*SSLA*, **31**, 175–198. Printed in the United States of America.
doi:10.1017/S0272263109090275

# THE EVOLVING CONTEXT OF THE FUNDAMENTAL DIFFERENCE HYPOTHESIS

Robert Bley-Vroman
*University of Hawai'i*

Foreign language learning contrasts with native language development in two key respects: It is unreliable and it is nonconvergent. At the same time, it is clear that foreign languages are languages. The fundamental difference hypothesis (FDH) was introduced as a way to account for the general characteristics of foreign language learning. The FDH was originally formulated in the context of the theory of rich Universal Grammar, and this theory has guided much foreign language acquisition research over the past two decades. However, advances in the understanding of language have undermined much of the supporting framework.

The FDH—indeed all of SLA research—must be rethought in light of these advances. It is proposed here that (a) foreign language grammars make central use of patches, which are also seen as peripheral phenomena in native languages; (b) non-domain-specific processes are used in foreign language acquisition, but that these are also employed—although more effectively because they are integrated into the language system—by native language development; and (c) foreign language online processing relies heavily on the use of shallow parses, but these are also available in native language processing, although less crucially.

Two decades have passed since the set of proposals commonly spoken of as the fundamental difference hypothesis (FDH; Bley-Vroman, 1988,

This article has benefited from useful discussions with colleagues, especially William O'Grady, Bonnie D. Schwartz, Roumyana Slabakova, Laurent Dekydtspotter, Silvina Montrul, Lydia White, and Henry Bley-Vroman. I would also like to thank two anonymous *SSLA* reviewers who provided exceptionally detailed and challenging comments.

Address correspondence to: Robert Bley-Vroman, Department of Second Language Studies, University of Hawai'i, Honolulu, HI 96822; e-mail: vroman@hawaii.edu.

© 2009 Cambridge University Press 0272-2631/09 $15.00 

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

1989, 1990) were made.[1] Those proposals address two theoretical areas of SLA:

(1) a. the explanatory burden of SLA
b. the nature of the SLA system and of second language (L2) knowledge.

The term *FDH* has a broad and a narrow meaning. Narrowly, and originally, it refers just to a set of related alternative proposals for (1b). However, because it is proposed that SLA is different in fundamental ways, both in the explanandum (1a) and in the explanans (1b), the term can appropriately be used broadly to comprise both.

Much empirical research over the past two decades has taken place with either implicit or explicit reference to the FDH. Because both the explanandum and the explanans are formulated within the generative theory of Universal Grammar (UG), the abundant research that addresses the place of UG in SLA is broadly relevant. The present article addresses these same issues again, from a point of view two decades later. The focus of this article, as that of the original work, is chiefly conceptual and programmatic. It does not attempt a review of the specific research findings, except as they impinge directly on the conceptual issues.[2]

First, the logical problem (1a), focusing on two properties of foreign language learning—namely, (un)reliability and (non)convergence—will be considered. The original arguments still hold in this case. Second, the problem of explanation (1b) is tackled; developments in learning theory, linguistics, psychology, mathematics, and philosophy require that the FDH be reconceived in radical ways.

## THE LOGICAL PROBLEM OF FOREIGN LANGUAGE LEARNING

The explanatory burden of the theory of native language development is to account for how acquisition of the language can succeed, given the data to which the learning device is exposed. It is against this background that all serious native language acquisition theories are constructed, independent of whether such theories are based in mainstream generative grammar or some other framework. Debates about the adequacy of theories are always in terms of whether they will work reliably and consistently.

Bley-Vroman (1990) gave ten characteristics with respect to which native language development contrasts with foreign language learning. Two of these are crucial for the logical structure of the theory. Using the terminology of Pullum and Scholz (2002), native language development is characterized by *reliability* and *convergence*. These are informally defined by Pullum and Scholz as in (2):[3]

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

(2) Reliability: Children always succeed at language learning.
Convergence: Children end up with systems that are so similar to those of others in the same speech community. (p. 10)

In the terminology of the original "Logical Problem of Foreign Language Learning," these correspond approximately to success-failure and uniformity-variability.[4]

## Reliability

There are, of course, various ways of modeling the concept of reliability in formal learning theory. One can require that the grammar learned is (eventually, in the limit) identical to the grammar that generated the input. One can require that the grammar learned (eventually) generate the same set of sentences that the target grammar generates. As in more recent approaches to mathematical learning theory, one can require that the grammar learned be probably approximately correct.[5] However, regardless of how implemented and regardless of the particular operationalization of success, native language learning theories aim to learn the target; they strive to be reliable.

Reliability includes that a theory be robust across possible input sequences, even in the presence of a certain amount of noise; for example, one would not consider a learning theory reliable that was dependent on encountering the right exemplars in the right order, or that would exhibit great fragility in the face of an occasional misleading or incorrect exemplar, or that required that unusual examples occur frequently.[6]

Some kind of reliability (even as an idealization) is uniformly adopted, often implicitly, as a standard by those working in (native) language acquisition theory, natural language processing, and mathematical learning theory. The very term *learnability* reflects this standard: We want to show that a language is in fact learnable, given a theory of learning—that is, that the theory works reliably. No theorist of native language development has ever advocated a theory on the basis of its unreliability, or because it is less likely to learn the target than a competing theory, or because it can be seduced into error, or because its success is dependent on careful structuring of the input.

Ironically, what makes a theory a good theory of native language learning—that it works reliably—makes it a bad theory of foreign language learning. This is because of the first basic property of foreign language learning, the *unreliability property* defined in (3):

(3) Foreign language learning is not reliable.

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

### Convergence

For our purposes, the second crucial property of native language acquisition is that it is convergent, as defined in (2); see Pullum and Scholz (2002). In Chomsky's (1965) seminal formulation, there is a "striking uniformity of the resulting grammars" (p. 11).

Convergence is a relationship among the acquired grammars of learners exposed to the same target language. Reliability, in contrast, is a relationship between the target grammar and the acquired grammar. Reliability and convergence are obviously related. Reliability leads to convergence, but the converse is not true. An unreliable acquisition system can nonetheless lead to convergent grammars; for example, a foreign language acquisition theory might hypothetically (and counterfactually) predict that learners of English invariably would arrive at a grammar in which relative clauses had resumptive pronouns, or in which *the* encoded specificity rather than definiteness, or in which subject-auxiliary inversion occurred only with pronoun subjects. In these cases, the target is not being reliably learned, but there is a grammar—a nontarget grammar to be sure—on which the learning system converges. Because theories of native language development assume reliability as a standard, convergence follows a fortiori.

The fact, of course, is that in foreign language learning, learners exposed to the same target language may arrive at very different interlanguage grammars. The theory must permit everything from the so-called near-native cases like Julie (Ioup, Boustagui, Tigi, & Moselle, 1994) to cases in which the acquired grammar is dramatically different from that of the input, like Schmidt's (1983) Wes or Schumann's (1978) Alberto. This leads to the second essential property of foreign language learning, the *nonconvergence property* defined in (4):

(4) Foreign language learning is not convergent.

## THE (NARROW) LINGUISTIC THEORY OF FOREIGN LANGUAGE LEARNING

Addressing the logical problem of foreign language learning requires, in short, postulating an acquisition system that does not work reliably and does not converge. It must permit, as it were, the full range from Wes to Julie. At this level, the theory is not predictive or deterministic; that is, it cannot say for sure what grammar will be arrived at given the input. Rather, the theory is permissive. Let us use the term *the (narrow) linguistic theory of foreign language learning* to refer to a theory that relates the target language input to a set (not necessarily finite) of interlanguage grammars.

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

This set will, no doubt, have some sort of internal structure, with some grammars having greater likelihood than others, and this likelihood structure may be related to the extent of exposure (modeled perhaps as the size of the input set). Finally, the structure of the native language grammar will influence the range of possibilities, especially their likelihood and the relationship of likelihood to amount of exposure. The (narrow) linguistic theory of foreign language learning is only part of the full picture of foreign language learning. Precisely because it is neither reliable nor convergent and hence cannot be fully predictive of outcome, it leaves a natural place for factors that go beyond the narrowly linguistic. In this way, it has the potential to fit in well with the rich research programs underway in other areas of L2 studies.

What would such a linguistic theory of foreign language learning look like? We are now moving from the explanandum (1a) to the explanans (1b). It would not look much like government binding/principles and parameters (GB/PP) theory.[7] The consequence of unreliability and nonconvergence for foreign language learning models is both simple and severe. Because virtually all work in generative native language acquisition theory takes reliability and convergence as desiderata for the theory, any foreign language learning theory that takes native language acquisition theory as a basis is at a serious disadvantage.[8] Researchers who take a broadly generative approach to language are in an especially difficult position. The thrust of generative acquisition theory, especially during the 1990s, and even continuing today, has been to develop a model of language that would specifically permit acquisition to take place in a reliable and convergent manner. UG as embodied in GB/PP models in particular, which form the background against which the FDH was originally formulated, incorporate extensive systems of built-in knowledge, thus minimizing what would need to be acquired and radically constraining the range of hypotheses that the learning device would need to consider. Parameters (i.e., switches) need to be set, admittedly, but the number of these is generally supposed to be manageable, and—most important—the data needed to set them are widely supposed to be readily available. Such a system is highly reliable and different learners exposed to the same target will converge on systems of striking uniformity, as native language acquisition theory appropriately requires.

## THE FUNDAMENTAL DIFFERENCE HYPOTHESIS

What would an adequate theory look like then? The early FDH suggested, following ideas current in generative grammar at the time, that there are two conceptual components to an acquisition theory: (a) a definition of possible language (i.e., a hypothesis space) and (b) a

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

learning procedure. GB/PP theory provides these components for native language development.

Bley-Vroman (1990) proposed that the native language, rather than UG itself, shapes the initial hypothesis space and is also the chief source of initial expectations about the likely character of the target language: The learning system will expect something of the same general character as the native language. Indeed, the initial hypothesis may be that the target resembles the native language quite closely. Parameter setting as the learning process is rejected by the FDH. Instead, learning processes are utilized separately from their integration into a specialized language faculty. Elimination of UG and parameter setting should dispose of reliability and convergence—a welcome result. Additionally, transfer is predicted to occur, but in a much looser and more variable way than the one predicted by theories that propose that the initial state of the learner is exactly the native language grammar with parameters waiting to be reset, as in the full transfer/full access hypothesis (Schwartz & Sprouse, 1994, 1996). Acquisition is expected to be piecemeal and dependent on what can be noticed in the input. This permits the FDH to articulate with the noticing hypothesis (Schmidt, 1990), as part of a much more encompassing framework, with indirect ties (chiefly through noticing) to cognitive science more generally and even to language teaching.

One of the most direct statements of the FDH is Chomsky's (1997) observation about the nature of foreign language learning:

> Like other kinds of growth, language acquisition happens easily at a certain age, but not later. There comes a time when the system doesn't work anymore. There are individual differences […] but for most people, after adolescence, it becomes very hard. The system is just not working for some reason, so, you have to teach the language as something strange. (p. 128)

Chomsky's "just not working" and "something strange" (p. 128) in fact captures quite directly the two central tenets of the FDH as originally formulated. "Just not working anymore" (p. 128) corresponds to the FDH's proposal that the domain-specific cognitive system of UG no longer guides acquisition. Treating "language as something strange" (p. 128) is one way of formulating the FDH's suggestion that language is being handled not in this domain-specific faculty of mind, to which language is not strange, but rather in a domain-general central system, to which language is strange.

## THE SHIFTING CONTEXT OF THE FUNDAMENTAL DIFFERENCE HYPOTHESIS

Whereas the essential elements of the logical problem (i.e., the unreliability property and the nonconvergence property) are unaffected by

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

developments in linguistics and related fields over the past years, the same is not true of certain other specific proposals of the original FDH.

As a preliminary matter, however, one more property of foreign languages must be introduced explicitly. This property was implicit in earlier formulations of the FDH; it is the *language property* defined in (5).

(5) Foreign languages are languages.

*Language* is used here in a very broad sense. Foreign languages, like native languages, map between human conceptual-intentional structures (i.e., meaning) and a physical expression realized using the speech apparatus. Elements can be combined with other elements to make larger units, and this can go on recursively. Like native languages, foreign languages are processed in the human brain in real time. They are used to communicate, for private thinking, and for all functions of native languages. The FDH has never claimed that foreign languages are not languages in this sense.

Under the FDH, to say that foreign languages are languages is not to say that foreign languages are governed (totally or partially) by a GB/PP-type UG; for example, one could find evidence that subjacency (i.e., a principle that disallows movement across more than one bounding node) does not reliably constrain movement in a particular foreign language grammar but that grammar still would map between meaning and sound. In fact, all of the highly productive UG-in-SLA work over the past decades has operated within this assumption. This includes research that concludes that particular components of UG do not work in foreign languages in the same way or that the interface with forms is affected by maturation (Hawkins & Chan, 1997; Lardiere, 1998a, 1998b; Prévost & White, 2000).

The important consequence of the language property is that features of a foreign language that relate to its being a language need not (necessarily) invoke UG as an explanation. To take a trivial example, you do not need UG to explain the fact that the tongue may be used to make foreign language sounds, or, you do not need UG to explain the fact that foreign languages may distinguish present from past time: The human conceptual system does make such distinctions, and one end of the sound-meaning map is that conceptual system.

The original context of work on the place of UG in SLA, and the context in which most research has taken place over the past two decades, was GB/PP theory (Chomsky, 1981a, 1981b, 1982, 1986a, 1986b). Along with mainstream GB/PP, the FDH assumed innate structural principles of language and processes of language acquisition that are elaborate and sui generis; these were clearly distinct from non-domain-specific principles of structural organization and of learning (general problem solving in the terminology current at the time). Four developments in

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

the science of language require a reconceptualization of UG-based L2 research.

First, an elaborate and highly structured UG has given way to a minimal UG. The thrust of research from the extended standard theory through GB/PP was that UG was rich and highly articulated. There were subtheories with their own special properties (bounding, binding, control, case theory, theta theory, etc.); UG provided specific parametric options and ways to set abstract parameters based on concrete experience. Jackendoff and Pinker (2005) were not exaggerating much when they characterized the UG of GB/PP as "baroque" (p. 219). I will use the term *rich UG* for this trend. The very richness and abstractness of the necessary theoretical constructs and the difficulty of seeing how they could be learned from input gave strong motivation for positing an innate UG. Furthermore, the principles and processes of rich UG seemed impossible to derive in any obvious way from principles and processes outside the language faculty. They were "special principles specific to this domain" (Chomsky, 1980, p. 44). For SLA research, the baroque richness, abstractness, and especially the domain-specificity of rich UG yielded testable predictions about the role of UG in SLA. The constructs of rich UG gave clear empirical content to the FDH, and because of the highly articulated structure of the overall theory, with different subcomponents, principles, and parameters, these theoretical constructs also made it possible to formulate hypotheses wherein some particular part(s) of a rich UG would fail to function normally in foreign language learning (Beck, 1998; Clahsen & Hong, 1995; Eubank, 1993/1994, 1994, 1996; Hawkins & Chan, 1997; Tsimpli & Roussou, 1991). An elaborate and articulated innate rich UG also underlies a full transfer/full access view (Schwartz & Sprouse, 1994, 1996).

Over the past ten years, and especially in the last few years, the rich UG concept of the language faculty has been replaced with a vision in which the specifically linguistic innate endowment (narrowly conceived) is very sparse. This is particularly clear in the Minimalist Program (Chomsky, 1995, 2000, 2004, 2005). Jackendoff (2005) succinctly summarized the Minimalist goal: "to eliminate the richness of Universal Grammar as conceived of in Principles and Parameters theory" (p. 191). If the Minimalist Program turns out to be headed in the right direction, there is no rich UG, and a theory of UG can be developed in which all one needs to say is that language maps between sound and meaning, that elements from a lexicon can be combined, that units can combine recursively to form larger units, and that there is some sort of displacement property that accounts for movement. All of what was previously accounted for by the elaborate domain-specific (sub)systems of rich UG can be accounted for by the interaction of this recursive system with meaning, on the one hand, and linearized physical expression, on the other.

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

The consequences of Minimalism for the FDH—indeed for any research program that takes UG access in foreign language learning to be an empirical issue—are serious. What would it mean to say that UG is not available in foreign language learning? The essence of UG under Minimalism is really just the language property itself. However, there is no version of the FDH, or indeed of any major foreign language learning theory, that would deny the language property. It does not even need to be stipulated (as it does in linguistic theory). In the FDH, it follows directly from the hypothesis that the native language provides information about the notion of possible language, "by making the very conservative assumption that the foreign language is not an utterly different sort of thing from the native language" (Bley-Vroman, 1989, pp. 51–52). As Dekydtspotter and Hathorn (2005), citing Hale (1996), noted, "given the level of generality of principles of Universal Grammar in Minimalist understanding, all grammatical principles are instantiated in all languages. The distinction [between the first language (L1) and UG] becomes vitiated" (p. 317).

The empirical content of full transfer/full access virtually disappears under strong Minimalism, because full access reduces to taking the target language to be a human language (rather than, say, a system that does not involve human cognition and recursion and displacement, but one that perhaps functions by mapping, animal-like, directly from a finite number of vocalizations to entities in the world or one that consists of things like analog danger warnings and mating calls). So far, the discussion has concentrated on Minimalism. However, similar considerations arise in other trends in contemporary linguistics. The family of approaches that one might call constructionist is also minimal in its own way: Simpler syntax is probably the best worked-out representative (Culicover & Jackendoff, 2005; Jackendoff, 2002; Pinker & Jackendoff, 2005). Jackendoff (2005) indeed spoke of alternative minimalist visions. Construction grammar and item-based learning (Fillmore & Kay, 1999; Fillmore, Kay, & O'Connor, 1988; Goldberg, 1995, 2006; Tomasello, 1998, 2003) have much in common with simpler syntax.[9] In general, such approaches require syntactic structures that are much less abstract, hence more directly learnable on the basis of experience. To consider the implications of this family of approaches would take us too far afield, and our focus here will remain on mainstream generative grammar as represented by the Minimalist Program. Still, it is worth emphasizing that an elaborate system of sui generis innate linguistic knowledge—rich UG—on which much UG-in-SLA research has been built, is coming under question from both ends of the theoretical spectrum. As rich UG falls away, the FDH, and indeed all UG-based L2 research, must rethink its basic assumptions.[10]

Second, the typological theory of principles and parameters, at least as originally imagined in linguistic theory and adopted in L2 research, has shown itself to be untenable. The original conception assumed a

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

small number of parameters, with languagewide settings, that could unify apparently disparate superficial phenomena in a language. A single abstract (meta)parameter setting could derive a cluster of apparently disparate phenomena. Furthermore, the parameter values could be acquired by exposure to readily available input—triggers. The possible values, triggers, and deductive consequences are a part of rich UG. A full discussion of the demise of rich-UG parameter theory is beyond the scope of this article. See Newmeyer (2004, 2005) and the references cited there for an extensive discussion of the reasons why this approach to typology cannot work and has been generally abandoned: "Metaparameters do not exist" (Pica, 2001, cited in Newmeyer, 2004, p. 209). Additionally, in 20 years of SLA research, not a single study has convincingly demonstrated the sort of triggering and clustering that might have been expected. Quite the opposite: Studies have uniformly showed that aspects of a parameter (i.e., the individual phenomena) are learned separately; there is little evidence of strong deductivity. The work on the verb-movement parameter provides a particularly clear example (Ayoun, 2002; White, 1990/1991, 1991, 1992). In the Minimalist Program, it may turn out that abstract switches on lexical items may survive as a descendant of parameter theory, but their status seems quite different. Certainly, in keeping with Minimalist philosophy, these abstract switches must derive from the very nature of language as giving expression to the conceptual-intentional system. Definitely, they will not have the same consequences for L2 research as rich-UG metaparameters. This is a second reason to rethink the foundations of UG-based SLA research.

Third, there is now increasing evidence that many facts of language originally thought to be formally grammatical and attributed to the workings of rich UG may actually be the consequences of extragrammatical factors. The principle of subjacency, for example, has been widely investigated as a touchstone for the availability of UG in foreign language learning. However, there is now very good reason to believe that subjacency should be thought of as having its roots in some combination of processing and semantics-discourse structure (Hawkins, 1999, 2004; Kehler, 2002; Kluender, 1998; Kluender & Kutas, 1993; Levine & Sag, 2003). Hawkins's work is particularly persuasive in showing how a wide range of grammatical facts can be understood in this way.[11] It is, of course, possible that the grammar of a particular language may reflect, through grammaticization, such functional motivations. However, this does not mean that rich UG is required; nor can one definitely conclude, on the basis of foreign language learners' sensitivity to such functional factors, that they have access to rich UG.

Fourth and finally, it is no longer possible to make a clear distinction between processes (especially data-processing acquisition processes) used by the language faculty and those used more generally. The FDH

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

depended on this distinction. Current linguistic thinking explores the possibility that many of the processes of the language faculty (broadly conceived) are not exclusive to language. This position is most strikingly evident in Chomsky's current work: "The modular view of learning of course does not entail that the component elements of the module are unique to it" (2005, p. 5), and "we need no longer assume that the means of generating structured expressions are highly articulated and specific to language. We can seriously entertain the possibility that they might be reducible to language-independent principles" (p. 9). Note the striking contrast to "special principles specific to this domain" (Chomsky, 1980, p. 44) posited by rich UG. Chomsky cited as a concrete example the work of Gambell and Yang (2003, published in Gambell & Yang, 2004), who showed that general principles of data analysis will give reasonable results in identifying word boundaries. One can clearly go farther than this. Work that shows the potential of general data analysis processes is not restricted to that couched in generative terminology. Indeed, such research includes—in fact is dominated by—that of scholars who are decidedly skeptical of the generative enterprise. There are now significant advances in machine learning and mathematical linguistics that point to the potential of general data-processing methods for language acquisition (see Lappin & Shieber, 2007, for a review).[12] To take but one example, Clark and Eyraud (2006, 2007) showed how structure-dependent auxiliary fronting in *yes*/*no* questions can be learned by distributional learning.

Work along these lines makes simplifying assumptions; for example, Gambell and Yang (2003, as cited by Chomsky, 2005) adopted the assumption that words are associated with a single stress; this in effect permits the system to say "there's a word here somewhere; where are the edges?." Clark and Eyraud (2006, 2007) assumed that the system to be learned has the mathematical property of substitutability, which is essential to show the learnability of structure dependence. At this stage of the research, such simplifying—and almost certainly false—assumptions should not be disturbing. What is important is the nature of the assumptions themselves: They are linguistic. There is an interaction between features that relate to the nature of human language (e.g., stress-word associations or substitutability) and principles of data analysis surely also used in other domains. In other words, language-specificity of processing resides in the nature of the data rather than in the inherent language-specificity of the procedure.

If the data-processing acquisition processes of the language module are not unique to language, L2 research cannot use the distinction between domain-specific and domain-general processes to articulate the UG-access question or to formulate the FDH. To say that the processes of the language faculty are not unique to the language faculty is not to say that a language module does not exist. The

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

language faculty is a system of processes, and the fact that the processes used by the faculty may also be shared by other systems does not show that the module, qua system, does not exist. The faculty of language is a "subcomponent of (mostly) the brain" (Chomsky, 2004, p. 104), but "its elements might be recruited from or used for other functions" (p. 124). So, the question of whether foreign language is handled by the language faculty does not reduce to the question of whether the processes used by foreign languages are the same as those used in native languages.

To illustrate this point, an analogy may be helpful. Special-purpose computer programs—a word processor, a spreadsheet, or a concordancer—may well recruit the same routines (perhaps provided by the operating system); for example, on my computer, almost all application programs use the same text-entry and text-editing systems as well as the same file-management systems. The generalized regular expression processor of Unix is recruited in a wide range of programs, with minimal adaptations. The fact that two programs may use the same routines does not make them the same program. The user can use different programs to perform a similar function, and these different programs may use the same routines in performing those functions. In a set of corpus files, I may want to find a linguistic pattern and list the lines in which it occurs. For this, I may use a command-line pattern matcher (such as grep in Unix), or I could write a small special-purpose script (using a language like perl or python), or I can use a specialized text-processing program with a multifile search capability, or a special corpus-linguistics concordancer. These different systems all use essentially the same basic method of pattern matching; these pattern matching methods in the different systems are not sui generis. Nonetheless, the systems are different systems: The command-line grep is not the same as the concordancer. Additionally—this is important—using one program-system may be more efficient than another, depending on a more encompassing goal. A system is more than the sum of its parts: The manner in which the parts are integrated matters.

To return to the FDH: In the context of the original FDH, it was possible to say that child language development used a rich UG and that foreign language learning relied, instead, on the L1. However, rich UG may not exist. Additionally, if all UG principles—minimal as they are—are instantiated in all languages, then the distinction between access to UG and access to UG through L1 is without content.

In the original context of the FDH, it was also possible to say that child language development used domain-specific data-processing and learning procedures and that foreign language learning relied, instead, on domain-general learning and data processing. However, it is possible that the elements used by the language faculty are "recruited from or used for other functions" (Chomsky, 2004, p. 124). If there is

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

no distinction between sui generis language processes and other processes, then that distinction cannot be used for the FDH.

## TOWARD A NEW FORMULATION

Here is the dilemma: Without rich UG and language-specific processes, we are left without a framework within which to formulate an account of the properties of unreliability and nonconvergence. Essentially, the property of language, if taken seriously and combined with contemporary views of the nature of language, deprives us of an explanation for nonconvergence and unreliability. The next subsections propose a way forward: The leading idea is that elements of the language faculty itself provide the seeds of nonconvergence and unreliability.

### Patches as a Bridge

Consider the consequences of this quote from Wodehouse's (1946) *Joy in the Morning* reproduced in (6):

(6) "I balanced a thoughtful lump of sugar on the teaspoon." (chapter 5)

and these (also from Wodehouse) from *Uneasy Money* (1917) and *Jeeves in the Offing* (1960) in (7) and (8), respectively:

(7) "He was now smoking an uneasy cigarette and waiting for the blow to fall." (chapter 9)
(8) "His eyes widened and an astonished piece of toast fell from his grasp." (chapter 2)

The sugar is, of course, not thoughtful; nor is the cigarette uneasy; nor is the piece of toast astonished. Whereas Wodehouse's examples are particularly striking, the device is actually not that unusual, rather it is the well-known *transferred epithet* in traditional rhetoric: There is the "restless night." It intrigued linguists early in the development of generative grammar—and was first noted by Hall (1973). It is of interest because it is a novel way of associating grammatical structure with conceptual structure that does not inevitably follow from general principles; Hall spoke of the generalization as being "on the semantic plane" (p. 94). Although it is no doubt rooted in the mechanisms of metaphor (Warren, 1983) and one can understand it readily, it can become conventionalized—certainly, it became conventionalized for Wodehouse—and one can learn it.

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

Of course, there is nothing really new in the observation that there are idiosyncratic structures and mappings from syntax to conceptual structure in a language. The *what is X doing Y* (WXDY) construction, as in *What is that fly doing in my soup?* for instance (Fillmore & Kay, 1999), is a notable example that has been an inspiration to construction grammar. Knowledge of WXDY is part of the grammar of (many, although probably not all) native speakers of English.

Morgan (1972), in a study of verb agreement in English, noted that speakers of English sometimes boggle at examples like (9) or (10).

(9) ??Are/Is (either) John or his parents here?
(10) ??Are/Is (either) John's parents or his wife here?

Morgan observed:

> The speaker learns a relatively simple principle of agreement which somehow fails to extend to complex cases. This sometimes leads to patching the principle by adding subsidiary principles, and [verb agreement] VA ends up as a (possibly hierarchical) set of principles. The details of the patching vary considerably from speaker to speaker, and some speakers apparently fail to patch at all for certain cases. (pp. 284–285)

Morgan suggested that patching happens when, for some reason, the general mechanisms of grammar learning fail. Learning fails because a single general principle cannot be derived from the available data and extended to all of the relevant cases. In cases for which the usual general language acquisition processes cannot apply, the learner is driven to develop a *patch*—to use Morgan's terminology—for the resulting hole caused by acquisition failure. Patches have a couple of characteristics of particular interest to those who are concerned with SLA. First, there is considerable variability among speakers (i.e., *diversity* in Morgan's terminology). This contrasts with the complete unanimity that one expects to find in UG-driven development—what we are calling convergence. Second, there is often uncertainty within a speaker: Morgan said some speakers "just don't know" (p. 284).

Emonds (1986) noted the considerable interspeaker variability (and uncertainty) in the mastery of prestige rules of prescriptive grammar. These, too, are not derived by the normal acquisition system. In this case, Emonds noted, the variability is exploited socially as a device for "arbitrarily signaling membership" in the "business-government-legal-religious-academic—the high culture 'community' (i.e. class)" and for "re-enforcing exclusion from it" (pp. 123–124). The TOEFL exploits the variability and uncertainty in the foreign language knowledge to effect similar inclusions and exclusions.

McCawley (1988) extended the notion of patch to cases like *he can't seem to do it right* (where *can't* clearly does not have scope over *seem*),

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

*the usual men were drunk* (*men* are not *usual*), and similar cases. These are reminiscent of the Wodehousean transferred epithet (as McCawley himself noted). McCawley suggested that

> a patch is a way of employing the syntactic means provided by the language so as to allow one to say things of the same sort that can normally be expressed by a given construction in some class of cases that the normal syntactic means of the language do not provide for. (p. 746)

One might understand "normal syntactic means" (p. 746) as the means developed in the operation of normal acquisition, to use McCawley's terminology.

The importance of such phenomena has become apparent to contemporary mainstream generative grammar. Lasnik and Sobin (2000) developed a theory of what they called *viruses*. (They explored the properties of the rules for *whom*.) Viruses—the concept of which seems very similar to patches—are superficial and parasitic on existing structures: These are shallow principles. Sobin's earlier (1997) study of viruses, published in *Linguistic Inquiry*, included tables of grammaticality judgments that could have been taken right out of an article published in *Studies in Second Language Acquisition*. The grammaticality judgments presented in Sobin's tables show the same sort of nonconvergence and uncertainty that we expect to find in SLA research.

The crucial point in all of these cases is this: Although *can't seem*, *the usual men*, *what's that fly doing in my soup*, *for you and I*, and *the person whom I am told is responsible* are not part of the core system that arises by the interaction of the principles of UG and input during development, there is no reason to believe that they are not language and that the structures cannot interact with the language module. Lasnik and Sobin (2000) put it very clearly: "Viruses comprise a subtheory of the Minimalist Program, distinct from the core system, though interactive with it" (p. 369). Although viruses are not part of the grammar proper, they are linguistic. With foreign language learning in mind, consider this observation:

> Grammatical viruses mimic the original phenomenon, but without complete success, since grammatical viruses are not of the same descriptive power as the devices of the grammar proper. Viruses lack generality and the possibility of normal interactivity with the core grammatical system. Such mimicry via grammatical viruses leaves the sort of telltale discrepant behavior which we have discussed here. (p. 370)

How much can patches do? Lasnik and Sobin (2000) aimed to constrain patches tightly: "Virus Theory claims that such manipulations are not manipulations of the true, abstract system, but are quite superficial surface manipulations of limited features of the objects that the

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

system produces" (pp. 367–368). Nonetheless, the system clearly has access at least to syntactic category information and to concepts like subject and theta-assigning verb. Moreover, the system can contain statements such as *check ACC (on a particular element)*, which shows that abstract features and at least the basic ideas of a morphological mechanism are available. Lasnik and Sobin are concerned to limit the extent to which linguistic-theory-internal concepts can be referred to in a patch. (It is clear that there are dangers to Fodorian modularity if the patch system, which seems to be affected by conscious awareness and may be amenable to explicit instruction, can manipulate objects deep within the core.) In this vein, Lasnik and Sobin proposed, for example, that "subject status may be determined here by considering the theta role or meaning of elements 1-2 (that is, *whom*) and 3, something that a speaker must be cognizant of to understand the sentence" (p. 359). Note that the elements that can be used in viruses are those that the speaker can be cognizant of (although it is a bit difficult to see that the learner can be cognizant of the abstract feature-checking mechanism). To the extent that foreign language learners are creating patches, cognizance puts restrictions on the abstractness of the rules that can be proposed. (For foreign language learning, the connection to noticing theory [Schmidt, 1990] is obvious.)

However, the tight constraints that Lasnik and Sobin (2000) proposed may not be inherent limitations of the patching system. It may simply be that the core acquisition system of native language development is robust enough so that holes are tiny as well as few and far between. Who knows what patching might be capable of? What if the core acquisition system simply does not work (to paraphrase Chomsky, 1997) and the holes are many and large?[13]

## Data Processing Outside the Language System

Can "is just not working" (Chomsky, 1997, p. 128) be characterized within contemporary conceptualizations of the relationship of domain-general and domain-specific systems? Suppose, following Chomsky (e.g., 2005) and much current work in mathematical learning theory and natural language processing, that the elements recruited by the language system are not specific to language. Suppose the language system itself is not working properly. This does not mean that these processes, which are not restricted to language, are not available. The whole system is greater than the sum of its parts. To return to the computer program analogy, a specialized concordance program—of the type that might be used in a corpus linguistics class—is basically just a packaging of otherwise available multipurpose routines into an efficient and

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

easy-to-use special-purpose package. If you do not have such a specialized program, you can still do the homework, by pulling in individual components, cobbling something together. However, it will be less efficient than an integrated system and probably more prone to error. (Additionally, different students in the class are more likely to get different results.)

## Shallow Processing

In addition to patches or viruses and the use of data-processing techniques apart from their integration into the language faculty, recent discoveries about language processing give us another piece of the puzzle. There are two parts to this language processing picture: The first involves native language and the second involves foreign language.

First, there is increasing reason to believe that (native) language processing employs both what one might call shallow mechanisms and deep mechanisms. Townsend and Bever (2001)—who accept mainstream Minimalism—proposed that "while grammatically defined representations appear to be computed during language behavior, the grammatical rules that define them may not be used" (p. 45). Processing often uses relatively superficial patterns that reflect the statistical structure of language.

Second, for L2 processing, Clahsen and colleagues have reviewed existing research and reported neurophysiological work using measures such as event-related potentials (see Clahsen & Felser, 2006b, and the references cited there). The evidence thus far is not completely conclusive. However, Clahsen and Felser tentatively proposed the shallow structure hypothesis: "For nonnative language processing, we suggest that the syntactic representations adult L2 learners compute for comprehension are shallower and less detailed than those of native speakers" (p. 32). Of course, if Townsend and Bever are right, then native speakers also can and do compute shallow representations, perhaps using them in parallel with grammatically defined representations. If this is true, then nonnative speakers in computing shallow representations are employing mechanisms of the language faculty, but they rely on the shallow route exclusively or primarily, perhaps because grammatically defined representations are unavailable or not fully definite (i.e., patched).[14]

There is an increasing consensus among processing researchers that some kind of dual-mechanism model will have a place in accounting for differences between native language acquisition and adult foreign language learning. Ullman's (e.g., 2001, 2005, 2006) application of

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

the declarative-procedural distinction seems particularly promising. Under Ullman's (2001) model, "linguistic forms whose grammatical computation depends upon procedural memory in L1 are posited to be largely dependent upon declarative/lexical memory in L2" (p. 105). "L2 learners should tend to rely heavily on declarative memory, even for functions that depend on the procedural system in L1" (Ullman, 2006, p. 99); that is, in L2 there is a "shift from procedural to declarative memory" (Ullman, 2001, p. 110) of grammatical processing. It is quite possible to relate this idea to the concept of patches. One might conceptualize patches as words, which would thus be handled by declarative-lexical memory in L2. Indeed, Ullman (2006) suggested that L2 learners may rely largely on "stored schemas or constructions" (p. 99), referring to Fillmore et al. (1998) and L2 productivity may involve "the use of 'rules' learned in declarative memory" (Ullman, 2006, p. 99); thus a patch could be thought of as such a "rule"—although not as a rule (i.e., without quotes). Paradis (e.g., 1994, 2004) has made a rather similar proposal, but he also focused on the distinction between implicit and explicit knowledge—these are related to level of conscious awareness.

Although there are differences among the proposals now being made in the psycholinguistics and neurolinguistics of L2 processing, there seems to be a general compatibility between them and the proposals made here in the context of the FDH. Still, the correspondence between the family of dual-processing proposals and the FDH is not completely straightforward. Much depends on working out the knotty relationship between "getting good at something" (through practice), the status of different memory and processing systems, or both, and the problem of conscious awareness.

## SUMMARY AND CONCLUSION

The proposals outlined here for recasting UG-in-SLA research in light of the evolving linguistic-theoretic context share a common thread. In characterizing the fundamental nature of foreign languages, these proposals make use of possibilities that are inherent in the language module but that, for native languages, may not be utilized in the same way or to the same extent. Patches or viruses, for example, are peripheral in the case of native languages: They are a minor factor—more like a wrinkle in the system—but they are there as a potentiality, awaiting more extensive exploitation. In foreign languages, they may carry great weight. The data-processing elements recruited by the native language acquisition system are not specific to language; therefore, these elements can be used even if the language acquisition system is no longer working properly, although they may not work as well for

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

language apart from their integration in the system. The shallow representations that foreign language employs in online processing are not absent in native language processing, but, in foreign language processing, they are relied on more heavily (or even exclusively).

The research program that would arise under this conception would have three thrusts. First, we need a much greater understanding of the nature and limits of patches and prescriptive rules, both in native languages and in foreign languages. Second, because we now understand that the data-processing elements of the language module are not sui generis, we can begin to integrate discoveries from machine learning, mathematical linguistics, probabilistic reasoning, and the like into foreign language acquisition theory. Crucial to the newly recast FDH is the speculation that such processes can work even outside a (hypothetically impaired) dedicated language acquisition system but will be less reliable when not integrated into the specialized system. Third, psycholinguistic research into online processing is required to explore further the interplay of rough-and-ready shallow processing and full grammar-driven processing.

To conclude, let us return to Chomsky's (1997) statement of the FDH. The key notions were "just not working" and "something strange" (p. 128). We can change "just not working" (p. 128) to "just not working in the same way": The core system is not working (or not working well), so the patching system is taking the burden. To the foreign language learning system, language is neither more nor less strange than the funny order of *can't* and *seem* in *he can't seem to get it right*. The language faculty has means of dealing with strangeness and uncertainty, and these are heavily employed in foreign language learning.

## NOTES

1. The article "The Logical Problem of Foreign Language Learning" was written and accepted for publication in *Linguistic Analysis* in late 1986. It circulated widely in manuscript form in 1987. It did not, however, appear in the journal until 1990. In the meantime, it was included in two different abridged forms in collections (Bley-Vroman, 1988, 1989). The version most frequently cited is probably that in *Linguistic Perspectives on Second Language Acquisition* (Schachter & Gass, 1989). However, despite its later date, the article in *Linguistic Analysis* is the original and the most complete; it is the version referred to here.

2. There are several useful summaries of relevant research (Braidi, 1998; Hawkins, 2001; Herschensohn, 2000; Montrul, 2000; Slabakova, 2006; White, 1989, 2003).

3. Pullum and Scholz (2002) noted seven putative properties of the child's accomplishment: speed, reliability, productivity, selectivity, underdetermination, convergence, and universality. Pullum and Scholz are somewhat skeptical, especially about underdetermination (the so-called poverty of the stimulus).

4. I have tried to avoid the terms *success* or *failure*. These terms are too loaded, too evaluative, too likely to be thought of as applying to struggling human learners rather than to acquisition models. Flesh-and-blood people are quite able to operate successfully, even with systems that do not reliably acquire language, for instance, in the technical

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

learnability sense. To take a famous example, it would be wrong to characterize Wes (Schmidt, 1983) as a failure.

5. On the concept of probably approximately correct, see Valliant (1984).

6. This is not to say that the system is able to learn reliably even in the face of misleadingly unrepresentative input; for example, one minimal characteristic of the input might be that the distribution of the input reflects the probabilities assigned by the grammar, given a stochastic grammar. See Clark and Thollard (2004), among others, for the relevant notion of partially distribution free.

7. Although the focus here is on generative grammar, especially GB/PP theory, it is important to emphasize that these considerations apply not just to generative or nativist theories taken as a basis for foreign language learning theory; they apply equally to other formal approaches, even of radically different camps, for all are designed to achieve something analogous to reliability and convergence.

8. This is something of an oversimplification, or at least an idealization. To account for language change, the grammar of one generation must be able, in principle, to deviate from the grammar of a successor generation. However, such deviations are highly constrained and arguably result not from unreliability of the acquisition system itself but rather to shifts in the input data. It would be possible, in principle, to explore the idea of using language change mechanisms to account for the unreliability of foreign language learning. However, even if that were possible, there is still the problem of nonconvergence, because the change process does not destroy the "striking uniformity of the resulting grammars" (Chomsky, 1965, p. 11). Speech communities still exist. Indeed, the creolization process, which clearly does not replicate the language of the predecessor (pidgin) generation, operates reliably and convergently, in effect creating a grammar-sharing speech community.

9. In this same general family, one might also include O'Grady (2005, 2008). In O'Grady's account, grammatical phenomena emerge from processing—indeed, there is no grammar apart from a processor, again dispensing with an innate rich UG.

10. A great deal of the most recent work in syntax has been devoted to attempts to formulate accounts of well-known linguistic phenomena in Minimalist terms, dispensing with the direct use of notions that cannot readily be accommodated, such as government, m-command, d-structure, and X-bar theory, among others. Such reformulation is important, in order to show that the Minimalist Program has a chance of success. Additionally, to the extent that such reformulation is possible, it will likewise be possible to recast discussions of foreign language learning in Minimalist terms; for example, a discussion of the acquisition of English tense, auxiliaries, and modals can utilize a Minimalist description of the phenomenon just as well as any other. Thus, one might say that the consequences of Minimalism for SLA are not serious, but only to the extent that the investigation of UG in foreign language learning is itself not serious. Simple restatements of phenomena in a new framework are not likely to bring us to any deep understanding of the nature of foreign language learning.

11. A different sort of functional approach to the explanation of syntactic universals is to suggest that languages are the way they are because the brain learns them (Christiansen & Chater, 2008). Similar trends are evident in phonological research. Evolutionary phonology (Blevins, 2006) tries to show that many typological features of phonological systems attributed to the formal characteristics of a rich UG are best explained by constraints on form and function, phonetic motivation, and historical development, among others. Interestingly, Blevins occasionally adduced L2 learning to explain an otherwise odd phonological development. For Blevins, L2 learning is in the category of "non-natural," "external" factors (p. 120).

12. My own view is that Bayesian methods will be especially appropriate for foreign language learning (Berger, 1985; Oaksford & Chater, 2007).

13. Weiß (1998, 2001, 2004) suggested in his study of the characteristics of standard languages (especially of standard German) that there are two basic kinds of languages: languages subject to L1 acquisition (his N1 languages) and languages not subject to L1 acquisition (his N2 languages). Standard German was not acquired as a L1 for at least 500 years. Weiß found aspects of N2 languages, such as standard German, reminiscent of patches or viruses and, in effect, attributed such features to their being learned as foreign languages. One can perhaps view the N1-N2 distinction as a version of the FDH based on

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

patches or viruses (although Weiß did not put it that way). Weiß was primarily concerned to advocate caution in building theories of the basic nature of language on N2 languages.

14. Clahsen and Felser (2006a) clearly saw these implications of Townsend and Bever (2001):

> According to the integrated processing model proposed by Townsend & Bever (2001), for example, the L1 comprehension mechanism normally assigns two different kinds of representations to an input string, a rough-and-ready 'pseudosyntax' representation based on lexical information and statistical patterns, and a fully specified syntactic description. […] Let us assume, then, that the human language processing system makes available two different routes for computing sentence interpretations, which usually work in parallel. While the full parsing route is fed by the grammar (a system of symbolic rules and principles of structure building), shallow processing is guided by lexical-semantic and pragmatic information, world knowledge, and strong associative meaning or form patterns. (p. 117)

## REFERENCES


Ayoun, D. (2002). *Parameter-setting theory in first and second language acquisition*. New York: Continuum.

Beck, M.-L. (1998). L2 acquisition and obligatory head movement: English-speaking learners of German and the local impairment hypothesis. *Studies in Second Language Acquisition, 20*, 311–348.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (2nd ed.). New York: Springer Verlag.

Blevins, J. (2006). A theoretical synopsis of evolutionary phonology. *Theoretical Linguistics, 32*, 117–166.

Bley-Vroman, R. (1988). The fundamental character of foreign language learning. In W. Rutherford & M. Sharwood Smith (Eds.), *Grammar and second language teaching* (pp. 19–30). Rowley, MA: Newbury House.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 41–68). New York: Cambridge University Press.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis, 20*, 3–49.

Braidi, S. (1998). *The acquisition of second-language syntax*. Oxford: Oxford University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.

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

Chomsky, N. (1981b). Principles and parameters in syntactic theory. In N. Hornstein & D. Lightfoot (Eds.), *Explanation in linguistics: The logical problem of language acquisition* (pp. 123–146). London: Longman.

Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.

Chomsky, N. (1986a). *Barriers*. Cambridge, MA: MIT Press.

Chomsky, N. (1986b). *Knowledge of language: Its nature, origin and use*. New York: Praeger.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomksy, N. (1997). Questions: São Paulo. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, *13*, 123–128.

Chomsky, N. (2000). *New horizons in the study of language and mind*. New York: Cambridge University Press.

Chomsky, N. (2004). Beyond explanatory adequacy. In A. Belletti (Ed.), *Structures and beyond: The cartography of syntactic structures* (Vol. 2, pp. 104–131). Oxford: Oxford University Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry, 36*, 1–22.


https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31*, 489–509.
Clahsen, H., & Felser, C. (2006a). Continuity and shallow structures in language processing. *Applied Psycholinguistics, 27*, 107–126.
Clahsen, H., & Felser, C. (2006b). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3–42.
Clahsen, H., & Hong, U. (1995). Agreement and null subjects in German L2 development: New evidence from reaction time experiments. *Second Language Research, 11*, 57–87.
Clark, A., & Eyraud, R. (2006). Learning auxiliary fronting with grammatical inference. In *Proceedings of the 8th Conference on Computational Natural Language Learning (CoNLL-X)* (pp. 125–132). Stroudsburg, PA: Association for Computational Linguistics.
Clark, A., & Eyraud, R. (2007). Polynomial identification in the limit of substitutable context-free Languages. *Journal of Machine Learning, 8*, 1725–1745.
Clark, A., & Thollard, F. (2004, August). *Partially distribution-free learning of regular languages from positive samples*. Paper presented at the 20th International Conference on Computational Linguistics, Geneva.
Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford: Oxford University Press.
Dekydtspotter, L., & Hathorn, J. C. (2005). *Quelque chose*. *de remarquable* in English-French acquisition: Mandatory, informationally encapsulated computations in second language interpretation. *Second Language Research*, *21*, 291–323.
Emonds, J. E. (1986). Grammatically deviant prestige constructions. In M. Brame, H. Contreras, & F. Newmeyer (Eds.), *Festschrift for Sol Saporta* (pp. 93–129). Seattle, WA: Noit Amrofer.
Eubank, L. (1993/1994). On the transfer of parametric values in L2 development. *Language Acquisition, 3*, 183–208.
Eubank, L. (1994). Optionality and the initial state in L2 development. In T. Hoekstra & B. D. Schwartz (Eds.), *Language acquisition studies in generative grammar: Papers in honor of Kenneth Wexler from the 1991 GLOW workshops* (pp. 369–388). Amsterdam: Benjamins.
Eubank, L. (1996). Negation in early German-English interlanguage: More valueless features in the L2 initial state. *Second Language Research*, *12*, 73–106.
Fillmore, C. J., & Kay, P. (1999). Grammatical constructions and linguistic generalizations: The *what's X doing Y?* construction. *Language, 75*, 1–33.
Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomacity in grammatical constructions: The case of *let alone*. *Language*, *64*, 501–538.
Gambell, T., & Yang, C. D. (2003). *Scope and limits of statistical learning in word segmentation*. Unpublished manuscript, Yale University, New Haven, CT.
Gambell, T., & Yang, C. D. (2004). Scope and limits of statistical learning in word segmentation. In K. Moulton & M. Wolf (Eds.), *Proceedings of the 34th Northeastern Linguistic Society Meeting (NELS)* (pp. 29–30). Amherst, MA: GLSA.
Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
Goldberg, A. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
Hale, K. (1996). Can UG and L1 be distinguished in L2 acquisition? *Behavioral and Brain Sciences, 19*, 726–728.
Hall, R. A. Jr. (1973). The transferred epithet in P. G. Wodehouse. *Linguistic Inquiry*, *4*, 92–94.
Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language*, *75*, 244–285.
Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
Hawkins, R. (2001). *Second language syntax: A generative introduction*. Oxford: Blackwell.
Hawkins, R., & Chan, C. Y.-H. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis.' *Second Language Research*, *13*, 187–226.
Herschensohn, J. (2000). *The second time around: Minimalism and L2 acquisition*. Amsterdam: Benjamins.

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

Ioup, G., Boustagui, E., Tigi, M. E., & Moselle, M. (1994). Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition, 16*, 73–98.
Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.
Jackendoff, R. (2005). Alternative minimalist visions of language. *Chicago Linguistics Society (CLS)*, *41*, 189–226.
Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition*, *97*, 211–225.
Kehler, A. (2002). *Coherence, reference and the theory of grammar*. Stanford, CA: Center for the Study of Language and Information.
Kluender, R. (1998). On the distinction between strong and weak islands: A processing perspective. In P. Culicover & L. McNally (Eds.), *Syntax and semantics 29: The limits of syntax* (pp. 241–279). San Diego, CA: Academic Press.
Kluender, R., & Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and Cognitive Processes, 8*, 573–633.
Lappin, S., & Shieber, S. M. (2007). Machine learning theory and practice as a source of insight into Universal Grammar. *Journal of Linguistics*, *43*, 393–427.
Lardiere, D. (1998a). Case and tense in the 'fossilized' steady state. *Second Language Research, 14*, 1–26.
Lardiere, D. (1998b). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research, 14*, 359–375.
Lasnik, H., & Sobin, N. (2000). The who/whom puzzle: On the preservation of an archaic feature. *Natural Language and Linguistic Theory, 18*, 343–371.
Levine, R. D., & Sag, I. A. (2003). WH-nonmovement. *Gengo Kenkyu: Journal of the Linguistic Society of Japan, 123*, 171–220.
McCawley, J. A. (1988). *The syntactic phenomena of English*. Chicago: University of Chicago Press.
Montrul, S. (2000). Transitivity alternations in L2 acquisition: Toward a modular view of transfer. *Studies in Second Language Acquisition*, *22*, 229–273.
Morgan, J. L. (1972). Verb agreement as a rule of English. *Chicago Linguistics Society (CLS)*, *8*, 278–286.
Newmeyer, F. J. (2004). Against a parameter-setting approach to typological variation. *Linguistic Variation Yearbook, 4*, 181–234.
Newmeyer, F. J. (2005). *Possible and probable language: A generative perspective on language typology*. Oxford: Oxford University Press.
Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.
O'Grady, W. (2005). *Syntactic carpentry: An emergentist approach to syntax*. Mahwah, NJ: Erlbaum.
O'Grady, W. (2008). Language without grammar. In N. C. Ellis & P. Robinson (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 139–167). London: Routledge.
Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N.C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393–419). San Diego, CA: Academic Press.
Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: Benjamins.
Pica, P. (2001). Introduction. *Linguistic Variation Yearbook*, *1*, v–xii.
Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, *95*, 201–236.
Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, *16*, 103–133.
Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, *19*, 9–50.
Schachter, J., & Gass, S. M. (Eds.). (1989). *Linguistic perspectives on second language acquisition*. New York: Cambridge University Press.
Schmidt, R. (1983). Interaction, acculturation, and the acquisition of communicative competence. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 137–174). Rowley, MA: Newbury House.

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 17–46.
Schumann, J. (1978). *The pidginisation process: A model for second language acquisition*. Rowley MA: Newbury House.
Schwartz, B. D., & Sprouse, R. A. (1994). Word order and nominative case in non-native language acquisition: A longitudinal study of (L1 Turkish) German interlanguage. In T. Hoekstra & B. D. Schwartz (Eds.), *Language acquisition studies in generative grammar: Papers in honor of Kenneth Wexler from the 1991 GLOW workshops* (pp. 317–167). Amsterdam: Benjamins.
Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, *12*, 40–72.
Slabakova, R. (2006). Is there a critical period for semantics? *Second Language Research*, *22*, 302–338.
Sobin, N. (1997). Agreement, default rules, and grammatical viruses. *Linguistic Inquiry*, *28*, 318–343.
Tomasello, M. (1998). The return of constructions. *Journal of Child Language*, *25*, 431–442.
Tomasello, M. (2003). *Constructing a language*. Cambridge, MA: Harvard University Press.
Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.
Tsimpli, I.-M., & Roussou, A. (1991). Parameter resetting in L2? *UCL Working Papers in Linguistics*, *3*, 149–169.
Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*, 105–122.
Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative procedure model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition* (pp. 141–178). Washington, DC: Georgetown University Press.
Ullman, M. T. (2006). The declarative/procedural model and the shallow structure hypothesis. *Applied Psycholinguistics*, *27*, 97–105.
Valiant, L. (1984). A theory of the learnable. *Communications of the ACM*, *27*, 1134–1142.
Warren, B. (1983, May). *A nervous report: The transferred epithet and body language*. Paper presented at the Nordic Conference for English Studies, Hanasaari/Hanaholmen, Finland.
Weiß, H. (1998). *Syntax des Bairischen: Studien zur Grammatik einer natürlichen Sprache*. [The syntax of Bavarian: Studies on the grammar of a natural language.] Tübingen: Niemeyer.
Weiß, H. (2001). On two types of natural languages: Some consequences for linguistics. *Theoretical Linguistics*, *27*, 87–103.
Weiß, H. (2004). A question of relevance: Some remarks on standard languages. *Studies in Language*, *28*, 648–674.
White, L. (1989). *Universal Grammar and second language acquisition*. Amsterdam: Benjamins.
White, L. (1990/1991). The verb-movement parameter in second language acquisition. *Language Acquisition*, *1*, 337–360.
White, L. (1991). Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Second Language Research*, *7*, 133–161.
White, L. (1992). Long and short verb movement in second language acquisition. *Canadian Journal of Linguistics*, *37*, 273–286.
White, L. (2003). *Second language acquisition and Universal Grammar*. New York: Cambridge University Press.

https://doi.org/10.1017/S0272263109090275 Published online by Cambridge University Press