by ‘variable indices’ drawn from the latter part of the Roman alphabet (i.e., x, y, z). Thus, the reflexive pronoun *himself* has the representation below, with the index x representing the referential dependency.

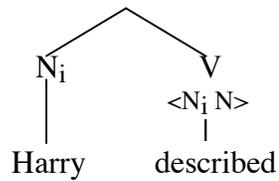
$$(8) \quad \begin{array}{c} N_x \\ | \\ \text{himself} \end{array}$$

Consistent with the Efficiency Requirement, this referential dependency must be resolved at the first opportunity. But when and how do such opportunities arise? The prototypical opportunity presents itself under the following circumstances:

- (9) The computational system has an opportunity to resolve a referential dependency when it encounters the index of another nominal.

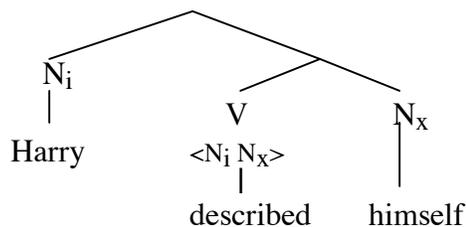
Consistent with the proposal outlined in section 1, the computational system initiates the formation of a sentence such as *Harry described himself* by combining the nominal *Harry* with the verb and copying its index into the verb’s argument grid, yielding the structure depicted below.

(10)



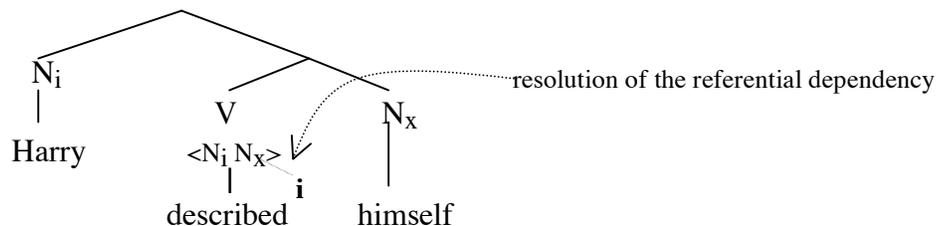
Next comes combination of the verb with its second argument, the reflexive pronoun *himself*, whose index is then copied into the verb's grid in the usual way.

(11)



This in turn creates an opportunity for the immediate resolution of the pronoun's referential dependency with the help of the index that is already in the verb's argument grid (i.e., the index of *Harry*). That is:

(12)



Given the Efficiency Requirement, no other result is possible. The verb has the opportunity to resolve its second argument dependency by combination with *himself*, so it must do so. And the reflexive pronoun has the opportunity to immediately resolve its referential dependency via the index already in the grid of the verb with which it combines, so it must do so. Anything else would be inefficient.

Now consider the unacceptability of sentences (7b) and (7c), repeated from above.

- (7) b. *[Harry's_i sister] described himself_i.
 c. *Harry_i thinks [_S Helen described himself_i].

In the case of (7b), the computational system proceeds as follows.

- (13) Step 1: Combination of *Harry* and *sister*

[Harry's_i sister]_j

Step 2: Combination of *Harry's sister* with the verb; the index of the argument phrase is copied into the grid of the verb.

[Harry's_i sister]_j described
 <N_j N>

Step 3: Combination of the verb with its second argument, the reflexive pronoun *himself*; resolution of the referential dependency by the index already in the grid of the verb.

[Harry's_i sister]_j described himself_x.
 <N_j N_x>
 ↓ resolution of the referential dependency
 *j

If the pronoun's referential dependency is not resolved by the index in the verb's grid in this manner, the Efficiency Requirement is violated. And if it is resolved in this way, the

sentence is semantically anomalous because of the gender mismatch between *himself* and *Harry's sister*. In either case, the sentence is unacceptable.

A similar problem arises in the case of (7c). Here, the first opportunity to resolve the referential dependency associated with the reflexive pronoun arises right after the computational system combines *himself* with the verb *describe*, whose argument grid contains the index of its subject argument *Helen*.

- (14) Combination of the embedded verb and its second argument, the reflexive pronoun *himself*, resolution of the referential dependency by the index already in the grid of the verb.

Harry_i thinks [Helen_j described himself_x]

<N_j N_x>

↓

resolution of the referential dependency

*j

If the index of *Helen* is used to resolve the referential dependency introduced by *himself*, a gender anomaly arises. If the index of *Helen* is not used, there is a violation of the Efficiency Requirement. Either way, the sentence is unacceptable.

3.1 Plain pronouns

But what of plain pronouns such as *him* and *her*? In the classic Binding Theory, they are subject to a constraint (Principle B) that ensures that they cannot have a c-commanding antecedent in the same clause—hence the unacceptability of sentences such as the following.

(15) *Harry_i described him_i.

The key observation is that there is no principled limit on the set of potential antecedents for a plain pronoun—*him* in (15) could refer to anyone who is made salient by the discourse and/or the background knowledge of the speaker and hearer. It is therefore evident that the interpretation of plain pronouns falls outside the domain of the sentence-level computational system whose drive for quickness limits it to the consideration of ‘local’ antecedents, as we have seen.

It is generally agreed that the interpretation of plain pronouns falls to a cognitive system—call it the ‘pragmatic system’ for convenience—whose primary concern is discourse salience and coherence (e.g., Kehler 2002). We can represent this intuition as follows, with ‘ $\rightarrow P$ ’ indicating that the interpretation of the referential dependency introduced by the plain pronoun is passed from the sentence-level computational system to the pragmatic system for resolution.

(16) Harry_i described him_{x \rightarrow P}

Why then can the pragmatic system normally not be used to select the salient nominal *Harry* as antecedent for *him* in (16)? Because, I propose, the pragmatic system—with its much wider range of options and its much larger domain—places a greater burden on working memory than does the sentence-level computational system, whose operation is far more locally focused. Plain pronouns are therefore COMPUTATIONALLY less efficient,

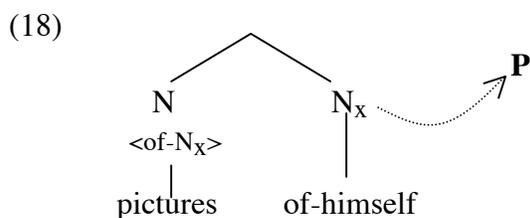
and their use is shunned where the more efficient alternative – a reflexive pronoun – is available. Thus (16) is unacceptable with *him* referring to Harry simply because the same interpretation could be achieved more efficiently via the reflexive pronoun, as in *Harry described himself*. (See Reinhart 1983:166 and Levinson 1987:410 for a similar suggestion from a pragmatic perspective.)

3.2 ‘Long-distance’ reflexives

A long-standing puzzle for theories of pronoun interpretation stems from the fact that reflexive pronouns may sometimes take a ‘long-distance’ antecedent. The pattern in (17) offers a typical example.

(17) John_i insisted that [pictures of himself_i] had appeared in yesterday’s newspaper.

As (18) below illustrates, immediate resolution of the referential dependency introduced by the reflexive pronoun is impossible in this case since the noun with which it combines has no other index in its argument grid. As a result, the computational system—which is compelled to act with alacrity, or not at all—passes the referential dependency to the pragmatic system for resolution there.



This creates the illusion that the anaphor is somehow ‘exempt’ (in the sense of Pollard & Sag 1992) from grammatical principles. In fact, no grammatical principles were ever in play; the phenomenon of long-distance anaphora simply reflects the inaction of the efficiency-driven computational system.

Because the domain of the pragmatic system is far broader than that of the sentence-based computational system, the eventual antecedent of the anaphor in a pattern such as (18)—selected with attention to discourse salience—may even lie in another sentence.

(19) Antecedent outside the sentence:

Larry_i had left his room in a terrible state. Pictures of himself_i lay on the floor, the dishes had not been washed, the bed was unmade...

The fact that reflexive pronouns in contexts such as this are dealt with by the pragmatic system and can therefore be associated with a distant antecedent dramatically reduces the computational advantage that ordinarily makes them preferable to plain pronouns. This opens the door for competition between reflexive and plain pronouns, as the following example illustrates. (O’Grady 2005:40ff considers a much broader range of cases.)

(20) Larry_i had left his room in a terrible state. Pictures of **himself/him**_i lay on the floor, the dishes had not been washed, the bed was unmade...

Table 1 summarizes the contrast between the two types of pronouns in the system I propose.

Table 1. Plain and Reflexive Pronouns in English

How the referential dependency is dealt with	Type of pronoun
Immediate resolution by the computational system	Reflexive pronoun is obligatory; plain pronoun is forbidden
No opportunity for immediate reso- lution by the computational system; recourse to the pragmatic system	Reflexive pronoun and plain pronoun may alternate with each other

In sum, there are no binding principles per se—that is, no autonomous grammatical constraints on coreference. The interpretive facts for which such principles have traditionally accounted emerge from more fundamental computational factors. As we have seen, the constraints embodied in Principle A simply follow from the Efficiency Requirement—reflexive pronouns are just words whose referential dependencies must be resolved at the first opportunity (immediately, if possible). And plain pronouns are just words whose referential dependencies escape the immediate interpretive action typical of the sentence-level computational system, relying instead on resolution by a pragmatic system that is sensitive to factors such as perspective and salience rather than the burden on working memory.

4. Control

Now let us consider the status of so-called ‘control structures’ such as (21), in which the subject argument of the embedded verb is not overtly expressed.

(21) Harry hopes [to succeed].

The key intuition here is that there are two ways to ‘project’ or express an argument requirement. On the one hand, it can be expressed as a categorial dependency—i.e., as a dependency that is resolved by combination with an overt nominal, as happens in the case of finite verbs (e.g., *Harry succeeded*, *Mary drank water*, etc.).

(22) V [+fin]: <N ...>

Alternatively, an argument requirement may be projected as a referential dependency (see the preceding section), as illustrated below.

(23) V [-fin]: <x ...>

This idea, which is similar in spirit to proposals found in Starosta (1988) and Sag & Pollard (1991), contrasts with the more commonly held view that subjects of infinitival verbs are expressed by PRO, a free-standing null pronoun. If we are on the right track, it should be possible to dispense with control theory, deriving its effects from more basic forces.

The two most important generalizations of traditional control theory are as follows (e.g., Chomsky 1981, Manzini 1983).

(i) The covert subject of an infinitival clause in complement position is coreferential with an argument of the immediately higher verb—with *Jean*, but not *Tim*, in in the following sentence.

(24) Tim_i thinks that [$Jean_j$ decided [$PRO_{j/*i}$ to leave].

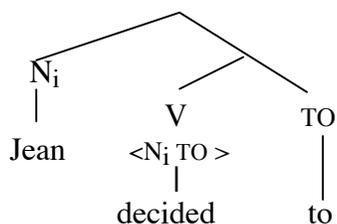
(ii) The covert subject of an infinitival clause in subject position can be interpreted pragmatically. Thus the sentence below can have the interpretations ‘for anyone to leave now,’ ‘for him to leave now,’ or ‘for us to leave now.’

(25) Tim thinks that [[PRO_i to leave now] would be impolite].

These generalizations follow automatically from the manner in which the efficiency-driven computational system seeks to resolve dependencies, namely at the first opportunity.

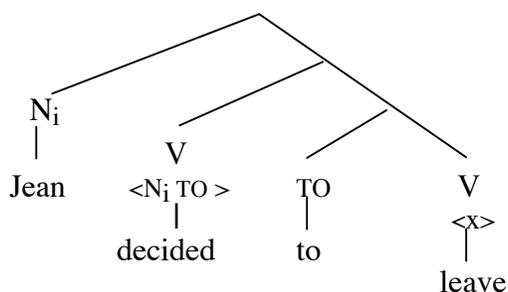
Let us begin with patterns such as *Jean decided to leave*, in which the unexpressed agent argument of the infinitival verb is obligatorily coreferential with the subject of the matrix verb. Just prior to the addition of the embedded verb, the sentence has the structure depicted below. (I assume that the infinitival marker *to* belongs to a single-member category that I will label ‘TO.’)

(26)



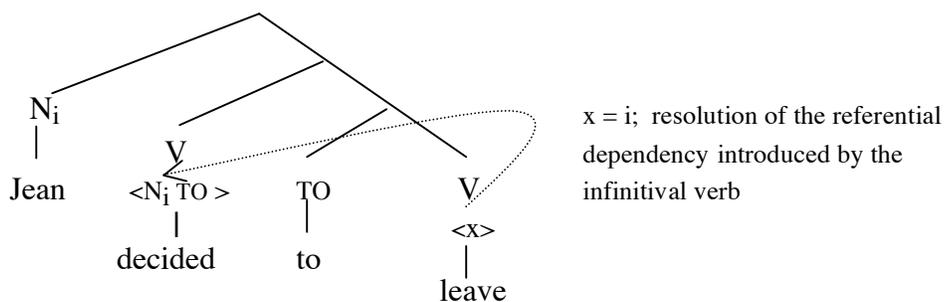
The embedded verb is then added, introducing a referential dependency (represented as x) that corresponds to its subject argument.

(27)



This referential dependency can be resolved instantly and locally, thanks to the presence of the index of *Jean* in the argument grid of the matrix verb *decide*.

(28)



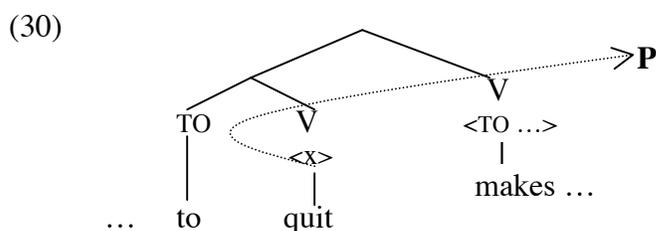
This is the only result compatible with the Efficiency Requirement; long-distance and sentence-external antecedents are thus automatically ruled out in this case.

Matters are quite different in patterns such as the following, in which the infinitival verb functions as first argument of the verb *make*.

(29) Jean said that [*to quit* makes no sense].

(='for Jean to quit now...', 'for us to quit now...', 'for anyone to quit now...')

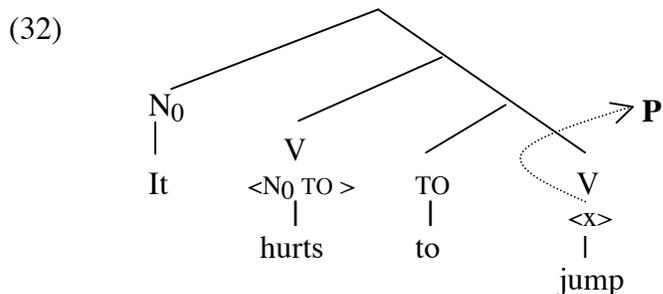
As illustrated below, *make* has no index in its argument grid at the point at which it combines with the infinitival phrase. In the absence of an immediate opportunity to resolve the referential dependency associated with the infinitival verb, it is transferred to the pragmatic system. This in turn opens the door for the observed range of nonlocal interpretations.



Now consider patterns such as (31), in which the infinitival combines with a verb whose only other argument is the expletive *it*.

(31) It hurts [*to jump*].

By definition, expletives do not have referents and thus cannot have referential indices — a property that I will represent by assigning them the ‘dummy’ index 0 . Sentence (31) therefore has the structure depicted below just after addition of the embedded verb.



Given the absence of a referential index in the argument grid of *hurt*, the referential dependency introduced by the infinitival verb cannot be satisfied by sentence-level computational mechanisms. It is therefore transferred to the pragmatic system for eventual resolution there, giving the desired generic and logophoric interpretations (‘It hurts when one jumps’ and ‘It hurts when I/you jump’).

In sum, the core properties of control theory appear to follow straightforwardly from the workings of the same computational system that is used to build syntactic representations and to resolve the sorts of referential dependencies associated with reflexive pronouns. The key idea is simply that the computational system seeks to resolve the referential dependency corresponding to the subject argument of an infinitival verb at the first opportunity. As we have seen, this gives the correct result in an important set of cases: the dependency is resolved by an index in the argument grid of the matrix verb when such an index is available (as in *Jean decided to leave*) and is otherwise resolved pragmatically, resulting in the generic or logophoric interpretation observed in the

examples considered above. O’Grady (2005, chs. 4 & 5) examines many other cases, including the contrast between control and raising.

5. Agreement

As a first approximation, English seems to require a match between a verb’s person and number features and those of its subject. (For the sake of exposition, I use Roman numerals and upper case for nominal features, and arabic numerals and lower case for verbal features.)

(33) Third person singular subject, third person singular verb form:

One remains.

IIISG 3sg

(34) Third person plural subject, third person plural verb form:

Two remain.

IIIPL 3pl

Agreement reflects the interaction of lexical and computational factors. On the lexical side, inflected verbs can introduce an ‘agreement dependency’—they carry person and number features that must be matched at some point with compatible features elsewhere in the sentence.

- (37) *We visits Harvey every day.
 IPL 3sg

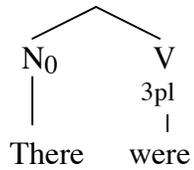
Because the presence of person and number features on the verb's first argument creates an opportunity to resolve the verb's agreement dependencies, either the computational system must bypass that opportunity, in violation of the Efficiency Requirement, or it must ignore the feature clash between the first person plural subject and the third person singular verb. Neither option can lead to an acceptable result.

The end result of all of this is that verbal agreement will be subject-oriented in all but one type of pattern. As illustrated in the following example, English verbs whose first argument is the featureless expletive *there* agree with their second argument.

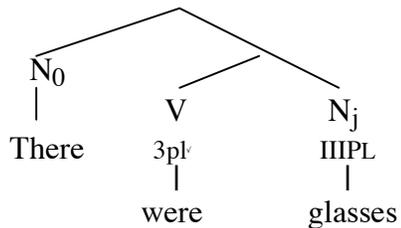
- (38)a. There was **glass** on the floor.
 b. There were **glasses** on the floor.

Our computational system offers a straightforward explanation for this: because the expletive *there* is featureless, it offers no opportunity for the verb to resolve its agreement dependencies. As illustrated in (39), the first opportunity to resolve these dependencies therefore arises at the point where the verb combines with its complement.

- (39) Step 1: Combination with *there* resolves the verb's first argument dependency, but offers no opportunity for resolution of its agreement dependencies.



- Step 2: Combination with *glasses* resolves the verb's second argument dependency and its agreement dependencies.



5.1 Agreement and coordination

A particularly striking agreement phenomenon arises in the case of coordinate structures such as the following, where the verb can sometimes agree with the first nominal inside a conjoined phrase.³

- (40) There **is** [paper and ink] on the desk.

The computational system builds this sentence as follows.

- (41) Step 1: Combination of the verb with its expletive subject. Because *there is* featureless, there is no opportunity to resolve the verb's agreement dependencies here.

[There is]
3sg

Step 2: Combination of the verb with the first conjunct of its second argument; resolution of the agreement dependencies

There [is paper]
3sg' IIISG

Step 3: Addition of the conjunction

There is [paper and]
3sg' IIISG

Step 4: Addition of the second conjunct

There is [paper [and ink]]
3sg' IIIPL

The key step is the second one, where an opportunity arises to resolve the verb's agreement dependencies with the help of the first conjunct of the coordinate noun phrase. Taking advantage of this opportunity, as demanded by the Efficiency Requirement, results in singular agreement even though later addition of the second conjunct ends up creating a plural argument.

As expected, the singular agreement option is impossible where the first conjunct is plural, in which case the verb must carry the plural number feature in order to satisfy the demands of the Efficiency Requirement.

(42) There *is/are [papers and ink] on the desk.

3pl IIIPL

As also expected, partial agreement is possible only when the coordinate NP follows the verb. Where it appears to the left, and is therefore fully formed before the verb is encountered, partial agreement is impossible.

(43) [Paper and ink] are/*is on the desk.

IIIPL 3pl

A variety of otherwise puzzling cases of agreement in English and other languages are considered by O’Grady (2005:96ff; 2008b,c).

In sum, the workings of verbal inflection in English reveal that efficiency, not grammatical relations, drives the agreement process. A verb agrees with its ‘subject’ only when this NP provides the first opportunity to resolve the agreement dependencies. In cases where the subject has no person and number features, the verb agrees with its second argument—as illustrated by patterns containing the expletive *there* (*There is a man at the door* vs. *There **are** two men at the door*). And in cases where that NP is a coordinate phrase, we see an even more radical manifestation of the Efficiency Requirement—agreement with the first conjunct.

(46) **What* were you wondering [*which clothes* to do with]?

(cf. I was wondering [which clothes to do something with].)

Kluender & Kutas (1993; see also Kluender 1998) suggest that the ungrammaticality of such patterns stems from the burden they create for working memory. Because holding a *wh* dependency is difficult, they argue, working memory balks at having to deal with more than one *wh* phrase per clause—as it must do in *wh* island patterns.

There must be more to it than this, however, since some *wh* island patterns are in fact quite acceptable, as (47) illustrates (e.g., Richards 1997:40).

(47) *Which clothes* were you wondering [*what* to do with]?

This is problematic both for the Kluender–Kutas account and for standard syntactic accounts. Why should there be such a contrast?

O’Grady (2005:118ff) suggests that the answer may lie in how working memory stores information. One commonly mentioned possibility (e.g., Marcus 1980:39, Kempen & Hoenkamp 1987:245) is that working memory makes use of push-down storage—which simply means that the most recently stored element is at the top of the ‘memory stack’ and therefore more accessible than previously stored elements.

In the case of a sentence such as (46), *what* appears first and therefore ends up being stored lower in the stack than the later-occurring *which clothes*.

(48) **What* were you wondering [*which clothes* to do with]?

- a. The computational system encounters and stores *what*:

What ...
 MEMORY STACK
 → *what*

- b. At a later point, *which clothes* is encountered and stored at the top of the stack:

What were you wondering [*which clothes* ...
 MEMORY STACK
which clothes ←
what

This is the reverse of what the computational system needs for this sentence. This is because the first opportunity to resolve a *wh* dependency arises at the verb *do*, which has an open argument position corresponding to its direct object. For the sake of semantic coherence, *what* should be associated with that position (cf. *do what with which clothes*), but this is impossible since it is ‘trapped’ at the bottom of the memory stack.

(49) **What* were you wondering [*which clothes* to do with]?

memory stack
which clothes
what X →

This places the computational system in an untenable position—it must either associate *which clothes* with *do*, yielding a semantically infelicitous result (cf. *do which clothes with what*), or it must spurn the opportunity to resolve a *wh* dependency, in violation of the Efficiency Requirement. Neither option is viable.

No such problem arises in the relatively acceptable sentence in (47), repeated below as (50). Here, *which clothes* is stored first and therefore ends up lower in the stack than the later occurring *what*.

(50) *Which clothes* were you wondering [*what* to do with]?

- a. The computational system encounters and stores *which clothes*:

Which clothes ...
 MEMORY STACK
 → *which clothes*

- b. At a later point, *what* is encountered and stored at the top of the stack:

Which clothes were you wondering [***what ...***]
 MEMORY STACK
what ←
which clothes

This is a felicitous result, since the computational system needs access to *what* first.⁴

(51) *Which clothes* were you wondering [***what*** to do with]?

memory stack
what
which clothes

The prospects for processing accounts of other island effects are excellent, and work in this area has been underway for some time (Kluender & Kutas 1993, Kluender 1998, Hawkins 2004, Hoffmeister et al. 2007), sometimes in combination with pragmatic analysis (e.g., Deane 1991, Kuno & Takami 1993).

7. Processing

So far, our discussion has focused on the claim that important properties of various core syntactic phenomena follow from the drive to minimize the burden on working memory, as embodied in the Efficiency Requirement. This is a necessary first step toward our goal of reducing the theory of grammar to the theory of sentence processing, but it takes us only half-way to our objective. In order to complete the task, we must establish that the computational system described here and the processor posited by psycholinguists are one and the same. More precisely, we need to show that the processor has the properties that we have been ascribing to the system that does the work of the grammar.

A defining feature of work on sentence processing is the assumption that syntactic structure is built one word at a time from left to right. As Frazier (1987:561) puts it, ‘perceivers incorporate each word of an input into a constituent structure representation of the sentence, roughly as [it] is encountered’ (see also Frazier 1998:126 and Pickering & Traxler 2001:1401, among many others). This is just what one expects of a cognitive system that has to deal with complex material under severe time constraints. As Frazier & Clifton (1996:21) observe, the operation of the processor reflects ‘universally present memory and time pressures resulting from the properties of human short-term memory.’ Humans, they note, ‘must quickly structure material to preserve it in a limited capacity memory’ (see also Deacon 1997:292-3 & 331 and Frazier 1998:125).

But what does the psycholinguistic literature say about the resolution of referential dependencies, agreement dependencies, and *wh* dependencies? Are they in fact all resolved at the first opportunity?

Nicol & Swinney (1989) make use of a cross-modal priming task to investigate the processing of English pronouns. Experiments of this type call for subjects to indicate whether they recognize probe words that are flashed on a screen at various points as they listen to sentences. The key assumption, validated in previous work, is that subjects make quicker decisions about probe words that are semantically related to words that they have recently accessed.

Now, if referential dependencies are in fact resolved at the first opportunity, as demanded by the Efficiency Requirement, the reflexive pronoun *himself* should reactivate *the doctor for the team* (its antecedent) in a sentence such as (52). This in turn should result in a shorter reaction time for a semantically related probe word such as *hospital* that is presented right after the reflexive is heard.

probe here
↓

(52) The boxer told the skier [that the doctor for the team would blame **himself** for the recent injury].

Nicol & Swinney's results bore out this prediction: probe words that were semantically related to *doctor* had accelerated reaction times after *himself*. This is just what one would expect if referential dependencies are interpreted at the first opportunity (immediately, in these patterns). More recent work (e.g., Sturt 2003, Runner et al. 2006) confirms the promptness with which the processor acts on referential dependencies.

There is also good evidence for immediate resolution of the referential dependencies corresponding to the unexpressed subject argument of infinitival verbs. A particularly

promising study in this regard was carried out for Spanish by Demestre et al. (1999), who exploited the fact that the gender of the adjective *educado/educada* ‘polite’ in the following patterns is determined by the (unexpressed) first argument of *ser* ‘be’ — *María* in (53a) requires the feminine form of the adjective and *Pedro* in (53b) requires the masculine.

(53)a. Pedro ha aconsejado a María ser más **educada/*educado** con los trabajadores.

Peter has advised to Maria to.be more polite-Fem/Masc with the employees.

‘Peter has advised Maria to be more polite with the employees.’

b. María ha aconsejado a Pedro ser más **educado/*educada** con los trabajadores.

Maria has advised to Peter to.be more polite-Masc/Fem with the employees.

‘Maria has advised Peter to be more polite with the employees.’

Drawing on ERP data,⁵ Demestre et al. found a significant wave form difference right after the adjective for the acceptable and unacceptable patterns, with the gender mismatch triggering a negative-going voltage wave. As the authors note, gender agreement errors could not have been identified so quickly if the computational system had not already interpreted the unexpressed subject argument of the infinitival verb. This is exactly what one would expect if the referential dependencies involved in control patterns are resolved at the first opportunity, as required by the Efficiency Requirement.

Agreement dependencies also seem to be resolved as promptly as possible. In an ERP study, Osterhout & Mobley (1995) had subjects read sentences such as (54) and then judge their acceptability.

(54) *The elected officials hopes to succeed.

The agreement mismatch triggered an almost immediate positive spike in electrical activity that peaked about 500 milliseconds after the violation—the usual response to syntactic anomalies on this sort of task. A similar finding is reported by Coulson, King, & Kutas (1998). This suggests an attempt to resolve the verb's agreement dependencies as soon as an argument carrying person and number features is encountered, as suggested in section 5.

Finally, there is compelling evidence that *wh* dependencies too are resolved at the first opportunity. One such piece of evidence comes from the measurement of ERPs to determine at what point speakers perceive the anomaly in the second of the following two sentences.

(55)a. The business man knew [**which customer** the secretary called _ at home].

b. *The business man knew [**which article** the secretary called _ at home].

If in fact the *wh* dependency is resolved at the first opportunity, then the anomaly in (55b) should be discerned right after *call*—whose open argument position should trigger action by the processor. Working with visually presented materials, Garnsey et al. (1989)

uncovered a significant difference in the wave forms for the two sentences immediately after the verb, suggesting that this is indeed the point where the *wh* dependency is resolved. Evidence for early resolution of *wh* dependencies also comes from studies of eye-tracking (Traxler & Pickering 1996; Pickering & Traxler 2001) and cross-modal priming (Swinney et al. 1988). Pulvermüller (2000) discusses the neurological correlates of push-down storage, the other key processing assumption underlying my account of *wh* island effects.

In sum, there is good reason to think that the efficiency-driven computational system that is required on independent grounds to account for the defining characteristics of phenomena ranging from pronoun interpretation to agreement to filler-gap dependencies is the same system at work in sentence processing. This in turn points toward the viability of the central idea that I have been outlining: the theory of grammar should be subsumed by the theory of sentence processing.

8. Why aren't all languages the same?

This brings us to a crucial and difficult question. If a general efficiency-driven computational system is in fact responsible for determining how language works, then why aren't all languages the same? The answer is simple: they are—with respect to the mandate to resolve dependencies in an expeditious manner. This still leaves very significant room for variation.

For one thing, languages differ from each other in terms of the properties of their lexical items. Although the processor is compelled by internal memory-related considerations to resolve dependencies at the first opportunity, language-particular

factors determine whether (for instance) the opportunity to resolve a verb's first argument dependency will occur to the left (as in English *I ran*) or to the right (as in Tagalog *Tumakbo ako 'ran I'*).

Or consider the subtler example of reflexive pronouns. As we have seen (section 3), English reflexive pronouns simply introduce a referential dependency, which the sentence-level computational system then tries to resolve at the first opportunity. In other languages though, so-called 'reflexive pronouns' carry more than just a referential dependency; they are used to encode subtle pragmatic contrasts involving logophoricity and perspective, as Huang's (1994) detailed study of the Mandarin system demonstrates. These are not the sorts of factors that the computational system is able to deal with. The interpretation of such pronouns must therefore be passed to the pragmatic system, whose priorities lead to a very different type of syntax, as illustrated in the following examples from Kao (1993:37-38).

(56)a. 'Reflexive pronoun' with a non-local antecedent:

Sam_i renwei [Lisi_j bu yinggai piping ziji_{i,j}].

Sam think Lisi not should criticize self

'Sam_i thought that Lisi_j should not criticize self_{i,j}.'

b. 'Reflexive pronoun' with a non-c-commanding antecedent:

Lisi-de_j jiaoao hai le ziji_j.

Lisi-Gen pride hurt Asp self

'Lisi's_i pride hurt self_i.'

Languages may also vary with regards to the size of the burden on working memory that they are willing to tolerate. The syntax of *wh* dependencies offers a revealing example of how this works. As Hawkins (2004:192ff) documents in considerable detail, languages differ in a principled way with respect to the computational complexity of filler-gap relationships that extend across a clause boundary. Some languages (e.g., Russian) permit extraction only from reduced (infinitival) embedded clauses.

(57) the cucumber which I promised [INF to bring _]

Others (such as English) allow extraction from more elaborate tensed embedded clauses.

(58) the cucumbers which I promised [S that I would bring _].

And still others (e.g., Swedish) tolerate extraction from a tensed clause that is embedded inside an NP.

(59) a bone which I see [NP a dog [S which is gnawing on _]]

The crucial generalization seems to be this: if a language permits the computationally more demanding pattern, then it must also permit its less difficult counterparts. Thus, English permits not only (58) but also (57), and Swedish allows not just (59), but also (57) and (58).

This state of affairs is reminiscent of what we find in phonology, where it is recognized that considerations relating to ease of production and perception motivate a variety of processes— word-final devoicing, vowel harmony, intervocalic voicing, and so forth. At the same time, there is no expectation that all languages will have intervocalic voicing or manifest vowel harmony. Rather, it is understood that the effects of articulatory and perceptual pressures are manifested via an implicational logic—if the more ‘difficult’ sound is permitted in a particular position, then so is its less difficult counterpart. Thus if a language allows voiced obstruents in word-final position (English, but not German), it must also permit voiceless obstruents in that position. Comparable asymmetries, motivated by processing difficulty, are pervasive in syntax and make up a good deal of the principled variation found across languages.

9. The problem of language acquisition

Space does not permit an adequate discussion of language acquisition, which is seen by many as the premier explanatory challenge for linguistics, but a few brief remarks are in order.

For many years, the field has been split over the question of whether children’s early experience with language provides enough information to support induction of the knowledge needed to speak and understand a language (the so-called ‘poverty of stimulus’ claim). Essentially without exception, opponents of Universal Grammar hold that the input suffices, while proponents of UG adhere to the opposite position, maintaining that without access to innate grammatical principles it would be impossible to learn the intricate details of language.

I disagree with both views. As explained in detail in O’Grady (2005, 2008a), I do not believe that induction from experience can be the whole story: the facts are simply too complex, the input too sparse, mastery too rapid, and errors too infrequent. On the other hand, I do not see UG as the answer either. Rather, I hold that the gap between experience and the intricacies of language is bridged with the help of the processor, which directs learners to particular options that are not evident from information available in the input—the association of reflexive pronouns with antecedents in particular positions, intricate patterns of agreement, certain scopal preferences (O’Grady 2008a), and so on.

The key idea is simply that children are born with an efficient brain—that is, a brain that seeks to carry out computations in a way that minimizes the burden on working memory. Early in life, language learners are exposed to adult utterances that contain lexical items with particular properties arranged in a particular order. As input of this sort is encountered and processed over and over again, routines consisting of particular computational operations and series of operations develop. Language acquisition takes place as these routines are automatized through repeated use. (As noted by Paradis 2004:28 and Bybee & McClelland 2005:382, a side effect of this automatization is that processing is speeded up, helping to give language use its characteristic effortlessness.)

Some computational routines, such as the ones summarized in Table 2, are direct instantiations of the propensity to resolve dependencies at the first opportunity. As such, they do not need to be learned per se—they are simply associated with the appropriate lexical item(s). We therefore expect more or less error-free development, which is in fact the case for the phenomena in question (O’Grady 2005:193ff).

Table 2 Some efficiency-driven computational routines

Routine	Effect
Resolve the referential dependency introduced by <i>X-self</i> at the first opportunity.	<i>himself</i> = <i>John</i> , not <i>Bob</i> , in <i>Bob thinks John mistrusts himself</i> .
Resolve the referential dependency introduced by an infinitival verb at the first opportunity.	A local controller in patterns such as <i>Mary decided to leave</i> .

Other routines develop in response to language-particular facts that are neutral with respect to computational burden. For example, given repeated exposure to the word order of English, children will develop routines in which determiners look to the right for their nominal argument, while verbs look to the left for their first argument and to the right for their second argument, and so on. Table 3 lists some simple routines of this sort.

Table 3 Some language-particular computational routines

Routine	Effect
Looking to the right for the noun associated with a determiner	<i>the book</i> , not <i>*book the</i>
Looking to the left for a verb's first argument	<i>John ran</i> , not <i>*Ran John</i>
Looking to the right for a verb's second argument	<i>eat the apple</i> , not <i>*the apple eat</i>
Resolve a verb's agreement dependencies at the first opportunity.	Agreement with the first argument in <i>A man is here</i> , and with the second argument in <i>There is a man at the door</i> .

As computational routines of both types become increasingly fixed and rigid over time, they restrict the options available to the processor when it operates on the words and morphemes of English. This gives a result identical in its effects to having a grammar that imposes a locality requirement on reflexive pronouns, has a subject agreement rule, orders determiners before nouns, places verbs before direct objects, and so forth. But what appears to be a grammar is just a processor that has become set in its ways.

If this processor-based view of language acquisition is on the right track, then we would expect to find independent evidence for processing effects in the way in which language development unfolds. This matter is discussed at some length in O'Grady

(2005, 2008a), and I will draw briefly on that discussion here to consider the case of pronoun interpretation.

As explained in section 3, the interpretation of reflexive pronouns in English follows directly from the Efficiency Requirement—the choice of antecedent is determined in most cases simply by the need to resolve the referential dependency at the first opportunity. Given that efficiency is an inborn computational imperative, we predict early mastery of reflexive pronouns—allowing of course for the time needed to identify the lexical items that function in this capacity. Plain pronouns have the potential to be somewhat more troublesome, however, since their interpretation falls outside the domain of the sentence-level computational system and requires a sensitivity to non-local factors such as discourse salience, coherence, and so forth. Indeed, it is known that the interpretation of plain pronouns is associated with increased processing difficulty in at least some cases (Sekerina et al. 2004).

The developmental facts are also very intriguing in this regard. As established in a variety of comprehension studies (e.g., Wexler & Chien 1985), children typically interpret reflexive pronouns correctly more than 95% of the time from a very early age (as young as age 3). In contrast, performance on plain pronouns during the same period is usually significantly lower, hovering between 50% and 85%, depending on how carefully the supporting contexts are constructed (Conroy et al. 2008). Moreover, typical errors involve interpreting the pronoun in patterns such as *Donald Duck washed him* as coreferential with the subject—i.e., as if it were a reflexive pronoun. This suggests an initial preference for referential dependencies that can be resolved at the least cost—through the mediation of sentence-level mechanisms that are sensitive to immediacy and

locality. This is just what we would expect if the language faculty is shaped by the need to minimize the burden on working memory.

10. Does grammar exist?

The theory that I have been describing is emergentist in the sense outlined at the outset: it seeks to attribute the defining properties of language to more basic non-linguistic forces, particularly efficiency-related processing considerations. Such an approach raises the question of whether there is still a need for the type of cognitive system that we traditionally think of as a grammar. Two clarifications are in order.

First, I have no argument with the existence of a lexicon that contains information about category membership and selectional properties—in fact, I make use of just such a lexicon. My doubts about grammar pertain to the character of the combinatorial system that is responsible for the many aspects of a sentence’s form and interpretation that do not follow from lexical properties—the fact that reflexive pronouns require antecedents in particular positions, that verbs exhibit particular patterns of agreement, or that certain types of *wh* island patterns are unacceptable, for example.

The traditional view is that the rules and principles that regulate these phenomena must be distinguished from processing mechanisms. As Jackendoff (2002) observes, grammatical rules describe patterns of elements (p. 57); they say what the structure is, not how it is built (p. 31). According to this view, the grammar and the processor interact, but they are not the same thing (Fodor 1989:177ff, Frazier & Clifton 1996:9 & 25, Frazier 1998:126). Moreover, it is widely held that the processor is subordinate to the grammar: ‘the most basic assumption about the nature of the human sentence processor,’ Ferreira et

al. (2001:13) write, is ‘that it obeys the fundamental principles of grammar when constructing interpretations.’ All of this is of course incompatible with the view I have adopted.

A second clarification has to do with the two very different senses in which the term ‘grammar’ is used by those committed to its existence. Sometimes, ‘grammar’ is used to refer to the putative system of inborn constraints that gives the human language faculty its unique character (i.e., Universal Grammar). And sometimes it is used for the system of rules that account for the particular patterns and facts that distinguish one language from another (i.e., ‘language-particular grammars,’ as when someone talks about writing the grammar of Balinese or the grammar of Turkish).

My position with respect to UG is that it does not exist; a simple efficiency-driven linear processor lies at the heart of the human language faculty and carries the burden of explaining why language has the particular properties that it does and how those properties are acquired with such apparent ease by children. As outlined in earlier sections of this chapter, core properties of sentence structure, binding, control, agreement, extraction and other phenomena that have long been offered as *prima facie* evidence in support of UG seem to follow from the processor’s commitment to minimizing the burden on working memory by resolving dependencies at the first opportunity.

My position with respect to language-particular grammars is that they too do not exist, but here the point is a subtler one, since many ‘low-level’ processing routines embody the sorts of generalizations that could just as easily be stated by a grammatical rule: determiners look to the right for a noun, verbs look to the left for their first argument, and so forth. But processing routines are not just grammatical rules under

another name. They are real-time PROCESSES, not static statements about how elements are arranged. Moreover, they are independently required—speech, after all, has to be processed. Rules, if they exist, must do something more than restate facts about language that follow from the operation of the processor. Fodor (1978:470) puts it this way:

... there *must* be psychological mechanisms for speaking and understanding, and simplicity considerations thus put the burden of proof on anyone who would claim that there is more than this. To defend the more traditional view, what is needed is some sign of life from the postulated mental grammar.

I see no such signs.

The proposals that I have outlined in this chapter barely merit consideration—coverage is offered for only a tiny body of data, attention has been focused on a single language, apparent counterexamples abound, and so forth. But the central thesis, however ill fated it may be, is perhaps at least clear: a single efficiency-driven computational system offers a solution to the classic problems confronting both the theory of grammar and the theory of sentence processing. A sentence's design reflects the way it is built, not the other way around—there are no architects, just carpenters.

References

- Bates, Elizabeth, and MacWhinney, Brian. (1988). 'What is Functionalism?', *Papers and Reports on Child Language Development* 27: 137-52.
- Bybee, Joan, and McClelland, John. (2005). 'Alternatives to the Combinatorial Paradigm of Linguistic Theory Based on Domain General Principles of Human Cognition', *Linguistic Review* 22: 381-410.
- Carpenter, Patricia; Miyake, Akira; and Just, Marcel. (1994). 'Working Memory Constraints in Comprehension: Evidence from Individual Differences, Aphasia, and Aging', in M Gernsbacher (ed.), *Handbook of Psycholinguistics*. San Diego: Academic Press, 1075-1122.
- Chomsky, Noam. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- Conroy, Anastasia, Takahashi, Eri; Lidz, Jeffrey; and Phillips, Colin. (2008). 'Equal Treatment for All Antecedents: How Children Succeed with Principle B', Manuscript, Department of Linguistics, University of Maryland.
- Coulson, Seanna; King, Jonathon; and Kutas, Marta. (1998). 'Expect the Unexpected: Event-related Brain Response to Morphosyntactic Violations', *Language and Cognitive Processes* 13: 21-58.
- Deacon, Terrence. (1997). *The Symbolic Species*. New York: Norton.
- Deane, Paul. (1992). *Grammar in Mind and Brain: Explorations in Cognitive Syntax*. New York: Mouton de Gruyter.

- Demestre, Josep; Meltzer, Sheila; García-Albea, José; and Vigil, Andreu. (1999). 'Identifying the Null Subject: Evidence from Event-Related Brain Potentials', *Journal of Psycholinguistic Research* 28: 293-312.
- Ferreira, Fernanda; Christianson, Kiel; and Hollingworth, Andrew. (2001). 'Misinterpretations of Garden-Path Sentences: Implications for Models of Sentence Processing and Reanalysis', *Journal of Psycholinguistic Research* 30: 3-20.
- Fodor, Janet Dean. (1978). 'Parsing Strategies and Constraints on Transformations', *Linguistic Inquiry* 9: 427-73.
- Fodor, Janet Dean. (1989). 'Empty Categories in Sentence Processing', *Language and Cognitive Processes* 4: 155-209.
- Frazier, Lyn. (1987). 'Sentence Processing: A Tutorial Review', in M. Coltheart (ed.), *Attention and Performance XII: The Psychology of Reading*. Mahwah, NJ: Erlbaum, 559-86.
- Frazier, Lyn. (1998). 'Getting There (Slowly)', *Journal of Psycholinguistic Research* 27: 123-46.
- Frazier, Lyn, and Clifton, Charles. (1989). 'Successive Cyclicity in the Grammar and the Parser', *Language and Cognitive Processes* 4: 93-126.
- Frazier, Lyn, and Clifton, Charles. (1996). *Construal*. Cambridge, MA: MIT Press.
- Garnsey, Susan; Tanenhaus, Michael; and Chapman, Robert. (1989). 'Evoked Potentials and the Study of Sentence Comprehension', *Journal of Psycholinguistic Research* 18: 51-60.
- Hawkins, John. (2004). *A Performance Theory of Word Order and Constituency*. Cambridge, UK: Cambridge University Press.

- Hoffmeister, Philip; Jaeger, T. Florian; Sag, Ivan; Arnon, Inbal; and Snider, Neal. (2007). 'Locality and Accessibility in *Wh*-Questions', in S. Featherston & W. Sternefeld (eds.), *Roots: Linguistics in Search of its Evidential Base*. Berlin: Mouton de Gruyter, 185-206.
- Huang, Yan. (1994). *The Syntax and Pragmatics of Anaphora: A Study with Special Reference to Chinese*. Cambridge, UK: Cambridge University Press.
- Jackendoff, Ray. (2002). *Foundations of Language*. Oxford, UK: Oxford University Press.
- Kao, Rong-Rong. (1993). *Grammatical Relations and Anaphoric Structures in Mandarin Chinese*. Ph.D dissertation, University of Hawai'i at Manoa.
- Kehler, Andrew. (2002). *Coherence, Reference, and the Theory of Grammar*. Stanford University: Center for the Study of Language and Information.
- Kempen, Gerard, and Hoenkamp, Eric. (1987). 'An Incremental Procedural Grammar for Sentence Formation', *Cognitive Science* 11: 201-58.
- Kluender, Robert. (1998). 'On the Distinction between Strong and Weak Islands: A Processing Perspective', in P. Culicover & L. McNally (eds.), *The Limits of Syntax (Syntax and Semantics 29)*. San Diego: Academic Press, 241-79.
- Kluender, Robert, and Kutas, Marta. (1993). 'Subjacency as a Processing Phenomenon', *Language and Cognitive Processes* 8: 573-633.
- Kuno, Susumu, and Takami, Ken-Ichi. (1993). *Grammar and Discourse Principles*. Chicago: University of Chicago Press.

- Levinson, Stephen. (1987) 'Pragmatics and the Grammar of Anaphora: A Partial Pragmatic Reduction of Binding and Control Phenomena', *Journal of Linguistics* 23: 379-434.
- Manzini, Maria Rita. (1983). 'On Control and Control Theory', *Linguistic Inquiry* 14: 421-46.
- Marcus, Mitchell. (1980). *A Theory of Syntactic Recognition for Natural Language*. Cambridge, MA: MIT Press.
- Nicol, Janet, and Swinney, David. (1989) 'The Role of Structure in Coreference Assignment during Sentence Comprehension', *Journal of Psycholinguistic Research* 18: 5-19.
- O'Grady, William. (1997). *Syntactic Development*. Chicago: University of Chicago Press.
- O'Grady, William. (2005). *Syntactic Carpentry: An Emergentist Approach to Syntax*. Mahwah, NJ: Erlbaum.
- O'Grady, William. (2008a). 'Does Emergentism Have a Chance?', in H. Chan, H. Jacob, and E. Kapia (eds.), *Proceedings of 32nd Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 16-35.
- O'Grady, William. (2008b). 'Innateness, Universal Grammar, and Emergentism', *Lingua* 118: 620-31.
- O'Grady, William. (2008c). 'Language without Grammar', in P. Robinson & N. Ellis (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York: Routledge, 139-67
- O'Grady, William. (2008d). 'The Emergentist Program', *Lingua* 118: 447-64.

- Osterhout, Lee, and Mobley, Linda. (1995). 'Event-Related Potentials Elicited by Failure to Agree', *Journal of Memory and Language* 34: 739–773.
- Paradis, Michel. (2004). *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins.
- Pickering, Martin, and Traxler, Matthew. (2001). 'Strategies for Processing Unbounded Dependencies: Lexical Information and Verb-Argument Assignment', *Journal of Experimental Psychology* 27: 1401-10.
- Pollard, Carl, and Sag, Ivan. (1992). 'Anaphora in English and the Scope of Binding Theory', *Linguistic Inquiry* 23: 261-303.
- Pulvermüller, Friedemann. (2000). 'Syntactic Circuits: How does the Brain Create Serial Order in Sentences?', *Brain and Language* 71: 194-99.
- Reinhart, Tanya. (1983). *Anaphora and Semantic Interpretation*. Chicago: University of Chicago Press.
- Richards, Norvin. (1997). *What Moves Where When in Which Language?* Ph.D. dissertation, MIT.
- Runner, Jeffrey; Sussman, Rachel; and Tanenhaus, Michael. (2006). 'Processing Reflexives and Pronouns in Picture Noun Phrases', *Cognitive Science* 30: 193-241.
- Sag, Ivan, and Wasow, Thomas. (1999). *Syntactic Theory: A Formal Introduction*. Stanford University: Center for the Study of Language and Information.
- Sag, Ivan, and Pollard, Carl. (1991). 'An Integrated Theory of Complement Control', *Language* 67: 63-113.

- Sekerina, Irina; Stromswold, Karin; and Hestvik, Arild. (2004). 'How do Adults and Children Process Referentially Ambiguous Pronouns?', *Journal of Child Language* 3: 123–152.
- Sobin, Nicolas. (1997) 'Agreement, Default Rules, and Grammatical Viruses', *Linguistic Inquiry* 28: 318-43.
- Starosta, Stanley. (1988). *The Case for Lexicase*. New York: Pinter.
- Sturt, Patrick. (2003). 'The Time-Course of the Application of Binding Constraints in Reference Resolution', *Journal of Memory and Language* 48: 542-62.
- Swinney, David; Ford, Marilyn; Frauenfelder, Ili; and Bresnan, Joan. (1988). 'On the Temporal Course of Gap-Filling and Antecedent Assignment during Sentence Comprehension', in B. Grosz, R. Kaplan, M. Macken (ed.), *Language Structure and Processing*. Stanford University: Center for the Study of Language and Information.
- Traxler, Matthew, and Pickering, Martin (1996). 'Plausibility and the Processing of Unbounded Dependencies: An Eye-Tracking Study', *Journal of Memory and Language* 35: 454-75.
- Ullman, Michael. (2001). 'The Neural Basis of Lexicon and Grammar in First and Second Language: The Declarative/Procedural Model', *Bilingualism: Language and Cognition* 4: 105-22.
- Wexler, Kenneth, and Chien, Yu-Chin. (1985). 'The Development of Lexical Anaphors and Pronouns', *Papers and Reports on Child Language Development* 24: 138-49.

Footnotes

*This paper was first written in early 2001 as a summary of my then forthcoming book (O'Grady 2005). It has been revised and updated for appearance in this volume. I am grateful to Kevin Gregg and Mark Campana for extensive discussion, to Hsiu-chuan Liao for her comments on an earlier version of this paper, and to Miho Choo for her invaluable editorial assistance.

1. I leave open the possibility that categorial contrasts are reducible to a semantic base, perhaps along the lines proposed in O'Grady (1997:312ff).

2. X c-commands Y if the first category above X contains Y.

3. In a survey of twelve speakers conducted by Sobin (1997), the pattern with the plural form of the verb (*There are paper and ink ...*) received a mean rating of just .81 out of 5, compared to 3.58 for the pattern with the singular form of the verb.

4. This gives a nested dependency in the sense of Fodor (1978).

5. Event-related potentials are voltage peaks that arise in the course of sentence processing and are sensitive to various linguistic factors, including agreement mismatches.