Explaining the "Natural Order of L2 Morpheme Acquisition" in English: A Meta-analysis of Multiple Determinants

Jennifer M. Goldschneider and Robert M. DeKeyser
University of Pittsburgh

This meta-analysis pools data from 25 years of research on the order of acquisition of English grammatical morphemes by students of English as a second language (ESL). Some researchers have posited a "natural" order of acquisition common to all ESL learners, but no single cause has been shown for this phenomenon. Our study investigated whether a combination of 5 determinants (perceptual salience, semantic complexity, morphophonological regularity, syntactic category, and frequency) accounts for the variance in acquisition order. Oral production data from 12 studies, together involving 924 participants, were pooled to obtain weighted accuracy scores for each of 6 grammatical functors. Results of a multiple-regression analysis showed that a large portion of the total variance in acquisition order was explained by the combination of the 5 determinants. Several of these determinants, it was argued, can be seen as part of a broad conceptualization of salience.

Robert M. DeKeyser and Jennifer M. Goldschneider, Department of Linguistics.

The authors gratefully acknowledge the advice and feedback from Alan Juffs, Margaret MacEachern, Sarah Thomason, Nick Ellis, and several anonymous reviewers, and the statistical help and advice from Elaine Rubinstein.

Correspondence concerning this article may be sent to Robert DeKeyser, Department of Linguistics, University of Pittsburgh, Pittsburgh, PA 15260. Internet: RDK1@pitt.edu

Since the article was originally published, a number of meta-analyses have appeared in the applied linguistics literature (e.g., Masgoret & Gardner, 2003; Norris & Ortega, 2000; Ortega, 2003), and a book on meta-analysis in applied linguistics research is forthcoming (Norris & Ortega, in press). Meanwhile, research on the acquisition of Hebrew as a second language has begun to provide cross-linguistic evidence for how different aspects of salience contribute to ease or difficulty of second language acquisition (DeKeyser, Alfi-Shabtay, Ravid, & Shi, 2005) and how salience interacts with age of learning (DeKeyser, Ravid, & Alfi-Shabtay, 2005).

This paper reports on a meta-analysis of previous research on the order of acquisition of certain grammatical functors[1] of English by students of English as a second language (ESL). Some researchers in this area have posited a "natural" order of acquisition common to all ESL learners irrespective of age,[2] learning environment, or prior languages learned (Dulay & Burt, 1973, 1974; Bailey, Madden, & Krashen, 1974). However, previous attempts to ascribe a single causality to these findings have failed (cf. Gass & Selinker, 1994). This study investigates whether a combination of contributing factors can provide a better account of the phenomenon. These factors include perceptual salience, semantic complexity, morphophonological regularity, syntactic category, and frequency.

The data on the use of grammatical functors is drawn from 12 primary studies reported between 1973 and 1996. These 12 studies were chosen on the basis of specific criteria (see below) from 25 candidate studies. This process makes it possible to examine the acquisition orders of over 900 subjects, far more than would be possible in an original study using oral production data. Pooling the studies brings together, for the first time, data from speakers of 29 first languages, both children and adults.

Additionally, because all of the included studies share a common scoring approach, i.e., tallying suppliance in obligatory contexts (SOC), it is possible to combine their results, in spite of the fact that different instruments were used, including the Bilingual Syntax Measure (BSM) (Dulay & Burt, 1973), the Second Language Oral Production English (SLOPE) test (Fathman, 1975), and the MAT-SEA-CAL Oral Proficiency Test (Mace-Matluck, 1977). For the functors in common across these studies, i.e., progressive *-ing*, plural *-s*, possessive *-s*, articles *a*, *an*, *the*, third person singular present *-s*, and regular past *-ed*, significant and high correlations in functor acquisition order have consistently been found (see Appendix A for rank orders). A great deal of criticism of the "morpheme order studies" has accumulated over the past 20 years (see especially Cook, 1993; Ellis, 1986; Hatch, 1978; Larsen-Freeman & Long, 1991; Long & Sato, 1984), e.g., that the group of functors is too linguistically heterogeneous, that the scoring methods did not take function into account and missed over-suppliance, that the elements of grammar studied constitute a very small part of the language, and that there is no cross-linguistic generalization. Nevertheless, Larsen-Freeman and Long (1991, p. 92) have concluded:

> [T]he morpheme studies provide strong evidence that ILs [interlanguages] exhibit common accuracy/acquisition orders. Contrary to what some critics have alleged, there are in our view too many studies conducted with sufficient methodological rigour and showing sufficiently consistent general findings for the commonalities to be ignored. As the hunter put it, "There is something moving in the bushes."

The purpose of this study was to determine the extent to which a combination of proposed predictors would predict the observed order of acquisition[3] of grammatical functors in ESL. This is an important question, because if the order of functor acquisition is best predicted by a combination of factors, then this must be accounted for in an overall theory of second language acquisition and may end the quest for a single cause

underlying this phenomenon. This quest has been carried on for decades, from Larsen-Freeman (1976) to Pienemann (1998), and has become somewhat of a Holy Grail of SLA research. In particular, a number of publications have suggested that the L2 morpheme acquisition order is indicative of some sort of innate blueprint, only hinted at in broad terms by, e.g., Dulay, Burt, and Krashen (1982), and elaborated in more detail, drawing on various developments in syntactic theory, in publications such as those of Zobl (1995) and Pienemann (1998).

Others, however, believe that the order of acquisition can be explained by the interaction between characteristics of the elements to be acquired and general cognitive principles of inductive learning. Ellis and Laporte (1997, p. 64), e.g., state:

> Sequences of development are as much, or even more, a consequence of epistemology, of the structure of knowledge in the relevant problem-space, as they are a consequence of the learners' biological processing capacity and neural development. Invariant developmental sequences of language acquisition are essentially interesting because they inform us about the informational content of language and how more complicated structures arise from simpler, more basic forms. They are as consistent with empiricist as with linguistic nativist theories of language.

In an attempt to document how characteristics of the very elements to be acquired determine order of acquisition, the multiple regression analysis in this study examines the effect of the aforementioned five predictors on the dependent variable: scores for accuracy of use in obligatory contexts of selected grammatical functors. The research question is as follows:

> How much of the total variance in ESL functor acquisition order can be accounted for by the combination of five determinants: perceptual salience, semantic complexity, morphophonological regularity, syntactic category, and frequency?

Our prediction was that a large portion of the variance can be explained by the five proposed factors. This finding would mean that the observed order in which functors are acquired could, in large part, be explained by the cumulative effect of the five determinants. The design and methodology of this study were chosen, therefore, to investigate the following hypothesis:

> A statistically significant portion of the variance in accurate use of grammatical functors in oral production data by ESL learners (measured by percentage of correct suppliance in obligatory contexts) will be accounted for by a combination of five factors: perceptual salience, semantic complexity, morphophonological regularity, syntactic category, and frequency.

Review of the Literature

Since the early days of SLA research, a question of particular interest has been whether the processes involved in first and second language acquisition are the same or different. To explore this question, it can be productive to replicate a line of research conducted on first language (L1) acquisition within the context of second language (L2) acquisition. This was the aim of the early SLA research concerning the order of acquisition of grammatical functors.

First Language Acquisition Studies

Roger Brown's (1973) longitudinal study of three American children acquiring English as their first language is most often recognized as the starting point for acquisition order research. Of course, Brown's work did not exist in a vacuum. Cook (1993) points out that Brown's research, and earlier child grammars of the 1960s, were founded on the "independent grammars assumption" (Cook, 1993, p. 14), according to which children acquiring their first language are speakers of their own personal languages rather than defective speakers of the adult

language. Once children acquiring their L1 were thought of as speakers of unknown and undocumented languages, researchers needed detailed written grammars of these languages. Brown (1973) was unique, however, in presenting an organizational framework for the grammar that yielded new insights into the stages of language development, and particularly the strikingly similar order in which his subjects acquired fourteen grammatical functors (see Appendix B).

According to Brown, what made the similar orders particularly interesting was that the pattern of development was still clear despite the fact that the data were drawn from spontaneous speech, so that the specific instantiations of the functors were not alike and the topics of the children's conversations were unrelated. As he put it, "Some factor or some set of factors caused these grammatical morphemes to evolve in an approximately consistent order in these children" (Brown, 1973, p. 272). Much of Brown's chapter on grammatical functors is an exploration of what these factors might be. His ideas have shaped much of the later research, and these ideas will be discussed in more detail under Methods and Procedures.

Brown's analysis also set the standard for scoring and comparing data on functor acquisition order. For example, on the basis of the observation that when the data were represented graphically, the curve of performance on a particular functor was initially erratic, but once a curve had passed above the 90% line for several consecutive samples it tended to remain above 90%, Brown set his cutoff point of "acquisition" at 90%. (Fortunately, he also scored and ranked those functors that did not attain this arbitrary criterion level, making later comparisons possible.)

Perhaps the most significant benchmark set by Brown (1973) is the concept of suppliance in obligatory context (SOC). Brown (1973, p. 255) chose to look at grammatical functors specifically because they could be more easily quantified than other sorts of structures:

> [G]rammatical morphemes are obligatory in certain contexts, and so one can set an acquisition criterion not simply in terms of output but in terms of output-where-required. Each obligatory context can be regarded as a kind of test item which the child passes by supplying the required morpheme or fails by supplying none or one that is not correct. This performance measure, the percentage of morphemes supplied in obligatory contexts, should not be dependent on the topic of conversation or the character of the interaction.

This method for scoring the data was adopted by many later studies and indeed is used in nearly all the studies included in this meta-analysis (Rosado, 1986, being the only exception). SOC also played a role in the selection of which 14 morphemes Brown studied, since his primary requirements for including a functor were that obligatory contexts could be identified for it and that it was used frequently enough to provide continuous data across speech samples (Brown, 1973, p. 270). (The use of SOC has been severely criticized on several grounds, and a discussion of these criticisms may be found below under Data Selection.)

Brown suggested and tested several possible determinants of the order of functor acquisition, including frequency in parental speech, semantic complexity, and grammatical complexity. After a thorough discussion of the problems inherent in operationalizing these determinants, he concluded that frequency was not a significant variable, and that semantic and grammatical complexity did play a role. However, he noted that the two types of complexity were confounded and that new definitions would be needed to separate the two. Despite the lack of an explanation for the observed order, the discovery of "an approximately invariant order of acquisition for the 14 morphemes" (Brown, 1973, p. 379) set in motion a series of new studies, which came to be known as the "morpheme order studies."

Even before Brown's book was published, a partial replication of his study was underway. Jill de Villiers and Peter de Villiers (1973) conducted a cross-sectional study of 21 English-speaking

children and used Brown's 14 functors and his coding rules for identifying obligatory contexts. Their results were very similar to Brown's longitudinal study. Additionally, de Villiers and de Villiers introduced a new ranking method: Instead of looking for a 90% criterion level of acquisition, they simply ranked the functors according to relative accuracy of use. Some variation of this method of rank ordering on the basis of percentage of functors correctly supplied in obligatory contexts has been employed in most of the L2 studies.

Second Language Acquisition Studies

Prior to the emergence of the "morpheme studies" in SLA, a primarily behaviorist view of language learning held that learning a second language meant acquiring a second set of speech habits through practice and ongoing exposure to stimulus-response patterns (summarized in, e.g., Cook, 1993; Gass & Selinker, 1994; cf. Bloomfield, 1933; Lado, 1957). A related theory, the Contrastive Analysis Hypothesis, assumed that all learning of a second language was done with reference to the first language, so that "difficulty and ease in learning are determined respectively by differences and similarities between the two languages in contrast" (Gass & Selinker, 1994, p. 60). In the 1960s, the study of language acquisition moved away from these concepts toward generative ideas of syntactic rules and children constructing grammars on the basis of the input they receive (Cook, 1993). All of this set the stage for Dulay and Burt's 1973 study, which in part "was designed to provide systematic empirical data on the 'habit formation' vs. 'creative construction' nature of second language learning by children" (p. 246).

Dulay and Burt began with the question: "Is there a common sequence with which children acquiring English as a second language learn certain structures?" (Dulay & Burt, 1973, p. 252). They predicted that if a common sequence were found for L2 children, it would be different from the order Brown (1973) found for L1, since "the older L2 learner need not struggle with

the same kinds of semantic notions already acquired in earlier childhood" (Dulay & Burt, 1973, p. 252). Their study looked at the performance of 151 Spanish-speaking children, aged 5 to 8, from two locations in California and one in New York. Instead of collecting spontaneous speech samples, Dulay and Burt utilized data elicited from the children with an instrument called the Bilingual Syntax Measure (BSM).

The BSM was not specifically designed to test order of acquisition. It is a test of L2 proficiency designed for young children. The test consists of 7 cartoon pictures and 33 questions designed to elicit some type of response from the child, so that a degree of proficiency may be calculated for the structures the child offers (Dulay & Burt, 1973). Dulay and Burt's speech corpus from the BSM for the three groups of children included 8 of the 14 functors from Brown (1973), namely present progressive, plural, past irregular, possessive, articles, third person singular regular, contractible copula, and contractible auxiliary. Each obligatory occasion for a functor was scored according to the following schema[4] (from Dulay & Burt, 1973, p. 254):

No functor supplied: = 0 (she's dance_)

Misformed functor supplied: = 0.5 (she's dance*s*)

Correct functor supplied: = 1.0 (she's danc*ing*)

The acquisition score for each functor was then calculated as a ratio of the sum of the scores for each obligatory occasion of that functor over the total number of obligatory occasions for that functor across the whole group. For the examples above, the ratio would be $1.5/3 = 50\%$.

Dulay and Burt found that, although there were differences in how accurately the functors were used by each group of children, the overall rank order of the functors was similar across the three groups, and as the authors had predicted, that order differed from the L1 order found in Brown (1973). These findings were confirmed by their next study (Dulay & Burt, 1974), which expanded the sense

of a "common" order of acquisition by looking at groups of children with two different L1s. Very similar orders were found for Spanish-speaking and Chinese-speaking children aged 6 to 8.

Once Dulay and Burt had reported a strikingly consistent order for children, the next logical extension of the question was: Is there a consistent order among adult ESL learners, and if so, does it differ from the child ESL order? Bailey, Madden, and Krashen (1974) delved into this new area by administering the BSM to 73 adults, aged 17 to 55. There were 33 Spanish speakers and 40 speakers of 11 languages other than Spanish. This study found that the orders for Spanish and non-Spanish speakers were significantly correlated ($r = .926$, $p < .005$), and that the order for the adults was extremely similar to the child order found by Dulay and Burt.

Larsen-Freeman (1975) took the study of adult ESL order of acquisition a step further by investigating the order obtained by using tasks other than speech production (i.e., reading, writing, listening, and imitating). She found that within tasks, the functor orders were very consistent across different language backgrounds, although she also identified variation in the functor orders across the different tasks. Most importantly, Larsen-Freeman introduced an important new point in functor order research: the question of an explanation for the observed order. After a brief discussion of possible explanations, including semantic and syntactic complexity, phonological form, perceptual salience, and frequency of occurrence, Larsen-Freeman concluded that "[a] single explanation seems insufficient to account for the findings" (1975, p. 419). This is the first hint in the literature that there may be multiple factors behind the common functor order. Larsen-Freeman (1976) did make an attempt at a single explanation, tentatively concluding that the major cause of the order was frequency in the input to the learner. This explanation was refuted by Dulay and Burt (1978), who cited neologisms and systematic errors in learners' speech, as well as the fact that grammatical functors occur more frequently in English than common content words, but are acquired later, as evidence

against frequency as a determinant of the order of functor acquisition.

As the number of studies concentrating on "morpheme order" grew, the methods for testing the order became more diverse. Fathman (1975) introduced the Second Language Oral Production English (SLOPE) test, which was designed to test 20 different morphological and syntactic items, including 7 of the functors used by Dulay and Burt (1973). Evidence from this test in children (Fathman, 1975; Kjarsgaard, 1979) and in adults (Krashen, Sferlazza, Feldman, & Fathman, 1976; Fuller, 1978) showed that for structures common to the SLOPE test and the BSM, there was a similar order despite the differences in the task and the scoring method. The MAT-SEA-CAL Oral Proficiency Test (Mace-Matluck, 1977) also yielded a functor order similar across language groups, both within Mace-Matluck's study and when compared with Dulay and Burt (1973) and Fathman (1975). Mace-Matluck also contributed to the discussion of causality, citing frequency of occurrence, perceptual salience, and L1 influence as possible determinants of the L2 order.

Rosansky (1976) pointed out the importance of individual variability in the accurate use of the functors, and cautioned against statistical processes that would obscure this level of variation. Andersen (1978) proposed an implicational model for scoring functor order data, to capture more of that variability. Hakuta (1976) reported on a longitudinal study of one Japanese child acquiring English as L2, and it was found that the difficulty order he observed did not correlate with any of the orders from the cross-sectional studies, which launched a debate over which method was best for studying functor acquisition order.

Throughout this flurry of research, there seems to have been very little concern about explaining the order findings. Most of the attention in that early period was focused on establishing whether the functor order was real, and under what conditions the order was found. In the 1970s, only Larsen-Freeman (1976) attempted to test an explanation of the morpheme order phenomenon. In the 1980s, however, more discussions appeared concerning what might cause the observed order. Pica (1983) investigated the effects of formal

classroom instruction versus learning in a naturalistic environment and found that "different conditions of exposure to English L2 do not significantly alter the accuracy order in which grammatical morphemes are produced" (1983, p. 465). Rosado (1986) examined the role that the monitor (Krashen's [1977] conscious linguistic device for editing output) might play in determining the order of functor acquisition, and he concluded that either the monitor did not affect the order of functor acquisition or his study did not successfully invoke the monitor.

Pienemann and Johnston (1987) did not test acquisition orders directly, but rather proposed a "theoretical basis for independently explaining the chronology of acquisition" (1987, p. 112). They proposed a specific explanation based on their theory of speech-processing constraints that are gradually and progressively eliminated during language development. According to this model, learners pass through a series of stages of development, with the move from one stage to the next occurring when the learner overcomes a particular constraint. The stages form an implicational sequence, such that "mastery of rules at a particular stage entail[s] mastery of the rules characterizing earlier stages" (Pienemann & Johnston, 1987, p. 74). The order of functor acquisition is thus predicted, because a functor will not be mastered until any constraints that might be "blocking" it are overcome. As constraints are overcome sequentially, the functors then appear sequentially.

Larsen-Freeman and Long call Pienemann and Johnston's theory "an explanation with predictive power" (1991, p. 284), but Pienemann and Johnston give very little empirical evidence to support their theory and give no explanation at all in their model for the mechanism of acquisition it seems to assume (Ellis, 1994, p. 388). Pienemann (1998) elaborated the multidimensional model into "Processability Theory," which is tied in with Levelt's 1989 model of speech production and Kaplan and Bresnan's 1982 lexical-functional grammar, but the main idea remains the same: Both morphological and syntactic phenomena develop in a certain order because of constraints on processability that gradually disappear.

Other interesting questions were posed by J. D. Brown (1983) in his study of morpheme-group interactions. He found, as had Andersen (1978) before him, that insights could be gained from grouping the functors on the basis of qualities inherent to the functors themselves, such as the free/bound distinction and the verb/noun phrase (V/NP) distinction. In light of these findings, Brown suggested that we should look at what kind of interactions among qualities of the functors might have an effect on acquisition order. He observed (Brown, 1983, pp. 37–38) that:

> [S]imilarities in overall acquisition orders may be due to systematic relationships between morphemes within each group [e.g., free NP, bound V, etc.] . . . while dissimilarities are being caused by as yet unpredictable interactions between the morpheme groups. . . . Could it be that the 3rd person singular -s morpheme, for example, is made up of verbness, boundedness, number agreement, and person agreement? If so, we are perhaps looking at morphemes incorrectly.

The possibility that the order of acquisition of grammatical morphemes is determined to a large extent by properties of the functors themselves forms the foundation for the present meta-analysis. This study examines a combination of features of the functors: perceptual salience, semantic complexity, morpho-phonological regularity, syntactic category, and frequency. Importantly, we do not wish to exclude the possibility that other factors external to the functors, such as L1 transfer, also contribute to the observed order. If, however, a large portion of the order findings can be explained on the basis of these five properties of the functors themselves, then we will be much closer to explanatory adequacy for the order-of-acquisition findings.

Methods and Procedures

"Meta-analysis is a non-experimental technique that uses previously reported research findings as its 'subjects'" (Driscoll, 1984), either by pooling data to assess the combined effect on

significance or by estimating the magnitude of the experimental effect across studies (Cooper & Hedges, 1994; Glass, McGaw, & Smith, 1981; Hedges & Olkin, 1985; Wolf, 1986). Meta-analysis is performed most frequently in the fields of medical and educational research. This type of research integration is an alternative to the traditional method of systematic review, in which an overview narrative summarizes a selection of studies one by one.

In the medical and social science literature, meta-analysis usually refers to the synthesizing of results from some number of studies, such that pooled results are reported quantitatively, but without posing any new research question. We expand the term in this study to refer to the process of extracting and pooling data from multiple studies and then using those data for the purpose of testing a new research question (cf. Wolf, 1986).

The quantitative advantages sought by clinical researchers are no less desirable in other disciplines, including the study of language acquisition. Meta-analysis can provide a larger and more representative body of data than would be practical in any single study.

Along with the benefits, meta-analysis presents special challenges. There are many difficulties inherent in the process of identifying, obtaining, and evaluating relevant studies, and in distilling usable data from them. Many of these will be discussed in detail in this section (a) to explain how the data for this study were obtained, and (b) to provide a model for this technique within the framework of language acquisition research.

Data Selection

A meta-analysis should pool as many data as possible in order to maximize the generalizability of the results. However, the need for numbers of studies has to be balanced by the focus of the research question and the need to merge results that are directly comparable. In selecting data for this study, for instance, we did not attempt to include studies on the order of acquisition found in

languages other than English. Because we wanted to look at the order of acquisition of grammatical functors, we did not include studies on the order of acquisition of other grammatical structures, such as Wh-questions or negatives. With this as a starting point, a systematic search of computer databases (including Current Contents, ERIC, PsycLit and MLA) and the indices for Dissertation Abstracts identified 25 studies, involving 2,567 subjects and performed between 1973 and 1996, as possible candidates for inclusion in this meta-analysis (see Appendix C).

We chose to limit our research question to the acquisition of English as a second language; therefore, studies on the acquisition of English as a foreign language (that is, in a setting where English is not the primary language of the learner's environment) were excluded (Andersen, 1977; Lightbown, 1983; Lightbown, Spada, & Wallace, 1980; Makino, 1979).

The next criterion for inclusion concerned the type of language data elicited. The earliest studies (e.g., Dulay & Burt, 1973, 1974; Bailey, Madden, & Krashen, 1974) all used oral production data in order to facilitate comparison with the L1 oral data from Brown (1973). In later L2 studies, especially Larsen-Freeman (1975), researchers began looking at data from other types of tasks (reading, writing, listening, repetition) to determine whether the orders obtained were similar to the order for oral production.[5] For pooling the data, it was crucial to combine only comparable types, and since more studies have used oral data, we decided to restrict our selection to those studies or parts of studies that involved oral production data. Another of the candidate studies, J. D. Brown (1983), was therefore eliminated because it dealt solely with written data.

In addition to limiting the scope in this study to oral production data, it was important to consider certain characteristics of the learners as described in the studies. We decided to combine data from children and adults, since the research has shown that the two groups manifest very similar acquisition orders (for children, Dulay & Burt, 1973; Dulay & Burt, 1974; Mace-Matluck, 1977; for adults, Bailey, Madden, & Krashen, 1974; Rosado, 1986). Similarly, we

were not concerned with the type of instruction or exposure to English received by the learners, since previous studies have suggested that type of instruction has no influence on the order of acquisition of grammatical functors (Pica, 1983; Perkins & Larsen-Freeman, 1975). However, we did have to consider the issue of length of residence (LOR) of the subjects in the study by Ball (1996). In most prior acquisition-order studies, the subjects were still in the process of actively acquiring English as an L2, either in ESL programs or in a naturalistic setting. Ball, however, was in part looking for critical period effects in her study and therefore tested a variety of subjects on the basis of their age of arrival (AOA) in the United States. Her data therefore may reflect "fossilization" of subjects who arrived as adults, as well as the near-native performance of those who arrived as young children. Because of the risk of a ceiling effect on performance on the functors, and because this group was so unlike all the others, the decision was made to exclude this study (with 102 subjects) from the meta-analysis.

The next consideration in selecting studies was the scoring or coding system used to calculate the accurate use of the functors. As described previously, Brown (1973) introduced the concept of suppliance in obligatory contexts (SOC): count the number of obligatory contexts in which a native speaker would have to use a functor and order the functors on the basis of how accurately they were supplied in those contexts.[6] Dulay and Burt (1973) expanded on this method by awarding "half-credit" to situations where a functor was required and an incorrect one was supplied. The SOC score is generally calculated as follows (Pica, 1983, p. 474):

$$\text{SOC} = \frac{\begin{matrix}n \text{ correct suppliance in} \\ \text{obligatory context}\end{matrix} \times 2 + \begin{matrix}n \text{ misformations in} \\ \text{obligatory context}\end{matrix} \times 1}{\text{total obligatory contexts} \times 2}$$

While scoring based on SOC does provide a method for quantitatively measuring what Brown calls "output-where-required" (1973, p. 255), it introduces the problem of the

Comparative Fallacy (Bley-Vroman, 1983). According to Bley-Vroman, "work on the linguistic description of learners' languages can be seriously hindered or sidetracked by a concern with the target language" (1983, p. 2). Measuring the accurate use of grammatical functors by an ESL student by comparing it to an ideal of the target language risks denying the internal logic of the student's interlanguage. What is "accurate" in the target language may have nothing to do with what is accurate in the student's own grammar at that point in time. As Bley-Vroman observes, "any study which classifies interlanguage (IL) data according to a target language (TL) scheme or depends on the notion of obligatory context or binary choice will likely fail to illuminate the structure of the IL" (1983, p. 15).

Another problem with SOC scoring is that it completely ignores the potential for overgeneralization and overuse of the functors (Lightbown, Spada, & Wallace, 1980; Pica, 1983). A student may appear to be completely accurate in the use of particular functors because s/he supplies them in all obligatory contexts, but may also use them inappropriately in other contexts. Pica (1983) compared SOC scoring with another scoring method, called Target-Like Use (TLU), which partially addresses this problem (although the dependency on obligatory or "native" use remains). TLU scoring is as follows (Pica, 1983, p. 474):

$$\text{TLU} = \frac{n \text{ correct suppliance in obligatory contexts}}{(n \text{ obligatory contexts}) + \begin{matrix}(n \text{ suppliance in} \\ \text{nonobligatory contexts})\end{matrix}}$$

Since most of the "natural order" studies relied on SOC scoring methodology, we had little choice in this study but to continue the practice. Therefore, in order to be included in this meta-analysis, the candidate studies had to evaluate the functors based on suppliance in obligatory context.

A related criterion for selection of studies for the "pool" involved the availability of actual scores. The group of

candidate studies comprises several types of material, namely journal articles, journal "research notes," book chapters, and PhD dissertations. Publication standards and practices vary widely over time and among publications. In many cases, notably Dulay and Burt (1973, 1974), only rank orders of the functors are given, with none of the accuracy percentages on which the rank orders are based. Not only can this be misleading, because the rank orders give no indication of the relative distance between the percentages for each functor, but it does not provide sufficient information for developing a meaningful pool of data. Of course, there is too little space in most sources to include raw data or individual scores, so we set the minimum criterion for inclusion in the meta-analysis as follows: the study must provide (or make derivable mathematically) the percentage of correct suppliance in obligatory context for each functor. On these grounds, a further five studies were eliminated (Dulay & Burt, 1973, 1974; Larsen-Freeman, 1975; Hakuta, 1976; Pak, 1987).

The final criterion for inclusion concerned the number of functors in common across the remaining 14 studies. Brown (1973) looked at 14 functors, and some combination of these is included in most studies of acquisition order, but unfortunately, the exact numbers and combinations vary from study to study. It is crucial to the statistics of the meta-analysis that each functor included be represented in most of the studies, such that there are few holes or gaps in the data to skew the results.[7] This created a regrettable trade-off between the need to maximize the data pool and the need to get comparable data. We found that there were six functors (present progressive *-ing*; plural *-s*; possessive *-'s*, articles *a*, *an*, *the*; 3rd person singular present *-s*; regular past *-ed*) largely in common across 12 of the studies (see Table 1). These 12 studies were retained for use in the analyses, and 2 (Kessler & Idar, 1979; O'Dowd, 1991) were not used at all because there were simply not enough functors in common with the other studies.

Table 1

Accuracy percentages for studies included in meta-analysis

Study Code	2	7	8	10	12	13	14	18	22	23	24	25
*N**=	73a	120c	80a	22a	45c	22a	66a	422c	18a	26c	24a	6m
-ing	84.38	67.33	93.75	97.00	78.67	84.00	88.67	97.87	96.33	99.04	85.58	91.00
plural	78.63	67.67	93.33	75.00	64.33	71.00	92.33	73.06	80.33	54.31	59.90	80.00
poss.	34.63	34.33	63.75	75.00	44.00	–	40.67	81.99	–	23.08	35.71	74.00
art.	78.38	62.67	85.83	82.00	76.67	69.00	81.33	73.55	89.66	96.08	87.66	89.00
reg past	–	20.67	71.25	61.00	50.67	64.00	64.67	53.93	51.00	–	61.71	64.00
3rd sing.	39.88	07.67	55.00	60.00	18.33	36.00	25.67	66.61	36.66	15.39	44.38	45.00

Code	Study	Status	Reason for Exclusion
1.	Andersen, 1977	Excluded	EFL
2.	Bailey et al., 1974		
3.	Ball, 1996	Excluded	Length-of-residence
4.	James Dean Brown, 1983	Excluded	Written data
5.	Dulay & Burt, 1973	Excluded	No scores available
6.	Dulay & Burt, 1974	Excluded	No scores available
7.	Fathman, 1975		
8.	Fuller, 1978		
9.	Hakuta, 1976	Excluded	No scores available
10.	Houck et al., 1978		

Table 1 (continued)

Code	Study	Status	Reason for Exclusion
11.	Kessler & Idar, 1979	Excluded	Too few functors
12.	Kjarsgaard, 1979		
13.	Krashen et al., 1977		
14.	Krashen et al., 1976		
15.	Larsen-Freeman, 1975	Excluded	No scores available
16.	Lightbown, 1983	Excluded	EFL
17.	Lightbown et al., 1980	Excluded	EFL
18.	Mace-Matluck, 1977		
19.	Makino, 1979	Excluded	EFL
20.	O'Dowd, 1991	Excluded	Too few functors
21.	Pak, 1987	Excluded	No scores available
22.	Pica, 1983		
23.	Riddle, 1993		
24.	Rosado, 1986		
25.	Rosansky, 1976		

*In further describing samples, a = adults, c = children, m = mixed adults and children, – = no data reported.

In this section we have described the criteria used to select studies for inclusion in this meta-analysis. From a total of 25 candidate studies, 12 qualified for use in the meta-analysis. The next section provides a detailed description of the proposed predictors of functor acquisition order.

Identifying and Quantifying Predictors

The literature contains many suggestions for possible determinants of functor acquisition order in ESL. Most of these do not appear in empirical tests, but have simply been mentioned in discussions of the phenomenon of "natural order" and its implications. For the purposes of this meta-analysis, these ideas were grouped into six potential predictors: perceptual salience, semantic complexity, morpho-phonological regularity, syntactic category, frequency in the input, and L1 transfer. For reasons detailed below, a combination of the first five of these factors is investigated in this study, while the sixth is excluded.

To test the degree to which this study's combination of five determinants predicts the "order of acquisition" of grammatical functors, it was necessary to develop a score for each functor on each determinant. For example, the present progressive *-ing* had to be scored on how perceptually salient, semantically complex, etc., it is. This is a challenging enterprise, since in most cases no general scales or measures for these features exist in the literature. The following discussion details how predictors were identified and how scores were developed for each predictor in this study (for a complete list of all scores, see Appendix D).

Perceptual Salience

Perceptual salience refers to how easy it is to hear or perceive a given structure. Dulay and Burt (1978) pointed out that "[p]erceptual salience is an input factor that has not as yet been precisely defined" (1978, p. 73), and that difficulty persists

today, although the term is used often. Brown clarifies his use of the term somewhat by breaking it down into "such variables as amount of phonetic substance, stress level, usual serial position in a sentence, and so on" (1973, p. 409) and points out that perceptual salience probably plays a role in the "telegraphic" speech production of a child acquiring English as L1. Already vague, the definition of salience can be further confused because salience can refer solely to the characteristics of the input itself, or to other factors that cause some parts of the input to *become* salient (Dulay & Burt, 1978). In this paper, only the former sense of the term is intended.

Several authors argue that perceptual salience may play a role in the order in which grammatical functors are acquired. Brown makes the important point that, at least for L1 acquisition, "*some* role for salience is guaranteed; the child will not learn what he cannot hear" (1973, p. 410). Slobin (1971) suggested "operating principles" in child L1 acquisition that make word-final morphemes particularly salient. Pye (1980) found that perceptual salience, "defined in terms of susceptibility to word and sentence stress and lack of disjuncture caused by a syllable boundary" (1980, p. 58), was the best predictor of children's order of acquisition of person markers in Quiche Mayan ($\rho = .59$, $p < .05$). Larsen-Freeman (1976) considered salience as a possible predictor (although she eventually rejected it in favor of frequency in the input). Gass and Selinker (1994) also posit some role for perceptual saliency. In all of these cases, the prediction is that the more perceptually salient a functor is, the earlier it will be acquired.

The score for perceptual salience is composed of three subfactors: the number of phones in the functor (phonetic substance), the presence/absence of a vowel in the surface form (syllabicity), and the total relative sonority of the functor. Two other possible subfactors, stressed/unstressed and serial position, were not considered here because all six of the functors in this study were unstressed, and, with the single exception of articles, all came at the end of the word. Scores for

the three relevant subfactors were converted to *z*-scores for the statistical analysis.

Number of phones. Assumption: The more phones in a functor, the more perceptually salient it should be.

The number of phones was calculated by adding up the number of phones in all the allomorphs of a given functor and dividing by the number of allomorphs. This averaging was needed because the studies do not differentiate among the allomorphs, and so there was no way of determining the absolute frequency of each allomorph in the data. So, for example, the [s, z, əz] alternations of the plural *-s* contain a total of four phones, divided by three allomorphs, yielding an average of 1.33 phones for that functor.

Syllabicity.[8] Assumption: Functors containing a vowel in the surface form should be more perceptually salient than those without a vowel.

The scoring for syllabicity was based on a value of zero (0) if a functor sequence never contains a vowel and one (1) if it always contains a vowel. Only two of the six functors in this study, present progressive *-ing* and the articles, contain vowels in all phonological environments. The other four have three allomorphs each, only one of which contains a vowel (for the *-s* functors the [əz] form and for the regular past *-ed* the [əd] form). Those four factors were scored as .33, that is, as containing a vowel one third of the time. Again, because use of allomorphs is determined by the learner's choice of content words, and no allomorph frequencies are available, this scoring represents an average for those functors with allomorphic variants.

Sonority. Assumption: Functors that are more sonorous should be more salient.

The sonority score for each functor was calculated by using the sonority hierarchy given by Laver (1994, p. 504). A numerical value was assigned to each level of the hierarchy (see Table 2). This value represents the number of points to be scored for each type of sound. The sonority score for a functor is a total of all the points for all of the phones. For example, for present

Table 2

Sonority hierarchy with point values

Range	Points	Description
Most sonorous	9	low vowels
	8	mid vowels
	7	high vowels
	6	glides
	5	liquids
	4	nasals
	3	fricatives
	2	affricates
Least sonorous	1	stops

progressive *-ing*, the high front vowel gets 7 points and the nasal gets 4 points, for a total score of 11. For those functors with allomorphs, the scores for each allomorph were again averaged to yield one score for the functor.

Semantic Complexity

Semantic complexity is a measure of how many meanings are expressed by a particular form. For example, the plural *-s* expresses number, whereas the third person singular *-s* expresses person, number, and present tense. Today, as for Brown in 1973, "There is no general theory of semantic complexity that makes it possible to assign complexity values" (1973, p. 369). However, Brown was able to establish a complexity hierarchy, which predicted that forms with more meanings should be more difficult to learn and later acquired than forms with fewer meanings. Semantic complexity is also mentioned as a possible determinant by Larsen-Freeman (1976), Andersen (1978), Krashen, Madden, and Bailey (1975), Pye (1980), and Gass and Selinker (1994).

There is clearly a difficulty in using standard glosses to count meanings for each functor. For example, the articles are

defined in Brown (1973) as expressing a particular degree of "definiteness," as in "*A* (indefinite) beaver builds a dam with mud and sticks," and "*The* (definite) beaver in my stream is building a dam." However, in English, it is also possible to use the so-called definite article when referring to generics, as in "*The* beaver has a flat tail and uses it to build dams." Arguably, the generic is neither definite nor indefinite, and could thus constitute a second meaning for this functor. The meanings assigned to the other functors may be similarly open to debate.[9] With this reservation, in this analysis we have followed Brown's (1973) assignment of meanings to the functors (cf. Lehmann, 1978, for a formalization of semantic complexity as meaning specificity).

Each meaning was assigned one point, and the points were totaled to create a complexity score. It is important to note that this method represents only a comparison of cumulative complexity, but says nothing about which of two functors with the same complexity score should be acquired first. Additionally, since a higher score is predicted to correlate with later acquisition, a *negative* correlation for semantic complexity is expected.

Morphophonological Regularity

Morphophonological regularity refers to the degree to which the functors are (or are not) affected by their phonological environment. Conditioned phonological variation (allomorphy) and contractibility (where conditioning factors are syntactic) are the main ways in which the functors may be affected, although redundancy is sometimes considered as well. This predictor is the least well-defined in the literature; we included it in this study mainly owing to criticisms by Cook (1993) and others that most of the functor-order studies failed to take allomorphy into account. The prediction is that the more phonologically regular a functor is, the earlier it should be acquired.

The score for morphophonological regularity comprises two subfactors: the number of phonological alternations, and homophony with other grammatical functors.[10] Here, too, a high score means that the functor should be acquired later; therefore, a negative correlation is expected.

Number of phonological alternations. Assumption: Functors with more alternations should be acquired later.

This score is a simple tally of the number of alternations found for each functor. The most commonly cited in the literature are the [s, z, əz] allomorphs of plural *-s*, possessive *-'s*, and third person singular *-s*, and the [t, d, əd] allomorphs of the regular past *-ed*. However, for the articles, *a* also alternates with *an*, and [ðə] alternates with [ði] before vowels.

Homophony with other grammatical functors. Assumption: Functors that are homophonous with other grammatical functors should be acquired later than functors that are not homophonous with other functors.

This score is based on a scale where 1 = no homophony with other grammatical functors and 2 = homophony with other functors.[11] No attempt can be made to rank *relative* degrees of homophony here, because for the six functors in question, three are not homophonous with any other functors (present progressive *-ing*, regular past *-ed*, articles), and the other three (plural *-s*, possessive *-'s*, third singular *-s*) are all homophonous with each other as well as with the contracted allomorphs of auxiliary (BE) and the copula.

Other subfactors. Another possible subfactor for this determinant would be redundancy; that is, to what extent the use of the functor is required to convey meaning. Many grammatical functors are redundant in at least some situations, such as plural marking on the noun in the phrase 'three apples,' when the noun phrase includes a quantifier that signals plurality (Kalnitz, 1978). For this meta-analysis, however, it is impossible to determine where redundancy did or did not exist in the learners' speech, because the original

utterances (and the extralinguistic context in which they were made) are not available in the published studies.

Syntactic Category

Syntactic category, for purposes of this analysis, refers to the characteristics of each functor from the point of view of Functional Category theory. This represents a great divergence from the ways in which syntax has been considered in the past as a predictor of acquisition. Brown (1973) claimed that "grammatical complexity" was a significant determinant of the order of acquisition, and he based his values for grammatical complexity on a model of syntactic representations by Jacobs and Rosenbaum in their 1968 book *English Transformational Grammar*.

Syntactic theory has evolved considerably since that time, and there is no longer support for the notion that there is a direct correspondence between number of syntactic transformations and degree of difficulty or ease of acquisition. However, Krashen, Madden, and Bailey (1975) reported interesting results from examining (NP)-related and (V)-related[12] morphemes separately, finding that most of the differences between L1 and L2 orders for the functors were due to differences in V-related functors, while NP-related items were acquired in very similar orders. This observation has been revised again by Zobl and Liceras (1994; cf. also Zobl, 1995), who propose category membership as a way of accounting for functor acquisition order. They observed that by grouping the functors by syntactic category (lexical/functional), and then further subdividing them according to the free/bound distinction, a pattern emerged in which lexical items appear to be acquired before functional items, and within each of these groups, free morphemes are acquired before bound ones.

We developed a scoring method for this determinant based on that of Zobl and Liceras (1994).[13] Points were assigned to each level of the hierarchy of acquisition they propose (see

Table 3). The highest point score (4 points) was assigned to the forms predicted to be acquired earliest, and the lowest score (1 point) was assigned to the forms which should be acquired latest[14] (see Appendix D for individual functor scores).

Frequency

Frequency in the input, the second most popular of the suggested causes of the L2 functor acquisition order (after L1 transfer), refers to the number of times a given structure occurs in speech addressed to the learner. The prediction is a straightforward one: The more frequent a grammatical item is in the input to the learner, the more easily and quickly that item should be acquired. Frequency is mentioned in nearly every discussion of possible determinants of acquisition order in L2, despite the fact that Brown (1973, p. 368) found "no clear evidence at all that parental frequencies influence the order of development of the forms" in L1. Larsen-Freeman and Long (1991, p. 91) even claim that, of all the determinants of acquisition order, "only input frequency has much empirical support to date." That the issue of frequency persists in the L2 literature (even though it is of less interest in L1) is in part due to Larsen-Freeman, who found that "morpheme frequency of occurrence in native-speaker speech is the principal determinant for the oral production morpheme order of second language learners" (1976, p. 132). Larsen-Freeman clearly

Table 3

Points assigned to syntactic categories

Lexical	
Free	4 points
Bound	3 points
Functional	
Free	2 points
Bound	1 point

stated, however, that this conclusion was tentative (the primary problem was that she was correlating Brown's original parental frequencies with the production data from 24 adult ESL students). Dulay and Burt (1978) argued against frequency as an explanation on the grounds that common grammatical functors occur more frequently in English than even the most common lexical items, and yet functors are acquired late compared with content words. This argument does not take into account, however, that, while lexemes are required earlier than grammatical morphemes, in spite of their respective frequencies, frequency may still be a predictor of acquisition order within the set of grammatical morphemes.

Finding a good operationalized measure of input frequency is not easy. On the one hand, Brown's 1973 parent-to-L1-learner frequencies cannot be completely representative of the topics/registers that would be used with adult ESL learners or the input (parental or otherwise) that would be available to children learning an L2. The studies included in this meta-analysis were conducted over a period of 25 years in different parts of the country, using subjects in naturalistic and classroom settings. On the other hand, however, the morphemes under scrutiny are such a basic part of the language that not too much variation can be expected from one input situation to another. Therefore we decided to use Brown's (1973, p. 358) frequency data, pooling the frequencies for three sets of parents. This operationalization probably implies a slight underestimation of the frequency effect, but it seems preferable to excluding frequency altogether.

L1 Transfer

The last of the six potential predictors is L1 transfer or interference. The role played by L1 transfer in L2 acquisition has been a subject of some controversy over the past 25 years. Dulay and Burt (1973) reported that only 3% of errors made by children in their study were due to L1 interference. Andersen

(1977) posited a major role for L1 interference in the acquisition of articles and the possessive *-'s* by native Spanish speakers, and the next year he concluded (Anderson, 1978, p. 267) that L1 transfer "clearly is a factor that must be taken into consideration as one of the factors that could interact with morpheme acquisition and accuracy orders." In recent years, the concept of L1 transfer has become increasingly sophisticated and diversified, and it is generally agreed that L1 plays an important role in L2, but often in more abstract ways than earlier conceptualizations would suggest (see especially Gass & Selinker, 1994).

In this meta-analysis, data are included from native speakers of 28 languages belonging to diverse language families (see Appendix E). The individual studies do not give data on error types broken down by native language. This is not surprising, since L1 transfer is not the research question those studies were designed to investigate; however, it does make it impossible to make comparisons between L1s or across studies. Therefore, L1 transfer is not tested in this analysis as a possible determinant of L2 acquisition order. Fortunately, the possibility of skewed results due to L1 transfer is significantly reduced by the number and diversity of L1s represented in the pooled data.

In summary, the combination of potential determinants tested in this study includes perceptual salience, semantic complexity, morphophonological regularity, syntactic category, and frequency. L1 transfer was excluded from this analysis. This section has described prior discussion of these predictors in the literature and the manner in which they were operationalized in this meta-analysis. The next section describes the multiple regression analyses that were performed to test the predictive power of this set of five proposed determinants.

Multiple Regression Analysis

A multiple regression analysis was conducted to determine how much of the total variance in ESL functor order can be

accounted for by a combination of five factors: perceptual salience, semantic complexity, morphophonological regularity, syntactic category, and frequency

Because this is a meta-analysis, the "subjects" for the study are actually group mean performance percentages (accuracy scores), with one score from each study for each functor.[15] The analysis used 68 data points (6 functors $\times$ 12 studies, with 4 cases of missing data, that is, lack of information for a given functor in a given study). In other words, each 'case' in the analysis corresponds to a given functor in a given study, with the percentage correct for that functor in that study, and the values for the five predictors for the functor.

To take into account the fact that in the original studies there were different numbers of individual subjects, each primary study was weighted by number of subjects. These weighted percent-correct scores became the dependent variable against which the combination of predictors was tested.

Results

Table 4 provides a correlation matrix for the one dependent and the five independent variables. The table shows that phonological salience and syntactic category have the highest correlation with percentage correct, while the predictive power of frequency, semantic complexity, and morphophonological regularity is somewhat lower. These results are to be interpreted with caution, however, because the table also shows a very high degree of intercorrelation among the predictors, especially between phonological salience, on the one hand, and frequency, morphophonological regularity, and syntactic category, on the other. The high degree of intercorrelation, combined with the unavoidable difference in reliability between the different predictors (perfect, for instance, for syntactic category, and clearly subject to measurement error for frequency), means that a high degree of predictive validity of a given variable could be largely due to its correlation with

Table 4

Intercorrelations for all independent and dependent variables

Variable	1	2	3	4	5	6
1. Frequency	—	.62***	−.44***	.16	.27*	.44***
2. Phonological salience		—	−.12	−.59***	.57***	.63***
3. Semantic complexity			—	−.26*	−.36**	−.41**
4. Morphophonological regularity				—	−.37**	−.41**
5. Syntactic category					—	.68***
6. Percentage correct						—

Note. $n = 68$; *$p < .05$; **$p < .01$; ***$p < .001$.

another variable and to its high reliability. Determining which of the five variables is the most important 'causal factor,' therefore, is very difficult.

The main goal of our analysis, however, was to determine to what extent the five predictors together would predict percentage of accuracy. A multiple regression analysis was performed to test the effect of the combination of the five predictors (perceptual salience, semantic complexity, morphophonological regularity, syntactic category, and frequency).

The results of the multiple regression analysis were checked against the research hypotheses. As shown in Table 5, the value for R^2 obtained in all three analyses was very high ($R = .84$; $R^2 = .71$; $p < .001$); therefore, the hypothesis that a large portion of the variance is explained by these five factors is supported.

Weighted Versus Unweighted Accuracy Scores

For the purposes of the multiple regression analysis, the accuracy scores for each functor from each study were weighted according to the number of subjects in each study.[16] This was

Table 5

Summary of multiple regression analysis for variables predicting percentage of accuracy

Variable	*B*	*SE B*	β
Frequency	−0.04	0.03	−0.23
Phonological salience	15.20	6.94	0.64
Semantic complexity	−13.90	3.57	−0.36
Morphophonological regularity	0.10	7.51	0.00
Syntactic category	8.18	3.14	0.30

Note. $n = 68$; $R = .84$; $R^2 = .71$.

done to balance the extremes in the primary studies, where the number of subjects ranged from 6 (Rosansky, 1976) to 422 (Mace-Matluck, 1979). For comparison purposes, the analysis was repeated, using exactly the same inputs but without weighted accuracy scores. The values for R (.80) and R^2 (.64) were somewhat lower than those obtained using the weighted accuracy scores, but they still show that a significant portion of the variance is attributable to the five predictors combined in this study. Clearly, weighting the scores in the first analysis helped to improve the reliability of the accuracy scores, and thereby the predictive validity of the functor characteristics, but not by much.

Discussion

The results of this meta-analysis indicate that the combination of perceptual salience, semantic complexity, morphophonological regularity, syntactic category, and frequency does account for a very large portion of the total variance in the accuracy scores for grammatical functors. These results obtain most clearly when weighted accuracy scores are used, but are also found with unweighted scores. The amount of variance accounted for, while high, still leaves room for other contributing factors, such as L1 transfer, which was not included in this study.

If these variables could be appropriately operationalized and tested in combination with the five factors from this study, the new combination of factors would probably account for an even larger percentage of the total variance.

At this point, the question inevitably arises of whether the five factors examined have something in common or whether they constitute a completely heterogeneous set. We will argue the former, in the sense that perceptual salience, morphophonological regularity, semantic complexity, syntactic category (as defined here), and frequency all constitute aspects of salience in a broad sense of the word.

The first factor, perceptual salience, is, of course, the clearest case. It fits even the narrow definition of Ravid (1995, p. 117): "[S]aliency is the property of a structure that is perceptually distinct from its environment." Number of phones, syllabicity, and sonority all contribute to this factor.

Morphophonological regularity leads to a more abstract form of salience, i.e., salience of the form–meaning relationship. Both allomorphy and homophony obscure the one-to-one relationship between form and meaning that would make this link easier to acquire. "Homonymy within a set of functors may be similar to fusion in that learners will have a hard time discovering the separate functions of all the numbers of the set; this may result in omission or in inappropriate production" (Peters, 1995, p. 480).

Semantic complexity adds further to the difficulty of establishing a one-to-one relationship between form and meaning. Third person *-s* conflates person, number, tense, and aspect. Only when these four independent components of meaning happen to coincide does this morpheme (or more precisely, morph) make its appearance.

Syntactic category as defined here, following Zobl and Liceras (1994), also reflects salience. Everything else being the same, free morphemes, because of their distributional variation, are going to be more noticeable than bound ones, and lexical morphemes are, of course, more salient than functional ones.

Finally, the frequency of a morpheme obviously contributes to its salience.

Thus, the five factors that have been shown in this study to account for a large percentage of the variance in order of acquisition constitute phonological, morphological, syntactic, semantic, and numerical aspects of salience. It is possible, therefore, that just one variable, salience, is the ultimate predictor of the order of acquisition. It is also possible, as Richard Schmidt suggests (personal communication, March 11, 2000), that two aspects of salience (salience of form and salience of meaning) operate independently. Alternatively, one may think along more connectionist lines and see the various aspects of the functors as multiple cues in a competition model (cf. MacWhinney, 1997). Whichever of these interpretations one may prefer, it is clear that a number of objective characteristics of the morphemes themselves, in all likelihood all related to salience, predict order of acquisition with a high degree of success. Therefore, no appeal to any innate blueprints or specific syntactic models is required to explain order of acquisition.

Conclusions and Implications

The results of this meta-analysis suggest that a considerable portion of the order of acquisition of grammatical functors by ESL learners (as reflected by percentage of functors supplied correctly in obligatory contexts) can be predicted by the combination of five factors: perceptual salience, semantic complexity, morphophonological regularity, syntactic category, and frequency.

A small amount of previous research (e.g., Larsen-Freeman, 1976; Pienemann, 1998; Zobl & Liceras, 1994; Zobl, 1995) has attempted to show that some single factor was responsible for the "natural order." None of the attempts satisfactorily resolved the basic question: Why do ESL learners of different ages, with different types of exposure to English and different L1s, appear to acquire certain grammatical functors in very nearly the same

order? Although the question remained unanswered, SLA research moved on to other issues. Even Krashen, who used the "natural order" as one of the underpinnings for his Monitor Model (Krashen, 1977), did not try to account for why the order occurred.

The present study has shown that several of the previously considered factors together do account for a very large percentage of the variance. Moreover, we have argued that these five factors are not a completely heterogeneous set, but can all be seen as aspects of salience in a broad sense of the word, and that this salience at various levels (phonological, morphological, syntactic, semantic, and numerical) facilitates the process of induction of grammatical structure from elements of the input.

As for a computer program or a cooking pot, the quality of a meta-analysis is affected by the quality of what goes into it. On the one hand, merging data from multiple studies limits the potential impact of methodological problems, scoring errors and other inconsistencies. On the other hand, relying on existing studies also limits the availability of information to what is included in the published study or dissertation. This can be particularly frustrating when key studies which are the cornerstone of a given area of research cannot be used in a meta-analysis because they do not include original score data (as was the case with five candidate studies in this meta-analysis).

The "morpheme order" studies have been the target of much criticism themselves.[17] Critics have claimed that the group of functors studied is too heterogeneous, or alternatively that grammatical functors represent too small and too trivial a part of English. The scoring system has been questioned, since it does not take into account oversuppliance of the functors in non-obligatory contexts. Additionally, Long and Sato observed that "[a]ccurate suppliance of a target language form does not necessarily mean the learner knows its function(s)" (1984, p. 260). Bley-Vroman (1983) questioned the validity of relying on any measurement involving obligatory contexts as defined by some ideal of the target

language. What this means for this meta-analysis is that we have inherited, and reanalyzed, a body of work known to have certain problems.

The combination of factors used in this study does not account for all the variance in accuracy scores. More research would be needed to test the importance of other predictors, including L1 transfer. It would also be interesting to try to tease apart the individual determinants' effects on acquisition in order to establish with more certainly whether the combination of factors accounts for the order through a cumulative effect or through the interaction of the factors. Additionally, this study has investigated ESL only, so the results are not generalizable to other languages. Research on acquisition orders and possible predictors in other languages is needed to determine to what extent, if any, these factors make universal predictions about what will be acquired early or late in a given language. Data from other languages would also reduce or eliminate the confounding among the predictors. The various predictors may not intercorrelate in other languages, or at least the intercorrelations may show different patterns, which would make it easier to determine which characteristic(s) of the functors show consistently high predictive validity across languages, and therefore would appear to be the ultimate causal factor.

If these findings on the predictors of functor acquisition order could be confirmed and refined through future research, they would have important implications for second language teaching and syllabus design. Recognizing the causes underlying the order, teachers could make the predictors work for them and could potentially increase the rate of acquisition by presenting material on functors in a way that capitalizes on these causes. If the order of acquisition is indeed largely a function of salience, then a primary task for teachers is to make the functors more salient in an attempt to bring them to the learner's consciousness, especially for adults. As Ellis and Laporte state (1997, p. 64), such an empiricist view "does not

deny the question of effect of instruction on route of acquisition as an important empirical issue, but it does weaken the logical role of any null answer in either denying any involvement of consciousness in language acquisition (L1 or L2) or implying innate language acquisition devices." In more pedagogical terms, "For forms that are frequent in the input and yet still seem to lack salience for learners, it may be that other means are required to induce learners to notice" (Doughty & Williams, 1998, p. 220), e.g., typographical enhancement (Jourdenais, Ota, Stauffer, Boyson, & Doughty, 1995; Leeman, Arteagoitia, Fridman, & Doughty, 1995; White, 1998).

Finally, at the level of SLA research methodology, this study may be useful as a first attempt to use meta-analysis to investigate a question that has proved resistant to other types of research. We have tried to provide a detailed description of the procedures involved and the difficulties encountered, so that it may serve as a point of reference for further applications of meta-analysis to language acquisition research.

Revised version accepted 23 March 2000

Notes

[1]The grammatical units considered here (including noun and verb inflections, articles, copulas, and auxiliaries) have traditionally been referred to in the literature as grammatical morphemes, following Brown (1973); however, since the term morpheme is not technically applicable to distinctions such as past regular vs. past irregular, we will use the less theory-burdened term "functor."

[2]Children under age 5 are excepted here, since they are probably acquiring English simultaneously with another language, rather than sequentially, and therefore are probably not ESL learners in the usual sense of the term.

[3]This phenomenon is also referred to in the literature as an accuracy order (Larsen-Freeman, 1976) or performance order (Andersen, 1977). We will use "order of acquisition" here, in keeping with the majority of primary studies pooled in this meta-analysis.

[4]In Dulay and Burt (1974) and later replications (e.g., Pica, 1983), this point system was revised to assign 2 points for each correct functor, 1 point for each misformed functor, and 0 points when no functor was

supplied. The total number of obligatory contexts in the denominator was then multiplied by 2 before the accuracy ratio was calculated.

[5]Larsen-Freeman (1975) found a different order for her written task, which Krashen (1977) attributed to the intrusion of that conscious part of the grammar that he calls the Monitor.

[6]For our purposes, Brown's arbitrary decision to use 90% correct suppliance in obligatory contexts as the point of actual acquisition is irrelevant. Since he ordered all of the functors he studied, not just the ones that achieved his criterion level, his results still reflect accuracy of use in obligatory contexts for all the functors.

[7]Special care must be taken to detect situations in which a functor is initially included in a study, but no obligatory contexts are reported for it, either because the instrument did not successfully elicit the structure or it did not come up in spontaneous speech. This was the case for possessives in O'Dowd's study (1991), where a 0% accuracy rate on the functor is given because no obligatory contexts for it ever came up. This kind of data cannot be included in a meta-analysis, because it can erroneously suggest that the functor was attempted, but was used with 0% accuracy. For all such cases, we have simply recorded that no results were available for the affected functor(s).

[8]We use the term here to refer to the presence or absence of a vowel in the surface form of the functor. This differs from the phonological definition of the syllable, but is consistent with the SLA literature on perceptual salience.

[9]The functor commonly referred to as present progressive *-ing* presents a related difficulty. This functor should be called simply "progressive," since the tense is determined by the form of BE used by the learner. However, due to the manner in which the data were tabulated in the original studies (where all occasions of progressive *-ing* included present tense BE), this functor is assigned 1 point for present tense and 1 point for ongoing action, for a total score of 2 points on the semantic complexity scale.

[10]Note that homophony with *lexical* items is not considered here, as this has not been shown to have an effect on the acquisition of *grammatical* structures.

[11]In order not to limit consideration of homophony to the six functors in this study, all 14 of Brown's (1973) grammatical functors (see Appendix A) were considered here. Of course, Brown's group does not represent an exhaustive list of the grammatical functors of English, and it is possible that more occasions of homophony among functors exist than are described here.

[12]The NP vs. V distinction is taken directly from Krashen, Madden, and Bailey (1975). Although it lacks the symmetry of NP vs. VP, the authors argued that VP is not a unit of perception, while NP is. This argument hints at the complementizer phase/inflection phase distinction made 20 years later by Zobl and Liceras (1994).

[13]Those authors note that "this framework does not make specific predictions about order of acquisition" (Zobl & Liceras, 1994, p. 175), so the adaptation of their work to this scoring hierarchy is strictly ours.

[14]Zobl and Liceras categorize plural *-s* and present progressive *-ing* as "lexical inflections" (1994, p. 172), or "lexical affixes" (1994, p. 173), and these two functors are therefore assigned three points each in our scale. The authors draw a distinction between these forms and "syntactic affixes" (1994, p. 173) such as third person singular *-s*. While this distinction is crucial to their argument, we have failed to find adequate justification for their position.

[15]Scores on the predictor (independent) variables are "clustered"; that is, the same scores for a given predictor are used repeatedly in the multiple regression equation. This is not a problem when scores on the independent variables are clustered. There would be a problem for the analysis if the criterion (dependent) variable did not represent independent observations, but this is not the case. Each accuracy score is specific to one functor within one study. See, e.g., Pedhazur (1982, p. 33). A multi-level analysis would be preferable from certain points of view, but would required a larger *n* (in this case, number of studies).

[16]The studies were also weighted by the amount of variance in the dependent variable (method of weighted least squares; see Neter, Wasserman, and Kutner, 1990). In other words, the weight in the SPSS analysis was the sample size divided by the variance.

[17]For a review of the debate, cf. *inter alia* Hatch (1978), Long and Sato (1984), Ellis (1986), and Cook (1993).

References

Andersen, R. W. (1977). The impoverished state of cross-sectional morpheme acquisition/accuracy methodology (or: The leftovers are more nourishing than the main course). *Working Papers on Bilingualism*, *14*, 47–82.

Andersen, R. W. (1978). An implicational model for second language research. *Language Learning*, *28*, 221–282.

Anderson, J. I. (1978). Order of difficulty in adult second language acquisition. In William C. Ritchie (Ed.), *Second language acquisition research: Issues and implications* (pp. 91–108). New York: Academic Press.

Bailey, N., Madden, C., & Krashen, S. D. (1974). Is there a "natural sequence" in adult second language learning? *Language Learning*, *24*, 235–243.

Ball, J. (1996). *Age and natural order in second language acquisition*. Unpublished doctoral dissertation, University of Rochester, Rochester, NY.

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning*, *33*, 1–17.

Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart & Winston.

Brown, J. D. (1983). An exploration of morpheme-group interactions. In K. M. Bailey, M. H. Long, and S. Peck (Eds.), *Second language acquisition studies* (pp. 25–40). Rowley, MA: Newbury House.

Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.

Burt, M. K., & Dulay, H. C. (1980). On acquisition orders. In S. W. Felix (Ed.), *Second language development: Trends and issues* (pp. 265–327). Tübingen, Germany: Gunter Narr.

Cook, V. (1993). *Linguistics and second language acquisition*. New York: St. Martin's Press.

Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

DeKeyser, R. M., Alfi-Shabtay, I., Ravid, D., & Shi, M. (2005). The role of salience in the acquisition of Hebrew as a second language. Manuscript in preparation.

DeKeyser, R. M., Ravid, D., & Alfi-Shabtay, I. (2005). Cross-linguistic evidence for the nature of age effects in second language acquisition. Manuscript in preparation.

de Villiers, J., & de Villiers, P. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research*, *2*, 267–278.

Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty and J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197–261). New York: Cambridge University Press.

Driscoll, M. P. (1984). Alternative paradigms for research in instructional systems. *Journal of Instructional Development*, *7*, 2–5.

Dulay, H. C., & Burt, M. K. (1973). Should we teach children syntax? *Language Learning*, *23*, 245–258.

Dulay, H. C., & Burt, M. K. (1974). Natural sequences in child second language acquisition. *Language Learning*, *24*, 37–53.

Dulay, H. C., & Burt, M. K. (1978). Some remarks on creativity in language acquisition. In W. C. Ritchie (Ed.), *Second language acquisition research: Issues and implications* (pp. 65–89). New York: Academic Press.

Dulay, H. C., Burt, M. K., & Krashen, S. D. (1982). *Language two*. New York: Oxford University Press.

Ellis, N., & Laporte, N. (1997). Contexts of acquisition: Effects of formal instruction and naturalistic exposure on second language acquisition. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 53–83). Mahwah, NJ: Erlbaum.

Ellis, R. (1986). *Understanding second language acquisition*. Oxford, England: Oxford University Press.

Ellis, R. (1994). *The study of second language acquisition*. Oxford, England: Oxford University Press.

Fathman, A. (1975). Language background, age and the order of acquisition of English structures. In M. K. Burt and H. C. Dulay (Eds.), *On*

TESOL'75: New directions in second language learning, teaching and bilingual education (pp. 33–43). Washington, DC: TESOL.

Fuller, J. K. (1978). *An investigation of natural and monitored morpheme difficulty orders by non-native adult students of English*. Unpublished doctoral dissertation, Florida State University, Tallahassee, FL.

Gass, S. M., & Selinker, L. (1994). *Second language acquisition: An introductory course*. Hillsdale, NJ: Erlbaum.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Hakuta, K. (1976). A case study of a Japanese child learning English as a second language. *Language Learning*, *26*, 321–351.

Hatch, E. (1978). Acquisition of syntax in a second language. In J. C. Richards (Ed.), *Understanding second and foreign language learning: Issues and approaches* (pp. 34–69). Rowley, MA: Newbury House.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Houck, N., Robertson, J., & Krashen, S. (1978). On the domain of the conscious grammar: Morpheme orders for corrected and uncorrected ESL student transcriptions. *TESOL Quarterly*, *12*, 335–339.

Jacobs, R., & Rosenbaum, P. (1968). *English transformational grammar*. Waltham, MA: Blaisdell.

Jourdenais, R., Ota, M., Stauffer, S., Boyson, B., & Doughty, C. (1995). Does textual enhancement promote noticing? A think-aloud protocol analysis. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 183–216). Honolulu: University of Hawai'i Press.

Kalnitz, J. (1978). *Universals in second language acquisition: Focus on English*. Unpublished master's thesis, Northeastern Illinois University, Chicago.

Kaplan, R. M., & Bresnan, J. (1982). Lexical-functional grammar: A formal system for grammatical representation. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 173–281). Cambridge, MA: MIT Press.

Kessler, C., & Idar, I. (1979). Acquisition of English by a Vietnamese mother and child. *Working Papers on Bilingualism*, *18*, 66–79.

Kjarsgaard, M. M. (1979). *The order of English morpheme category acquisition by Vietnamese children*. Unpublished doctoral dissertation, Arizona State University, Phoenix.

Krashen, S. (1977). The Monitor Model for adult second language performance. In M. Burt, H. Dulay, and M. Finocchiaro (Eds.), *Viewpoints on English as a second language* (pp. 152–161). New York: Regents.

Krashen, S., Houck, N., Giunchi, P., Bode, S., Birnbaum, R., & Strei, G. (1977). Difficulty order for grammatical morphemes for adult second language performers using free speech. *TESOL Quarterly*, *11*, 338–341.

Krashen, S. D., Madden, C., & Bailey, N. (1975). Theoretical aspects of grammatical sequencing. In M. K. Burt and H. C. Dulay (Eds.), *On TESOL'75: New directions in second language learning, teaching and bilingual education* (pp. 44–54). Washington, DC: TESOL.

Krashen, S. D., Sferlazza, V., Feldman, L., & Fathman, A. K. (1976). Adult performance on the SLOPE test: More evidence for a natural sequence in adult second language acquisition. *Language Learning*, *26*, 145–151.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.

Larsen-Freeman, D. E. (1975). The acquisition of grammatical morphemes by adult ESL students. *TESOL Quarterly*, *9*, 409–419.

Larsen-Freeman, D. E. (1976). An explanation for the morpheme acquisition order of second language learners. *Language Learning*, *26*, 125–134.

Larsen-Freeman, D. E., & Long, M. H. (1991). *An introduction to second language acquisition research*. New York: Longman.

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

Leeman, J., Arteagoitia, I., Fridman, B., & Doughty, C. (1995). Integrating attention to form with meaning: Focus on form in content-based Spanish instruction. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 217–258). Honolulu: University of Hawai'i Press.

Lehmann, C. (1978). On measuring semantic complexity: A contribution to a rapprochement of semantics and statistical linguistics. *Georgetown University Papers on Languages*, *14*, 83–120.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lightbown, P. (1983). Exploring relationships between developmental and instructional sequences in L2 acquisition. In H. Seliger and M. H. Long (Eds.), *Classroom oriented research in second language acquisition* (pp. 217–243). Rowley, MA: Newbury House.

Lightbown, P., Spada, N., & Wallace, R. (1980). Some effects of instruction on child and adolescent ESL learners. In R. C. Scarcella and S. D. Krashen (Eds.), *Research in second language acquisition* (pp. 162–172). Rowley, MA: Newbury House.

Long, M. H., & Sato, C. J. (1984). Methodological issues in interlanguage studies: An interactionist perspective. In A. Davies, C. Criper, and A. P. R. Howatt (Eds.), *Interlanguage* (pp. 253–279). Edinburgh, UK: Edinburgh University Press.

Mace-Matluck, B. J. (1977). *The order of acquisition of certain oral English structures by native-speaking children of Spanish, Cantonese, Tagalog, and Ilokano learning English as a second language between the ages of five and ten*. Unpublished doctoral dissertation, University of Texas at Austin.

Mace-Matluck, B. J. (1979). The order of acquisition of English structures by Spanish speaking children: Some possible determinants. In R. W. Andersen (Ed.), *The acquisition and use of Spanish and English as first and second languages* (pp. 75–89). Washington, DC: TESOL.

MacWhinney, B. (1997). Second language acquisition and the competition model. In A. M. B. de Groot and J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 113–142). Mahwah, NJ: Erlbaum.

Makino, T. (1979). *English morpheme acquisition order of Japanese secondary school students*. Unpublished doctoral dissertation, University of New Mexico, Albuquerque.

Masgoret, A.-M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning*, *53*, 123–163.

Neter, J., Wasserman, W., & Kutner, M. H. (1990). *Applied linear statistical models: Regression, analysis of variance, and experimental designs* (3rd ed.). Homewood, IL: Irwin.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.

Norris, J. M., & Ortega, L. (Eds.). (in press). *Synthesizing research on language learning and teaching*. Amsterdam and Philadelphia: Benjamins.

O'Dowd, E. (1991, April). *Pragmatics and morphosyntactic acquisition: A case study*. Paper presented at the fifth annual meeting of the International Conference on Pragmatics and Language Learning, Champaign, IL.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24*, 492–518.

Pak, Y. (1987). *Age differences in morpheme acquisition among Korean ESL learners: Acquisition order and acquisition rate*. Unpublished doctoral dissertation, University of Texas at Austin.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and prediction*. New York: CBS College Publishing.

Perkins, K., & Larsen-Freeman, D. (1975). The effect of formal language instruction on the order of morpheme acquisition. *Language Learning*, *25*, 237–243.

Peters, A. M. (1995). Strategies in the acquisition of syntax. In P. Fletcher and B. MacWhinney (Eds.), *The handbook of child language*. Oxford, England: Blackwell.

Pica, T. (1983). Adult acquisition of English as a second language under different conditions of exposure. *Language Learning*, *33*, 465–497.

Pienemann, M. (1998). *Language processing and second language development*. Amsterdam, The Netherlands: Benjamins.

Pienemann, M., & Johnston, M. (1987). Factors influencing the development of language proficiency. In D. Nunan (Ed.), *Applying second language acquisition research* (pp. 45–141). Adelaide, South Australia, Australia: National Curriculum Resource Centre.

Pye, C. (1980). The acquisition of person markers in Quiche Mayan. *Papers and Reports on Child Language Development*, *19*, 53–59.

Ravid, D. D. (1995). *Language change in child and adult Hebrew*. New York: Oxford University Press.

Riddle, P. J. W. (1993). *The relationship of the bilingual syntax measure to the English morpheme development measure*. Unpublished doctoral dissertation, Temple University, Philadelphia.

Rosado, L. A. (1986). *The role of the Monitor in the acquisition sequence of twelve English morphemes by adult English as a second language learners from four different language groups*. Unpublished doctoral dissertation, Texas A & I University, Kingsville, TX.

Rosansky, E. J. (1976). Methods and morphemes in second language acquisition research. *Language Learning*, *26*, 409–425.

Slobin, D. I. (1971). Data for the symposium. In D. I. Slobin (Ed.), *The ontogenesis of grammar* (pp. 3–14). New York: Academic Press.

White, J. (1998). Getting the learners' attention: A typographical input enhancement study. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 85–113). New York: Cambridge University Press.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage.

Zobl, H. (1995). Converging evidence for the "acquisition-learning" distinction. *Applied Linguistics*, *16*, 35–56.

Zobl, H., & Liceras, J. (1994). Functional categories and acquisition orders. *Language Learning*, *44*, 159–180.

Appendix A

Rank orders of accuracy scores for six functors across the twelve studies pooled in this meta-analysis (scores from Dulay & Burt 1973 and Dulay & Burt 1974 included for comparison)

Study Code	D&B '73	D&B '74	2	7	8	10	12	13	14	18	22	23	24	25
-ing	2	2	1	2	1	1	1	1	2	1	1	1	2	1
plural	1	3	2	1	2	3	3	2	1	4	3	3	3	3
articles	3	1	3	3	3	2	2	3	3	3	2	2	1	2
reg. past	4	4	–	5	4	5	4	4	4	6	4	–	4	5
poss.	6	5	5	4	5	4	5	–	5	2	–	4	6	4
3 sg	5	6	4	6	6	6	6	5	6	5	5	5	5	6

Study Code	Study
D&B '73	Dulay & Burt, 1973
D&B '74	Dulay & Burt, 1974
2	Bailey, Madden, & Krashen, 1974
7	Fathman, 1975
8	Fuller, 1978
10	Houck, Robertson, & Krashen, 1978
12	Kjarsgaard, 1979
13	Krashen, Houck, Giunchi, Bode, Birnbaum, & Strei, 1977
14	Krashen, Sferlazza, Feldman, & Fathman, 1976
18	Mace-Matluck, 1977
22	Pica, 1983
23	Riddle, 1993
24	Rosado, 1986
25	Rosansky, 1976

Appendix B

Fourteen grammatical functors reported in Brown (1973, p. 271)—Mean order of acquisition across three children

1. present progressive *-ing*
2. in
3. on

4. plural *-s*
5. past irregular
6. possessive *-'s*
7. uncontractible copula
8. articles *a*, *the*
9. past regular *-ed*
10. 3rd person singular regular *-s*
11. 3rd person singular (e.g., *does, has*)
12. uncontractible auxiliary
13. contractible copula
14. contractible auxiliary

Appendix C

Studies reviewed for possible inclusion in this meta-analysis (see References for complete bibliographic information)

1. Andersen, 1977
2. Bailey, Madden, & Krashen, 1974
3. Ball, 1996
4. Brown, J., 1983
5. Dulay & Burt, 1973
6. Dulay & Burt, 1974
7. Fathman, 1975
8. Fuller, 1978
9. Hakuta, 1976
10. Houck, Robertson, & Krashen, 1978
11. Kessler & Idar, 1979
12. Kjarsgaard, 1979
13. Krashen, Houck, Giunchi, Bode, Birnbaum, & Strei, 1977
14. Krashen, Sferlazza, Feldman, & Fathman, 1976
15. Larsen-Freeman, 1975
16. Lightbown, 1983
17. Lightbown, Spada, & Wallace, 1980
18. Mace-Matluck, 1977
19. Makino, 1979
20. O'Dowd, 1991
21. Pak, 1987
22. Pica, 1983
23. Riddle, 1993
24. Rosado, 1986
25. Rosansky, 1976

Appendix D

Scores for each functor on each determinant

1	pres. progr.	plural *-s*	poss. *-s*	articles	reg. past	3 sg. *-s*
2	2	1.33	1.33	1.66	1.33	1.33
3	1.7996	−.5959	−.5959	.5840	−.5959	−.5959
4	1	.33	.33	1	.33	.33
5	1.2910	−.6455	−.6455	1.291	−.6455	−.6455
6	11	5.33	5.33	10.33	3.66	5.33
7	1.3683	−.4922	−.4922	1.1485	−1.0402	−.4922
8	2	1	1	1	2	3
9	2	3	3	4	3	3
10	−1.5811	0	0	1.5811	0	0
11	1	2	2	1	1	2
12	−.9129	.9129	.9129	−.9129	−.9129	.9129
13	3	3	1	2	1	1
14	160	147	71	552	44	89

Rows

1. functor name
2. raw score on number of phones
3. *z*-score on number of phones
4. raw score on syllabic/nonsyllabic
5. *z*-score on syllabic/nonsyllabic
6. raw score on sonority
7. *z*-score on sonority
8. score on semantic complexity
9. raw score on number of phonological alternations
10. *z*-score on number of phonological alternations
11. raw score on homophony
12. *z*-score on homophony
13. score on syntactic category
14. frequency

Appendix E

Languages in the meta-analysis

1. Breakdown by study

Code No./Study	Languages	Speakers
2. Bailey et al., 1974	Spanish	33 adults
	Non-Spanish (Greek, Persian, Italian, Turkish, Japanese, Chinese, Thai, Afghan, Hebrew, Arabic, Vietnamese)	40 adults
7. Fathman, 1975	Korean	60 children 6–14
	Spanish	60 children 6–14
8. Fuller, 1978	Danish	1 adult
	Farsi	11 adults
	French	5 adults
	German	1 adult
	Kurdish	2 adults
	Portuguese	4 adults
	Spanish	16 adults
	African languages	2 adults
	Arabic	15 adults
	Japanese	7 adults
	Korean	1 adult
	Mandarin	5 adults
	Tae-chew Chinese	1 adult
	Thai	2 adults
	Taiwanese	6 adults
	Vietnamese	1 adult
10. Houck et al., 1978	Arabic	8 adults
	Farsi	3 adults
	Japanese	5 adults
	Korean	2 adults
	Thai	1 adult
	Portuguese	1 adult
	Indonesian	1 adult
	Spanish	1 adult

12. Kjarsgaard, 1979	Vietnamese	45 children 7–14
13. Krashen et al., 1977	Farsi	6 adults
	Japanese	5 adults
	Arabic	7 adults
	Spanish	2 adults
	Indonesian	1 adult
	Korean	1 adult
14. Krashen et al., 1976	Spanish	18 adults
	Greek	14 adults
	Chinese	7 adults
	Persian	7 adults
	French	3 adults
	Russian	2 adults
	Korean	2 adults
	Italian	2 adults
	Polish	2 adults
	Hebrew	2 adults
	Thai	2 adults
	Japanese	2 adults
	Czech	1 adult
	Armenian	1 adult
	Hungarian	1 adult
18. Mace-Matluck, 1977	Cantonese	175 children 5–10
	Spanish	168 children 5–10
	Tagalog	49 children 5–10
	Ilokano	30 children 5–10
22. Pica, 1983	Spanish	18 adults
23. Riddle, 1993	Spanish	26 children 6–10
24. Rosado, 1986	Chinese	6 adults
	Farsi	6 adults
	Malay	6 adults
	Spanish	6 adults
25. Rosansky, 1976	Spanish	2 adults, 2 adolescents, 2 children

2. Breakdown by number of subjects

Spanish	354
Cantonese	175
Korean	66
Tagalog	49
Vietnamese	46
Non-Spanish (no breakdown by L1 but includes Greek, Persian, Italian, Turkish, Japanese, Chinese, Thai, Afghan, Hebrew, Arabic, Vietnamese)	40
Arabic	30
Ilokano	30
Farsi	26
Chinese	19
Japanese	19
Greek	14
French	8
Persian	7
Taiwanese	6
Malay	6
Thai	5
Portuguese	5
Indonesian	2
Russian	2
Italian	2
Hebrew	2
Polish	2
Kurdish	2
African languages	2
Czech	1
Armenian	1
Hungarian	1
Danish	1
German	1
Total Number of Subjects	924