Emergentism and Second Language Acquisition[*]

William O'Grady, Miseon Lee, & Hye-Young Kwak

1. Introduction

Language presents us with many puzzles. Why does it have the particular properties that it does? Why does it vary and change in certain ways, but not others? How is it acquired so quickly and with so little effort by pre-school children despite its apparent complexity? And why is the acquisition of a second language so difficult for adults, despite their intellectual sophistication and their access to carefully designed educational programs?

An attractive feature of approaches to language based on Universal Grammar (UG) is that they offer an integrated account of these puzzles—an inborn system of grammatical categories and principles gives language its defining properties, places limits on the ways in which it can vary and change, and explains how even the most complex phenomena are acquired with such ease by children. With the help of additional assumptions, it even appears possible to offer insights into why the acquisition of a second language proves so challenging.

Yet the UG-based program has encountered deep suspicion and resistance from many quarters during the half century that it has dominated explanatory work on language. For a significant segment of the professional linguistic community, it simply does not ring true. The objections vary with the commentator—UG principles are too abstract (Tomasello 2003:3-7), the type of nativism that UG seems to presuppose is biologically implausible (Elman et al. 1996, MacWhinney 2000), a focus on faculty-specific principles distances the study of language from the rest of cognitive science (Jackendoff 1988, 2002:xi-xii), the phenomena purportedly accounted for by UG theories are better explained in other ways (Hawkins 2004, O'Grady 2005, Haspelmath to appear), and so forth. But is there an alternative?

In recent years, much of the opposition to the UG program has coalesced around a loosely associated set of ideas that have come to be grouped together under the rubric of emergentism. Despite the very considerable diversity of emergentist thought, there seems to be at least one central thesis to which all of its various proponents adhere: the complexity of language must be understood in terms of the interaction of simpler and more basic non-linguistic factors.

In the case of language, it has been suggested that those factors include features of human physiology (the vocal tract, for instance), the nature of the perceptual mechanisms, the effect of pragmatic principles, the role of social interaction in communication, the character of the learning mechanisms, and

limitations on working memory and processing capacity—but not inborn grammatical principles.

The earliest emergentist work focused on the important problem of how children acquire a language in response to the sorts of experience typical of childhood. More recently, there has been growing interest in the relevance of emergentism to understanding second language acquisition as well, marked in part by the fact that three major journals have devoted special issues to the examination and evaluation of emergentist work on SLA—*Applied Linguistics* 27/4 (2006), under the editorship of Nick Ellis and Diane Larsen-Freeman, *Lingua* 118 (2008), edited by Roger Hawkins, and *The Modern Language Journal* 92/2 (2008), under the editorship of Kees de Bot. Although modest in comparison to work in the UG framework, the emergentist literature on SLA offers analyses for a representative range of phenomena, including grammatical morphology (Ellis 2006b), competition-based processing (MacWhinney 2008 and the references cited there), quantifier scope (O'Grady 2007), and *want to* contraction (O'Grady et al. 2008). See Ellis & Larsen-Freeman (2006) for some general discussion.

2. Emergentist approaches to language acquisition

Emergentist approaches to language acquisition can be divided into two types, depending on the dominant strategy that they adopt. On the one hand, there is a very influential and impressive body of research that focuses on the importance of the input (or usage) for understanding how language acquisition works. Ellis (2002, 2006a) provides a far-reaching discussion of this approach. On the other hand, a smaller body of research explores the role of the processor–working memory interface in language acquisition, addressing problems of learnability and development that have traditionally been the exclusive domain of UG-based work. O'Grady (2008a,c) offers an introduction to this approach. Let us consider each program in more detail.

2.1 Input-based emergentism

One of the earliest examples of a systematic input-based approach to language learning is the Competition Model put forward by Brian MacWhinney (MacWhinney 1987, Bates & MacWhinney 1987). This approach, which remains highly influential, offers a theory of how language learners come to identify and prioritize the various competing cues (word order, animacy, case, agreement, and so on) that are relevant to sentence comprehension. The key variables, MacWhinney suggests, are to be found in the input: how often the cue is present when a particular pattern is being interpreted (cue availability), and how often it points to a particular interpretation (cue reliability). In the case of English, for instance, word order is a highly available and reliable cue for identifying a sentence's subject—which almost always occurs preverbally. In contrast,

agreement is highly reliable (only subjects trigger agreement), but is often unavailable since there is so little inflection in English. The situation could well be reversed in a free word order language, where agreement (or case) might be both more available and more reliable than word order.

Pioneering work of a different sort in the input-oriented tradition has been carried out by Jeffrey Elman (e.g., 1993, 2002, 2005), who has employed the techniques of connectionist modeling to investigate the language acquisition process. This has led to a number of intriguing findings, including Lewis & Elman's (2001) demonstration that a Simple Recurrent Network (SRN)[1] can simulate the acquisition of agreement in English from data similar to the input available to children.

A focus on the input is also characteristic of many other scholars working in an emergentist framework. A recurring intuition is that the frequency with which particular phenomena are encountered plays a key role in shaping the developmental process. One of the strongest advocates of this view is Nick Ellis (e.g., 2002, 2006a,b), who holds that language learning is, in essence, 'the gathering of information about the relative frequency of form-function mappings' (2006a:1)—with provisos concerning perceptibility (see below), attention, interference from other languages, and the like.

Frequency also has an important role to play in the view of language acquisition associated with Construction Grammar and other usage-based theories. As Tomasello observes (2003:327), language acquisition in such theories 'depends crucially on the type and token frequency with which certain structures appear in the input'—an idea that has been put forward and developed in promising ways by Goldberg (1999), Goldberg & Casenhiser (2008), and Ambridge et al. (2006), among many others. A similar attention to frequency guides research in other areas of cognitive science too, including psycholinguistics and neurolinguistics (e.g., Dick et al. 2001, Ferreira 2003, Chater & Manning 2006, Chang et al. 2006) as well as the study of language change (e.g., Bybee & Hopper 2001).

In defense of frequency factors

In considering the role of input frequency in language acquisition (first or second), it is vital to bear in mind a key point: what counts is not how many times learners hear a particular form—it is how many times they encounter *mappings* between a form and its meaning. An illustration of why this distinction is important comes from the English determiner *the*.

Although *the* is the most frequent word in the English language, it is mastered relatively late, both in first language acquisition and second language learning.

---

[1] An SRN processes patterns of sequentially ordered elements, producing outputs that are a function of both the current input and the SRN's own internal state. SRNs are especially good at noticing local co-occurrence relationships—given the word X, what's the likelihood that the next word will be Y?

How can this fact be reconciled with the claim that frequency has a major role in shaping the language acquisition process? Two observations are crucial.

First, as noted by Ellis (2006b:171) grammatical functors are frequently difficult to perceive on the basis of 'bottom-up' (purely acoustic) evidence. Herron & Bates (1997) had subjects identify semi-homophonous function words and content words (the pronoun *I* and the noun *eye*, the auxiliary verb *will* and the noun *will*, and so on) that had been spliced into various contexts—some neutral, some appropriate for just the content word, and some appropriate for just the function word. They found that even adult native speakers of English were able to recognize the word out of its normal context as little as half the time.

Second, contextual indeterminacy often undermines the association of a form with its intended meaning. As Ionin et al. (2008) observe, for example, it is no small matter to determine that the *the–a* contrast in English turns on definiteness rather than specificity. Language learners must confront two sorts of problems. Not only is the contrast between definiteness and specificity a subtle one, definites are often specific, as in the following example from Ionin et al.

(1)   [+definite, +specific]
        I want to talk to the owner of this store—she is my neighbor, and I have an
        urgent message for her.

Is *the* used here because of definiteness—the speaker is referring to someone whose existence and uniqueness are established by general world knowledge (typically a store has a unique owner)? Or is it because of specificity—the speaker intends to refer to someone known to her (a neighbor for whom she has an urgent message)?

This is not to say that the function of *the* as a definiteness marker is never clear. Sometimes it is (as in *I want to talk to the owner, whoever that is),* but as Ionin et al. observe (pp. 573-574):

    Given the subtlety of the discourse triggers related to speaker and hearer
    knowledge, generalizing from them is likely to be a fairly long and
    difficult process.

In the final analysis then, although *the* may indeed occur very frequently, the transparent mappings between form and meaning that are needed by the acquisition device occur far less often—perhaps even quite rarely.

In the case of second language acquisition, the learner's sensitivity to frequency effects can be obscured by yet another set of considerations—transfer from the native language. Ellis (2006b) develops this point at considerable length, examining the effects of L1-related cue competition, salience, attentional tuning, overshadowing and blocking.

2.2 Processor-based emergentism

The starting point for processor-based emergentism is the view, put forward by Hawkins (2004) and O'Grady (2005), that key properties of the syntactic phenomena that have long been used as support for UG-based approaches to language are in fact better explained in terms of processing factors. Hawkins develops this idea for a number of phenomena central to typology, such as Greenbergian universals and cross-linguistic variation in the syntax of filler-gap dependencies (including the body of relativization facts traditionally described by the Keenan-Comrie markedness hierarchy). O'Grady's work focuses more directly on the problem of language acquisition—hence its relevance to this chapter.

The central thesis of that work, which we also adopt here, is that a simple processor, committed to reducing the burden on working memory, lies at the heart of the human language faculty. Although such a processor makes no use of grammatical principles, its operation plays a key role in explaining the properties of many core syntactic phenomena—binding, control, agreement, island constraints, scope, and so forth. It is also crucial to the account of how those properties can be acquired in response to the limited sorts of experience available in the early years of life. We return to this idea and its implications for second language acquisition in section 3, where we present a detailed illustration of this approach.

It is vital to note that there is no inherent incompatibility between input-based research and the processor-based program. To the contrary, the proponents of input-based emergentism are strongly committed to the existence of a processor—their central point is in fact a claim about the processor, namely that it is highly sensitive to the relative frequency of form-meaning mappings and distributional contingencies (Seidenberg & MacDonald 1999, Chang et al. 2006).

Likewise, the processor-based approach does not deny the relevance of the input to understanding language acquisition and use. To the contrary, it insists that frequency of occurrence is extremely important (e.g., O'Grady 2008a,b)—although not more important than the calculus that assesses the burden that computational operations of various sorts place on working memory. The acquisition of relative clauses is a case in point: direct object relatives are more frequent in the input than are subject relatives (Diessel & Tomasello 2005:898-99), but are nonetheless mastered later due to their well-documented processing burden (e.g., Gibson 1998, Caplan & Waters 2002).

To the extent that there is a major point of dispute between input-based and processor-based approaches to language, it lies in the question of whether the input provides learners with enough information to support induction of the full language. The dominant view of those committed to input-based emergentism seems to be that the input is sufficient in this regard (e.g., MacWhinney 2004). This is of course consistent with the opposition of many emergentists to the long-

standing assumption, heavily promoted in the literature on formal syntax, that input deficiencies constitute a *prima facie* argument for Universal Grammar (the so-called 'poverty-of-stimulus' claim).

A different view is put forward by O'Grady (2005, 2008a), who holds that there is in fact a poverty-of-stimulus problem, but that it does not support the case for UG. Instead, he proposes, the gap between experience and a speaker's linguistic knowledge is bridged with the help of the processor, which directs learners to particular options that are not evident from information available in the input. We return to this point in sections 3 and 4.

In his critique of emergentist work on second language acquisition, Gregg (2003:122) raises a point with which we agree: emergentism is badly in need of a 'property theory' of linguistic competence—i.e., a theory of the language faculty and of language itself—that can rival UG-based theories. We believe that processor-based emergentism offers the promise of just such a theory in that it is able to address the three issues that rightly lie at the heart of contemporary linguistic theory:

i.   Why does language have the particular properties that it does?
ii. Why is typological variation involving those properties restricted in particular ways?
iii. How are those properties acquired by children, based on experience that is limited in particular ways?

We believe that the explanations needed to address these questions within an emergentist framework also shed light on the problems associated with understanding why second language acquisition follows the particular course that it does.

A program devoted to the exploration of these matters obviously requires very specific proposals—and very specific types of evidence in support of those proposals. With that in mind, we will devote the remainder of this chapter to a relatively detailed case study of a subtle phenomenon involving the interaction between negation and quantification exemplified in English and Korean constructions such as the following.

(2)   English
       Mary didn't read all the books.

(3)   Korean
       Mary-ka motun chayk-ul ilk -ci   anh-ass-ta.
       Mary-NOM all   book-ACC read-COMP NEG-PST-DECL
       'Mary didn't read all the books.'

We begin, in the next section, by considering how the processor goes about assigning interpretations to sentences of this type and how its operation contributes to an account both of language acquisition and of typological variation. Section 4 reports on new experimental work that we have done on the acquisition of English scope by native speakers of Korean, illustrating the sorts of insights that an emergentist approach can offer to the classic problems of second language learning.

3. Scope

Our starting point is the proposal outlined in detail by O'Grady (2005), which holds that the core properties of language are best explained by reference to an 'efficiency-driven' processor whose primary objective is not to implement grammatical rules but simply to minimize the burden on working memory.[2] In the case of scope, two simple ideas come into play. Because each implies a strategy that minimizes the burden on working memory by avoiding delays and revisions, we will refer to them as 'efficiency assumptions.'

(i)   As the processor works its way through a sentence, it immediately assigns each NP an interpretation, based on available clues such as position, determiner type, case marker, context, and so forth.

(ii)   The revision of a previously assigned interpretation is costly since it disrupts the normal linear operation of the processor, which forms and interprets sentences in real time under conditions that value quickness.

These assumptions can be represented schematically as follows.

(4)a. An NP is encountered and assigned an interpretation 'x,' based on its position and other local properties:

    NP
    [x]

  b. Based on the properties of a subsequently encountered element, the NP's interpretation is recomputed:

    NP . . . . .Z . . .
    [*] -->[y]

The latter procedure adds to the burden on working memory resources by requiring both the recovery of the earlier interpreted NP and its recomputation.

---

[2] Following Carpenter et al. (1994), we take working memory to be a pool of operational resources that both holds representations and supports computations on them. See also Jackendoff (2002:200).

Now let us consider a concrete example involving a universally quantified NP such as *all the men*. In a typical case, such an NP has a 'full set' interpretation, grouping together each and every man in the relevant discourse context so that some property can be attributed to them all. We can depict this as follows for expository purposes.
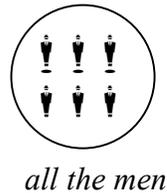

all the men

Figure 1: 'Full set' interpretation of *all the men*: the set includes each and every man in the relevant discourse context

Thus in a sentence such as *All the men arrived on time*, it is understood that the property of having arrived on time applies to the entire group of men—anyone in the relevant domain of discourse who is a man must have arrived on time. Likewise, in a pattern such as *The committee interviewed all the men*, it is understood that the property of having been interviewed by the committee holds of each man.

The negative operator *not* interacts with *all* in a variety of intricate ways. Perhaps the simplest interaction occurs in sentences such as the following, in which *not* combines directly with the quantified NP and unambiguously has scope over it. (That is, the interpretation of *all* is modified under the influence of the negative.)

(5)   [Not all the men] arrived on time.

Here the set denoted by the quantified NP is partitioned, so that the property of having arrived on time applies to only some of its members. We can depict this type of interpretation of the NP as follows. (The actual proportion of men in the excluded subset can of course vary.)
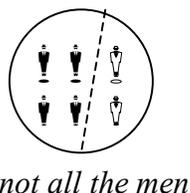

not all the men

Figure 2: Interpretation of *all the men* when it is in the scope of negation; the set of men is partitioned, so that the property of having arrived on time applies to only some of its members.

Although this interpretation requires a computational operation (the partitioning of the set) not found in the absence of negation, it at least does not run afoul of the efficiency assumptions outlined earlier. By the time the processor encounters the quantified NP, it has already come upon the negation. It is thus able to consider a partitioned set interpretation for the NP right away, without having to abandon or modify a *previously* computed interpretation.

Matters are different in the case of a sentence such as the following, in which the quantified NP serves as subject and the negative occurs to its right.
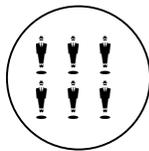
(6)   All the men didn't arrive on time.

Here, there are two interpretations. On the first interpretation, traditionally called the 'universal wide scope reading' (*all > not*), all members of the discoursally relevant set of men behave alike with respect to the property of having arrived on time—they fail to do so. We will henceforth refer to this as the 'full set interpretation.'

On the second interpretation, traditionally labeled the 'negation wide scope reading' (*not > all*), the set of men differ internally with respect to the property at hand—it is understood that some men arrived on time and some didn't. We will call this the 'partitioned set interpretation.'

The two readings differ with respect to their compliance with our efficiency assumptions. Consider first the full set interpretation, which can be derived as follows.
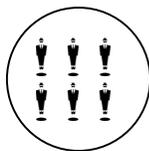
(7)a.   <u>First</u>: Formation of the NP *all the men* and assignment of the full set interpretation:

[All the men]

b.   <u>Subsequently</u>: Formation and interpretation of the rest of the sentence, with no change to the interpretation of the subject NP.
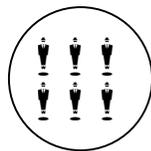
[All the men didn't arrive on time]
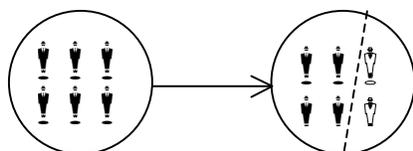

Now consider the partitioned set interpretation.

(8)a. <u>First</u>: Formation of the NP *all the men* and assignment of the usual full set interpretation:

[All the men]

b. <u>Subsequently</u>: The negative operator is encountered and assigned wide scope, forcing recomputation of the subject NP by partitioning the previously formed set.

[All the men didn't ...]

According to the second of our efficiency assumptions, this interpretation should be more difficult since the processor has to depart from its normal linear course and revise an earlier assigned interpretation—creating a significant extra burden on working memory.

If these ideas are on the right track, we would expect to find supporting evidence for the difficulty of the partitioned set interpretation both in the developmental profile observed in language acquisition and in the facts of typological variation.

Acquisition

Musolino & Lidz (2006) report on the results of a truth-value judgment task that they carried out with 20 five-years-olds, following up on earlier work by Musolino et al. (2000). The children watched an experimenter use props to act out a scenario in which two of three horses succeed in jumping over a fence. The children were then asked to judge the truth of *Every horse didn't jump over the fence*. Given the scenario presented by the experimenter, the sentence is true on the partitioned set interpretation (since not every horse jumped over the fence) and false on the full set interpretation (since two of the three horses did in fact jump over the fence).

Figure 3 Scenario used to test the truth of *Every horse didn't jump over the fence*

The children opted for the full set interpretation 85% of the time, accepting the sentences as true just 15% of the time.[3]

In the studies just mentioned, adults manifested no difficulty with the partitioned set interpretation. However, more recent work by Conroy & Lidz (2007) involving a different experimental paradigm suggests that they too prefer the full set interpretation.

Could the interpretive preferences manifested by children (and adults) be the result of exposure to the relevant sentences during the language acquisition process? It seems not. Based on an examination of maternal speech to a total of 42 children in the CHILDES data base, Gennari & MacDonald (2005/2006) report finding no instances of either *every* or *all* in the subject position of a negated sentence. Evidently then, children hear few if any sentences like *Every horse didn't jump over the fence*, with either interpretation.[4]

This is of course precisely the sort of situation typically used to make the case for Universal Grammar—a subtle and abstract syntactic fact is underdetermined by experience. However, processor-based emergentism offers an alternative: children are directed toward the full set interpretation of sentences such as *Every horse didn't jump over the fence* not by prior experience, but rather by a processor dedicated to minimizing the burden on working memory. As we have already seen (example (8) above), the partitioned set interpretation in sentences such as these requires the processor to depart from its normal linear course in order to revise the interpretation previously assigned to the quantified NP, creating a burden not associated with the full set interpretation.

Typology

The processing effects that contribute to interpretive preferences in scope seem to have typological consequences as well. Particularly relevant in this regard is Chinese, in which sentences such as the following have only the full set interpretation in which the property of not having jumped over the fence is attributed to the entire group of horses (Musolino et al. 2000:22).

(9)   Mei-pi ma dou mei tiao-guo langan.
        Every horse all not jump over fence
        'Every horse didn't jump over the fence.'

---

[3] The acceptance rate increased to as high as 60% when children were presented with 'contextual support' in the form of a contrastive sentence such as (i).

(i)   Every horse jumped over the log, but every horse didn't jump over the fence.

This suggests a dispreference for the partitioned set interpretation rather than its absolute rejection.

[4] Musolino & Lidz (2006:841-42) have collected examples of negated sentences with a universally quantified subject in adult-to-adult speech, but these sentence all have the partitioned set interpretation that children disprefer (e.g., *All the birds don't seem to be the same*)—the exact opposite of what the input-based account would predict.

Crucially, however, we know of no language in which the reverse situation holds and negation MUST have scope over a universally quantified subject NP to its left. That is, there are no languages that are just like English or Chinese in their syntax, except that a sentence such as *Every horse didn't jump over the fence* can ONLY mean 'Not every horse jumped over the fence.'

All of this makes sense if we assume, following Hawkins (2004) and O'Grady (2005), that processing considerations define degrees of computational difficulty and that individual languages can differ in terms of the burden they are willing to accept—with the proviso that if the interpretation that is harder to process is permitted, then the easier interpretation must also be allowed. This yields two possibilities:

i. The computationally costly pattern, in which the processor has to revise the interpretation of a previously interpreted NP, is disallowed; the pattern in which the NP retains its full set interpretation is permitted. Chinese works this way; as we have just seen, a sentence such as (9) permits only the full set interpretation for the quantified NP.

ii. The computationally costly pattern is allowed, but is disfavored compared to its less difficult-to-process counterpart. English works this way—a sentence such as *Every horse didn't jump over the fence* permits two readings, but the computationally easier full set interpretation is preferred, as we have seen.

Similar sorts of implicational relationships are widely acknowledged in phonology. For instance, to take a simple example, it is a matter of consensus among phonologists that a CVC syllable is articulatorily more difficult than a CV syllable. Predictably, this leads to two options:

i. languages in which syllable-final consonants are prohibited—e.g., Hawaiian

ii. languages in which a syllable-final consonant is permitted, but there are indications that it is articulatorily difficult—e.g., English, in which children initially drop syllable-final consonants

We propose that comparable asymmetries, motivated by processing difficulty rather than articulatory factors, are pervasive in syntax.

## 4. Scope in Korean-speaking ESL learners

Because Korean is an SOV language with negation adjacent to the verb, a universally quantified NP precedes the negative operator even when it functions as direct object.

(10)   John-i **motun** salam-ul   po-ci   anh-ass-ta.
       John-Nom   all   person-Acc see-Comp Neg-Pst-Decl
       'John didn't see all the people.'

The two potential interpretations of this sentence therefore differ with respect to the efficiency assumptions that we have proposed. Whereas the full set reading is built without modifying the interpretation that is initially assigned to the quantified NP, the partitioned set reading requires this interpretation to be revised, as illustrated below.

(11)a. <u>First</u>: Formation of the NP *motun salam* 'all people' and assignment of the default full set interpretation:

   John-i   **motun** salam-ul   po-ci   anh-ass-ta.
   John-Nom   all   person-Acc see-Comp Neg-Pst-Decl
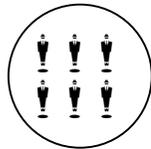
b. <u>Subsequently</u>: The negative operator is encountered and assigned wide scope, forcing recomputation of the quantified NP by partitioning the previously formed set.

   John-i   **motun** salam-ul   po-ci   anh-ass-ta.
   John-Nom   all   person-Acc see-Comp Neg-Pst-Decl
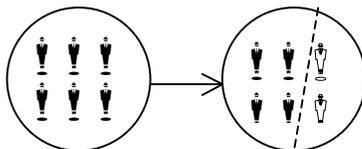
Consistent with the second of our efficiency assumptions, we expect the partitioned set interpretation to be accompanied by an increase in the burden on working memory as the processor abandons its normal linear course to revise an earlier assigned interpretation.

   We therefore predict that Korean speakers will manifest just two types of acceptability judgments: either they will permit only the full set reading, or they will permit both the full set reading and the more demanding partitioned set reading. Under no circumstances should they permit only the more difficult partitioned set reading.

Table 1   Predictions for scope interpretation in Korean

	full set reading	partitioned set reading
Possibility 1	accept	reject
Possibility 2	accept	accept (but dispreferred)
Impossible	reject	accept

These predictions seem to be at least partially borne out in the study conducted by Han et al. (2007). Using a truth-value judgment task, they found an acceptance rate by their adult subjects of 98% for the full set interpretation in the type of negative pattern we are considering. In contrast, the acceptance rate for the partitioned set interpretation was just 46%, with seven of the twenty subjects rejecting it on all the test items.[5]

A parallel preference was manifested among Han et al's child subjects (all four year olds), who accepted the full set interpretation 86.67% of the time, compared to just 33.33% for the partitioned set reading. Indeed, many child subjects rejected the partitioned set interpretation on all test items.

These preferences are very different from those associated with English patterns such as *John didn't see all the men* or *John didn't see everyone*, in which the partitioned set interpretation is dominant. As observed earlier, such a preference is fully consistent with our efficiency assumptions. This is because the occurrence of the negative to the left of the quantified direct object in English allows the processor to partition the set immediately, instead of first building a full set interpretation and then revising it.[6]

(12)   John didn't see all the men.

---
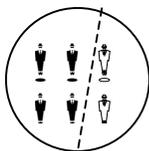
[5] Because of Han et al.'s between-participants design, their results do not allow us to ascertain the correctness of our prediction that no Koreans will permit ONLY the partitioned set interpretation.

[6] Of course, this does not explain why the partitioned set interpretation is not only possible, but also preferred in this case. Interestingly, Musolino & Lidz (2006) present evidence that five year olds accept the full set interpretation about 75% of the time, in contrast with adults whose acceptance rate in basic descriptive contexts is just 20%. We agree with Musolino & Lidz that these preferences can be traced to an implicature which children are slow to learn: use of the *not ... every/all* pattern rather than the *not ... any* pattern (e.g., *John didn't see anyone*) implicates that the stronger statement is inappropriate and that John must in fact have seen some of the people.

It is also important to note that the full set interpretation is actually preferred in certain cases, as in *Max didn't consider all the people who would be inconvenienced by his decision*, brought to our attention by Kevin R. Gregg.

This brings us to the question of what happens when a speaker of Korean learns English. In order to get at this issue, we conducted the experiment described below.

Participants:

Forty-two native speakers of Korean participated in our experiment, all of whom were students in a linguistics class at Hanyang University in Seoul, Korea. Based on their previous English-language courses, their proficiency in English was estimated to be at the intermediate or high intermediate level. They had received no formal training in semantics, but anecdotal reports suggest that at least some may have received instruction in high school or college English courses about the preferred interpretation for sentences such as (10). We return to this matter below.

Procedure and materials:

Subjects were presented with a total of 8 test items, preceded by two practice items and interspersed with 10 fillers. There were two conditions for the test items—one with a context favoring the full set interpretation (see below) and another with a context supporting a partitioned set reading. A Latin square design was employed, so that every subject was exposed to each of the eight test items, but no test item was encountered in more than one context.

The stories were presented orally (via a pre-made recording) as the subjects read a written version in their individual test booklets. A sample story favoring the partitioned set interpretation of the sentence *Tom didn't fix all the computers* follows:

Tom is at his uncle's repair shop.

Tom's uncle is about to go out for lunch. He asks Tom to fix three radios and three computers before he returns. Tom promises to do so.

Tom fixes the three radios easily.

Then, Tom examines the first computer. But, he can't fix it. He decides to wait until his uncle comes back.

Then, Tom looks at the second computer. There is something wrong with the sound, but he can't fix it.

Finally, Tom comes to the third computer. There is something wrong with the screen. Screens are very hard to fix. But, Tom manages to fix it.

The context favoring the full set interpretation begins in the same manner, but the final paragraph goes as follows:

Finally, Tom comes to the third computer. There is something wrong with the screen. He thinks that he can fix it quickly. However, after Tom works on it for a while, he gives up.

Each story was accompanied by a picture that summarized the end result—one repaired computer and two still-broken computers in the case of the partitioned set

context, and three still-broken computers in the case of the full set context.

At the end of the story, the subjects were asked to judge the truth of a summary sentence in the test booklet. Each such sentence contained a negated verb with a universally quantified direct object phrase of the form *all the N*.

(13)   Tom didn't fix all the computers.

Subjects were given ten seconds to make their selection before presentation of the next item.

In order to gather data relevant to the assessment of possible transfer effects and in order to ascertain how Korean speakers judge scope in their native language, subjects participated in both an English version of the experiment and a Korean version (written only). As a precaution against the possibility that rendering judgments about scope in Korean might influence the judgments given for English, all subjects were tested first on English and then on the corresponding Korean materials a week later.

The subjects were tested as a group in a classroom at their university under the supervision of one of the experimenters. Each session took approximately 30 minutes to complete.

Results and discussion:
Table 2 summarizes the results from the truth value judgment portion of our experiment for Korean.

Table 2 Percentage of 'true' responses (Korean version)

context favoring full set reading	context favoring partitioned set reading
97% (163/168)	21% (36/168)

As can be seen here, our subjects exhibit a very strong preference for the full set interpretation, accepting it as true in the matching contexts 97% of the time; in contrast, the partitioned set interpretation was judged to be true in contexts that favored it just 21% of the time. This difference is statistically significant (t (41) = 12.49, p < .05).

This is just what we would expect. As explained above, the partitioned set interpretation in Korean involves a higher level of computational difficulty than the full set reading and should therefore be less accessible.

Now consider the truth value judgments given by the same subjects on the English version of our experiment.[7]

---

[7] A control group of six native English speakers judged the test items to be true in the partitioned set context 100% of the time, compared to just 67% in the full set context. In a follow-up questionnaire, one of these subjects observed that although the test sentence was 'technically accurate' in the full set context, it was a 'slightly confusing way of describing what happened' and that it would be clearer to say (for example) *Mary didn't take out any boxes*.

Table 3 Percentage of 'true' responses (English version)

context favoring full set reading	context favoring partitioned set reading
93% (157/168)	28% (47/168)

Here again, the preference for the full set interpretation is statistically significant (t (41) = 9.06, p < .05).

As the contrast in table 3 makes clear, our subjects exhibited a very strong preference for the full set interpretation of the quantified NP in English. Indeed the preference is not significantly different from the one manifested in Korean, despite the subjects' relatively high proficiency in English and the fact that at least some had probably received instruction concerning the preferred partitioned set interpretation of the English test items. What should we make of this result?

It has sometimes been suggested, notably by Pienemann (1998), that processing considerations play a major role in determining whether and when particular properties of the L1 are transferred to the L2. We adopt a similar idea here, although we do not of course adopt the grammatical mechanisms that typically accompany such proposals. Our hypothesis can be stated as follows.

(14)   The preferred interpretation in the L1 will be favored in the L2 if and only if it does not have a greater processing cost in the L2.

As we have already seen, the full set interpretation of the universally quantified NP is preferred in Korean. Since that interpretation has a comparable low cost in English, our hypothesis predicts that it should also be favored in that language by Korean L2 learners—which is indeed what our results show.

An independent test for our hypothesis comes from the acquisition of Korean by native speakers of English—the reverse of the situation we have been focusing on. As already observed, native speakers of English prefer the partitioned set interpretation of *negated V + all N* patterns in their first language. Crucially, the corresponding interpretation in Korean has a much higher computational cost since the processor is forced to revise its initial interpretation of the quantified NP.

(15)   The partitioned set interpretation in Korean:
        John-i **motun** salam-ul   po-ci   anh-ass-ta.
        John-Nom   all   person-Acc see-Comp Neg-Pst-Decl
        'John didn't see all the people.'

    a. First: Formation of the NP *motun salam* 'all people' and assignment of the default full set interpretation:

        John-i **motun** salam-ul   ....
   John-Nom   all   person-Acc
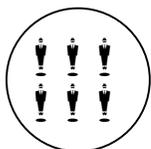
b. <u>Subsequently</u>: The negative operator is encountered and assigned wide scope, forcing recomputation of the quantified NP by partitioning the previously formed set.

John-i **motun** salam-ul   po-ci   anh-ass-ta.
John-Nom   all   person-Acc see-Comp Neg-Pst-Decl
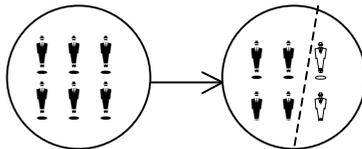
In contrast, the partitioned set reading of a direct object NP in English places no special burden on working memory, since it does not require the recomputation of a previously determined interpretation. (Because *not* precedes the quantified NP, the option of forming the partitioned set interpretation is available from the outset.)

(16)   The partitioned set interpretation in English:
       John didn't see all the people.
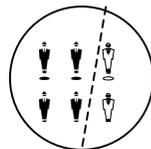
Which interpretation do English-speaking L2 learners prefer in Korean? Is it the partitioned set reading that is favored in English? Or is it, as our hypothesis predicts, the less costly full set reading?

In order to address this issue, we conducted a small pilot study involving five relatively advanced English-speaking learners of Korean as a second language by having them take the Korean version of our test. The full set interpretation in Korean was accepted 100% of the time, compared to just 50% for the partitioned set interpretation that is preferred in English. Three of the subjects accepted both interpretations in Korean (as do some native speakers of Korean), and two accepted only the full set reading. Crucially, no one showed a preference for the partitioned set reading that is favored in English.

Although no more than mildly suggestive, these results are consistent with what we predict: the preferred interpretation in the L1 is not carried over to the L2 when it has a higher processing cost in the second language. We hope that work currently underway with beginning and intermediate subjects will allow us to deepen our understanding of the role of processing cost in the acquisition of scopal patterns.

5. Concluding Remarks

After a brief introductory survey of emergentist work in SLA, we focused our attention on a phenomenon that exemplifies the challenges confronting contemporary linguistic theory. Negative-quantifier scope manifests a subtle form-meaning mapping that is underdetermined by the input and yet is acquired with apparent effortlessness. Moreover, although it exhibits significant variation across languages, that variation is strictly limited. For these reasons, the phenomenon is a classic candidate for treatment in a UG-type framework.

In fact though, such a treatment is unnecessary; there is an emergentist alternative. A simple efficiency-driven processor, committed to minimizing the burden on working memory, explains why scope has the properties that it does, how language learners come to acquire those properties, and why typological variation is constrained in the way that it is. The key insight, as we have tried to show, is simply that the processor resists revisions to previously computed interpretations—a procedure that requires the recovery of the earlier interpreted NP from memory and the use of additional working memory resources to support its reinterpretation. The aversion to such reinterpretations and to the computational cost that they incur offers insights into key typological and developmental facts—and into why second language acquisition follows the particular course that it does.

References

Ambridge, B., Theakston, A., Lieven, E., and Tomasello, M. (2006). The distributed learning effect for children's acquisition of an abstract syntactic construction. Cognitive Development **21**, 174-193.

Bates, E. and MacWhinney, B. (1987). Competition, variation, and language learning. *In*: B. MacWhinney (Ed.), Mechanisms of Language Acquisition, Erlbaum, Mahwah, NJ, pp. 157-193.

Bybee, J. and Hopper, P. (2001). Frequency and the Emergence of Linguistic Structure, John Benjamins, Amsterdam.

Caplan, D. and Waters, G. (2002). Working memory and connectionist models of parsing: A reply to MacDonald and Christiansen (2002). Psychological Review **109**, 66–74.

Carpenter, P., Miyake, A., and Just, M. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. *In*: M. Gernsbacher (Ed.), Handbook of Psycholinguistics, Academic Press, San Diego, pp. 1075-1122.

Chang, F., Dell, G., and Bock, K. (2006). Becoming syntactic. Psychological Review **113**, 234-272.

Chater, N. and Manning, C. (2006). Probabilistic models of language processing and acquisition. Trends in Cognitive Sciences **10**, 335-344.

Conroy, A., and Lidz, J. (2007). Seriality in LF processing. Ms. Department of Linguistics, University of Maryland.

Dick, F., Bates, E., Wulfeck, B., Aydelott Utman, J., Dronkers, N., and Gernsbacher, M. (2001). Language deficits, localization, and grammar: Evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. Psychological Review **108**, 759-788.

Diessel, H. and Tomasello, M. (2005). A new look at the acquisition of relative clauses. Language **81**, 882-906.

Ellis, N. (2002). Frequency effects in language processing. Studies in Second Language Acquisition **24**, 143-188.

Ellis, N. (2006a). Language acquisition as rational contingency learning. Applied Linguistics **27**, 1-24.

Ellis, N. (2006b). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. Applied Linguistics **27**, 164-194.

Ellis, N. and Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics. Applied Linguistics **27**, 558-589.

Elman, J. (1993). Learning and development in neural networks: The importance of starting small. Cognition **48**, 71-99.

Elman, J. (2002). Generalization from sparse input. Proceedings of the 38th Regional Meeting of the Chicago Linguistic Society, 175-200.

Elman, J. (2005). Computational approaches to language acquisition. *In*: K. Brown (Ed.), Encyclopedia of Language and Linguistics, 2nd ed., Elsevier, Oxford, UK.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996). Rethinking Innateness: A Connectionist Perspective on Development. MIT Press, Cambridge, MA.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. Cognitive Psychology **47**, 164-203.

Genarri, S. and MacDonald, M. (2005/2006). Acquisition of negation and quantification: Insights from adult production and comprehension. Language Acquisition **13**, 125-168.

Goldberg, A. (1998). The emergence of the semantics of argument structures. *In*: B. MacWhinney (Ed.), The Emergence of Language. Erlbaum, Mahwah, NJ, pp. 197-212.

Goldberg, A. and Casenhiser, D. (2008). Construction learning and second language acquisition. *In*: P. Robinson, N. Ellis (Eds.), Handbook of Cognitive Linguistics and Second Language Acquisition, Routledge, New York, pp. 197-215.

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

Gregg, K. (2003). The state of emergentism in second language acquisition. Second Language Research **19**, 95-128.

Han, C., Lidz, J., Musolino, J. (2007). V-raising and grammar competition in Korean: Evidence from negation and quantifier scope. Linguistic Inquiry **38**, 1-48.

Hawkins, J. (2004). Efficiency and Complexity in Grammars, Oxford University Press, Oxford, UK.

Haspelmath, M. to appear. Parametric versus functional explanations of syntactic universals. *In*: T. Biberauer (Ed.), The Limits of Syntactic Variation, Jon Benjamins, Amsterdam.

Herron, D. and Bates, E. (1997). Sentential and acoustic factors in the recognition of open- and closed-class words. Journal of Memory and Language **37**, 217-239.

Ionin, T., Zubizarreta, M., and Maldonado, S. (2008). Sources of linguistic knowledge in the second language acquisition of English articles. Lingua **118**, 554-576.

Jackendoff, R. (1988). Why are they saying these things about us? Natural Language and Linguistic Theory **6**, 435-442.

Jackendoff, R. (2002). Foundations of Language, Oxford University Press, Oxford, UK.

Lewis, J., and Elman, J. (2001). Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. Proceedings of the 26th Annual Boston University Conference on Language Development, 359-370.

MacWhinney, B. (1987). The competition model. *In*: B. MacWhinney (Ed.), Mechanisms of Language Acquisition, Erlbaum, Mahwah, NJ, pp. 249-308.

MacWhinney, B. (2000) Emergence from what? Comments on Sabbagh & Gelman. Journal of Child Language **27**, 727–733.

MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. Journal of Child Language **31**, 883-914.

MacWhinney, B. (2008). A unified model. *In*: P. Robinson, N. Ellis (Eds.), Handbook of Cognitive Linguistics and Second Language Acquisition, Routledge, New York, pp. 341-371.

Musolino, J., Crain, S., and Thornton, R. (2000). Navigating negative quantificational space. Linguistics **38**, 1-32.

Musolino, J. and Lidz, J. (2006). Why children aren't universally successful with quantification. Linguistics **44**, 817-852.

O'Grady, W. (2005). Syntactic Carpentry: An Emergentist Approach to Syntax, Erlbaum, Mahwah, NJ.

O'Grady, W. (2007). The syntax of quantification in SLA: An emergentist approach. *In*: M. O'Brien, C. Shea, J. Archibald (Eds.), Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006): The Banff Conference, Cascadilla Press, Somerville, MA, pp. 98-113.

O'Grady, W. (2008a). Does emergentism have a chance? *In*: H. Chan, H. Jacob, E. Kapia (Eds.), Proceedings of 32nd Annual Boston University Conference on Language Development. Cascadilla Press, Somerville, MA, pp. 16-35.

O'Grady, W. (2008b). Innateness, Universal Grammar, and emergentism. Lingua **118**, 620-631.

O'Grady, W. (2008c). The emergentist program. Lingua **118**, 447-464.

O'Grady, W., Nakamura, N., and Ito, Y. (2008). Want-to contraction in second language acquisition: An emergentist approach. Lingua **118**, 478-498.

Pienemann, Manfred. 1998. Language Processing and Second Language Development: Processability theory, John Benjamins, Amsterdam.

Seidenberg, M. and MacDonald, M. (1999). A probabilistic constraints approach to language acquisition and processing. Cognitive Science **23**, 569-588.

Tomasello, M. (2003). Constructing a Language: A Usage-Based Theory of Language Acquisition, Harvard University Press, Cambridge, MA.