3 Second Language Acquisition and Ultimate Attainment

DAVID BIRDSONG

3.1 Introduction

In second language acquisition (L2A) research, ultimate attainment refers to the outcome or end point of acquisition, and is used interchangeably with the terms final state, end state, and asymptote. “Ultimate” is not to be thought of as synonymous with “native-like,” although native-likeness is one of the observed outcomes of L2A.

Most L2A studies have focused on the initial state, stages in L2 development, and rates of acquisition. However, data from such studies do not directly speak to the potential of the learner, which is an inescapable consideration of L2A theory. As we will see, the study of ultimate attainment engages such core L2A issues as native language influence, access to Universal Grammar (UG), maturational effects, and fossilization.

Just as ultimate attainment is a fundamental consideration of L2A research, L2A itself is a central concern of Applied Linguistics (or, more precisely, following the distinction made by this volume’s editors, of Linguistics-Applied (L-A)). Since the mid-1950s, the understanding of how linguistic knowledge is acquired and represented mentally has been a cornerstone of linguistic inquiry. Starting with first language acquisition (L1A) and eventually embracing L2A, much of this inquiry has been guided by the heuristic of constraints.

By hypothesis, language acquisition is constrained epistemologically: learners’ hypotheses about the possible forms of language are finite, and are not inconsistent with the range of structural features of natural language grammars. Similarly, it is believed that language acquisition is constrained maturationally: if native-like grammars are to be acquired, the learning must begin at an early developmental stage. One of the basic missions of L-A is to provide empirical data that speak to the adequacy of these elemental premises of modern linguistic theory. The methods recruited for this purpose are varied, and are informed by research in cognitive neuroscience, linguistic theory, and experimental psychology.
We begin the chapter by outlining the reasons why researchers are interested in investigating L2A at the end state. This rationale will provide a context for discussion in the remainder of the chapter, where we will consider a variety of ongoing research efforts relating to ultimate attainment in L2A.

3.2 Why Study Ultimate Attainment?

Ultimate attainment data are invaluable for ongoing mainstream research in L2A theory, in that they afford unique perspectives on the limits of L2A. On the received view of late L2A, the upper limits of competence are not comparable to those of a native monolingual. “Success,” construed as attainment of native-likeness, is ruled out in principle by advocates of the Critical Period Hypothesis (e.g., Johnson & Newport, 1989; Long, 1990) and by those who argue that UG and associated learning mechanisms are not available to post-adolescent L2 acquirers (e.g., Bley-Vroman, 1989). Under these views, the typical, if not unique, outcome of L2A is “failure” or non-native-like competence. However, recent research has challenged the notion of universal or near-universal failure (see below, and Birdsong, 1999, for a review). It appears that native-likeness may not be so rare as to be “peripheral to the enterprise of second language acquisition theory” (Bley-Vroman, 1989; see Selinker, 1972). Clearly, for educators and social-policy-makers, as well as for theorists, it is of compelling interest to know more about the rate of native-like attainment. For this purpose, the data de rigueur are those from learners at the end state; data from any other acquisitional stage can, at best, address only indirectly the upper limits of attainment.

In the most general terms, L2A theory tackles the question of the resemblance of L2A to L1A. L1A is uniformly successful, with all normal children attaining full competence, whereas in L2A there are various outcomes. As we will see below, the mature grammar may be incomplete vis-à-vis the target grammar, or it may diverge from it. And, unlike L1A, certain L2 learner grammars have been characterized as non-deterministic, or probabilistic. Perhaps the most basic issue in L2A research is whether this difference in ends (i.e., final states) implies different means (i.e., learning procedures), as suggested by Bley-Vroman’s (1989) Fundamental Difference Hypothesis. The obverse question can also be posed, namely, whether, in those cases where native-like attainment is observed, L1A-like learning (with access to UG) is necessarily involved.

The age factor in L2A is another domain in which ultimate attainment data figure prominently. It is widely recognized that the age at which L2A begins is reliably the strongest predictor of level of ultimate attainment. At issue, however, is the nature of this function. If there is a linear decrement in performance over all ages of immersion, this suggests a general age effect, with the possibility that experiential factors covarying with age may be implicated.
If, on the other hand, the age effect ceases at a maturationally-defined developmental point, and is not predictive thereafter, this suggests a qualitative change in learning. Researchers also look for evidence of discontinuity in the age function, which would suggest the start of a decline from peak levels of sensitivity (e.g., Flege, 1999).

With end-state data one brings a privileged perspective to the perennial question of native language effects. In particular, one looks at the pairing of different L1s with a single L2 to determine if there is a corresponding varying incidence of native-likeness. In addition, there is the question of whether some areas of the L2 grammar, but not others, are ultimately mastered, and if this asymmetry is a function of the learner’s native language (Bialystok & Hakuta, 1999; Bialystok & Miller, 1999).

Ultimate attainment data are useful when investigating other linguistically-motivated distinctions as well. One may compare, for example, learner proficiency on low-level phonetic features, which are presumably learned in a data-driven, frequency-sensitive manner, to the acquisition of morphosyntactic features such as that-trace, which are deductive consequences of parameter resetting, and whose acquisition is not dependent on frequency in the input. A similar logic is applied to the acquisition of regular versus irregular verb past and noun plural morphology (Birdsong & Flege, 2001).

In the remainder of this chapter we will consider a number of topics that fall under the umbrella of ultimate attainment in L2A. The emphasis will be on late learners, who typically are defined in terms of having arrived in the target language setting at age 12 or later. (Most studies operationalize age of learning in terms of age of immersion or age of arrival (AOA) in the target country, not in terms of age of first exposure, which typically is brief or sporadic, and which may take the form of school study, watching films or TV, or vacations.)

### 3.3 Non-Native-Like Outcomes in L2A

As pointed out by Sorace (1993, pp. 23–4), learners at the end state may have a grammar of the L2 that lacks some property P of the target grammar; accordingly this grammar is said to be incomplete. Another type of non-native-like grammatical representation is divergence, whereby property P is instantiated but in a manner that is not consistent with that property of the target grammar. On the basis of grammaticality judgments of Italian unaccusative constructions, Sorace found that French learners of L2 Italian preferred avere to essere in instances where both auxiliaries are permitted. This outcome was considered divergent with respect to the target grammar. English natives, on the other hand, did not show a principled preference for either essere or avere, even in contexts where essere only was permitted (i.e., sentences with clitic-climbing). Sorace therefore concluded that their grammar was incomplete with respect to the range of features associated with unaccusativity in Italian.
A variant of incompleteness in grammatical representation is indeterminacy. Indeterminate or probabilistic grammars are characterized by variability in intuitions for grammaticality from Time 1 to Time 2. Such inconsistency was observed by Johnson et al. (1996) in a sample of Chinese-speaking adult learners of English at end state (mean length of residence = 6.45 years). By Johnson et al.'s reckoning, some 35 percent of the learner performance in their sample was attributable to guessing, response bias, or problems with retrieving the target form from memory.

As Johnson et al. point out, what appears to be indeterminacy may actually reflect optionality in the grammar, i.e., a representation that allows multiple surface realizations of a single construction, such as the choice of relatives in *There’s the boy (that/whom/0) Mary likes*. In L2A, a learner who accepts John *seeked Fred* at Time 1, then John sought Fred at Time 2, might be inconsistent not because of indeterminacy in the grammar, but because the grammar permits both forms optionally. (Observe that optionality in this instance would reflect a grammar that diverges from the English target grammar.)

Non-native-like outcomes have been examined in the context of UG. Schachter (1990) maintains that an L2 learner’s access to UG principles is incomplete, that is, it is restricted to those principles that are instantiated in the learner’s L1. Johnson and Newport (1991) suggest that subjacency “survives in a weak and probabilistic form” (p. 237) among Sinophone late learners of English. In addition to indeterminacy, Johnson and Newport (1991) find divergence in the form of non-compliance with UG, that is, learner structures that are not consistent with any known natural language (see also E. C. Klein, 1995).

### 3.4 A Closer Look at the Concept

First let us recall that, as we noted at the beginning of this chapter, “ultimate attainment” is not to be misunderstood as suggesting native-likeness. Rather, it refers to the end point of L2A, irrespective of degree of approximation to the native grammar. Moving beyond this clarification, let us try to pin down conceptually what is meant by the term. At a basic level of understanding, the notion of end state in L2A is no different from its counterpart in L1A, as both denote the mature grammar. However, as we have just seen, the end state of L2A may be non-deterministic, and thereby differ qualitatively from the L1A end state. As a result, the idealization of the mature grammar as a “steady-state grammar” must be finessed: compared with L1, the L2 steady state seems “unsteady,” as it admits more variability in surface realizations and more uncertainty of intuitions. This is the nature of an indeterminate end-state L2 grammar, and as such this outcome should not be confused with “backsliding” or ongoing grammatical re-representation, which would suggest learning still in progress (see Johnson et al., 1996, p. 336 for further discussion of this distinction). With this understanding, it should be clear that the labels “end state,” “final state,” “asymptote,” and “ultimate attainment” are not inappropriately
applied to the outcome of L2A. (Splitting hairs, the label “asymptote” is often associated with a learning function that continuously approaches but does not reach the horizontal asymptote. This conceptualization of the mature state (in both L1A and L2A) allows for incremental progress, and thus no absolute finality, in learning. This view would accommodate additions of novel lexical items (along with idioms, slang, dialectal variants, technical jargon, etc.) and occasional changes in surface morphological or phonetic forms, but not re-representation of the underlying grammar.)

We are still left with the matter of determining when the end state has been reached. For example, how do we know that the abstract features associated with functional heads have been set, permanently, to native-like or non-native-like values (see, e.g., Lardiere, 1998)? To a large extent, the answer to this question depends on the adequacy of our methods for probing learner grammars. That is, we need reliable data – ideally, convergent evidence from multiple elicitation methods – and sensible interpretation of these data. If the data were longitudinal, researchers would be better able to determine whether dissimilar performance at Time X and Time Y reflects ongoing learning or a probabilistic end-state grammar. Moreover, with a longitudinal approach one could safely conclude that similar performances over Times X, Y, Z, etc. reflect an asymptotic level of attainment. As it happens, however, most studies of ultimate attainment are one-shot observations. In such cases, researchers have arbitrarily, but not unreasonably, established a length-of-exposure proxy for the L2 end state. Thus, for example, Johnson et al. (1996) operationalized the asymptote as a minimum of five years of immersion in the US. In Flege, Yeni-Komshian, and Liu (1999), participants were required to have lived at least eight years in the US. Birdsong and Flege (2001) employed a criterion of ten years or more of residence as a proxy for L2 ultimate attainment.

Common-sense caveats apply. An immigrant with ten years’ residence in the target country, yet isolated socially from native speakers, may not have attained the levels of L2 competence he is capable of. Thrust into an immersion situation after these ten years of isolation, this hypothetical learner is likely to go on to higher levels of attainment. Relatedly, one cannot assume that learners with comparable lengths of residence, even if fully immersed, have comparable levels of proficiency. That is – mindful again of the distinction between ultimate attainment and native-likeness – the assumption that the L2 end state has been reached is independent of observable levels of L2 proficiency. With this understanding, one may distinguish conceptually as well as methodologically between L2A studies that refer to their subjects as “near-natives” and those that sample learners at the end state.

3.5 A Note on Fossilization

Since the term was popularized in the L2A context by Selinker (1972), “fossilization” has been understood in various ways, among them, as a process, as a
cognitive mechanism, and as a result of learning. Selinker and Han (2000) catalogue various learner behaviors that researchers have associated with fossilization. These include backsliding, low proficiency, errors that are impervious to negative evidence, and persistent non-targetlike performance. They also list a host of proposed explanations for these behaviors, such as simplification, avoidance, end of sensitivity to language data, and lack of understanding, acculturation, input, or corrective feedback.

Unquestionably, the study of various representational and acquisitional facts that might fall under the umbrella of fossilization has advanced our knowledge of L2A. But among researchers there is disagreement at the most basic level, for example, on whether fossilization is an *explanans* or an *explanandum*, whether it is a process or a product, whether its domain extends to L1A, and whether it refers to invariant non-native forms or variable non-native forms (Han, 1998). Fossilization appears to be a protean, catch-all term that fails to capture a unitary or even coherent construct. This being the case, one must recognize the limitations of attempts to characterize the nature of fossilization. For the sake of descriptive and explanatory precision, it may be more reasonable to investigate discrete products, processes, behaviors, and epistemological states of L2A. Imagine, for example, that a given learner at presumed L2A asymptote exemplifies Behavior A (e.g., use of the imperfective to encode progressive past aspect) and Behavior B (e.g., use of the imperfective in telic contexts); Behavior A is native-like and Behavior B is non-native-like. Imagine further that Behavior A appears to be unsystematic, perhaps reflecting a probabilistic grammar, while Behavior B is invariant, suggesting a stable divergent grammar. The unique character of each behavior makes each worthy of investigation in its own right. Trying to decide whether one or both behaviors qualify as “fossilization” is unnecessary. Moreover, such labeling would not meaningfully illuminate matters, and would be likely to provoke unhelpful disputes over “questions of semantics.” It is self-defeating to be so bound to a term – which to date has defied attempts at meaningful characterization – that fundamental descriptive and explanatory goals become obscured.

### 3.6 Ultimate Attainment and the Critical Period Hypothesis: The Age Function

A key feature of the Critical Period Hypothesis for second language acquisition (CPH/L2A) is the prediction that native-like attainment in a second language will not be possible if the start of L2A is delayed past a certain critical age. (For consideration of what the critical age might be, see Long, 1990; Moyer, 1999; and discussion below. For a review of the different formulations of the CPH/L2A, see Birdsong, 1999.) Because the CPH/L2A addresses the upper limits of attainment possible in L2A, the only evidence that is decisively relevant to the adequacy of the CPH/L2A comes from learners at the L2A end
state. (Surprisingly, this common-sense requirement is not met in all studies purporting to test the CPH/L2A.)

As a general rule, level of ultimate attainment in L2A is predicted by age of arrival in the target country. Note that other age-related factors such as age of initial exposure, particularly in classroom contexts, are not strongly predictive; see, e.g., Birdsong and Molis, 2001; Johnson and Newport, 1989. After age of arrival, the strongest predictor appears to be amount of L2 input and interaction (e.g., Birdsong & Molis, 2001; Flege, 1999; Flege, Frieda, & Nozawa, 1997; Flege, Yeni-Komshian, & Liu, 1999). For discussion of other endogenous and exogenous variables, see, e.g., Bialystok and Miller (1999); Hyltenstam and Abrahamsson (2000); W. Klein (1995).

Not all apparent age effects are maturational in nature. Johnson and Newport (1989), articulating the logic of a critical period for L2A, point out that attainment should correlate negatively with age of arrival (AOA), just in cases of learners whose AOA predate the end of maturation. However, under the CPH/L2A, correlations of AOA with attainment should not be observed in cases where the AOA is later than the end of maturation, since maturational factors could no longer be at play.

With a sample of Chinese and Korean learners of English assumed to be at asymptote (≥ 5 years’ residence), Johnson and Newport (1989) obtained exactly this type of result. Participants were asked to provide grammaticality judgments for 276 English sentences presented on an audiotape. Stimuli exemplified basic surface contrasts in English, for example, regular verb morphology:

(1) Every Friday our neighbor washes her car.
   *Every Friday our neighbor wash her car

irregular noun morphology:

(2) Two mice ran into the house this morning.
   *Two mouses ran into the house this morning

and particle placement:

(3) The horse jumped over the fence yesterday.
   *The horse jumped the fence over yesterday

Accuracy on the judgment task varied as a function of age for those subjects whose AOA was less than 16 years (r = −0.87), but not for later arrivals (r = −0.16). Birdsong and Molis (2001) conducted a replication study of Johnson and Newport (1989). Using the original materials and methods, but Spanish natives as their subjects, Birdsong and Molis obtained very different results. Learners with AOA ≤ 16 performed at ceiling (r = −0.23), while the performance of later arrivals was predicted by AOA (r = −0.69).
Other studies of late learners (e.g., Birdsong, 1992) have observed significant correlations of attainment with AOA. Interestingly, Bialystok and Hakuta (1994, p. 69), reanalyzing the Johnson and Newport (1989) data, found a significant correlation of age and performance among late arrivals if the lower end of the late arrival group was set at 20 years. Many studies, including Johnson and Newport (1989) have found correlations of performance with AOA when later and earlier arrivals are pooled (Bialystok & Hakuta, 1999; DeKeyser, 2000; Flege, 1999; Oyama, 1976). A correlation of AOA with declining performance past the end of maturation – indeed, over the entire life span – has been viewed as a priori evidence for falsification of maturational accounts of L2A (Pulvermüller & Schumann, 1994, p. 684).

Researchers (e.g., Bialystok & Hakuta, 1999; Flege, 1999) have also argued that a distribution of end-state performance, to be consistent with the CPH/L2A, should incorporate a point of inflection, an “elbow” corresponding to the start of a decline in learning ability, i.e., the offset of the period of peak sensitivity. Flege (1999, p. 104), finding no evidence for such non-linearity in studies of L2 pronunciation, states: “In my view, the lack of a non-linearity in the function relating AOA to degree of foreign accent is inconsistent with the view that a critical period exists for speech learning.” Further, a series of regression analyses performed by Birdsong and Molis (2001) on their data suggests that, if there is an inflection point in the age function, it occurs at a point past the end of maturation (> 18 years). That is, the observed decline begins at a developmental point where sensitivity should presumably be already at its lowest level. (For further discussion of the timing of age-related effects and its relevance to the CPH/L2A, see Birdsong & Molis, 2001; Elman et al., 1996, pp. 187-8; Moyer, 1999, p. 100. For consideration of biographical factors that may covary with AOA, and that are unrelated to maturation, see Bialystok & Hakuta, 1999.)

Birdsong (in press) argues that even if one ignores the timing of the age effects in L2A, the shape of the age function is inconsistent with standard conceptions of critical periods. According to Bornstein (1989), one of the characteristic features of a critical period is an end to enhanced receptivity or sensitivity. That is, after the peak of sensitivity, there is a decline – the beginning of the offset phase of the critical period – which culminates at a point of zero or baseline sensitivity marking the end of the offset phase. From this point on, sensitivity should not decline further. The overall age function should resemble a stretched ‘Z,’ as described by Johnson and Newport (1989, p. 79) and Pinker (1994, p. 293).

In contrast, a meta-analysis of L2A end-state studies (Birdsong, in press) reveals a consistent picture of ongoing declines in attainment over the span of AOA. These indefinitely-persisting age effects usually take the form of a simple straight-line decline or a stretched “7” shape, the bottom end pulled rightward. With no apparent end to the decline of sensitivity, the notion of a bounded time frame, or critical “period” of sensitivity, fails to match up with the ultimate attainment data.
3.7 The Incidence of Native-Like Attainment

Along with post-maturational age effects in ultimate attainment, native-likeness among late learners of L2 has been considered as a criterion for falsification of the CPH/L2A. In fact, Long (1990, p. 255) maintained that a single case of demonstrable native-like proficiency among late learners would be sufficient to refute the CPH/L2A.

Such a standard was not out of keeping with the Zeitgeist of the late 1980s and early 1990s. Non-native-likeness was the presumed end state of post-pubertal L2A, and there was little or no empirical evidence to the contrary (see the comprehensive review by Long, 1990). Estimates of the incidence of native-likeness ranged from near 0 (Bley-Vroman, 1989) to 5 percent (Selinker, 1972). Success in adult L2A was thought to be so rare as to be pathological, in the sense that the rate of native-like attainment could be compared to the rate of failure to acquire a first language (Bley-Vroman, 1989). Mainstream texts deemed a lack of mastery a basic characteristic of late L2A, a fact in need of an explanation (e.g., Towell & Hawkins, 1994).

Two studies in particular contributed to this view: Coppieters (1987) and Johnson and Newport (1989). Coppieters studied 21 near-native speakers of French from varying L1 backgrounds. All were late learners who had resided in France for at least five and a half years. Participants judged the grammaticality of 107 complex French sentences, some of which exemplified language-specific structures, such as the choice of subject pronoun in identificational constructions:

(4) Qui est Victor Hugo? Who is Victor Hugo?
C’est un grand écrivain.
He’s [identification function] a great writer.
*Il est un grand écrivain.
He’s [anaphoric function] a great writer.

Other items illustrated universal constraints or principles, for example, use and placement of the clitic en, which varies according to the predicate:

(5) Elle en aime l’auteur.
She likes its author.
*Elle en téléphone à l’auteur.
She telephones its author.

Twenty monolingual native speakers of French served as controls, and the judgments of both groups were compared to acceptability norms. In the Coppieters sample the observed incidence of native-likeness was zero.

A similar result was obtained in the Johnson and Newport (1989) study. Among their 23 late learners, the highest score was 254 out of 276. The lowest
score among native controls was 265. The researchers consider this depressed performance to be consistent with the idea of maturational constraints in L2A. Indeed, among late learners, non-native-likeness is thought to be an inevitable outcome: “for adults, later age of acquisition determines that one will not become native or near-native in a language” (Johnson & Newport, 1989, p. 81).

Since the publication of these two highly influential papers, replication studies have been carried out. With tasks and stimuli modeled on Coppieters (1987), Birdsong (1992) looked at the acquisition of French by 20 native speakers of English. All had been exposed to French post-pubertally (range = 11–28 years, mean = 14.9); all had been residing in France for at least three years (range 3–36 years, mean = 11.8 years). Mean age of arrival was 28.5 years (range = 19–48). On scalar grammaticality judgments, the performance of more than half of the 20 experimental subjects was within the range of performance of native controls.

Differences between the two studies are likely the result of variation in procedural controls and subject sampling, details of which are found in Birdsong (1992). It is unlikely that native-like levels of attainment are attributable to stimuli choice. The replication used many of the original stimuli (such as (4) and (5) above), and the additional stimuli exemplified subtle and complex features of the French grammar, for example, prenominal past participle:

(6) Le très-connu Marcel Proust vient d’arriver.
   The well-known Marcel Proust just arrived.

that-trace:

(7) *Qui crois-tu qui rendra visite à Marc?
   Who do you think [that] will visit Marc?

adjacency (verb raising):

(8) Les garçons regardent avec intérêt la télévision.
   The boys look with interest at the television.

Several other studies have attested native-like performance among late learners. For 20 Sinophone and 20 Francophone subjects, all late learners of English, Cranshaw (1997) studied the acquisition of tense and aspect in English. The study involved a series of production and judgment tasks. Over all tasks, three of the Francophones and one Sinophone performed like English native controls. Van Wuijtswinkel (1994), using a grammaticality judgment task based in part on the Johnson and Newport (1989) items, tested Dutch natives who began learning English after age 12. In one group van Wuijtswinkel studied, 8 of 26 participants performed like native English speakers, and in another group 7 of 8 were indistinguishable from natives. White and Genesee (1996) investigated the acquisition of subtle properties of English syntax. Their
subjects were Montréal Francophones, whose first significant exposure to English had taken place after age 12. Some 16 of the 45 participants had demonstrated English-native-like performance on various screening measures. These subjects performed like natives on production and judgment tasks involving wh-extraction, e.g., What did the newspaper report the minister had done? In another study where anglophone subjects were pre-screened for native-like performance (here, screening involved oral interviews and a proficiency test), Montrul and Slabakova (2001) studied the L2 acquisition of the Spanish preterit/imperfective distinction. Participants’ average age of exposure to Spanish was about 15 years (range = 12–24), and they were not living in a Hispanophone country at the time of testing. Across a variety of tasks and sentence types, 35 percent of the sample performed like native controls. A lower rate of native-likeness was observed by Birdsong and Molis (2001), using the Johnson and Newport (1989) instrument. Of the 32 late arrivals (AOA range = 17–44 years), only one scored within the native range of performance. However, 13 of these participants scored at a 92 percent level of accuracy or above.

To allow for meaningful extrapolations to L2 learning generally, the incidence of native-like attainment, expressed as a proportion of the participant sample, must not be established on the basis of a “stacked deck” – a group of subjects who have been pre-screened for demonstrably high attainment (or for having extraordinary motivation, input/interaction with natives, etc.). In unscreened samples of learners at presumed L2 asymptote, the levels of ultimate attainment represented are quite diverse, and desirably so. In this type of sampling, the observed rates of native-like attainment (usually from 5 to 15 percent of the sample) may be more safely generalized to broader populations.

Typically, native-likeness among late learners is observed less frequently in the area of pronunciation than in morphosyntax (e.g., Oyama, 1982; Patkowski, 1980; Scovel, 1988; see also Flege, 1999, for an overview). For example, in Flege, Munro, & MacKay (1995), a 6 percent incidence of unaccented pronunciation was found among late learners. However, none of the participants with AOA greater than 16 years had authentic pronunciation. Such results suggest not only that the incidence of native-like pronunciation is low, but also that the rate continues to decline with increasing AOA, even after the presumed end of maturation.

In contrast to the general pattern of accentedness observed in late L2A, Bongaerts (1999) has demonstrated that Dutch late learners of English and French (age of exposure > 12 years) can speak without accent, though the rate of native-likeness is lower for French L2 than for English L2. Pronunciation was sampled at the sentence level (e.g., My sister Paula prefers coffee to tea; Avec ce brouillard horrible j’allumerais mes phares). In addition, for the French study the complete range of nasal and oral vowels was sampled in CV (consonant-vowel) frames (e.g., /u/ in pou, tout, and loup). Relatedly, Birdsong (2001) performed an instrumental analysis of the pronunciation of late learners of French (AOA ≥ 12 years, mean AOA = 23 years) whose native language was English. Two of the 20 subjects were indistinguishable from native Parisian
controls in terms of voice-onset time (VOT) for word-initial consonants (e.g., le père), release of word-final obstruents (e.g., le cap), and word-final vowel duration (e.g., le dé).

To recap the preceding observations about the rate of native-likeness in late L2A: A significant incidence of native-likeness has been found in several studies (in addition to those mentioned above, other studies where native-likeness is observed include Bruhn de Garavito, 1999; Ioup et al., 1994; Juffs & Harrington, 1995; Mayberry, 1993; and White & Juffs, 1997); the rate of native-likeness appears to be lower for features of pronunciation than for morphosyntactic features; and varying rates of native-likeness may result from different L1-L2 pairings.

Some researchers in bilingualism and neurocognitive development dispute the a priori appropriateness of the native standard for the study of the L2 end state. For example, Cook (1997) and Grosjean (1998) note that an L2 learner can never be or become a native speaker. According to this line of thinking, it is ill conceived to peg success in L2A theory to native-likeness. This argument applies as well to social contexts, where immigrants are often stigmatized for non-native-like linguistic behaviors.

Further, one could argue that the criterion of native-likeness sets the bar too high, since late learners routinely attain quite impressive, if not native-like, levels of L2 proficiency and linguistic knowledge. Late L2 learners rarely resemble Genie, whose delayed L1 acquisition was characterized by profound deficits in syntactic and morphology at various levels of analysis (Curtiss, 1977). Nor do they exhibit the extreme pathology of another late L1 learner, Chelsea, whose output included violations of structure dependency, for example, determiners preceding finite verbs (Curtiss, 1989).

From the perspective of research in developmental psychology and language acquisition, however, the native competence level affords a benchmark for comparison that permits ready interpretation of experimental results (see also Mack, 1997). Perhaps most importantly, demonstrations of native-likeness represent dramatic counterpoints to received views of the upper limits of L2A, whereby the outcome of L2A is doomed to be inferior to that of L1A.

3.8 Initial State, End State, and Universal Grammar

Recent research in the UG/L2A framework has stressed the theoretical relationship of initial state competence to final state competence (e.g., Hardin, 2001; White, 2000). In the most basic terms, researchers make predictions about end-state competence based on a theorized initial-state grammar. Thus, for example, if the L2A initial state is not characterized by transfer from the L1, and there is full access to UG, then native-like competence at the end state
should be predicted. In contrast, if the initial state of L2A is the full L1 grammar, and there is no access to UG, then a failure to attain native-like competence at L2 ultimate attainment is expected. Under this approach additional aspects of the end-state grammar may also be anticipated, such as the nature of the grammar (e.g., incomplete, divergent, indeterminate), “rogue” or non-UG-compliant features, and effects of L1-L2 pairing.

Hardin (2001) examines in detail the relationships between initial and end states in L2A. Under the Full Transfer/No Access theory of initial state (e.g., Bley-Vroman, 1989; Schachter, 1990), for example, Hardin observes that the end-state grammar could be incomplete, divergent, and indeterminate; it may have optionality in contexts where the L2 does not; the grammar may not conform to constraints given by UG, and there are likely to be L1-L2 pairing effects. Native-likeness, if observed at all, is rare, and would not be attained via direct access to UG and associated domain-specific learning principles but by extraction of universal properties of grammar from the L1 and use of generalized learning principles. In contrast, by the No Transfer/Full Access theory of the initial state (e.g., Epstein, Flynn, & Martohardjono, 1996; Martohardjono & Flynn, 1995), the grammar at the end state should be native-like, at least with respect to the core grammar, and there should be no evidence of incompleteness, divergence, indeterminacy, non-native optionality, UG-non-compliance, or L1-L2 pairing effects. An intermediate position assumes full L1 transfer and complete access to UG (e.g., Schwartz, 1998; Schwartz & Sprouse, 1996). Hardin (2001, p. 113) points out that native-like ultimate attainment is not excluded under this approach, but that L1 influence is understood to persist throughout L2 development. In addition to L1-L2 pairing effects, under the Full Transfer/Full Access theory of the initial state one could expect any number of non-native-like outcomes (incompleteness, divergence, indeterminacy), but the grammar would conform to constraints of UG.

It is important to note that a native-like outcome in L2A does not necessarily imply that UG is accessed. The raison d’être of UG is to provide a solution to the logical problem of language acquisition. That is, grammatical knowledge at the end state is underdetermined by the linguistic evidence at the learner’s disposal, and it is hypothesized that the constraints on grammatical form given by UG fill in the epistemological gap. This argument applies equally to L2A and L1A. In both instances, to discern a role for UG one must demonstrate that there is in fact a logical problem that is solved by UG. Thus, if one wishes to attribute observed native-likeness at the L2A end state to UG, it must be shown that there was a logical problem in the first place and that the underlying grammatical competence could only have been gained by access to UG.

### 3.9 Dissociations and Asymmetries

Pinker (e.g., Pinker, 1999) proposes a dual-mechanism model for knowledge of regular inflectional morphology (e.g., verb pasts such as walk-ed; noun plurals
such as cup-s) versus irregular morphology (run – ran; child – children). Under this model, computation of regulars involves rule-based, or symbolic, processing of the compositional features stem + ending, whereas irregulars are accessed as individual units from associative memory. Unlike regulars, the representation of irregulars is sensitive to the items’ frequency in the input. Were verb pasts and noun plurals represented under a single-system connectionist model, on the other hand, then there would be no symbolic manipulation, and all retrieval would require accessing inflected forms from (frequency-sensitive) associative memory.

There is behavioral and neurofunctional evidence of dissociations between rule-based and lexical knowledge. For example, Jaeger et al. (1996) asked native English adults to produce past tense forms of regular, irregular, and nonce verb stems, and found significantly different reaction times for the three types of verbs. In addition, using positron emission tomography (PET) technology, the researchers observed that the regular and irregular computations were subserved by different areas of the brain and required different amounts of cortical activation. Although most research has involved L1 adults and children (e.g., Marcus et al., 1995; Pinker, 1999; Ullman et al., 1997), Beck (1997), Marzilli and O’Brien (2000), and others have obtained experimental evidence for regular-irregular dissociations in L2A short of asymptote (see, however, discussion of Brovetto & Ullman, 2001, below).

The end-state perspective was adopted by Flege, Yeni-Komshian, and Liu (1999), in their study of 240 Korean learners of L2 English. From the Johnson and Newport (1989) instrument they isolated a subset of 44 items to represent rule-based and irregular forms. For late learners (AOA > 12), the participants’ accuracy figures revealed a clear dissociation in performance as a function of age, with regulars much less affected by increasing AOA than irregulars.

Taking this finding as their point of departure, Birdsong and Flege (2001) hypothesized that input frequency should be a factor in knowledge of irregular, but not regular forms (e.g., Beck, 1997). Also, regular versus irregular differences should obtain across grammatical categories such as verbs and nouns (e.g., Marzilli & O’Brien, 2000). Most importantly, they expected to replicate the different age of arrival effects for regulars versus irregulars found by Flege, Yeni-Komshian, & Liu (1999). Finally, if (ir)regularity effects are universal, the predicted dissociations should be observed regardless of L1-L2 pairing. The researchers recruited a sample of educated Spanish (n = 30) and Korean (n = 30) natives at L2 asymptote (length of residence ranged between 10 and 16 years). The sample was broken down into groups of 10 based on age of arrival in the US (6–10 years; 11–15 years; 16–20 years). Participants performed a timed multiple-choice judgment task on 80 items exemplifying regularity vs. irregularity, high vs. low stem frequency, and noun plural vs. verb past tense morphology. The following items exemplified low frequency regular noun plural, and high frequency irregular verb past, respectively:
(9) There are five
   a. knuckli
   b. knuckle
   c. knuckles
   d. knackle
   e. knuckleses

(10) Yesterday the little girl
    a. swim
    b. swam
    c. swammed
    d. swims
    e. swammed

A series of Analyses of Variance (ANOVAs) was performed on both accuracy and response latency data. For both types of data, several significant effects were obtained. First, the effect of item frequency was found to be significantly higher for irregular items than for regulars. Also, in most respects, the results for noun plurals were comparable to those for verb pasts, the exception being that, among Korean natives, but not among Spanish natives, performance on noun plurals was depressed relative to that for verb pasts. (Birdsong and Flege suggest that this result reflects the fact that Korean typically does not inflect for plurals, but plurality is inferred pragmatically from contextual cues.) Finally, consistent with their principal prediction, Birdsong and Flege found that the accuracy decline and increased response time (RT) with increasing AOA were more pronounced for irregulars than regulars. In fact, no significant age-related declines at all were observed for the regular items. Accuracy and RT data for 20 additional items exemplifying phrasal verbs, e.g., “The student cannot come up with the correct answer” were comparable to performance on irregulars, suggesting that age of arrival effects apply to other varieties of idiosyncratic information in addition to morphological irregularities.

Interestingly, Brovetto and Ullman (2001) in a study of oral production of regular and irregular English pasts by 32 Spanish and 32 Chinese natives (AOA ≥ 17 years) with a minimum of three years’ US residence, found that performance on both irregulars and regulars was sensitive to frequency. To reconcile this result with Birdsong and Flege’s finding of greater frequency sensitivity for irregulars than regulars, one may hypothesize that, for learners at stages leading up to the end state, many if not most target language forms are bits of idiosyncratic information stored in declarative memory. However, by the L2 end state, computation may take on a more L1-like flavor. As Ullman (2001, p. 118) suggests, “an increasing amount of experience (i.e., practice) with a [second] language should lead to better learning of grammatical rules in procedural memory, which in turn should result in higher proficiency in the language.” Thus it may be that the course of attainment of proficiency in an L2 involves a transition from unitary associative L2 processing to a system that exploits both symbolic and associative processes, in procedural and declarative
memory respectively, with activation of the corresponding neural substrates. Clearly, this line of reasoning underscores the importance of studying learners at the end state (and, where appropriate, comparing them with learners not yet at asymptote).

One may speculate that declarative memory, which provides for learning and storage of facts, names, and arbitrary and irregular forms, is more susceptible to aging effects than the procedural memory system, which may be responsible for rule-based learning. Evidence of age-related declines in various types of declarative memory is found in the literature on cognitive aging (e.g., Salthouse, 1991), though much of the experimental work relates to short-term memory effects. Moreover, certain well-known histological features associated with cognitive decline over the course of normal aging appear to be concentrated in neural regions implicated in declarative memory. Specifically, neurofibrillary tangles appear mainly in the cortical pyramidal cells of the hippocampus and temporal association areas, and neuritic (senile) plaques are generally found in the hippocampus and second and third layers of the temporal/associative cortex (Scheibel, 1996). These degenerative features of normal neurologic aging – which, in high concentrations characterize the synaptic pathology of Alzheimer’s Disease – appear not only to corrupt existing cortical pathways but may impair the work of neurotransmitters, particularly acetylcholine, which are crucial to encoding and consolidation of memories; for overviews, see Hasselmo (1999) and Martin (1999). (This is not to suggest that the aforementioned etiologies and loci of cognitive decline are the only ones associated with aging. For example, declines in dopamine D2 receptors – on the order of 10 percent per decade after 20 years of age – are observed in the basal ganglia, hippocampus, frontal cortex, anterior cingulate cortex, and amygdala regions (Li, Lindenberger, & Sikström, 2001). The relevant issue is the extent to which age-related declines are more severe in some brain areas than in others, and the corresponding effects on specific types of cognitive functions involved in L2A.)

Moving beyond speculation as to the underlying causes of regular-irregular dissociations over the age function, it is clear that the question of age effects in late L2A cannot be approached monolithically. Future investigations should aim for sufficient granularity to take into account not only the moderator variable of L1-L2 pairing, but also principled questions of representation and processing such as those raised by the study of regular versus irregular inflectional morphology.

Another variable that should not be overlooked in the study of ultimate attainment is the participants’ dominant language. Interest in the dominance factor goes back to a study by Cutler et al. (1989), who observed an asymmetry in early French-English bilinguals’ ability to process spoken words into segments. Those whose self-reported dominant language was French were able to switch back and forth between syllable-based and non-syllable based segmentation routines as a function of the language being processed. In contrast, English-dominant bilinguals controlled only one segmentation strategy,
suitable to the processing of English words, and applied it to both French and English speech. Golato (1998) attempted to reproduce the results of Cutler et al. (1989), using late L2 learners at the end state. Twenty-one late bilinguals (English natives n = 10) participated. Overall, they had spent a mean of 8.5 years in an anglophone or francophone country, had an age of immersion ranging from 13 to 33 years, and had a mean age of 30 years at time of testing. Golato found that the English-dominant bilinguals commanded two syllable segmentation routines, and applied the English-appropriate strategy to English stimuli, and the French-appropriate strategy to French words. In contrast, the French-dominants used a single strategy indiscriminately for both French and English stimuli. Different operationalizations of language dominance were considered, and the asymmetry obtained for every operationalization. Provocatively, Golato’s results for learners at the L2A end state are opposite to those that Cutler et al. (1989) had reported for early bilinguals; follow-up replications are unquestionably warranted. It is also clear that, as with L1-based and regularity-based dissociations, the study of asymmetries promises to contribute significantly to a finer-grained understanding of the end state of L2A.

3.10 Ultimate Attainment and Cortical Function

Modern technologies such as Functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET), and Event-Related Brain Potentials (ERPs) allow L2A researchers to investigate the neural systems involved in language processing. In most cases, work does not address the L2A end state specifically, but is concerned with the age at which L2 acquisition was begun and L2 proficiency. A recurrent goal in this research is determining the degree to which L1 processing and L2 processing involve similar neural substrates. For example, Weber-Fox and Neville (1999), in a study of Sinophone learners of English with over five years’ immersion, found that the neural subsystems involved in language processing differ as a function of age of acquisition. Notably, for the processing of phrase structure violations, involvement of both the right and left hemispheres increased as the age of immersion in English was delayed. The researchers also compared the learners’ processing of grammatical features (closed-class words and syntactic anomalies) with their processing of semantic features (open-class words and semantic anomalies), and found that these types of activity are differentially affected by age of acquisition. However, an ERP study by Osterhout, Davis, and McLaughlin (in press) has revealed a confound of word length and open versus closed class: “Although the two word classes did elicit distinct ERPs, all of these differences were highly correlated with word length. We conclude that ERP differences between open- and closed-class words are primarily due to quantitative differences in word length rather than qualitative differences in linguistic function” (Osterhout, Davis, & McLaughlin, in press, p. 1).
In a study of highly proficient late L2 learners (mean initial exposure = 12.25 years of age), Illes et al. (1999) used fMRI to investigate the semantic processing of nouns by eight English-Spanish bilinguals. Comparing the scans of processing in the two languages, the researchers found no activity differences in either the left or right inferior frontal gyrus, and both languages seemed to be dominantly lateralized in the left hemisphere. At least with respect to vocabulary, Illes et al. (1999) suggest that, irrespective of the age of acquisition, increasing proficiency in the L2 leads to a common cortical representation of the two languages. An earlier PET study by D. Klein et al. (1995) had reached a similar conclusion for repetition and translation of single words, albeit with early learners. The highly fluent subjects in this instance were 12 anglophone learners of French whose mean age of acquisition was 7.3 years. The researchers determined that semantic processing in the two languages involved similar areas of the front left cortical regions, particularly in the left anterior frontal gyrus.

We note that in the Weber-Fox and Neville (1999) study, the late bilinguals’ L2 proficiency (both self-rated and based on standardized tests) was significantly below that of the early learners. In this instance, both late age of acquisition and low proficiency are predictive of processing differences in the relevant neural substrates. However, in the Illes et al. (1999) and D. Klein et al. (1995) research, high proficiency, but not early age of acquisition, is predictive of homotropic cortical representations for the L1 and L2. Thus, for some aspects of semantic processing, the evidence suggests that those late learners who are native-like or near-native-like at end state will have common cortical localization of activity for the L1 and the L2.

In an investigation designed to disentangle the factors of proficiency and age of acquisition, a similar conclusion was reached by Perani et al. (1998). This was a PET study involving monitoring of brain activity of subjects (nine proficient late learners, AOA > 10 years) listening passively to a story in the L1 (Italian) and the L2 (English). In this case, the cortical responses were not only similar in the L1 and L2, but were comparable to the L1 and L2 brain activity of 12 speakers of Catalan and Spanish who had learned both languages early in life. The researchers conclude that “these findings suggest that, at least for pairs of L1 and L2 languages that are fairly close, attained proficiency is more important than age of acquisition as a determinant of the cortical representation of L2” (Perani et al., 1998, p. 1841).

For other tasks, however, the story is somewhat different. Kim et al. (1997) asked bilinguals from various language backgrounds to silently recount events from the previous day. Using fMRI, Kim et al. found a common neural representation for L2 and L1 among early bilinguals, but for late bilinguals (mean age of acquisition = 11.2 years) distinct regions of Broca’s area were involved. In Wernike’s area, on the other hand, similar cortical regions served both the L1 and the L2. This similarity was observed for each of the various L2s represented in the sample, and across all ages at which L2A was begun. Comparing the results of Kim et al. (1997) with those of their own study, Perani et al.
(1998, p. 1846) ascribe the differences in brain activation to the differences in task. We note further that in Kim et al. (1997), the question of whether proficiency trumps age of acquisition was not addressed, as the proficiency levels of the late acquirers was not specified.

On the general question of cortical function in L2 versus L1, Sanders (2000) reviews several additional studies, which vary in terms of L1-L2 pairings, tasks, measurement techniques, and emphasis on proficiency and age of acquisition. Although none of the studies cited makes specific reference to the L2A asymptote, the early bilinguals in each instance are uncontroversially at the end state, and one can assume that the highly proficient late learners are near if not at the end state. This being the case, it appears that, depending on the task and the L1-L2 pairing, native-likeness among late L2 learners can be observed not only in linguistic behaviors but in cortical function as well. For a recent review of neuroimaging studies of cortical function in bilingualism, see Abutalebi, Cappa, and Perani (2001).

3.11 Conclusion

The study of learners immersed in an L2 for significant lengths of time has led to significant advances in the understanding of the nature of L2A. Researchers recognize that a range of variables – in particular, age of immersion, L1-L2 pairings, and quantity of input – may interactively determine the level of ultimate attainment. As we move forward, we are alert to the need for finer-grained investigation of the limits of bilingualism, as suggested by the discovery of asymmetries at the end state, and their relation to representational variables such as the learner’s dominant language. Granularity is further motivated by demonstrations that discrepant effects of AOA are associated with various features of the language, possibly reflecting principled cognitive distinctions such as declarativized versus proceduralized knowledge, or symbolic computation versus lexical retrieval.

A dozen or so years ago the study of ultimate attainment in L2A was in its infancy. The field is now entering adolescence (and, to paraphrase Oscar Wilde’s witticism, is no longer young enough to know everything). With additional nourishment from cognitive neuroscience, linguistic theory, and developmental psychology, there is every reason to believe that the spurts of growth – and sophistication – will continue.

See also 8 Language Attrition, 17 The Native Speaker in Applied Linguistics, 20 Language Learning, 21 Individual Differences in second Language Learning.
REFERENCES


Illes, J., Francis, W. S., Desmond, J. E., Gabrieli, J. D. E., Glover, G. H., Poldrack, R., Lee, C. J., &


