

16 Stabilization and Fossilization in Interlanguage Development

MICHAEL H. LONG

1 Introduction

A construct first introduced into the field of SLA by Selinker in 1972, fossilization has become widely accepted as a psychologically real phenomenon of considerable theoretical and practical importance. The literature reveals several problems, however. Fossilization is alternately explanandum and explanans. Its definition and alleged scope vary markedly across writers and studies. So do the research designs, subjects, data, and measurement criteria considered relevant. So do the explanations offered for it when fossilization is treated as product, not process. And while it is often said that fossilization is pervasive, especially in adult SLA, the evidence to date has been largely impressionistic.

After tracing the history and evolution of the idea, a review of empirical studies on fossilization, including some recent longitudinal work, will show almost all to have suffered from one or more of four problems: assuming, not demonstrating, fossilization; selecting inappropriate learners for study; basing findings on insufficient data; and using inadequate analyses. For a variety of reasons, most explanations for fossilization are equally unsatisfactory. It will be concluded, therefore, that fossilization of IL grammars may occur, but that until research shows convincingly that it does, researchers would do better to focus on describing and explaining the well-attested phenomenon of stabilization, a strategy with several advantages from a theory-construction perspective.

2 The Theory

“Fossilization” is a construct first introduced into the SLA literature by Selinker (1972), who appears to have seen it as a way of both characterizing and explaining the product of the SLA process in terms of what many observers consider one of its single most salient qualities (compared to L1A), that is,

relative failure. The end-state was viewed as a grammar which differed from that of the target-language variety, among other ways, in its permanent retention of deviant rules and forms, despite adequate opportunity for improvement – forms which persistently reappeared in L2 performance long after they were thought to have been supplanted, a phenomenon referred to by Selinker as “backsliding.” The permanent non-nativelike state was termed “fossilization” (as product), while “fossilization” (as process), constrained by L1 transfer, was viewed as part of the individual learner’s underlying psychological structure, a putative cognitive mechanism which could explain the failure. Fossilization was:

a mechanism which is assumed also to exist in the latent psychological structure . . . Fossilizable linguistic phenomena are linguistic items, rules, and subsystems which speakers of a particular NL will tend to keep in their IL relative to a particular TL, no matter what the age of the learner or the amount of explanation and instruction he [*sic*] receives in the TL . . . A crucial fact which any adequate theory of second language learning will have to explain is this regular reappearance or re-emergence in IL productive performance of linguistic structures which were thought to be eradicated. This behavioral reappearance is what has led me to postulate the reality of fossilization and ILs. (Selinker, 1972, p. 215)¹

Recognition of how widespread acceptance of the notion quickly became can be seen in its qualifying as an entry in a non-field-specific dictionary just five years later, a feat apparently accomplished by no other SLA term before or since:

fossilize 5. *Ling.* (of a linguistic form, feature, rule, etc.) to become permanently established in the interlanguage of a second-language learner in a form that is deviant from the target-language norm and that continues to appear in performance regardless of further exposure to the target language. (*The Random House Dictionary of the English Language*, 1987, p. 755)

And just five years after that, while admitting that definitions of fossilization varied widely, Selinker reported the existence of “literally hundreds of studies in the literature which claim to have shown a fossilized phenomenon, or speculate on a fossilizable phenomenon, or assume fossilization and speculate on its possible cause in the case under study” (Selinker, 1992, p. 250).

While permanence and deviance despite favorable conditions for change, and L1 transfer as a causal factor (see below), have been retained by Selinker as defining criteria over the years, other emphases have shifted somewhat, from a predominant focus on performance to one on underlying competence, and from fossilization as a global IL phenomenon to a more differentiated approach. Paradoxically, the changes have made the construct more restricted, yet less verifiable. Thus, Selinker and Lamandella wrote: “Fossilization is the permanent cessation of IL learning [*sic*]² before the learner has attained target

language norms at all levels of linguistic structure and *in all discourse domains* in spite of the learner's positive ability, opportunity or motivation to learn or acculturate into target society" (Selinker and Lamendella, 1978, p. 187, emphasis added). While increasing the power of the theory (a negative), implying that fossilization may occur in individual "domains" appears to make identification easier for the researcher than verifying that change has ceased everywhere in a grammar, yet really makes it harder, since, as detailed below, "discourse domain" remains a nebulous construct to this day. A year later, Selinker suggested that fossilization was not only domain-dependent, but context-dependent, and so could be evidenced by variability ("fluctuation") across contexts, not just by uniformity in performance across all contexts, and was meaningfully sought under conditions of natural exposure, that is, in second, as opposed to foreign, language settings. Again, "context" was undefined, and in practice difficult to operationalize. Selinker also attempted to deal with the obvious problem of what would constitute "permanence," and more to the point, what the lower bounds might be for an empirical test. Fossilization was now:

a situation in which the learner might produce a target language form correctly in one context but not in another, thereby evidencing a fluctuation in interlanguage performance. In order to qualify as fossilization, this fluctuation would have to have persisted in the learner's speech for an extended period of time (perhaps two to five years at the very least) – in spite of copious interaction with native speakers in an environment where the learner's L2 is spoken as a first language. (Selinker, 1989, p.c., cited in Bean and Gergen, 1990, p. 206)

Again, unless it is possible to specify where one "context" (and/or "discourse domain") ends and another begins, testing a claim that all or part of a grammar has fossilized becomes impossible. And if persistence of unvarying IL phenomena for from at least two to five years is required to qualify as evidence of fossilization, it should be noted right away that only three studies of putative fossilization in nearly 30 years (Han, 1998, 2000a; Lardiere, 1998a, 1998b; Long, 1997) have lasted that long.³

Finally, if, as seems uncontroversial, and as has been recognized by Selinker (Selinker and Han, 1996; Selinker and Lakshmanan, 1992), stabilization is the first sign of (putative) fossilization, and if the only difference between stabilization and fossilization is permanence (see, e.g., Bley-Vroman, 1989),⁴ then including persistent "fluctuation" as a legitimate index of fossilization creates another problem. The dictionary already quoted defines stabilization thus:

stabilize 2. to maintain at a given or *unfluctuating* level or quantity. (*The Random House Dictionary of the English Language*, 1987, p. 1852, emphasis added)

Fluctuation is not part of stabilization, yet stabilization is the precursor to fossilization, which can supposedly include fluctuation.

The various definitions of fossilization as process and product raise several methodological difficulties concerning, among other matters, testability, scope, learner age, unit of analysis, and deviance. First, where *testability* is concerned, a claim that something in a person's make-up is "permanent" is unfalsifiable during her or his lifetime, yet permanence is the only quality distinguishing fossilization from stabilization. Either an inevitably somewhat arbitrary minimum period must be specified as acceptably long for permanence to be inferred, therefore, or a claim of fossilization remains untestable. However, given that both U-shaped behavior and renewed language development after periods of plateau-like stability, some lasting for several years, are widely attested characteristics of normal child first and second language acquisition (see, e.g., Bowerman, 1982; Harley and Swain, 1984, respectively), understanding the causes of *stabilization* (and destabilization) would seem to promise as much for SLA theory as work on fossilization, and do so without fossilization's attendant theoretical and empirical baggage. Selinker recognizes the empirical problem, but not the potential implication: "at any point in time it is nonetheless very difficult, if not impossible, to tell, at a particular point in time, if a learner's *stabilized* IL is in fact fossilized. Thus it is common in SLA discussion to distinguish theoretically 'permanent fossilization' from 'temporary stabilization' of the IL" (Selinker, 1993, p. 16). The question, however, is not whether such a distinction can be made "theoretically," but whether it is useful for SLA theory construction to do so, and with what theoretical and empirical consequences.⁵ Also, the two processes might share the same surface characteristics, but differ in their underlying causes.⁶

Second, as noted above, the *scope* of putative fossilization remains unspecified. Learners do not fossilize, and neither do whole ILs or whole IL systems (syntax, phonology, etc.); rather, IL development within certain contexts and "discourse domains" – roughly, topics mediated by personal life history – supposedly does. Thus, according to Selinker and Douglas (1985, 1989), a structure can be fossilized in one discourse domain, while still developing in another. But contexts are often vague, defined by a host of sometimes rather nebulous sociolinguistic and social-psychological parameters (see Douglas, 2000, pp. 41–74), and discourse domains turn out to be even more elusive. Douglas writes:

Douglas and Selinker (1985) use the term *discourse domain* to refer to the [learner's] internal interpretation of context . . . Douglas and Selinker define discourse domain as a cognitive construct created by a language learner as a context for interlanguage development and use. Discourse domains are engaged when strategic competence, in assessing the communicative situation, recognizes cues in the environment that allow the language user to identify the situation and his or her role in it. If there are insufficient cues, if they are unrecognized by the language user, or if they are contradictory or ambiguous, the result will be uncertainty and stumbling around. (Douglas, 2000, p. 46)

Identification of discourse domains, that is to say, involves considerable ambiguity and risk of misinterpretation by both learner and researcher. Discourse

domains, moreover, are idiosyncratic (Selinker and Douglas, 1985), only identifiable for each learner empirically, a posteriori. This means not only that testing a fossilization claim is laborious, involving identification of discourse domains for that learner first, but that generalization and prediction are impossible.⁷

Third, Selinker repeatedly asserts that fossilization, resulting in non-target-like ultimate attainment, operates in learners regardless of *age*; for example, as quoted above, “no matter what the age of the learner” (1972, p. 215). While its appearance in *child*, as well as adult, SLA is necessary if fossilization is to qualify as a phenomenon characterizing *second*, as opposed to *adult*, language acquisition, or just *adult second* language acquisition (an issue to which we return), the fact is that no studies have shown fossilization in child L2 acquirers,⁸ and it is doubtful whether this would ever happen with children learning an L2 any more than with child L1A. On the contrary, given adequate opportunity, children appear to attain nativelike levels in a second language, just as they do in their first. A more likely scenario, albeit still a controversial one in some quarters (see, e.g., Birdsong, 1999), is that the ability to acquire either a first or a second language to nativelike levels is maturationally constrained: learners first exposed before the offset of one or more sensitive periods for language development can reach nativelike levels; those first exposed later cannot (for a comprehensive review of the literature on maturational constraints, see Hyltenstam and Abrahamsson, this volume).

Fourth, at what level does fossilization supposedly occur? What is the appropriate *unit of analysis*: the whole IL, the module, the linguistic rule, particular forms, words, meanings, collocations, form–function relationships, ranges of variation, all of these, or something else? Does fossilization halt IL development at the level of type or token? For instance, is it necessary to show that (target-like or non-target-like) plural *-s* marking remains the same on all noun phrases to support a fossilization claim, or just on particular NPs, perhaps always supplied accurately on some, but always omitted on others? Would a claim that fossilization has occurred be supported by proof of stability (within discourse domain X, context Y, and over time period Z) in a learner’s failure to use the regular past tense morpheme *-ed* in English appropriately on any verbs, that is, at the level of type, or on particular verbs, that is, at the level of token? What if, for example, a learner’s average target-like use (TLU) for regular past remained constant at around 50 percent (or varied only within a narrow range) over time, but the marking of individual verbs changed during that period? And even if conducting an analysis at the level of token (particular plural NPs, or particular verbs marked, or not, for past time), does the researcher further need to take precise linguistic contexts, collocations, and intended meanings into account when comparing multiple uses of the same tokens? Suppose, for example, that a learner invariably uses singular and plural forms of some “measure words” (*days, years, etc.*) and a few other lexical items (e.g., *ladies*) correctly, but marks plurality variably or not at all on some other NPs. Is one to conclude that plural *-s* (either the rule or the form) has fossilized

altogether, has fossilized in the case of some NPs but not others, some uses but not others, or that it has not fossilized at all?⁹ And if structures, such as English relative clauses (Schachter, 1974) or passives (Seliger, 1989), are produced with increasing accuracy over time (a matter of the system), but are persistently and consistently undersupplied, or “avoided,” by speakers of a particular L1 (a matter of norms), can it be said that while the structures are still developing, the uses have fossilized?

Fifth, is fossilization a matter of *deviance* only, or, as might reasonably be supposed, of correct, nativelike rules and forms, too?¹⁰ A cognitive mechanism that could differentiate nativelike from non-nativelike elements and apply only to the latter requires some imagination. Yet, given that many target-like, as well as non-target-like, rules and forms are acquired early, even by ultimately unsuccessful learners, and remain unchanged “permanently,” belief in such an uncannily sophisticated device is what acceptance of the construct entails. Conversely, positing that target-like forms fossilize, too, increases plausibility, but creates another problem, for what kind of cognitive mechanism could simultaneously apply and not apply to different structures, “freezing” grammatical ones while allowing ungrammatical ones to continue to develop, or as noted above, simultaneously apply and not apply to the same structure in different discourse domains?

3 The Evidence

In light of the widespread acceptance of fossilization as a force in SLA,¹¹ or at least the pervasive casual use of the term in the SLA literature,¹² the scarcity – until recently, the complete absence – of even potentially supporting evidence is surprising, to say the least. Numerous studies over the past 30 years or so have purported to demonstrate and/or explain fossilization, but each finding may be questioned, often on multiple grounds.¹³ Common problems include, but are not limited to: (i) assuming, not demonstrating, fossilization (or stabilization); (ii) selecting inappropriate learners for study; (iii) basing findings on insufficient data; and (iv) using inadequate analyses.

3.1 *Assuming, not demonstrating, fossilization*

A number of researchers start by asserting that various structures or whole ILs – and sometimes even learners, or whole groups of learners – have fossilized, a claim usually accompanied by speculations about the reasons why (see, e.g., Lin, 1995; Washburn, 1992). The most common justifications offered for such a priori classifications are that certain errors are frequent¹⁴ or that the informants have resided in a society in which the target language is widely spoken for what the researcher considers long enough for them to have learned more than they have in fact learned. Other factors occasionally invoked include

length and type of prior language-learning experience. Thus, in an interesting study of self-correction and incorporation of other-correction by eight Mandarin-speaking Taiwanese learners of Spanish, Lin (1995) compared two groups of four informants, one group supposedly fossilized, the other not. The allegedly fossilized group consisted of three men and one woman, all of low proficiency in Spanish, three of them restaurant workers, one a manager, who ranged in age from 36 to 53, and who, to qualify for the study, had to have lived in Spain for at least 10 years, have acquired Spanish naturalistically for the most part (starting Spanish after their arrival in Spain, with an average of about eight months of instruction), (for reasons not clear to me) not be married to a Spaniard, and have had continual contact with native speakers throughout their period of residence. The comparison group consisted of four women, two graduate students and two professionals, ranging in age from 26 to 33, all of whom had majored in Spanish at university in Taiwan, had received from three to four additional years of instruction in Spanish in Spain, had lived in Spain for from three to five years, and had attained considerably higher levels of proficiency in the L2 than members of the supposedly fossilized group. After studying transcripts of single conversations, lasting between 23 and 45 minutes, between each of the informants and one of four native speakers of Spanish (and in one case, two such conversations), Lin reported a clear difference between the two groups in their sensitivity to, and use, of self- and other-repair. The non-fossilized group incorporated 69 percent of other-corrections, compared with the allegedly fossilized group's 7 percent, and self-corrected seven times as often as the longer-term residents.

Lin expresses a belief in multiple causes of fossilization, and in different ones affecting different learners. Those he lists include time available for, and interest in, L2 study, basic educational level, and aptitude (Lin, 1995, p. 140), psychological and social distance, and felt communicative need (1995, p. 149). He also recognizes (1995, p. 143) that the (on numerous grounds, non-equivalent control group) design of his study precludes any assignment of causality. Lin nevertheless suggests that an important factor might also be a learner's metalinguistic ability, as evidenced by the two groups' differential rates of self-correction and sensitivity to negative feedback (Lin, 1995; Lin and Hedgcock, 1996). Lin may well be right, but his findings do not support such a conclusion for several reasons, most obviously the fact that one brief conversational sample from each informant precludes any assessment of the persistence of elements in that person's IL over time, and hence, any judgment as to whether that informant has stabilized, much less fossilized. Asserting that some informants were fossilized (or stabilized), and others not, on the basis of differing personal histories, language-learning profiles, and L2 proficiency is no substitute for longitudinal data. The findings on self-correction and negative feedback, valuable though they are, might be due to one or more of several differences between the two groups, including those mentioned by Lin, and might have nothing to do with fossilization, if such a thing exists.

3.2 *Selecting inappropriate learners for study*

A pervasive problem in fossilization studies involves selection of inappropriate informants. Given that even under optimal conditions, it takes several years to learn a second language, it is clearly essential to base findings, as Selinker has always rightly stressed, on learners who have had adequate ability, motivation, and opportunity to learn. It would be absurd to show that the ILs of classroom foreign language learners or of low-proficiency learners recently arrived in an L2 environment were still developing, and to claim, therefore, that learners do not fossilize (quite apart from the impossibility of proving a negative). But it is equally absurd to base a fossilization claim on such learners, for example, those in foreign language settings, who, motivated or not, could not have had adequate time or opportunity to acquire the target language, and when the researcher could not have studied the learner long enough to show lack of change persistent enough even to meet Selinker's lowest suggested minimum of at least two to five years. Even learners who have resided in a target-language community for periods of ten years or more may be unsuitable for study, at least before data are collected on their ability, motivation, and opportunity to learn during that time. Many such individuals spend considerable proportions of their lives in L1 linguistic ghettos. Others, while enjoying plenty of L2 exposure, have little need for, or interest in, acquiring the new language, perhaps due to use of the L1 at home and/or at work, their own social status, the relative sociolinguistic status of the L1 and L2 involved, or low intended (as opposed to actual) length of residence. Yet a number of fossilization claims have been based on studies of learners of these types (see, e.g., Agnello, 1977; Bean and Gergen, 1990; Bruzese, 1977; Mukattash, 1986; Sola, 1989; Thep-Ackrapong, 1990; Washburn, 1992).

Washburn (1992, 1994), for example, divided 18 undergraduate students enrolled in the same level of a writing course at a US university into two groups of nine, which she designated "fossilized" and "non-fossilized" on the basis of length of residence (LOR) and whether or not the student had ever failed an ESL course (thereby also making this another case of a study where fossilization was assumed, not demonstrated). Washburn writes: "Since there is no operationalized definition of fossilized speakers based on linguistic behavior, a working definition was employed" (1994, p. 72). LOR ranged from six months to four-and-a-half years for the "non-fossilized" group, and from five to seven years for the "fossilized" group. No students in the "non-fossilized" group had failed a previous ESL course; all students in the "fossilized" group had failed at least one. Students participated in three sessions for the research during the semester, each lasting about 45 minutes: an interview designed to elicit certain structures known to be problematic; a session where students completed a cloze test and a combined grammaticality judgment and imitation task (the latter based on deviant utterances from the particular informant's earlier interview); and finally, a short-term learning task in which each student received intensive corrective feedback on structures on which he

or she had continued to make errors in the grammaticality judgment and imitation task. The feedback took the form, in sequence, of correct models, repetition of the correct models with emphasis, breaking the utterance into smaller units, backwards build-up, and overt correction. The "treatments" during the second and third sessions were innovative and especially interesting because instead of focusing on arbitrarily chosen structures, they targeted items for each student which were more likely to be "learnable" in a processing sense (Pienemann, 1984), as suggested by that student's attempts at production during previous sessions.¹⁵

As Washburn predicted, the "non-fossilized" group improved in accuracy more quickly (measured in numbers of turns required) than the allegedly fossilized learners following the intensive corrective feedback. Interestingly – that is, when potential explanations of fossilization are considered – Washburn's allegedly fossilized learners, like those in several other studies (e.g., Lennon, 1991a, 1991b; Lin, 1995; Lin and Hedgcock, 1996; Mukkatash, 1986; Thep-Ackrapong, 1990), seemed less sensitive to negative feedback. Students in the two groups were not distinguishable by the errors they made, many of these being the same, although the quantity of errors was higher in the "fossilized" group. Rather, it was in two patterns of errors across tasks that (again, quantitative) differences emerged. First, students in the "fossilized" group exhibited statistically significantly less stability in their production of correct forms during the feedback sessions, providing them one moment, and then what Washburn calls "regressing" the next. Second, their TLU across the interview and elicitation tasks was consistent for only 36.5 percent of the target structures, accuracy being lower during the interview, compared with 52 percent consistency for the "non-fossilized" group (a clear, if statistically non-significant, trend). How these two indications of greater *instability* among the allegedly "fossilized" group are to be reconciled with fossilization is unclear. It could presumably be argued that "regression" was suggestive of Selinker's "backsliding," and so an indication of fossilization, and that the instability across tasks was due to their constituting different contexts or discourse domains. However, one would then want to know why the supposedly "non-fossilized" learners exhibited the same patterns, if less markedly. An alternative, simpler interpretation of Washburn's results is that students in the supposedly "fossilized" group performed less accurately and more unstably because they were of lower proficiency, as suggested by their cloze test scores, which Washburn reports (1994, p. 73) were considerably (and statistically significantly) lower, and that neither group was fossilized, as shown by the ability of both to benefit from corrective feedback. A period of from six months to seven years is insufficient for most adults to acquire a new language. However intriguing Washburn's findings, this means that the initial classification of informants as "fossilized" or "non-fossilized" chiefly on the basis of LOR (with no data on L2 exposure and use during that period), as in several other studies, was arbitrary and by assertion. In all probability, none of the students was an appropriate choice for a fossilization study.

3.3 *Basing findings on insufficient data*

A surprising number of studies have purported to investigate fossilization using cross-sectional designs. Some, for example, Lin (1995), based their findings on a single sample of learner speech or writing. Some, for instance Bean and Gergen (1990), gathered data from the same learner(s) on two or more tasks, but at one time. Some collected two or more (sometimes many more) samples, but over too short a period for fossilization to be ascertained, especially if a five-year-minimum period of observation is required; for example two samples in six weeks (Mukkatash, 1986), three samples during a semester (Washburn, 1994), 16 samples in six months (Lennon, 1991a, 1991b), multiple samples in nine months (Han and Selinker, 1997), and three samples in 18 months (Thep-Ackrapong, 1990). Such studies can provide useful insights on stabilization, and often have, but arguably not on fossilization. This is so even if they are otherwise methodologically sound, unless Selinker's suggested two-to-five-year criterion and what is already known about the normal irregular pace of SLA are disregarded, in which case a distinction between stabilization and fossilization would no longer be sustainable. As Selinker and Mascia put it:

only with longitudinal interlanguage data in the context of positive evidence to the learner where there exists the motivational criterion are we able to show instances of fossilization. Otherwise, we just do not believe fossilization can be demonstrated. That is, cross-sectional studies by definition just do not capture what is happening to individual learners, the necessary locus of fossilization. (Selinker and Mascia, 1999, p. 257)

Possibly emulating Schumann's (unsuccessful) attempt to teach Alberto negation at the end of a study of common simplification processes underlying early naturalistic SLA and pidginization (Schumann, 1978), another popular approach to buttressing fossilization claims is to show that learners persist with errors despite attempts to "correct" them through instruction and/or negative feedback of one or more kinds.¹⁶ Mukkatash (1986), for example, argued that 80 Jordanian fourth-year college English majors with an average of 11 years of prior classroom EFL instruction had fossilized when explicit grammatical explanations and error correction failed to improve their written production of various constructions, including *be*-deletion, confusion of simple present and simple past forms, and retention of pronominal copies in relative clauses (allowed in Arabic), as evidenced by two written assignments over a six-week period. Similarly, Thep-Ackrapong (1990) collected a speech sample from Lin, a Chinese speaker, tutored her for four months, collected a second sample, and then collected a third one year later. Errors with infinitival complements and related structures were frequent and persisted in all three samples, leading Thep-Ackrapong to claim Lin had fossilized. Obvious potential problems with this approach include the inadequate time allowed for

improvement, failure to ascertain whether some or all of the targeted structures are “teachable” and “learnable” for the students concerned,¹⁷ use of non-comparable data over time, analysis at the level of type, not token, a variety of well-known methodological difficulties in measuring the effects of any kind of instructional intervention (Mellow, Reeder, and Forster, 1996), and the possibility that the instruction or error-correction is inadequate.

More reasonably, some, such as Kellerman (1989) and Schouten (1996), have employed pseudo-longitudinal, panel designs, buttressed by evidence from typological studies and diachronic language change. The reasoning is that if single samples obtained at one point in time from groups of progressively more advanced learners with the same L1 show widespread persistence of the same errors (e.g., use of ‘would’ in the protasis of hypothetical conditionals, as in **If he would be taller, he would be a better player*, by most Dutch learners of English), especially when all the groups are highly proficient, then it is safe to assume that such structures are at the very least vulnerable to fossilization. This may well be true, but it is not the same (and Kellerman or Schouten do not claim it is) as showing that the structure concerned is stabilized or fossilized in any individual’s IL (where it may still be improving, even though still not target-like), and it is the individual IL that is the appropriate unit of analysis when advancing a claim of fossilization in IL development, meaning that longitudinal studies of one or more individuals are required. The well-documented tendency of speakers of various L1s to retain common errors in their L2 is good *prima facie* evidence of transfer, but not, alone, sufficient evidence of fossilization, since some speakers of those L1s do *not* persist with those particular errors. Thus, while many Spanish-speaking learners of English operate with pre-verbal negation for long periods (see, e.g., Schumann, 1978; Stauble, 1984), many Spanish speakers can be found who control a fully analyzed English negation system – and the same appears to be true for any well-attested common error, allegedly L1-influenced or not. Like their pure cross-sectional counterparts, pseudo-longitudinal studies can be useful sources of hypotheses about fossilization, but a claim to have demonstrated fossilization must be supported (among other things) by evidence of lack of change in an individual IL over time. Again, true longitudinal studies are needed.

Data are also sometimes inadequate not only because of the single time at which they were collected, and/or the short period over which they were collected, and/or the type of informants from whom they were collected, but due to the kind collected. Preference will usually be accorded spontaneous speech (supplemented by elicited spoken data and data on comprehension as needed). Speech data will be closer to the vernacular, and hence, more likely to offer a window on whatever is systematic in the IL concerned – and systematicity, as opposed to variability, is potentially a key indication of stabilization. Conversely, test scores, especially if grouped across linguistic features and/or informants, and written data of any kind (as in Mukkatash, 1986), especially if from formal genres, such as academic papers (see, e.g., Han, 1998), are more vulnerable to various well-known sources of variability, such

as transfer and monitoring. That is, they are more permeable, and so likely to be less useful (although still potentially useful in some cases if such sources of systematic variation are taken into account during the analysis). In addition, when (minimally) two sets of data are available on the same (appropriate) informants, and gathered over an adequate time period, they need to be comparable – not, say, supplied-in-obligatory-context (SOC) morpheme test scores at time 1 and SOC morpheme scores from free speech at time 2, or target-like-use (TLU) scores for past time reference in informal conversation at time 1 and in a picture-strip narration at time 2. All the usual sociolinguistic parameters of speech or writing need to be considered, given their well-documented roles as potential sources of variation (see, e.g., Tarone, 1988), and hence, in concealing stabilization or fossilization, especially in light of Selinker's caveats concerning fossilization's possible sensitivity to context and discourse domain.

In sum, rather than one factor alone determining the worth of a fossilization claim, it is a combination of appropriate informant(s) and adequate data that is required. A five-year longitudinal study, using (in various senses) adequate data, of a learner who had already lived in the target-language community for 20 years when the study began, with good motivation and opportunity to acquire, could be more valuable than a 10-year study of a learner in a foreign language setting or of a learner who had only recently arrived in the target-language setting and started SLA when the study began. Conversely, the shorter study could be more useful if it involved multiple samples of comparable free speech, while the longer study relied exclusively on test scores or translation. All other things being equal, a study involving advanced learners is more likely to be successful in identifying persistent errors, simply because errors remaining in the ILs of advanced learners are more likely to be potentially permanent problems than errors found in the ILs of less proficient learners, which will include a greater variety and number, but many that will disappear with increasing proficiency. Similarly, errors known to be common in highly proficient speakers of a particular L1 background are more likely to include persistent ones, or they would not be more common with advanced learners from a particular L1 background. A constellation of methodological factors needs to be evaluated, in other words. The problem is that very few studies indeed have managed to avoid all of the pitfalls described, and some almost none of them, meaning that in addition to its theoretical problems, fossilization suffers from a paucity of credible supporting evidence. Moreover, data to be presented below (section 3.4.) suggest that the level at which many, possibly all, analyses are conducted needs to include not just type, but token, something found in no published study of fossilization to date.

3.4 *Using inadequate analyses*

Given the present state of ignorance about fossilization – not least, uncertainty as to whether such a thing exists – it is difficult to be sure how to analyze data

appropriately. It is possible to be sure that many methods are *inappropriate*, however, and the analyses employed in virtually every study of fossilization have arguably been flawed in one or more ways. Common problems have included (i) use of group means (e.g., for a whole class of children) instead of individual scores, meaning that changes in some informants' ILs risk canceling out changes in others, giving the false appearance of overall "fossilization" at the level of the group; (ii) use of pooled data (e.g., mean SOC percentages for 10 morphemes) within an individual, meaning that changes in some morphemes risk canceling out changes in others, again giving the false appearance of overall "fossilization" at the level of morphology; (iii) use of accuracy or accuracy ranges (e.g., SOC or TLU measures) instead of stability/change measures, regardless of whether or not the rules or elements studied are target-like; and (iv) conducting analyses at the level of types, not tokens. All such analyses are also likely to miss changes in form–function relationships over time, zig-zag developmental curves, and U-shaped behavior.

By way of illustration, Bean and Gergen (1990) sought to determine whether the ILs of fossilized speakers (*sic*) varied across tasks in the L2, and whether fossilized ILs (*sic*) varied among individuals with the same L1 when they performed similar tasks in the L2. The subjects were two young women, Jean and May. Jean was an ethnic Chinese Malaysian, aged 33, with a LOR of about 10 years and a bachelor's degree in business from a US university, working at a bank in Los Angeles at the time of the study. May was from Hong Kong, aged 21, with a LOR of three years, in her last year of a business degree at a university in Los Angeles, working in a related area, and intending to stay in the US. Both women had been raised in upper-middle-class families, with Cantonese L1 spoken in the home. Both had been exposed to English from the age of 5 at school in their countries of origin, but reported having had minimal opportunity or need to speak the language in or out of school until coming to the US (see Bean and Gergen, 1990, p. 216n.3). There were several obvious problems, in other words, with the initial choice of informants for such a study.

Data consisted of speech production on three tasks – an informal interview, a picture story narration, and an oral morphology test (the Solomonick-Williams Morpheme Test, which uses pictures and sentence completion) – gathered at one point in time. There was really no possibility of determining whether the informants had stabilized, in other words, much less fossilized, regardless of what the analysis revealed.¹⁸ For each informant, Bean and Gergen calculated percentage accurate SOC across the three tasks for 11 morphemes: progressive *-ing*, regular noun plural, copula (*is, am, are, was, and were*), auxiliary *be*, modal auxiliaries, two regular auxiliaries (*have* and *do*), articles (*a, an, and the*), regular past, third person singular *-s*, and possessive *s*. Drawing an innovative but unmotivated distinction between "what appears to be fossilized and what may actually be acquired . . . and thus not eligible for analysis as fossilized morphemes" (1990, p. 211), they then assigned morphemes to one of three categories for each informant:

Table 16.1 SOC percentages for “fossilized” morphemes

<i>Jean</i>				<i>May</i>			
<i>Morpheme</i>	<i>I</i>	<i>N</i>	<i>T</i>	<i>Morpheme</i>	<i>I</i>	<i>N</i>	<i>T</i>
Copula	66	96	100	<i>-ing</i>	100	100	60
Article	83	100	65	Plural	81	67	100
Regular past	16	18	30	Auxiliary <i>be</i>	100	50	60
				Regular past	26	50	80
				3rd person sing. <i>-s</i>	20	5	70

Source: based on Bean and Gergen (1990)

(1) those that appear to be acquired (that is, 80 percent or more accurate on all three tasks), (2) those that are candidates for the designation “fossilized” (that is, those showing *wide variation in accuracy within or across tasks*), and (3) those which occur too infrequently for analysis (that is, fewer than 10 occurrences in two of the three tasks). (Bean and Gergen, 1990, p. 211, emphasis added)

Table 16.1 shows the morphemes Bean and Gergen classified as fossilized, together with the percentage accuracy scores across the three tasks: interview (I), story narration (N), and test (T).

Aside from the impossibility of ascertaining whether IL development has ceased on the basis of data from a cross-sectional study, this analysis raises a number of questions. Not least, if (i) acquired items are considered irrelevant (despite the potential support that their relative stability both within and across tasks might be thought to lend a fossilization claim) – that is, if fossilization is assumed to be able to affect some items *within* a task, discourse domain, or context, while leaving others untouched, as well as the same items across tasks, domains, or contexts – and if (ii) “wide variation in accuracy within or across tasks” is considered evidence of fossilization, what is left that could constitute counter-evidence? Bean and Gergen concluded, “This study’s findings demonstrate that fossilization entails idiosyncratic, task-based variation in interlanguage morphological production” (1990, p. 215). But if considerable variation is evidence of fossilization, what would constitute evidence of development? Moreover, how can a researcher tell that, say, Jean’s 66 percent accuracy with copula in her interview, compared with her near-perfect (96 percent accurate) and perfect (100 percent) suppliance during the story narration and on the test, respectively, or May’s perfect accuracy with *-ing* during the interview and story narration, but 60 percent score on the test, are indications of fossilization and not simply of acquisition of copula and *-ing* being incomplete, but still in progress (especially given the relatively short period over

which either woman had actually spoken English – about three years in May's case)? In any case, how could the same linguistic item simultaneously be acquired perfectly and not acquired, depending on task? And if acquired items, that is, those supplied with 80 percent accuracy or better across all three tasks, are not to be considered fossilized, how can items like May's *-ing*, supplied with 100 percent accuracy on two tasks and 60 percent on the third, be considered fossilized, not acquired?¹⁹

In addition to these problems, accuracy percentage for morphemes in spontaneous spoken or written data (even if longitudinal) is potentially a highly unreliable measure of whether development of an item has ceased, since the analysis is conducted at the level of type, such as third person singular *-s* or regular past, not token. To illustrate (using round numbers for ease of exposition), a learner on task 1 or at time 1 might produce 20 tokens of third person singular verbs, of which 10 were marked accurately with word-final *-s*, and 10 not, yielding an average accuracy score of 50 percent. Ten of 20 tokens on task 2 or at time 2 might also be accurately marked, again giving an overall accuracy score of 50 percent. Yet this result could just as well indicate development as stabilization (or fossilization), depending on which verbs were accurately and inaccurately marked in each case. Leaving aside additional problems, such as the fact that error rates can rise with increased development (see, e.g., Meisel, Clahsen, and Pienemann, 1981), meaning that similar SOC percentages over time may not indicate stability over time, or the way intended meanings of the same form sometimes change from one use to the next, or the way new functions are sometimes attempted with a given form, of 10 verbs common to both samples, seven might be unmarked for past time on the first occasion, and marked accurately on the second. Meanwhile, of a second group of 10 different verbs, each appearing in one sample only, seven might be accurately supplied in the first, whereas only three were accurately supplied in the second. That might justifiably be taken as evidence of development, and certainly of change, not stability. Similarly, outcomes and interpretations might easily be influenced by the presence of several instances (tokens) of the same verb (type), marked correctly or incorrectly, in one sample, but not another, say as a result of particular topics discussed. These and other problems (for additional examples, see Long and Sato, 1984) are threats to the validity not just of SOC analysis of morphemes, but of any analysis conducted at the level of type, not token.

In sum, while widely taken to be a proven universal feature of IL development, the empirical evidence for fossilization in the 30 or so years since the construct's first appearance in the SLA literature has been vanishingly small. Studies offered in support of fossilization claims turn out simply to have assumed, not demonstrated, fossilization; to have used inappropriate subjects, insufficient data, and inadequate analyses; or in many cases to have been marred by two or more of these flaws. Partly in response to this state of affairs, three ongoing longitudinal studies have been undertaken, finally offering to determine whether fossilization is myth or reality.

4 Three Longitudinal Studies

The need for longitudinal empirical studies of IL stabilization and putative fossilization is clear (for discussion, see Selinker and Han, 2001), and the first work of this kind is under way. A study by Han (1998, 2000) just meets the lower bound of Selinker's minimum requirement of from two to five years to substantiate a fossilization claim, and studies by Lardiere (1998a, 1998b, 2000a, 2000b) and Long (1997) comfortably exceed the upper bound.

While recognizing (1998, p. 89, and elsewhere) that not all stabilization is a precursor to, or an indication of, fossilization, Han (1998) views stabilization and fossilization as two parts of a continuum. She conceptualizes fossilization as a cognitive process, properly inferable only from long-term stabilization, demonstrable only by longitudinal studies, occurring at the level of IL subsystems rather than the entire system, and manifesting itself in three ways: invariant appearance of IL forms over time, backsliding over time, and stabilized variations over time (Han, 1998, p. 87).

Two Chinese speakers, F and G, aged 32 and 36, served as informants for Han's research. Both had studied English in the People's Republic of China, and both had lived in an English-speaking country for two years when the study began. A LOR of only two years made them questionable subjects for a fossilization study, but this was offset somewhat by their high level of prior L2 attainment, each having achieved TOEFL scores of over 600 ten years earlier, and by Han's focusing on a typical IL construction for Chinese-speaking learners of English. F was first a post-doctoral civil engineer at a British university, and subsequently an engineer in a computer software company in Australia; G was a researcher in astrophysics at a US university. Both needed to research and publish as part of their work, and both were motivated to improve their English. Data consisted primarily of drafts of academic papers, and formal and informal letters, supplemented by results from translation, grammaticality judgment and correction tasks, and a cloze test. In a detailed analysis that involved both type and token, Han focused on three related constructions: (i) pseudo-passives, such as "The letter about graphics file has not received," a common error in Chinese-English IL – due, among other reasons, she claims, to its matching the topic-comment structure of Chinese; (ii) a subset of target-like passives; and (iii) cases of "over-passivization," that is, passivized unaccusatives, such as "This problem is originated from some numerical error" (1998, p. 168).

There were three main findings. First, in their writing, the informants produced pseudo-passive sentences, such as "Fanta's software can use to model processing procedure" and "The reference keeps at the central surface" (1998, pp. 101–2), throughout the period of observation, even though the pseudo-passives occasionally featured in backsliding. Such errors, that is, novel unaccusatives, were more common in informal letters (perhaps because the writer's attention was focused proportionately more on message than form in

that genre), were rare in the research papers, and did not occur in data from the experimental tasks. In addition, the pseudo-passives involved only a small set of verbs, which usually appeared or reappeared when the context favored function-to-form transfer, that is, when the pseudo-passive was used to express what would have been a null-subject topic-comment structure in Chinese; for example, "I also received a card that my health check-up has already sent to the office" (p. 136) and "The letter about graphics file has not received" (p. 139). It is those persistent, L1-influenced, "non-developmental" novel unaccusatives that Han considers vulnerable to fossilization, as opposed to the transitional, "developmental" novel unaccusatives like those found in child language acquisition, such as "The stupid Nintendo unplugged" and "The table knocked over" (p. 126), which result mostly from incorrect lexical entries.

Second, a subset of target-like passives, such as "Your email message was received" and "My reply will be sent to you following this mail," which Han took to be a monitored form of the pseudo-passive, appeared invariably, again in informal writing. Han noted that these target-like passives were essentially driven by the same type of L1 topic-comment influence that induced the pseudo-passives; pragmatically, they differ from true English passives but are identical to the IL pseudo-passives.

Third, passivised unaccusatives, such as "The reflection 'hump' could be disappeared" and "Cough is almost disappeared" (1998, p. 149), appeared in variation with non-passivized ones, such as "We notice that the 'hump' disappear from the composite spectrum" and "My teeth pain almost disappeared" (1998, p. 149) throughout the observation, that is, they manifested stabilized variation. Han suggested that in this case, it was the dual factors of input and learnability, not L1 influence as in the case of pseudo-passives, that will be implicated in fossilization. Thus, all three characteristics of fossilization posited by Han and Selinker (1999) were observed: backsliding over time, invariant appearance of IL forms over time, and stabilized variation over time.

Lardiere (1998a, 1998b, 2000a, 2000b) reports findings from a thus far nearly 10-year study of grammatical knowledge in what appears to be the end-state in the acquisition of English by Patty, a native speaker of Chinese who arrived in the USA at the age of 22. The study began when Patty had already lived in the target-language environment for 10 years. She was immersed in English throughout the observation period – for nearly 20 years by the time of the later recordings – so had had plenty of opportunity to acquire the target language. The data consist of three relatively short conversations between Patty and the researcher,²⁰ supplemented by two grammaticality judgment tasks administered 18 months apart. Lardiere's focus is not fossilization per se, but a claim by some UG theorists that a contingent relationship exists between the acquisition of verbal morphological inflection and underlying syntactic knowledge – a claim which Lardiere rejects, arguing instead for a dissociation between morphology and syntax even in end-state grammars – and the broader question of whether the underlying abstract syntactic knowledge posited to be involved in child language acquisition is available to the adult L2 acquirer.

Patty does indeed seem to provide evidence of a dissociation between morphology and syntax, and of continued adult access to innate syntactic knowledge. This can be seen in the supposed relationship between the abstract syntactic property of finiteness and pronoun morphology. Patty's past tense marking on verbs has remained low and very stable, at close to 34 percent SOC, over the entire period of the study. Nominative case marking on pronominal subjects, conversely, has been perfect (100 percent SOC) throughout the same period, and clearly a function of finiteness, since only subjects in finite contexts receive subject case, at the same time as subjects in non-finite contexts are also always correctly marked (100 percent SOC) for object case. Patty's grammatical knowledge, that is, includes the functional category T(ense), specified for [+/-finiteness], even though her tense marking on verbs is relatively poor. Patty's grammar is described in detail in an ongoing series of papers summarized by Lardiere for this volume (see box 16.1).

Despite its somewhat different principal focus, Lardiere's research is of great interest for the light it throws on fossilization. Patty's LOR and history of plentiful L2 exposure make her an appropriate informant. The duration of the study (nearly 10 years to date) and the use of comparable samples collected over that period mean that the data constitute a legitimate basis for a potential fossilization claim. The study's motivation by a detailed linguistic theory helps guide data collection and analysis and means that the researcher has a coherent explanation for her findings. In addition to the stability in past tense marking mentioned above, Lardiere has reported that Patty's production of third person singular *-s* on thematic (lexical) verbs has remained stable throughout

Box 16.1 Lardiere (1998a, 1998b, 2000a, 2000b)

Research question: Does a morphological deficiency in production data reflect a corresponding deficit in the abstract representation of functional features and phrase structure in the syntax?

This ongoing study focuses on the nature of grammatical knowledge in the "fossilized" end-state of adult SLA. The results so far indicate a dissociation between morphology and syntax; in other words, the contingent relation often argued in the literature to hold between the acquisition of verbal morphological inflection and underlying syntactic knowledge is not supported. The long-term goal of the study is to revisit the question of access to UG in adult SLA by considering how and/or whether the scope of UG extends to the often highly complex procedures for mapping from abstract grammatical features in the syntax to language-specific morphophonological forms. (For the situating of this study in its larger theoretical context, see Lardiere 2000a.)

Methodology: The findings are based on a detailed longitudinal case study comprising naturalistic L2 production data collected in three audiotaped recordings spanning nearly nine years and, more recently, on elicited task-based data from Patty, a native

Chinese speaker who arrived in the US at the age of 22. Data collection began after Patty had already been living continuously in the US for about 10 years. From the beginning of data collection, Patty was immersed totally and virtually exclusively in the target language environment, English, spoken by native speakers. Fossilization cannot, therefore, be due to any relative paucity of input in quantity or quality, or to lack of assimilation into the target culture.

Results: Although Patty's morphological marking on verbs has apparently fossilized at a production rate well below the usual criteria typically assumed throughout the literature for "acquisition," we can nonetheless find alternative types of evidence suggesting knowledge of the functional categories and features associated with verbal inflection. Three kinds of evidence have been investigated to date: (i) pronominal case on subjects (indicating abstract knowledge of finiteness); (ii) the position of verbs with respect to negation and adverbs (indicating knowledge of feature strength and/or UG general economy principles prohibiting overt verb raising in English); and (iii) the extensive presence in the data of *wh*-questions and embedded clauses, many with overt complementizers (indicating the representation of a CP functional category and therefore, presumably, all lower functional projections as well). Taking a brief look at each of these in turn:

i In English (as in many languages), there is a relation between finiteness – an abstract feature of the grammar – and the form of subject pronouns. Within both Minimalist and pre-Minimalist approaches to generative grammar, if the functional category I(nfl) or T(ense) is specified as [+finite], the pronominal subject will require nominative or subject case; otherwise it will receive the default case marking for English, object case. Lardiere (1998a) examined the suppliance of past tense marking on verbs in Patty's data and found it to be stable and low over the entire period of data collection, at only about 34 percent suppliance in obligatory contexts. Nonetheless, the distribution of subject case marking on Patty's pronominal subjects in the same contexts was absolutely perfect, at 100 percent suppliance over the entire period of data collection. Moreover, pronominal case marking on subjects was clearly a function of finiteness: only subjects in finite contexts received subject case, whereas all subjects in non-finite contexts (such as infinitive, ECM, and small clauses) were correctly produced in the object case form. A few first person examples of the latter follow:

- (1) he make *me*, uh, spending money
- (2) that doesn't have anything to do with *me* leaving home
- (3) it's, uh, best for *me* to stay in Shanghai
- (4) she didn't tell me to . . . like let *me* know that there's nothing going on in China

These findings indicate that Patty's grammatical representation of English includes the functional category T(ense), specified for [\pm finiteness], despite the relative impoverishment of tense marking on verbs.

ii This study again departs from recent claims in the theoretical and acquisition literature that posit a contingent relation between (the acquisition of) the verbal morphological paradigm for agreement (in English, 3sg -s) and the parameterized possibility of verb raising in the target language. The relevant abstract feature in English is "weak," thereby prohibiting thematic verb raising past adverbs and

negation, and necessitating *do*-support in the case of negation. Thus, the following are ungrammatical in English:

- (5) *they drink not beer
- (6) *they drink frequently beer

Knowledge of the “weak” specification of English should result in evidence that the learner knows thematic verbs do not raise in English and will reject sentences such as those above as ungrammatical. The data from Patty, including both naturalistic production data and elicited grammaticality judgments, unequivocally show this, despite the fact that Patty produces 3sg *-s* agreement marking on thematic verbs in only about 4 percent of obligatory contexts. Lardiere (1998b) examined all possible contexts for verb raising over negation and/or adverbs in the production data and found that verb raising does not occur and does not appear to be an option. Additionally, Lardiere (2000b) reported the results of two grammaticality judgment tasks administered 18 months apart assessing the acceptability of verb raising over adverbs (the second test included 25 native speaker controls). Both yielded identical results: Patty correctly rejected all ungrammatical sentences involving verb raising past adverbs, a finding completely convergent with the production data. Again, the data suggest a total dissociation between morphological inflection and abstract featural knowledge; that is, even though Patty has *never* acquired verbal agreement affixation, she was still able to determine the status of verb raising in the target L2.

iii Finally, the presence of a CP, the highest functional projection in the clause, is claimed within nearly all models of language acquisition to implicate the presence of the lower functional categories as well. These include the categories of IP, such as Tense and Agreement which are typically associated with verbal inflectional morphology. Lardiere (1998a, 2000a) observes that, despite the very low suppliance rates for tense and agreement marking on verbs, Patty’s grammatical representation of English nonetheless clearly includes a CP projection, indicating the presence of fully extended clausal phrase structure. The data provide abundant evidence for a CP projection in the form of embedded clauses with various complementizers, relative and free relative clauses, and *wh*- and yes-no questions involving subject–aux inversion. A few examples follow:

- (7) why do you want me to go?
- (8) he have the inspiration to say what he want to say
- (9) something that have to show the unbeliever that you are in spirit
- (10) can I have onion?

In sum, Patty’s representation of English phrase structure appears to be complete, and clearly not contingent on the acquisition of verbal inflectional morphology.

Conclusion: Taken together, the results support the modularity of grammatical domains, and suggest that some domains are more susceptible to fossilization than others. In Patty’s case, the mapping from morphosyntactic features to morphophonological spell-out appears to be particularly vulnerable; in contrast, her knowledge of finiteness and feature strength and the development of extended phrase structure in English all seem quite nativelike.

the period of observation, and very low, at around just 4 percent SOC. Related abstract syntactic knowledge – parameterized knowledge of verb raising, shown by correct placement of verbs with adverbs and in negation – however, is again intact. Lardiere suggests that such findings in grammatical subsystems in Patty’s IL support the notion that adult L2 acquirers have continued access to innate knowledge of abstract syntactic features, but not to the complex procedures for mapping from those features to language-specific morpho-phonological forms, procedures which may lie outside the scope of UG. She concludes:

In sum, Patty’s representation of English phrase structure appears to be complete, and clearly not contingent on the acquisition of verbal inflectional morphology . . . Taken together, the results support the modularity of grammatical domains, and suggest that some domains are more susceptible to fossilization than others. (Lardiere, box 16.1, p. 506)

In a very recent, ongoing study conducted within the same linguistic framework, White (2002) reports on what may also turn out to be the end-state grammar of SD, a 50-year-old adult Turkish woman whose family emigrated to Montreal when she was 40. SD speaks Turkish at home, but has otherwise been exposed to a considerable amount of English over the past decade through a college course in interior design and subsequent work in English-speaking environments. She is a fluent, “advanced” speaker, as judged by her score of 93 percent on a University ELI placement test, but makes some errors, particularly with articles, of which she is well aware. Data were obtained from four interviews conducted over a two-month period, as well as a series of communication tasks targeting various morphological items, and several written tasks. Since there appeared to be few or no changes over the four interviews, the data were collapsed for the initial analysis.²¹

Production of several morphological inflections in obligatory contexts was found to be variably accurate, 60 percent or better on definite article, indefinite article, plural -s, third person singular -s, all persons aux + cop, lexical past verbs, and past aux + cop. In the realm of syntax, however, SD had moved from her subject and object pro-drop L1 to the English system virtually perfectly, and made no case errors at all: her choice of nominative and accusative pronoun forms was always appropriate, with subject pronouns nominative even when the verb was uninflected. White notes that these findings reveal SD’s unconscious knowledge of certain syntactic requirements in English; for example, that subjects must be overt, and subject pronouns must be marked nominative. Similarly, SD’s accuracy with pronouns and with definite and indefinite articles on an elicitation task suggest that the +/- definite feature is intact, and the lack of verb raising also argues against “inert” feature strength (cf. Eubank, 1995). White concludes that her results suggest missing surface inflection, and support Lardiere’s findings of access problems, rather than representational deficits (cf. Hawkins, 2000).

Long (1997) has reported preliminary findings from a thus far 16-year study of "Ayako," a Japanese woman, born in 1926, now 75, who immigrated to Hawai'i in 1948, aged 22. A "war bride," Ayako came as the wife of a local third-generation Japanese-American man who had served as an interpreter in the US army of occupation in Japan. He was a blue-collar worker (now retired) and native speaker of Hawai'i Creole English (HCE). They have been happily married ever since. Ayako had already lived in the L2 environment for 37 years when the study began in 1985. She is very popular and has a wide circle of English-speaking and some Japanese-speaking friends in Honolulu. Except for the first three or four years after her arrival, English has been the main language at home, a fact made necessary, among other things, by the need to communicate with her three children, their friends, and neighbors, and later by her two jobs, the first working in a florist's for four years, the second as a salesperson at the local PX store for 16 years before retiring in 1988. While Ayako still uses both languages for a variety of purposes, she has used English more frequently for most of the past 52 years – she estimates about 75 percent of the time with her husband, for instance, and more than that outside the home. She is, and considers herself to be, highly acculturated, and often says she much prefers life in Hawai'i to what she would have experienced, especially as a woman, had she stayed in Japan.

Data collection began in 1985, when Ayako completed a battery of six oral production tasks designed to elicit a variety of narrative and expository discourse:

- i a semi-structured interview, during which she spoke freely in response to eight intentionally broad, open-ended questions, such as "In as much detail as possible, would you please tell me about your childhood?" and "What differences do you think there are between Japan and America and between Japanese and Americans?";
- ii a picture description of a detailed street scene showing a serious traffic jam, followed by her reply to the question, "What do you think caused the traffic jam in the first place?";
- iii a 20-item repetition test, using pre-recorded stimulus sentences, each designed to reflect one of six differing degrees of processing complexity as (then) specified in work by Meisel et al. (1981, and elsewhere);
- iv a second picture description, this time of a six-frame cartoon strip story about a young boy, John, being knocked off his bicycle by a careless motorist, the narrative started by the researcher with the prompt "One day last year . . .," and recounted by Ayako without the pictures after a two-minute period of silent study and planning with the pictures present;
- v a second, 60-item repetition test, using a pre-recorded stimulus tape, designed to probe a wide range of grammatical features;
- vi a brief, open-ended discussion of Ayako's reflections on her experience completing tasks (i)–(v).

Cue cards with written Japanese translations of the eight questions and of the instructions for each task were available when needed. The whole session lasted approximately one hour.

In the interest of comparability, exactly the same procedure was repeated 10 years later, in 1995, and except for the two repetition tests, which proved too difficult for Ayako, repeated again in 1996, 1998, and 2000. The 1995 and subsequent sessions usually lasted two hours or longer, despite removal of the two repetition tasks from 1996 on, due to Ayako's volunteering more information and the researcher's improved use of follow-up prompts. One question in the informal conversation each year asks, "How important is English in your life?" In addition to eliciting speech data, this, plus follow-up questions, serves to help keep track of any changes in Ayako's language attitudes and use, and her perceptions thereof. The two highly controlled picture-description tasks are intended to provide sub-samples over time that are not only exactly comparable in terms of content and procedure, but in which almost everything Ayako intends to say is unambiguous. That is often not the case in spontaneous speech with lower-proficiency learners, sometimes making analyses difficult.²² Data from these sessions have been supplemented by a few audio-recordings at informal family gatherings over the years, by occasional notes on interesting spontaneous utterances written down verbatim when they occurred, and in 2000 by a written version of the bicycle story, completed after the usual spoken version.²³

Ayako has had both motivation and opportunity to learn English for the past 52 years, and at first sight appears to have been quite successful. She speaks and understands HCE fluently in face-to-face conversation with familiar topics and people in partly routinized informal situations, such as at mealtimes, when discussing family matters, out shopping, or during social visits with friends. However, communication on most such occasions is made easier not only by the routinization factor, but by the fact that frequent interlocutors are tuned into her English, and she to theirs. Ayako and many of her family members and friends use a variety of strategies to pre-empt and repair trouble. She can have difficulty in less familiar situations, especially when they involve more impersonal expository discourse dealing with displaced time and space. Transcripts of her speech show numerous lexical gaps, little complex syntax, and many persistent morphological errors.²⁴ For example, plural *s*-marking, which varies across tasks and time, was supplied correctly only 71 percent of the time in obligatory contexts in free conversation in 1985, and 48 percent of the time in free conversation in 1995, while reference to past time, which also varies greatly, was marked accurately in fewer than 50 percent of obligatory contexts in the bicycle story narrative in both 1985 and 1995.

What is most noticeable about Ayako's speech is that while performing far short of nativelike levels, with pervasive and persistent errors despite ample opportunity to acquire the target language, and so constituting an apparently

perfect candidate for a fossilization claim, her interlanguage exhibits extensive amounts of variation, both synchronic and diachronic. Some of this variability may turn out to be systematic, but much of it appears not to be. Therefore, unless considerable and unpredictable synchronic and diachronic IL change are acceptable indices of fossilization, which would surely be to bleed the construct of any remaining meaning,²⁵ some subsystems in Ayako's IL, at least, are not, in fact, stabilized, much less fossilized.

To illustrate, consider Ayako's plural marking during her responses to questions 1 (about her childhood) and 2 (about Japanese-American differences) in the informal interview in 1985 and 1995, shown in table 16.2. Within just a few lines, she supplies and fails to supply plural -s in identical linguistic contexts, with identical referents and identical intended meanings. This kind of variability is seen both synchronically, within and across "tasks," or what Selinker would probably call "discourse domains" (here, responses to the two questions), as in *ten month/s old*, *sister/s*, and *(girl)friend/s* in 1985, and *sister/s* and *friend/s* in 1995, and diachronically, within the same task or discourse domain (here, responses to the two questions) over time, as in *month/s*, *sister/s*, and *friend/s*. This variability is seen with these and other tokens throughout the transcripts, and with a variety of features, not just plural -s. Thus, while recounting the bicycle story in 1985, and due to a procedural error, twice in

Table 16.2 Ayako's plural -s marking across tasks and time

Question	1985	1995
1	Ten months old (85, 1, 3) Ten month_ old (85, 1, 5) Five sisters My sisters Three sister_ Seven years old A couple of years	Ten months old (× 3) Da older sisters (95, 8, 22) Older sister_ (95, 9, 28) My older sister_ (95, 9, 30) My sisters all (95, 9, 38) Six years old Four years Olden days time
2	Those things (× 2) Stories Book_ My friend_ (85, 2, 4) My girlfriends (85, 2, 5)	Her friends My friend_ (× 3) We become close friend_ Retirees Sometime_ Trip_ News

1995 (once with the strip-story pictures present, and then a day later, without the pictures), past time marking shows similar variability, again sometimes with the same verbs, within identical linguistic contexts, and with identical referents and intended meanings; for example *fix/fixed his bike, he _/was riding, he see/saw a car, he's trying to/was try to/try to fix the car*. Elsewhere in the data, Ayako shows that plurality and past time reference, and the forms used to mark each morphologically in English, are both known to her and are used correctly on some noun phrases and verbs fairly consistently, while consistently being omitted on others. Pairs like *lady/ladies, child/children, day/days, and year/years*, for instance, and *have/had, go/went, doesn't/didn't, cannot/couldn't*, are almost invariably used correctly and productively, whereas some other noun phrases, such as *horn, noun, textbook, and sentence* (generally, but not only, those involving lower-frequency items for Ayako), are rarely or never marked for plural, and some verbs, such as *look, pass, stop, try, and want*, are rarely, if ever, marked for past.²⁶

This combination of (i) relatively stable suppliance of appropriate marking on certain nouns and verbs, (ii) relatively stable omission of the same marking on others, and (iii) highly variable, unpredictable performance, that is, free variation, on still others, seems different in kind from the "free" variation often reported in the SLA literature, and is here termed *volatility*. *Free variation* is commonly defined as cases where two or more forms or variants of a form occur interchangeably with the same meaning or function in the same linguistic, discursive, and situational context, and with no evidence of difference in the degree of attention to form during their production (see Ellis, 1999, and elsewhere).²⁷ Ellis claims that free variation is the result of items having been learned, but *not yet integrated* into an IL system for lack either of internal linguistic or external communicative pressure for the integration to occur. This would cover cases under (iii), above, where Ayako's suppliance is seemingly random, were it not for the fact that cases under (i) show that she *has* learned and integrated the forms in question for some nouns and verbs, if not others. Accordingly, I have tentatively defined *volatility* as cases where:

a target-like or non-target-like form has been learned and integrated into the grammar with a target-like or non-target-like form–function relationship, yet where that form still also occurs interchangeably with one or more other forms or variants with different lexical types in the same linguistic, discursive and situational context to express the same meaning or function, with no evidence of different degrees of attention during its production. (Long, 1997, p. 4)

What makes volatility of particular methodological relevance in a study of stabilization or fossilization is that a case where a learner seems to have integrated productive (in this case, also target-like) rules into his or her underlying grammar, but only applies them with certain subsets of the classes of items to which they should apply, will only be discernible through an analysis conducted at the level of token, not type. This means that claims of stabilization

or fossilization in IL framed in terms of SOC or TLU accuracy percentages, or of other measures at the level of type, may obscure considerable synchronic or diachronic change, and so be unfounded. Despite the volatility described above, for example, Ayako's SOC percentages for past time reference in the two comparable 1995 and 1996 renditions of the bicycle story (without the pictures present) were 48 percent and 45 percent, respectively, giving the illusion of stabilization.²⁸

In sum, two of the three longitudinal studies briefly summarized here, those by Han and Lardiere, appear to provide the strongest evidence to date for fossilization as *product*, and each comes accompanied by (very different) interesting putative explanations for the findings. In addition, Patty's (and SD's) data suggest that if fossilization occurs, it operates locally, not globally throughout an IL. Fossilization would not simply be the same thing as general non-nativelike L2 attainment by adult starters, in other words. In this context, it is important to note, however, that to substantiate such a claim and show that fossilization affects specific modules or structures, or consistent access to them, it is necessary to provide evidence that the items concerned have ceased to develop while other IL subsystems continue to make progress. Failing that, it is unjustified to argue for fossilization of particular items as distinct from maturational constraints on the whole system – a separate issue. No study to date has sought, much less provided, such evidence.

It is too early to say, on the other hand, whether parts of Ayako's IL have fossilized. The evidence so far suggests that they have not, and that the two small grammatical domains reported on above, at least, may not even have stabilized, in spite of the fact that Ayako's speech is far from nativelike after plenty of motivation and opportunity to have advanced further. Should this be the eventual finding, it will not show that fossilization does not exist (it is impossible to prove a negative, of course), but it may serve as a note of caution for those purporting to show fossilization in learners with less optimal profiles than Ayako, using cross-sectional designs, less complete data, and more superficial analyses.

5 Explanations for Stabilization and/or Fossilization

Whether or not fossilization is a psychological reality is a question of how the construct is defined, and whether or not cases of fossilization have been documented depends not only on the definition, but very much on one's evaluation of the methodology employed in the search. Already assuming fossilization to be a proven reality, however, researchers have advanced a variety of explanations, some as well founded as many of the fossilization claims themselves, some more serious. Surprisingly, no one seems to have considered the possibility that if fossilization is, as Selinker (1972) claimed,

a cognitive mechanism producing the non-target-like end-state also called “fossilization,” there is no need for other explanations, or conversely, that if L1 transfer, learnability, markedness, etc., or some combination of linguistic and psycholinguistic factors is responsible, there is no need for “fossilization” as an explanation.

There seem to be two problems. First, for many, “fossilization” has simply become a general, non-technical name for non-target-like ultimate attainment, that is, a performance descriptor, a broad-brush method of characterizing what a learner did not do, not a competence issue, a matter of what he or she could not do, which is what made the original claim interesting. In Selinker’s original formulation, fossilization (as product) was supposedly a technical term for a special state of permanent non-target-like ultimate attainment that was due to a change in an individual’s underlying capacity for SLA, also known as “fossilization” – a process which, it has since been suggested, appears to affect particular structures, modules, domains, etc., rather than whole ILs. The second problem is that, even in Selinker’s original definition, reference was made to the process manifesting itself in “linguistic phenomena . . . which speakers of a particular NL will tend to keep in their IL relative to a particular TL” (Selinker, 1972, p. 215). In other words, Selinker himself was positing the existence of a cognitive mechanism, fossilization, responsible for fossilization as product, but simultaneously suggesting that the mechanism was in turn controlled or constrained by other factors, such as L1 transfer. On that view, fossilization (as process) is not itself an explanation, but really a cover term for one or more causal variables in SLA, such as transfer, that is, a process itself in need of explanation. Selinker has continued to elaborate on his belief in a central role for transfer ever since (see, e.g., Selinker, 1992; Selinker and Lakshmanan, 1992), and others have joined him in the search for an explanation for the explanation.

Factors proposed as causes of fossilization²⁹ include (but are not limited to) the following: lack of negative feedback on error, both external, and internal in the form of self-monitoring, and/or provision of positive feedback on successful communication despite error (Higgs and Clifford, 1982; Vigil and Oller, 1976; Yorio, 1994), especially when the latter co-occurs with unavailability of negative evidence in natural L2 input (White, 1987); insensitivity to negative feedback (Lin, 1995); age-related loss of sensitivity to language data, caused by learners reaching the steady state in the L1 (Schnitzer, 1993); maturational constraints (Seliger, 1978); lack of access to various components of UG, either computational resources, with mapping problems between the lexicon and syntax (Lardiere, 1998b; White, 2002; and others), representational resources (Beck, 1998; Eubank, 1995; and others), or representational resources not instantiated in the L1 (Hawkins, 2000; and others); loss of access to UG altogether (Bley-Vroman, 1989; Clahsen, 1988; Meisel, 1991, 1997); L1 transfer (Selinker, 1972; and others); idiosyncratic transfer of L1 elements which particular learners (as opposed to all learners from that L1 background) perceive as equivalent to elements in the L2, so as to avoid duplicating them in

the new language (Nakuma, 1998); a combination of L1 transfer and one or more other factors, such as perceived typological markedness or a desire for symmetry, converging on the same error (Kellerman, 1989; and others), as expressed in the weak form of the Multiple Effects Principle (MEP), in which L1 transfer is a privileged factor (Selinker and Lakshmanan, 1992); the strong form of the MEP, in which L1 transfer is a necessary factor, in combination with one or more other factors (Selinker and Lakshmanan, 1992); processing constraints (as distinct from lack of grammatical knowledge) producing fossilized random variation, especially of semantically light morphology (Schachter, 1996); failure to acculturate (Schumann, 1978); a variety of social-psychological variables (Preston, 1989);³⁰ premature communicative pressure (Higgs and Clifford, 1982); automatization of incorrect forms and rules, with resulting errors more likely to appear in casual than careful style due to less attention to form being exercised there (Hulstijn, 1989); satisfaction of communicative needs (Corder, 1967; and others); the ease of using what learners may know is a simplified system, but one that handles their basic communicative needs (Klein, 1986); communication breakdown, leading to avoidance of contact with native speakers, and hence to early fossilization (Perdue, 1993); inability to notice input-output discrepancies, that is, the Matching Problem Hypothesis (Klein, 1986); unwillingness to risk restructuring (Klein and Perdue, 1993); and ungrammatical input from native speakers (Gass and Lakshmanan, 1991) or non-native speakers (Harley and Swain, 1978). What almost all explanations on offer have in common is that they do not work – at least, not for fossilization, even when they may for stabilization, and not for some learners or for some supposedly fossilized features of L2 performance.

As in any area of SLA theory construction, one way of evaluating proposals to account for stabilization and/or fossilization is to subject them to empirical test. Short of other problems, any that can survive such testing are candidate explanations. Any that cannot are probably not. What is sought of an explanation is predictive power,³¹ not an ability after the fact to describe cases where the proposed causal factors (supposedly) did work, while ignoring those where they did not. That would be to return to the pseudo-explanations of the Error Analysis period.³² Does the explanation potentially apply to all learners and all supposedly fossilized structures? It loses credibility if it can be shown that it only applies to some learners, and/or only to some structures alleged to have fossilized, and not to others, or predicts stabilization or fossilization which does not occur.³³ Is there counter-evidence, in other words?

To illustrate, a claim that stabilization is caused by transfer operating in tandem with one or more additional factors, such as typological markedness, perceptual saliency, or general cognitive constraints underlying developmental sequences, has plenty of empirical support (Andersen, 1983; Harley and Swain, 1984; Jain, 1974; Wode, 1981; Zobl, 1982). Zobl (1982), for example, reviews evidence from a number of studies showing, among other things, that learners tend to persist longer with an interlingual structure, such as *No V*

negation in ESL, when it is the same as, or similar to, one with the same function in their L1. Thus, Spanish speakers stay with pre-verbal negation longer than Japanese speakers, whose L1 has post-verbal negation. A claim that fossilization is caused by transfer operating in tandem with one or more additional variables is equally obviously unfounded, however. While many Spanish speakers, and some Japanese speakers, as shown by Stauble (1984), never progress beyond the *No V* stage, many do. The claim cannot survive the universality test, in other words – it simply does not work for all learners; indeed, it fails for a large proportion of them.

One need look no further than the same findings to show that the MEP, too, cannot be correct, in either its strong or weak form. The four-stage development of negation in ESL is probably the single best-documented developmental sequence in SLA to date, and has been shown to occur in the ILs of speakers from every L1 background yet studied (for review, see, e.g., Schumann, 1979), including those like Japanese, Swedish, and Turkish (Hyltenstam, 1977), whose post-verbal L1 systems mean that the two initial stages (*No V* and *Don't V*) cannot be the result of L1 influence (probably not even in the cases of speakers of L1s which do have pre-verbal negation), and so must be due to other factors. Here, then, is a clear case where L1 and one or more other factors combine, but where the fact that many learners progress beyond *No V* negation shows that the L1 + X combination cannot predict fossilization. The MEP, too, fails the universality test. It is potentially a more accurate predictor of learning *difficulty* than transfer alone, but not of fossilization.

Quite apart from the poor empirical track record of transfer and several other factors in the above list of putative explanations for fossilization, very few of the many suggested even have the *potential* to predict fossilization, due to the simple, but crucial, fact that they concern either universal human characteristics or pervasive qualities of the linguistic environment, whereas fossilization, according to Selinker, is supposedly a process constrained by L1 properties, but manifesting itself idiosyncratically at the level of the individual. Factors which are immutable and the same for everyone could only work as explanations for the entire population of L2 learners and for all structures if they worked at all. They involve: (i) unchanging facts about L1–L2 relationships (e.g., the MEP); (ii) cognitive abilities and processes which are presumably universal, or at least vary only in degree, not kind (e.g., processing constraints, automatization of incorrect forms or rules, ease of using simpler IL systems); (iii) changes in language-learning ability (e.g., loss of sensitivity to language data, complete or partial loss of access to UG, and other effects of putative maturational constraints), which are supposedly part of the human biological inheritance, and so universal; or (iv) pervasive characteristics of language use (e.g., the absence of negative feedback and/or presence of positive feedback on error in non-instructional talk, the ungrammaticality of natural speech, communication breakdown, and unwillingness to risk restructuring), which, again, are presumably roughly the same for everyone (if extremely hard, or even impossible, to measure in some cases). A few

supposed universals or constants are serious candidates (some far more plausible, and with much stronger empirical credentials) for explaining putatively universal non-nativelike ultimate attainment in general, but arguably should not be considered as explanations for fossilization at the level of the individual. They can explain neither differences among individuals – why one IL stabilizes or fossilizes, but not another, given learners with basically the same genetic endowment, the same cognitive abilities, similar input, and so on – nor differences within individuals – why some structures but not others are affected.

Conversely, because they themselves can and do vary from one individual to another, a second set of factors in the above list might appear to have the *potential* to account for fossilization (but less so universal non-nativelike ultimate attainment, if that is indeed the end-state for all adult starters, as many researchers maintain). They include satisfaction of communicative needs, social-psychological variables, (in)sensitivity to feedback (including internal self-monitoring), and (in)ability to notice mismatches between input and output. In practice, however, the first two fail empirically. First, it is well known that language development continues to progress in many individuals long after they are capable of satisfying their communicative needs (just as it does in children doing L1A). Second, despite unsupported assertions to the contrary (see Schumann, 1993), various arrays of social and psychological factors have repeatedly failed to account for age-related success and failure in SLA at the level of individuals (see, e.g., Schmidt, 1983) and groups (see, e.g., Schumann, 1986), and have no obvious potential, either, for explaining differential success within the same individual at the level of linguistic domain or grammatical structure.

This leaves only (in)sensitivity to feedback (including internal self-monitoring), and (in)ability to notice mismatches between input and output, which are clearly very similar proposals. If it were only deviant structures that stabilized or (supposedly) fossilized, one might propose sensitivity to (negative) feedback as an explanation. For reasons discussed earlier, however, correct rules and structures stabilize, and must be subject to the same mental processes as incorrect ones. It is highly unlikely, moreover, that the same individual would be differentially (in)sensitive to positive and negative input, as opposed to input in general. Accordingly, while several factors predict stabilization, including L1–L2 and typological markedness relationships, and various combinations of social-psychological factors, just one factor, *sensitivity to input*, is the most likely explanation for fossilization (as product, in the sense of a frozen end-state grammar), if fossilization itself turns out to be a reality. It would, of course, also be a predictor of stabilization, which certainly is a reality.

Common input characteristics, such as occasional ungrammaticality, typically vary very little from one setting to another, and with the exception of comprehensibility, such variance as has been studied appears to affect first or second language acquisition very little, either.³⁴ Individual learners' sensitivity

to input, conversely, can vary a great deal, and beyond the importance for acquisition in general of “noticing,” in the sense of registering the existence of items in the input (see Schmidt, 1995), there are several hints in the literature as to the possible importance of individual differences in this ability. One well-known example is the discussion of possible reasons for Wes’s poor rate of development (Schmidt, 1983). Another, perhaps less obvious, case may be the solid empirical track record of language aptitude as a predictor of success in SLA (for review, see Skehan and Dörnyei, this volume). Sensitivity to input is arguably a key component of aptitude, tapped, for instance, in both the spelling clues and words in sentences subtests of the Modern Language Aptitude Test (Carroll and Sapon, 1959). In fact, three of the four components which Carroll proposed made up language aptitude could be viewed as involving input sensitivity: phonetic coding ability, grammatical sensitivity, and inductive language learning ability. Grammatical sensitivity, for instance, supposedly concerns the ability to recognize the grammatical functions of words or other linguistic elements in sentence structures. As noted earlier (with due caveats about methodological aspects of the studies concerned), there are reports (e.g., Lin, 1995; Lin and Hedgcock, 1996) within the fossilization literature itself that learners whom the researchers considered fossilized showed relatively low sensitivity to feedback. Indeed, as also reported earlier, several researchers have (unjustifiably) treated the apparent failure of learners to destabilize following corrective feedback of various kinds as a key indicator that they had fossilized. A thorough test of the current hypothesis would require a validated measure of sensitivity to input, with scores predicted to be lower for learners whose ILs revealed longer periods of stabilization.

An obvious problem for input sensitivity as an explanation for stabilization or fossilization is the question of why, if it is a general characteristic of an individual’s language learning ability (or aptitude), only some structures are affected, and not others.³⁵ Hence, an adequate account of stabilization or fossilization will also need to recognize the importance of various characteristics of target structures in the input, especially perceptual saliency, which is in turn often related to frequency, communicative value, semantic weight, and so forth. In sum, the interaction of *input sensitivity* (a constant within the individual, but varying across individuals) with *perceptual saliency* (which varies across structures) has the potential to account for stabilization or fossilization of some structures, but not others, observed in some individuals, but not others.³⁶

If this proposal is correct, it should predict accurately which classes of linguistic elements are more likely to stabilize (or fossilize) than others. In a valuable contribution on this issue (see also Kellerman, 1989), Todeva (1992, pp. 232–9) suggested that three high-risk categories (HRCs) of linguistic features are especially prone to fossilization: (i) categories lacking a straightforward form–function relationship, such as articles; (ii) semi-productive rules, whose exceptions (unlike, say, irregular English past tense forms) do not constitute clearly defined sets, such as English negative prefixation, dative

alternation, and stress shift in verb-to-adjective formations (e.g., *analyze/analyzable*, *present/presentable*, but *admire/admirable*); and (iii) units of a highly arbitrary nature, such as prepositions, collocations, and gender assignment. A similar search for classes of linguistic features potentially vulnerable to maturational constraints on language acquisition (Long, 1993b) involved a survey of findings from studies of a range of situations in which language is developed, lost, or impaired late in life or under other abnormal circumstances, including pidginization, aphasia, and first and second language acquisition by older children and adults. Morphology was found to be more vulnerable than syntax, inflections more at risk than free morphemes, and exceptional cases within a language-specific paradigm especially problematic. Counterexamples to every generalization were not hard to find, however. Given that language development, not language, is the object of study, a better strategy, it was suggested, should be to combine linguistic classifications with psycholinguistically relevant qualities, such as frequency, regularity, semantic transparency, communicative redundancy, and perceptual saliency. In other words, a processing dimension is needed, one which combines cognitive factors with input characteristics. It is not the case that all inflectional morphology is vulnerable to maturational constraints – or, in the present context, likely to stabilize, or if such a thing exists, fossilize – but perhaps non-salient, irregular inflections, for example, or ambiguous, optional pragmatic rules, are the items that even good learners are most likely to miss and which are especially problematic for learners with low input sensitivity. This is obviously an area where some painstaking research is needed. Meanwhile, it is possible to assess the findings to date on fossilization as process and product, along with implications for the role of the construct in SLA theory.

6 The Status of Fossilization in SLA Theory

Fossilization has been beset with definitional and methodological ambiguities from the outset, not least as to whether it is (i) a term used to *describe* the permanent end-state of IL development (in some subsystems and/or discourse domains, for some learners), (ii) a term used to *explain* permanent cessation of learning short of the target, despite ample opportunity, motivation, and ability to acquire the target language, or (iii) *both*. In other words, fossilization has sometimes been explanandum, the phenomenon to be explained, sometimes explanans, the putative explanation, and sometimes explanandum *and* explanans.

Even its use simply as a descriptor of the product of learning has become vaguer over time, with an increasing tendency in the SLA literature (and outside it) to equate fossilization and general non-nativelike attainment (not necessarily with permanent connotations).³⁷ This is a mistake. The original, narrower use of the term involves a potentially interesting claim that the current level of development is the permanent end-state because a learner *cannot*

progress any further (in one or more IL subsystems), a claim about a *loss of capacity to acquire*. The latter is simply an observation about the level of one or more individuals' L2 proficiency – a statement about what they have not (perhaps, simply, have not yet) accomplished.

A similar dilution of the construct has begun to occur at the process level, as well, with fossilization sometimes being offered as an explanation for general age-related differences in the capacity for language learning. The link became clear in Selinker's work in 1996:

fossilization is the process whereby the learner creates a cessation of interlanguage learning [*sic*], thus stopping the interlanguage from developing, it is hypothesized, in a permanent way . . . The argument is that no adult can hope to ever speak a second language in such a way that s/he is indistinguishable from native speakers of that language. (Selinker, 1996, cited in Han, 2000b, p. 5)

Again, conflating fossilization in SLA and general maturational constraints on (all) language acquisition is a mistake. Few dispute that ultimate attainment in child L1A and adult L2A – native in the former, non-native (nearly always markedly so) in the latter – is one of the most salient differences between the two processes, although disagreement persists as to the principal underlying cause(s), commonly, but not universally, held to be age of onset and/or a constellation of linguistic factors dealt with under the general rubric of “L1 transfer.” If fossilization is to have value as a construct in SLA theory, it must refer to something other than this general age-related decline in the capacity to acquire any language, first or additional, that is, to a loss of ability to acquire a *second* (including foreign) language. Put another way, the important questions both for SLA theory and for a variety of practical matters are (i) whether typically poor adult L2 attainment is due to circumstantial environmental and personal factors (inadequate opportunity to acquire, lack of motivation, etc.), that is, simple *failure to acquire*, or to a qualitative or quantitative *loss of ability to acquire* even when conditions are optimal; and (ii) whether the factor(s) underlying failure are peculiar to L2A, as opposed to language acquisition, in general.

Two broad bodies of research findings speak to the first issue: that on putative universal maturational constraints on the human capacity to learn languages, including work on so-called “sensitive periods” (for review, see Hyltenstam and Abrahamsson, this volume), and that, reviewed above, on fossilization, an allegedly localized loss of capacity supposedly affecting individual second language acquirers differently (although all learners eventually, on some accounts), not necessarily age-related and not necessarily system-wide in its effects. General maturational constraints, on the one hand, and fossilization (as cognitive mechanism), on the other, are supposedly very different in pervasiveness, scope, timing, and more, in other words, and the two should not be equated. As Hyltenstam (1988, p. 69) points out, young second language learners provide the test case on the second issue. If fossilization

only occurs (if it occurs at all) in adult starters, it should be seen as an age-related learning phenomenon. If it is found in the ILs of child starters, as well, it will be seen as constituting a pure second language, as opposed to first language, phenomenon.³⁸

As product, unless fossilization and mere non-nativelike proficiency are clearly distinguished, and as process, unless fossilization remains a claim about what is possible in adult language learning separate from a general belief in maturational constraints, the construct can be expected to disappear from SLA theory (if not from colloquial pedagogic parlance) for being redundant in each case. With the more restricted and thus potentially theoretically interesting meanings, therefore, how does fossilization fare as description and/or explanation?

To assess its *descriptive* value, the relevant question is whether there is evidence of the phenomenon to be thus described, that is, evidence of permanently immobilized IL grammars, or parts thereof, which cannot undergo further development. In light of the research findings to date, there would appear to be little compelling evidence that IL grammars fossilize. Only two studies, those by Lardiere and Han, have obtained results potentially interpretable as evidence of fossilization, and their findings on this issue (as opposed to others they speak to) may be questioned methodologically: in Han's research, on the basis of the kind of (primarily planned, written) data employed, and the study's limited duration, and in both cases due to the lack of evidence that one or more other aspects of the informants' ILs were still developing, and the level (type, not token) at which analyses were conducted. All other studies to date have suffered from one or more serious problems invalidating their findings as far as the fossilization issue is concerned (although, as noted earlier, many remain interesting and very valuable for other reasons): to reiterate, assuming, not demonstrating, fossilization, selecting inappropriate learners for study, basing findings on insufficient data, and using inadequate analyses. In other words, while fossilization may yet turn out to exist, as the studies by Lardiere (1988a, and elsewhere) and White (2002), in particular, suggest, there is little evidence that it does thus far, and hence, currently little or nothing to explain.

It would certainly be premature to dismiss fossilization as an empty construct at this stage, however, just as it is quite unwarranted to assume its reality. Fossilization may very well occur in some ILs. The fact is that the very problems that have beset almost all the empirical work to date mean that the notion remains largely unexplored. Several methodological improvements needed in future research were outlined above. To recapitulate, to have any potential for substantiating a claim of fossilization (as product), the subject(s) chosen for study need to have had the ability, motivation, and opportunity to acquire the L2 for many years (perhaps 10 or more) before the study begins. Then, accompanied by evidence of continuing ability, motivation, and opportunity, repeated comparable observations are required over time (perhaps five years or more), ideally involving ample samples of the spoken vernacular,

supplemented where appropriate by elicited data of various kinds. Analyses should be carried out at the level of token, as well as type, with a rational account provided of the analyst's treatment of the inevitable synchronic and diachronic variation.³⁹ If a resulting fossilization claim is specified to apply to certain IL subsystems, data should be provided to show that one or more other subsystems continue to develop. If a claim is specified to apply within a certain discourse domain, context, task, or other unit, the unit(s) concerned need to be defined operationally before the analysis begins, and data need to be provided to show that the linguistic elements covered by the fossilization claim continue to progress in one or more other discourse domains, etc. These are stringent requirements, but requirements needing to be met if a case of fossilization is to be distinguished from the more general one of an IL grammar affected by general maturational constraints, or even from an IL grammar that is still developing uninhibited by either.

Whatever the current or future verdict on its validity as a description of the end-product of at least some cases of SLA, as an *explanation* fossilization clearly fails. Left to stand on its own, it is a "black box," no more revealing than saying that learners cannot progress any further because of "Force X." Alternatively, if itself to be explained by other factors, as seems to be the consensus even among true believers, it is redundant: if the MEP, input sensitivity, or whatever, is the reason for linguistic rigor mortis having set in, then that is the reason, not Force X. In fact, however, in the absence to date, at least, of convincing evidence of fossilization as product, the more relevant object of study for researchers becomes *stabilization*, not fossilization, and explanations for that. From a theory-construction perspective, too, such a shift in focus has several advantages: (i) the existence of stabilization is not in doubt; (ii) it avoids the methodologically problematic "permanence" issue; (iii) it makes an additional subset of claims empirically testable; and (iv) unless and until solid evidence appears of the psychological reality of fossilization, it lightens the burden of SLA theory and theories by one variably operationalized and as yet empirically unsubstantiated construct.

ACKNOWLEDGMENT

My initial realization of the urgent need to re-examine the whole notion of fossilization was triggered by a superb presentation on the subject by Malcolm Johnston at the Australian Association for Applied Linguistics conference in Adelaide in 1986. For their contributions to the sections of the present chapter describing their studies, I thank ZhaoHong Han and Donna Lardiere. For constructive feedback on the whole manuscript, I am grateful to Georgette Ioup, Malcolm Johnston, and Larry Selinker. None is responsible for the errors of fact or interpretation which no doubt remain.

NOTES

- 1 Selinker regards fossilization as having been foundational for SLA: "It could be argued that the field of second language acquisition was spurred into existence by the phenomenon usually labeled 'fossilization.' That is, the idea that no matter what the learner does, the learner will always 'be stuck' in the second language at some distance from the expected target. The phenomenon of 'being stuck' in the L2 seems to occur to most if not all learners even at the most advanced stages. This phenomenon seemed to force early SLA researchers, who believed they were working in a contrastive analysis framework (e.g., Briere, 1966; Nemser, 1971; and Selinker, 1966) into positing intermediate linguistic systems that in some serious sense did not seem to change. These systems were thought to be 'intermediate' between and, importantly, different from, the native language and from the target language, an 'approximative system' in Nemser's terms. What is interesting is that until the late 1960s none of these researchers knew about the others' work and each discovered the phenomenon independently." (Selinker, p.c., September 27, 2000)
- 2 "Interlanguage learning" is an unfortunate term. Each interlanguage is an idiosyncratic variety of the target language created by a particular learner. Each interlanguage is unique, the incomplete product of the L2 learning process, not an existing variety available to be learned (except, probably, in the rare case where another non-native speaker might be a learner's sole model).
- 3 More recently, Selinker appears to favor the higher figure:

we often get asked how much time is enough to show fossilization? At the end of the day, we believe the number will be arbitrary. What we mean by a "substantial period" of time in any case must veer towards the years side of the continuum, perhaps a minimum of five years. (Selinker and Mascia, 1999, p. 258)

A five-year-minimum requirement would rule out Han's study, it should be noted, leaving just two potential cases, Patty and Ayako (discussed below), in the entire literature.
- 4 It has long been noted that foreign language learners reach a certain stage of learning – a stage short of success – and that learners then permanently stabilize at this stage. Development ceases, and even serious conscious efforts to change are often fruitless. Brief changes are sometimes observed, but they do not "take." The learner backslides to the stable state. (Bley-Vroman, 1989, pp. 46–7)
- 5 A variety of unfortunate terms and unwarranted prescriptions surrounding fossilization have appeared in the SLA and (especially) pedagogical literatures over the years (see, e.g., Calve, 1992; Johnson, 1993; Valette, 1991). If fossilization is, by definition, permanent, "permanent

- fossilization" is tautologous, "temporary fossilization" an oxymoron, and "defossilization" impossible. Yet Selinker himself urges research to answer the question: "Is it possible for a person to 'de-fossilize' at some point and, if so, under what conditions, internal/external to the learner?" (Selinker, 1993, p. 18). Despite assertions to the contrary (see, e.g., Graham, 1981; Johnson, 1993; Linn and Sucher, 1995; Wales, 1993), the answer *must* be negative, or fossilization is an empty construct. By contrast, "de-stabilization of previously considered fossilized forms" (Selinker and Mascia, 1999, p. 258) is conceptually coherent.
- 6 In this light, some clarification of statements like the following would be useful: "in terms of the logic of fossilization, if we can demonstrate at any one time that highly [?] stabilized forms are cognitively present, then the case is closed and the forms are permanently stabilized and we can call them 'fossilized'" (Selinker and Mascia, 1999, p. 258).
 - 7 Differential performance across discourse domains also raises the specter of variable rules, an unfortunate import from group-level variationist sociolinguistics to theorizing at the level of the individual in SLA – an example of the ecological fallacy. Todeva (1992, p. 220) suggests that differential performance across discourse domains is a question of control, not knowledge.
 - 8 Early reports of fossilized errors in the speech of sequential child bilinguals (Naiman, 1974; Selinker, Swain, and Dumas, 1975) were based on insufficient data and inadequate analyses (see sections 3.3. and 3.4). Plann (1976, 1977) discussed apparent three-year plateaus in the Spanish of anglophone children in the Culver City immersion program in terms of fossilization and the emergence of a classroom L2 dialect, but her claims were supported by a lack of evidence of significant progress across grade levels in the average morpheme accuracy scores in pooled data on different groups of children, as opposed to data on the same individuals over (sufficient) time. Also, Plann suggested that whatever lack of progress she had documented might have been due to the relatively low status of Spanish in California and to the fact that the immersion children lacked native-speaking Spanish peers with whom to bond. If that is true, the children were not a test case for fossilization, given Selinker's stipulation of lack of progress in the face of adequate ability, *motivation*, and *opportunity* to learn. Similarly, and again using cross-sectional data on groups of school-age children, Harley and Swain (1984) reported plateaus as long as four years, from grades 1 to 4, in the L2 development of French by anglophone youngsters in two immersion programs in Canada, but noted that this was followed by renewed, often substantial, progress by grade 10. Harley and Swain recognized that there was "to date no evidence of fossilization in any particular L2 domain at any particular level" (1984, pp. 301–2). A third study involving young learners, this time six children in a Canadian French immersion program (Pellerin and Hammerly, 1986), concluded that various errors had indeed fossilized, this despite the authors' data showing considerable improvement from time 1 to time 2 in three of five

- grammatical domains examined. The study was flawed in several other respects, as well, including its inadequate duration for a fossilization claim, and the use of mean accuracy scores at the level of type (prepositions, verb forms, gender, pronouns, and reflexive pronouns), not token. All five categories, most obviously “verb forms,” potentially concealed development among a miscellany of forms and structures falling under those headings (for useful discussion, see VanPatten, 1988, pp. 248–9, 256n.3). In fact, evidence of the need for caution in such grade-level comparisons of French immersion data had already been provided by Harley (1979) with respect to the development of gender marking from grades 2 to 5.
- 9 These and other cases will be returned to below in the discussion of data on Ayako.
 - 10 Todeva (1992, p. 221) argues that it is easy to demonstrate fossilization of correct structures, as well, by showing that very advanced speakers consistently overuse correct structures in contexts where native speakers of the target language use different ones, that is, by identifying persistent deviations from the norms (for use), as opposed to deviations from the system. One well-attested example she cites is Bulgarian and Russian speakers’ use of correctly formed relative clauses in place of attributive infinitives at ratios of 17:1 and 23:1, respectively.
 - 11 In an encyclopedia entry on interlanguage, for instance, the reader is informed, “A central characteristic of any interlanguage is that it fossilizes – that is, it ceases to develop at some point short of full identity with the target language” (Tarone, 1994, p. 1715). Tarone provides no evidence for her assertion, but if she is right, and fossilization not only exists, but is inevitable in all cases of SLA, equally unsubstantiated pedagogic recipes for preventing it (see, e.g., Valette, 1991) must be doomed to failure.
 - 12 Casual use of the term is something I have been guilty of myself; for instance “Japanese acquirers (with post-verbal L1 negation) also pass through a *No V* stage in English . . . some Japanese–English ILs appearing to fossilize at that stage” (Larsen-Freeman and Long, 1991, p. 260).
 - 13 Few of the more than 40 investigations of fossilization to date of which I am aware will be cited here – and then only the better ones – as there is no value, or advantage to the field, in dwelling upon flawed studies by named researchers. Rather, the aim should be to identify what is and is not known about fossilization, what sort of data and analysis permit what kind of claim, and, in general, to improve future research in the area. It should also be noted that some of the studies cited critically in what follows are useful in other ways, and that the present focus is exclusively on what they show, or do not show, about fossilization, and how the researchers went about it. It should also be pointed out that Selinker is not responsible for methodologically inadequate work on fossilization conducted by third parties.
 - 14 To illustrate:

The criterion used to determine whether some specific error types could be considered fossilized was their frequency across subjects and speech modes. If that error was made frequently by all

the subjects of the study in both free-elicited speech and writing, then the error in question may be attributed to the fact that the rules controlling its production have fossilized. (Sola, 1989, p. 63)

- An obvious problem with this approach is that many successfully mastered aspects of a L2 (or L1) were once prone to frequent errors.
- 15 A related, but vaguer, construct utilized by sociocultural theorists is Vygotsky's "Zone of Proximal Development." Malcolm Johnston (p.c.) considers the uncertain validity of most measures utilized in fossilization studies to be one of their greatest, largely unrecognized, weaknesses.
- 16 This reasoning also appears to underlie the "pedagogic corollary" to the Multiple Effects Principle (see below) advanced by Selinker and Lakshmanan (1992): apparently fossilized structures will not become open to destabilization through consciousness-raising strategies when multiple effects apply.
- 17 For example, errors with some relative clause constructions and with nominal and pronominal copies have been found to persist in the Italian-English ILs of both instructed and naturalistic acquirers, even though Italian licenses neither type of copy (Pavesi, 1986).
- 18 Bean and Gergen write: "While the present study relies on a cross-sectional analysis of fossilized interlanguage, the benefits of longitudinal data are not to be overlooked [*sic*]. Ideally, a more comprehensive study of fossilization would involve a longitudinal, comparative analysis of many speakers of the same L1 who have fossilized in the same L2" (1990, p. 215n.1). Later, however, they defend a different aspect of their methodology, the use of only two informants in cross-sectional research, with the comment, "However, the design of the study has been most efficacious" (1990, p. 209).
- 19 Bean and Gergen write: "Future studies of fossilization will need to contend with the issue of what counts as 'fossilized.' For the sake of space and time, we have chosen not to engage in this debate here" (1990, p. 216).
- 20 The first conversation lasted 34 minutes, the second, eight years later, lasted 75 minutes, and the third, two months after that, 31 minutes.
- 21 White (p.c.) reports that a second set of data, gathered 18 months after those reported on in her paper, will soon be analyzed, thereby helping to determine whether SD's grammar has really reached end-state.
- 22 For discussion of strengths and weaknesses of a range of elicitation tasks used with adult second language learners, see Doughty and Long (2000) and Chaudron (this volume).
- 23 It is hoped that this and other written samples, possibly later supplemented by grammaticality judgment data, will help determine which of certain persistent problems are due to production constraints (for example, on word-final consonants and consonant clusters) interfering with overt marking of some grammatical relations, including plurality and past time reference, as distinct from lack of knowledge. Writing in English is hard for Ayako, however, and something she rarely does, save for the occasional greeting card message, and grammaticality

- judgments are often problematic for learners of this kind, too (for discussion of this problem, see Long, 1993a).
- 24 Care has to be taken when analyzing Ayako's speech to distinguish what would be errors in "standard" spoken English in Hawai'i from what are perfectly grammatical constructions in HCE. Whether or not a given form is deviant is relatively unimportant in a fossilization study compared to whether or not it is supplied consistently. However, determining consistency is sometimes made difficult by the fact that, like most native speakers of the local variety, Ayako's command of HCE is not limited to one level, but allows her to shift up and down within a certain range on a creole continuum according to such factors as topic and, especially, interlocutor. In some contexts, for example, variation in Ayako's suppliance or omission of copula, morphological markers of past time reference, and other forms can be due to a shift toward or away from "standard" spoken English in Hawai'i rather than to variation in her suppliance of the item within a variety. Ambiguous cases are eliminated from the analysis, as, of course, are all instances where suppliance of a targeted item is unclear acoustically, such as past time /t/ or /d/ preceding an initial consonant on the following word.
- 25 If considerable and unpredictable synchronic and diachronic IL change were acceptable indices of fossilization, what would constitute counter-evidence for a fossilization claim? See the earlier discussion of Bean and Gergen (1990).
- 26 This is probably for phonological reasons in some cases. Japanese has CV syllable structure and disallows all word-final consonants except /n/, and all consonant clusters – a constraint known to affect adversely production of English past tense marking by speakers of Vietnamese (Sato, 1984, 1985, 1990). Ayako does, however, produce word-final consonants in many words, including /d/ and /t/, for example, *child, died, that, not, childhood, polite, and eight*, and even some word-final consonant clusters, for example, *it's, that's, raised, and passed*. Written data and grammaticality judgment data collected from Ayako will help clarify this matter.
- 27 IL variation initially pronounced to be "free" has sometimes turned out to be systematic when more carefully analyzed, as shown, for instance, by Berdan (1996), and that may yet turn out to be the case with Ayako's data.
- 28 Malcolm Johnston (p.c.) notes that if a totally systematic fossilized IL is at least a hypothetical possibility, but something always obscured in practice by the kinds of variation inevitable in samples of performance, a proper definition of fossilization would have to rule such a case out, thereby making the construct even more circumscribed than it already is or need be. He writes:
- My feeling is that, if fossilization were to exist, it would have to be something like the case of Ayako, where in a general context of cessation, there is still a kind of "head-banging" variation, i.e., the learner has been stopped by a "wall," but still continues to move back and forth laterally, oscillating as he or she "tries" to breach the

- obstacle at different places. So variation would be fundamental to any postulated definition of fossilization.
- 29 Some of the factors that follow in the main body of the chapter – specifically, diminished access to UG, failure to reset parameters, and maturational constraints – were originally discussed by some of the authors concerned (but not others) as potential explanations of (allegedly universal) failure to acquire an L2 to nativelike levels in general, rather than of fossilization of particular subsystems within individual IL grammars. Others, such as failure to acculturate and automatization, were discussed in both contexts.
- 30 Preston (1989) claims that fossilization in Selinker’s sense, which Preston terms “social fossilization,” is caused by “the social and psychological make-up of the learners, their relationship to other learners, especially shared L1, and their feelings toward their reception in the L2 community” (1989, p. 254). He suggests (p. 255) that a second, symbolic kind of fossilization, which he terms “socio-linguistic fossilization,” can occur when learners deliberately retain variability in their ILs as a marker of their identity in the speech community. While an interesting claim in its own right, it is doubtful whether it meets Selinker’s criteria for fossilization, given that this variability is supposedly under the learner’s control.
- 31 The requirement of predictive power rules out proposals like Nakuma’s. On that account, fossilized structures are the result of entirely idiosyncratic perceptions of interlingual identification of L1–L2 equivalents, with consequent avoidance of L2 forms. Even if testable, this proposal could never predict the future course of development (or arrested development) even for the learner under study, much less for any other learner.
- 32 For discussion and illustrations of the hollowness of such accounts, see Long and Sato (1984, pp. 255–8). Selinker (1972, p. 24) recognizes the danger when he rejects overgeneralization as a potential explanation for fossilization on the grounds that some learners recover from overgeneralizations, whereas others do not, and that some learners recover from some overgeneralizations but not others. He endorses the search for an explanation with predictive power, not descriptive power. Unfortunately, however, as explained below, lack of predictive power is a problem for several candidate explanations, including the MEP.
- 33 Some might claim that while very few, if any, variables can successfully predict stabilization or fossilization, a variety, like those listed earlier, can genuinely account for individual cases after the fact. Put another way, stabilization and/or fossilization might be caused by different factors in different individuals and/or grammatical subsystems and/or discourse domains. Quite apart from the unwelcome enormous increase in the power of the theory such a stance allows (in this as in any other domain of SLA), with data potentially being “explicable” after the fact by any variables the theorist likes – anything goes, and no claim is falsifiable – to take this stance is to forget that fossilization is supposedly a cognitive mechanism

- affecting all L2 learners, albeit manifesting itself in a variety of linguistic domains in different learners.
- 34 Variation in the interactional structure of conversation, on the other hand, does appear to affect both comprehension and acquisition (for review, see Gass, this volume; Long, 1996).
- 35 Klein (1986) implicitly recognizes this problem in his brief discussion of the relevance of "rule criticalness," "confirmation index," "target heterogeneity," and "reflection," in learners' failure to notice input-output mismatches.
- 36 The fact that there is nothing perceptually non-salient about structures like those marking the four stages of ESL negation means that at least some stabilization, for example at the *No V* stage, must simply reflect failure to develop, not loss of capacity to do so. That is, it must reflect factors associated with general success and failure in SLA, like impoverished input, rather than a change in underlying competence, such as that envisaged when fossilization sets in.
- 37 The extent of the problem can be seen in the opening paragraph of a recent encyclopedia entry on fossilization:
- Fossilization is the term used to describe incomplete language learning. This is identified by certain features of the learner's language being different from the speech of the target population, marking the point when progress in that aspect of the target language stops and the learner's language becomes fixed at an intermediate state. This is considered to occur
- because the learner's internalised rule system differs from that of the target system. Fossilization can take a number of forms, such as fossilized accent or syntax, in which case it might approximate to pidginization. Fossilization would normally be judged in relation to native speaker skills and would be seen as a permanent feature of the learner's language, although some authorities (Brown, 1980) describe it as "relatively permanent." (Daniels, 2000, p. 218)
- 38 Hyltenstam (1988) found that, while near-native in most respects, the speech and writing of 24 17- and 18-year-old Swedish high school students, native speakers of Spanish or Finnish who had begun acquiring Swedish between the ages of 4 and 12, still exhibited a number of lexical errors that might indicate fossilization in child SLA. Hyltenstam (1988, pp. 82–3) was careful to note, however, that the cross-sectional nature of his data precluded a definitive answer. The phenomena in question might have reflected permanent problems of the kind associated with fossilization; alternatively, they might simply turn out to have concerned late-acquired items, errors that disappeared at an extremely low rate, that is, incompleteness (see Schachter, 1988) and/or processing capacity restrictions in bilinguals. For further insightful discussion of the difficulty in distinguishing problems of competence or control in such data, see Hawkins (2000), Hyltenstam (1992), and White (2001).
- 39 The status of variation, in particular, clearly remains a major unresolved

issue. Can rules or structures which exhibit wide within-task or cross-task (context, discourse domain, etc.) variation be said to have stabilized, let alone fossilized? If so, is this not to immunize the claim against falsification? How much variation is permissible? Where does one task, context, discourse domain (and what other units?) end and the next

begin? Why should variation (especially, but not only, within-task variation) be taken as evidence for, or consistent with, a claim of fossilization, as opposed to an *unstable* IL or, indeed, of the very opposite, that is, *IL development*? Free variation, after all, is claimed to play a catalytic role in some theories of SLA (see, e.g., Ellis, 1985, 1999).

REFERENCES

- Agnello, F. 1977: Exploring the pidginization hypothesis: a study of three fossilized negation systems. In C. Henning (ed.), *Proceedings of the Los Angeles Second Language Research Forum 1977*. Los Angeles: TESL Section, Department of English, University of California, 246–61.
- Andersen, R. W. 1983: Transfer to somewhere. In S. M. Gass and L. Selinker (eds), *Language Transfer in Language Learning*, Rowley, MA: Newbury House, 177–201.
- Bean, M. and Gergen, C. 1990: Individual variation in fossilized interlanguage performance. In H. Burmeister and P. Rounds (eds), *Variability in Second Language Acquisition. Proceedings of the Tenth Second Language Research Forum. Vol. 1*. Eugene, OR: Department of Linguistics, University of Oregon, 205–19.
- 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 (4), 311–48.
- Berdan, R. 1996: Disentangling language acquisition from language variation. In R. Bayley and D. Preston (eds), *Second Language Acquisition and Linguistic Variation*. Amsterdam: John Benjamins, 203–44.
- Birdsong, D. 1999: Introduction: whys and why nots of the critical period hypothesis for second language acquisition. In D. Birdsong (ed.), *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah, NJ: Lawrence Erlbaum Associates, 1–22.
- Bley-Vroman, R. 1989: The logical problem of second language learning. In S. M. Gass and J. Schachter (eds), *Linguistic Perspectives on Second Language Acquisition*. Cambridge: Cambridge University Press, 41–68.
- Bowerman, M. 1982: Starting to talk worse: clues to language acquisition from children's late speech errors. In S. Strauss and R. Stavy (eds), *U-Shaped Behavioral Growth*. New York: Academic Press, 101–45.
- Briere, E. 1966: An investigation of phonological interference. *Language*, 42, 768–96.
- Brown, H. D. 1980: *Principles of Language Learning and Teaching*. Englewood Cliffs, NJ: Prentice-Hall.
- Bruzzese, G. 1977: English/Italian secondary hybridization: a case study of the pidginization of a second language learner's speech. In C. Henning (ed.), *Proceedings of the Los Angeles Second Language Research Forum 1977*. Los Angeles: TESL Section, Department of English, University of California, Los Angeles, 235–45.

- Calve, P. 1992: Corriger ou ne pas courriger, ça n'est pas la question. *Canadian Modern Language Review*, 48, 458–71.
- Carroll, J. B. and Sapon, S. 1959: *The Modern Languages Aptitude Test*. San Antonio, TX: Psychological Corporation.
- Clahsen, H. 1988: Parameterized grammatical theory and language acquisition: a study of the acquisition of verb placement and inflection by children and adults. In S. Flynn and W. O'Neil (eds), *Linguistic Theory in Second Language Acquisition*. Dordrecht: Kluwer, 47–75.
- Corder, S. P. 1967: The significance of learners' errors. *IRAL*, 5, 161–9.
- Daniels, J. 2000: Fossilization. In M. Byram (ed.), *Routledge Encyclopedia of Language Teaching and Learning*. London: Routledge, 218–20.
- Doughty, C. J. and Long, M. H. 2000: Eliciting second language speech data. In L. Menn and N. Bernstein Ratner (eds), *Methods for Studying Language Production*. Mahwah, NJ: Lawrence Erlbaum Associates, 149–77.
- Douglas, D. 2000: *Assessing Languages for Specific Purposes*. Cambridge: Cambridge University Press.
- Douglas, D. and Selinker, L. 1985: Principles for language tests within the "discourse domain" theory of interlanguage: research, test construction and interpretation. *Language Testing*, 2, 205–26.
- Ellis, R. 1985: Sources of variability in interlanguage. *Applied Linguistics*, 6, 118–31.
- Ellis, R. 1999: Item versus system learning: explaining free variation. *Applied Linguistics*, 20 (4), 461–80.
- Eubank, L. 1995: Generative research on second language acquisition. *Annual Review of Applied Linguistics*, 15, 93–107.
- Gass, S. M. and Lakshmanan, U. 1991: Accounting for interlanguage subject pronouns. *Second Language Research*, 7, 181–203.
- Graham, J. 1981: Overcoming fossilized English. ERIC Document No. ED 209 924.
- Han, Z.-H. 1998: Fossilization: an investigation into advanced L2 learning of a typologically distant language. Doctoral dissertation. Birkbeck College, University of London, Department of Applied Linguistics.
- Han, Z.-H. 2000a: Persistence of the implicit influence of NL: the case of the pseudo-passive. *Applied Linguistics*, 21 (1), 55–82.
- Han, Z.-H. 2000b: Fossilization: the state of the art. Ms. Teachers College, New York.
- Han, Z.-H. and Selinker, L. 1997: Multiple effects and error persistence: a longitudinal case study. Ms. Birkbeck College.
- Han, Z.-H. and Selinker, L. 1999: Error resistance: towards an empirical pedagogy. *Language Teaching Research*, 3 (3), 248–75.
- Harley, B. 1979: French gender "rules" in the speech of English-dominant, French-dominant and monolingual French-speaking children. *Working Papers on Bilingualism*, 19, 129–56.
- Harley, B. and Swain, M. 1978: An analysis of the verb system by young learners of French. *Interlanguage Studies Bulletin*, 3, 35–79.
- Harley, B. and Swain, M. 1984: The interlanguage of immersion students and its implications for second language teaching. In A. Davies, C. Criper, and A. P. R. Howatt (eds), *Interlanguage*. Edinburgh: Edinburgh University Press, 291–311.
- Hawkins, R. 2000: Persistent selective fossilization in second language acquisition and the optimal design of the language faculty. *Essex Research Reports in Linguistics*, 34, 75–90.

- Higgs, T. and Clifford, R. 1982: The push toward communication. In T. Higgs (ed.), *Curriculum, Competence and the Foreign Language Teacher*, Skokie, IL: National Textbook Company, 57–79.
- Hulstijn, J. H. 1989: A cognitive view on interlanguage variability. In M. Eisenstein (ed.), *The Dynamic Interlanguage: Empirical Studies in Second Language Variation*. New York: Plenum Press, 17–31.
- Hyltenstam, K. 1977: Implicational patterns in interlanguage syntax variation. *Language Learning*, 27, 383–411.
- Hyltenstam, K. 1988: Lexical characteristics of near-native second-language learners of Swedish. *Journal of Multilingual and Multicultural Development*, 9 (1 and 2), 67–84.
- Hyltenstam, K. 1992: Non-native features of near-native speakers: on the ultimate attainment of childhood L2 learners. In R. J. Harris (ed.), *Cognitive Processing in Bilinguals*. Amsterdam: Elsevier Science, 351–68.
- Jain, M. P. 1974: Error analysis: source, cause and significance. In J. C. Richards (ed.), *Error Analysis*. London: Longman, 189–215.
- Johnson, H. 1993: Defossilizing. In J. Oller and L. Richards d'Amato (eds), *Methods that Work: Ideas for Literacy and Language Teachers*. Second edition. Boston, MA: Heinle and Heinle, 281–8.
- Kellerman, E. 1989: The imperfect conditional. In K. Hyltenstam and L. K. Obler (eds), *Bilingualism Across the Lifespan*. Cambridge: Cambridge University Press, 87–115.
- Klein, W. 1986: *Second Language Acquisition*. Cambridge: Cambridge University Press.
- Klein, W. and Perdue, C. 1997: The Basic Variety (or: couldn't natural languages be much simpler?). *Second Language Research*, 13 (4), 301–47.
- 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, 359–75.
- Lardiere, D. 2000a: Mapping features to forms in second language acquisition. In J. Archibald (ed.), *Second Language Acquisition and Linguistic Theory*. Malden, MA: Blackwell, 102–29.
- Lardiere, D. 2000b: On optionality and grammaticality in L2 knowledge. Paper presented at the 4th Meeting on Generative Approaches to Second Language Acquisition (GASLA). MIT, April 1.
- Larsen-Freeman, D. and Long, M. H. 1991: *An Introduction to Second Language Acquisition Research*. London: Longman.
- Lennon, P. 1991a: Error and the very advanced learner. *IRAL*, 19 (1), 31–44.
- Lennon, P. 1991b: Error elimination and error fossilization: a study of an advanced learner in the L2 community. *ITL*, 93–4, 129–51.
- Lin, Y.-H. 1995: Un análisis empírico de la estabilización/fosilización: la incorporación y la autocorrección en unos sujetos chinos. Doctoral thesis. Departamento de Didáctica de la Lengua y la Literatura, Universidad de Barcelona.
- Lin, Y.-H. and Hedgcock, J. 1996: Negative feedback incorporation among high-proficiency and low-proficiency Chinese-speaking learners of Spanish. *Language Learning*, 46 (4), 567–611.
- Linn, C. A. and Sucher, K. 1995: De-fossilization through memorization of the perfect paragraph. Paper presented at the International TESOL Convention. Long Beach, CA, March.
- Long, M. H. 1993a: Second language acquisition as a function of age:

- substantive findings and methodological issues. In K. Hyltenstam and A. Viberg (eds), *Progression and Regression in Language*. Cambridge: Cambridge University Press, 196–221.
- Long, M. H. 1993b: Sensitive periods in second language acquisition. Paper presented at the invitational seminar “SLA and the Philosophy of Science.” Washington, DC, National Institutes of Mental Health.
- Long, M. H. 1996: The role of the linguistic environment in second language acquisition. In W. C. Ritchie and T. K. Bahtia (eds), *Handbook of Second Language Acquisition*. New York: Academic Press, 413–68.
- Long, M. H. 1997: Fossilization: Rigor mortis in living linguistic systems? Plenary address to the EUROSLA 97 conference. Universitat Pompeu Fabra, Barcelona, May 22–4.
- Long, M. H. and Sato, C. J. 1984: Methodological issues in interlanguage studies: an interactionist perspective. In A. Davies, C. Cramer, and A. P. R. Howatt (eds), *Interlanguage*. Edinburgh: Edinburgh University Press, 253–79.
- Meisel, J. M. 1991: Principles of Universal Grammar and strategies of language learning: some similarities and differences between first and second language acquisition. In L. Eubank (ed.), *Point Counterpoint: Universal Grammar in the Second Language*. Amsterdam: John Benjamins, 231–76.
- Meisel, J. M. 1997: The acquisition of the syntax of negation in French and German: contrasting first and second language acquisition. *Second Language Research*, 13, 227–63.
- Meisel, J. M., Clahsen, H., and Pienemann, M. 1981: On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition*, 3 (2), 109–35.
- Mellow, D., Reeder, K., and Forster, E. 1996: Using time-series research designs to investigate the effects of instruction. *Studies in Second Language Acquisition*, 18 (3), 325–50.
- Mukkatash, L. 1986: Persistence in fossilization. *IRAL*, 24 (2), 187–203.
- Naiman, N. 1974: Imitation, comprehension, and production of certain syntactic forms by young children acquiring a second language. Doctoral dissertation. University of Toronto.
- Nakuma, C. 1998: A new theoretical account of “fossilization”: implications for L2 attrition research. *IRAL*, 36 (3), 247–56.
- Nemser, W. 1971: Approximative systems of foreign language learners. *IRAL*, 9 (2), 115–24.
- Pavesi, M. 1986: Markedness, discursial modes, and relative clause formation in a formal and an informal context. *Studies in Second Language Acquisition*, 8 (1), 38–55.
- Pellerin, M. and Hammerly, H. 1986: L’expression orale apres treize ans d’immersion francaise. *Canadian Modern Language Review*, 42, 592–606.
- Perdue, C. 1993: *Adult Language Acquisition: Cross-Linguistic Perspectives. Vol. 1: Field Methods*. Cambridge: Cambridge University Press.
- Pienemann, M. 1984: Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition*, 6, 186–214.
- Plann, S. 1976: The Spanish immersion program: towards nativelike proficiency or a classroom dialect? MA in TESL thesis. University of California, Los Angeles.
- Plann, S. 1977: Acquiring a second language in an immersion situation. In H. D. Brown, C. Yorino, and R. Crymes (eds), *On TESOL ’77*. Washington, DC: TESOL, 213–23.
- Preston, D. 1989: *Sociolinguistics and Second Language Acquisition*. Oxford: Blackwell.

- The Random House Dictionary of the English Language* 1987: Second edition, unabridged. New York: Random House.
- Sato, C. J. 1984: Phonological processes in second language acquisition: another look at interlanguage syllable structure. *Language Learning*, 34 (4), 43–57. Also in G. Ioup and S. Weinberger (eds), *Interlanguage Phonology*. Cambridge, MA: Newbury House, 248–60.
- Sato, C. J. 1985: Task variation in interlanguage phonology. In S. M. Gass and C. Madden (eds), *Input in Second Language Acquisition*. Rowley, MA: Newbury House, 181–96.
- Sato, C. J. 1990: *The Syntax of Conversation in Interlanguage Development*. Tubingen: Gunter Narr.
- Schachter, J. 1974: An error in error analysis. *Language Learning*, 27, 205–14.
- Schachter, J. 1988: The notion of completeness in second language acquisition. Paper presented at the Boston University Conference on Language Development.
- Schachter, J. 1996: Maturation and the issue of Universal Grammar in second language acquisition. In W. R. Ritchie and T. J. Bhatia (eds), *Handbook of Second Language Acquisition*. San Diego: Academic Press, 159–93.
- Schmidt, R. W. 1983: Interaction, acculturation, and the acquisition of communicative competence: a case study of an adult. In N. Wolfson and E. Judd (eds), *Sociolinguistics and Second Language Acquisition*. Rowley, MA: Newbury House, 137–74.
- Schmidt, R. W. 1995: Consciousness and foreign language learning: a tutorial on the role of attention and awareness in learning. In R. W. Schmidt (ed.), *Attention and Awareness in Foreign Language Learning*. Honolulu: University of Hawai'i Press, 1–63.
- Schnitzer, M. L. 1993: Steady as a rock: does the steady state represent cognitive fossilization? *Journal of Psycholinguistic Research*, 22 (1), 1–20.
- Schouten, E. 1996: Crosslinguistic influence and the expression of hypothetical meaning. In E. B. Kellerman, B. Weltens, and T. Bongaerts (eds), *EUROSLA 6: A Selection of Papers. Toegepaste Taalwetenschap in Artikelen [Applied Linguistics in Article Form]*, 55, 161–74.
- Schumann, J. H. 1978: *The Pidginization Process: A Model for Second Language Acquisition*. Rowley, MA: Newbury House.
- Schumann, J. H. 1979: The acquisition of English negation by speakers of Spanish: a review of the literature. In R. W. Andersen (ed.), *The Acquisition and Use of Spanish and English as First and Second Languages*. Washington, DC: TESOL, 3–32.
- Schumann, J. H. 1986: Research on the acculturation model for second language acquisition. *Journal of Multilingual and Multicultural Development*, 7, 379–92.
- Schumann, J. H. 1993: Some problems with falsification: an illustration from SLA research. *Applied Linguistics*, 14, 295–306.
- Seliger, H. W. 1978: Implications of a multiple critical periods hypothesis for second language learning. In W. Ritchie (ed.), *Second Language Acquisition Research*. New York: Academic Press, 11–19.
- Seliger, H. W. 1989: Semantic transfer constraints on the production of English passive by Hebrew–English bilinguals. In H. Dechert and M. Raupach (eds), *Transfer in Language Production*. Norwood, NJ: Ablex, 21–34.
- Selinker, L. 1966: A psycholinguistic study of language transfer. Ph.D. dissertation. Department of Linguistics, Georgetown University.
- Selinker, L. 1972: Interlanguage. *IRAL*, 10 (3), 209–31.
- Selinker, L. 1992: *Rediscovering Interlanguage*. New York: Longman.

- Selinker, L. 1993: Fossilization as simplification? In M. L. Tickoo (ed.), *Simplification: Theory and Application*. Singapore: Regional Language Centre, 14–28.
- Selinker, L. 1996: Research proposal for grant application submitted to the British Library. Ms.
- Selinker, L. and Douglas, D. 1985: Wrestling with “context” in interlanguage theory. *Applied Linguistics*, 6, 190–204.
- Selinker, L. and Douglas, D. 1989: Research methodology in contextually based second language research. *Second Language Research*, 5 (1), 1–34.
- Selinker, L. and Han, Z.-H. 1996: Fossilization: what we think we know. Paper presented at EUROSOLA 96. Nijmegen, May 30.
- Selinker, L. and Han, Z.-H. 2001: Fossilization: moving the concept into empirical longitudinal study. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, and K. O’Loughlin (eds), *Studies in Language Testing: Experimenting with Uncertainty*. Cambridge: Cambridge University Press, 276–91.
- Selinker, L. and Lakshmanan, U. 1992: Language transfer and fossilization: the multiple effects principle. In S. M. Gass and L. Selinker (eds), *Language Transfer in Language Learning*. Amsterdam: John Benjamins, 176–217.
- Selinker, L. and Lamendella, J. T. 1978: Two perspectives on fossilization in interlanguage learning. *Interlanguage Studies Bulletin*, 3 (2), 143–91.
- Selinker, L. and Mascia, R. 1999: Fossilization: trying to get the logic right. In P. Robinson (ed.), *Representation and process: Proceedings of the 3rd Pacific Second Language Research Forum. Vol. 1*. Tokyo: Pacific Second Language Research Forum, 257–65.
- Selinker, L., Swain, M., and Dumas, G. 1975: The interlanguage hypothesis extended to children. *Language Learning*, 25 (1), 139–52.
- Sola, D. 1989: Fossilized errors and the effect of formal instruction on highly proficient Hispanic learners of English. M.A. thesis. University of South Carolina, Linguistics Program.
- Stauble, A.-M. 1984: A comparison of the Spanish–English and Japanese–English interlanguage continuum. In R. W. Andersen (ed.), *Second language: A Cross-Linguistic Perspective*. Rowley, MA: Newbury House, 323–53.
- Tarone, E. 1988: *Variation in Interlanguage*. London: Edward Arnold.
- Tarone, E. 1994: Interlanguage. In R. E. Asher (ed.), *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon Press, 1715–19.
- Thep-Ackrapong, T. 1990: Fossilization: a case study of practical and theoretical parameters. Doctoral dissertation. Illinois State University. DAI-A 51/09.
- Todeva, E. 1992: On fossilization in (S)LA theory. In D. Staub and C. Delk (eds), *Proceedings of the Twelfth Second Language Research Forum*. East Lansing: Center for International Programs, Michigan State University, 216–54.
- Valette, R. 1991: Proficiency and the prevention of fossilization – an editorial. *Modern Language Journal*, 75 (3), 325–8.
- VanPatten, B. 1988: How juries get hung: problems with the evidence for a focus on form in teaching. *Language Learning*, 38 (2), 243–60.
- Vigil, N. A. and J. W. Oller, Jr 1976: Rule fossilization: a tentative model. *Language Learning*, 26 (2), 281–95.
- Wales, L. 1993: The benefits of literacy development for fossilized ESL learners. *ELT Journal*, 47 (2), 144–56.
- Washburn, G. 1992: Fossilization in second language acquisition: a Vygotskian analysis. Ph.D. dissertation. University of Pennsylvania.
- Washburn, G. 1994: Working in the ZPD: fossilized and nonfossilized nonnative

- speakers. In J. P. Lantolf and G. Appel (eds), *Vygotskian Approaches to Second Language Research*. Norwood, NJ: Ablex, 69–81.
- White, L. 1987: Against comprehensible input: the input hypothesis and the development of second language competence. *Applied Linguistics*, 8 (1), 95–110.
- White, L. 2002: Morphological variability in endstate L2 grammars: the question of L1 influence. In B. Skarabela, S. Fish, and A.H.-J. Do (eds), *Proceedings of the 26th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 758–68.
- Wode, H. 1981: *Learning a Second Language*. Tübingen: Gunter Narr.
- Yorio, C. 1994: The case for learning. In R. M. Barasch and C. Vaughn James (eds), *Beyond the Monitor Model*. Boston, MA: Heinle and Heinle, 125–37.
- Zobl, H. 1982: A direction for contrastive analysis: the comparative study of developmental sequences. *TESOL Quarterly*, 16 (2), 169–83.

