The problem of age in second-language acquisition: Influences from language, structure, and task*

ELLEN BIALYSTOK
BARRY MILLER
York University

Three groups of participants were given a grammaticality judgement test based on five structures of English grammar in both an oral and written form. The first group consisted of native speakers of Chinese, the second, native speakers of Spanish, and the third, native English speakers. The two learner groups were divided into those who had begun learning English at a younger (less than 15 years) or older (more than 15 years) age. Performance was measured for both accuracy of judgement and time taken to respond. The results showed that performance patterns were different for the two learner groups, that the linguistic structure tested in the item affected participants' ability to respond correctly, and that task modality produced reliable response differences for the two learner groups. Although there were proficiency differences in the grammaticality judgement task between the younger and older Spanish learners, there were no such differences for the Chinese group. Furthermore, age of learning influenced achieved proficiency through all ages tested rather than defining a point of critical period. The results are interpreted as failing to provide sufficient evidence to accept the hypothesis that there is a critical period for second language acquisition.

Stories abound of the extraordinary prowess children appear to display in acquiring a new language and the apparently hopeless failure encountered by adults attempting the same feat. The common wisdom is straightforward: children are more able to learn another language than adults are. The preferred explanation is equally simple: children are biologically prepared to learn languages but adults are not. This is the basis for the view that there is a critical period for second-language acquisition.

Overviews of the research into the issue of a critical period for second-language acquisition often equivocate about the conclusion that can be drawn (e.g., Harley and Wang, 1997; Singleton, 1995). Although these authors frequently do adopt a position – for example, Singleton endorses the notion of a critical period – they acknowledge nonetheless its tentativeness. Other scholars point dogmatically to conclusions. Gleitman and Newport (1995) and Danesi (1994) argue strongly for a critical period, while Epstein, Flynn and Martohardjono (1996) argue equally strongly against it.

The empirical studies that have been conducted to evaluate the issue are extraordinarily heterogeneous. They use different methodologies, engage different subject groups, administer different types of tasks, and assess different linguistic features. The lack of consensus in outcomes is hardly surprising. The studies that tend to sample across a range of subjects on a rather specific proficiency test usually find a relationship between language competence and the age of acquisition for that language (e.g., Coppieters, 1987; Johnson and Newport, 1989, 1991; Oyama, 1976; Patkowski, 1980, 1994; Schachter, 1990). Conversely, studies that either preselect participants according to some criterion or use a broader assessment of proficiency than a single objective grammar test often report that native-like proficiency is attainable irrespective of the age at which acquisition begins (e.g., Birdsong, 1992; Bongaerts, Planken, and Schils, 1995; Ioup, Boustagui, Tigi and Mosell, 1994; Snow and Hoefnagel-Hohle, 1978; White and

* The research reported in this paper was funded by Grant A2559 from the Natural Sciences and Engineering Research Council, Canada (NSERC) to the first author. We are grateful for the assistance of many people in conducting this study. Heather Jenkin wrote the computer programme, Judith Codd carried out the data analyses, Jane Herman read the sentences into the programme for the oral condition, Eddie Yew collected data from the Chinese bilinguals, Stefka Marinova-Todd collected data from the Spanish bilinguals, and Caroline Roncadin collected data from the native speakers. We also relied on two linguistic consultants. Professor Li Zeng advised us on the structure of Chinese and provided the tests of Chinese L1 competence, and Professor Ellen Anderson was our consultant for Spanish. In addition, the manuscript profited from comments by David Birdsong, Jim Flege, Kenji Hakuta, Eric Kellerman, Judy Kroll, and an anonymous reviewer.

Address for correspondence
Department of Psychology, York University, 4700 Keele Street. Toronto, Ontario M3J 1P3, Canada
E-mail: ellenb@yorku.ca

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

Genesee, 1996). These conclusions are different points, and they are not necessarily incompatible with each other. It can be the case both that age of learning is related to decline in eventual proficiency of specific grammatical structures and that native-like competence is possible for late learners. However, if these positions are both correct, then the age-related decline in learning cannot be caused by a critical period. Indeed, there is little doubt that older learners typically attain lower levels of proficiency than younger ones, but the negative correlation between age and achievement is not in itself evidence for a critical period.

Long (1990) identified four categories of reasons for age-related changes in acquisition – social, input, cognitive, and neurological. Although he claimed that all of these support the critical period hypothesis, that conclusion is based on an excessively broad interpretation of the term "critical period". Following standard definitions (e.g., Bornstein, 1989; Colombo, 1982), a critical period must be based in a biologically determined sensitivity of the organism. Hence, only the neurological factors provide a potential source of a critical period. Empirically, a biologically based explanation for the age-related decline in proficiency should reveal a different pattern of outcomes than would a decline based on the other potential factors.

Establishing the existence of a critical period for second-language acquisition is important because it can shed light on an even more intractable problem, namely, whether or not there is a critical period for first-language acquisition. Ultimately, it is the potential for first-language learning across maturational ages that is the more crucial issue in cognitive science, because it speaks more directly to the structure and functioning of human learning and intelligence. However, aside from a handful of unusual case studies that have limited credibility because of their bizarre circumstances, there is no reliable way of addressing the question. Research with deaf populations who learned American Sign Language as a first language at various ages (e.g., Mayberry, 1993; Newport, 1988), or special cases of language delay caused by hearing impairments, epitomised by the case of Chelsea (Curtiss, 1989), have been more promising because the participants have had normal social and cognitive histories.

The logic for studying second-language acquisition as an avenue to this issue is as follows: if there is a critical period for first-language acquisition, then there may or may not be a similar constraint on second-language acquisition. Although language learning may shut down after a critical period and restrict late second-language acquisition, it is also conceivable that learning any language within the time-bounds of the critical period leaves the system open to subsequent language learning. Both scenarios are plausible, so the problem of time-bounded second-language learning is unresolved by an answer to the question for first-language acquisition. The opposite, however, is not true. If there is a critical period for second-language acquisition, then there is certainly a similar constraint for first language acquisition. It is logically impossible to restrict the time period for learning a second language without similarly constraining the primary language. Therefore, a positive decision about age-related constraints on second-language acquisition applies equally to the first language, and hence identifies a general restriction on a significant human learning mechanism. At the same time, the absence of a critical period for second-language acquisition leaves open the possibility that such a restriction does operate in first language acquisition.

For this reason, the interpretation of the data regarding the presence of a critical period in second-language acquisition must be conservative. Specifically, the burden of proof should lie with the view that there is a critical period; the null hypothesis that there is no critical period should only be rejected with clear evidence. As always, this conservatism is required to avoid a Type I error (claiming an effect when none exists), but the consequence of a Type I error in this case would be to misrepresent a fundamental aspect of human learning.

If there is a critical period that constrains second-language acquisition, then it follows that a significant portion of language acquisition is controlled by a dedicated modular function described in formal theories of both first- and second-language acquisition (Chomsky, 1986, 1995; Pinker, 1984; White, 1989). If the evidence fails to support the existence of such a biological constraint on language acquisition, then the options for language acquisition are more diverse but are based on a much larger role for general cognitive mechanisms and environmental influences throughout all stages of language acquisition. One theoretical alternative is the competition model developed by Bates and MacWhinney (Bates and MacWhinney, 1989; MacWhinney, 1997) that is based on connectionist principles. These models rely on input features of the language, so language learning is not the same for children learning different languages. This provision may be useful in explaining how adult second-language learners experience different learning paths as a function of the first language. Still, language learning is not unconstrained nor is it left entirely to experience. Elman et al. (1996, p. 31) state: "almost all connectionist models assume innate

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

architectural constraints, and very few assume innate representations.'' The formal models that are consistent with a critical period rely on innate representations of abstract structure.

Accepting the conclusion that there is a critical period for second-language acquisition, therefore, must be justified by relatively stringent criteria. Three kinds of evidence would provide rigorous support for the hypothesis. The first is that the measure of learners' L2 proficiency should show a discontinuous function across ages of acquisition. Newport (1991) notes that the close of the critical period is rarely abrupt, but a curvilinear function that clearly marks the period of heightened sensitivity to learning and the decline of that ability to some plateau is an essential feature. A continuous linear function that reveals a negative relation between proficiency and age of acquisition could be produced by several factors that do not implicate a critical period. If the diminished success of older learners is attributable to such factors as length of time studying the language, educational opportunities in the language, cross-linguistic influences from the first language, or motivational and attitudinal factors, then the decline in proficiency with starting age should be gradual across a large age span and should vary across speakers. However, if the age-related decline is because of a change in the learning mechanism, in other words, caused by an endogenously triggered critical period, then there should be an identifiable change in the shape of the proficiency function at around the close of the critical period, normally considered to be puberty. A large decrease in proficiency should be observed just beyond the close of the critical period. A stronger version of this prediction is that the majority of that decrease would be evident soon after the close of the critical period, and increasing distance from that barrier through the lifespan should not be related to systematic decline in learning. In sum, a non-linear relation between age and outcome is minimal and essential evidence for a critical period. Moreover, Johnson and Newport (1989, p. 93) argue that these effects are universal and should be the same for all first languages and for all relations between the L1 and L2.

A second type of evidence is the nature of the effect of the first language on the acquisition of the second. If there is a critical period, then there should be an asymmetry in this effect. Specifically, the first language should have an appreciably greater impact on the competence attained in the second language for older learners than for younger learners. Since learners in the critical period have access to the mechanisms involved in the acquisition of the first language, these mechanisms should apply to the second language with the same facility as in their application to the first language and minimise the impact of the first language on the (ultimate) competence of the second language. Learners within the critical period would use both specific language learning mechanisms and knowledge of other languages to build up their knowledge of the new system. In contrast, learners outside the critical period only have access to general learning mechanisms and knowledge of other languages, making them more vulnerable to the influence from the structure of the first language. Note that learners inside the critical period could well use transfer strategies and show an influence of one language on the other; the point is that the effect of this influence would be significantly *greater* for learners outside the critical period because they only have general cognitive mechanisms to guide their language acquisition.

A problem in obtaining this second type of evidence is in determining where to seek the effects of linguistic structure. Based on the linkage between a critical period and formal, nativist approaches to language development and competence, the most appropriate place to look would appear to be aspects of language learning that directly implicate central, abstract UG notions. One such aspect of second-language learning is parameter setting in which the values for parameters of Universal Grammar (UG) are established. Several recent studies argue that all or some parametric values from the first language (L1) serve as starting values for the second language (L2) (Eubank, 1993/94, 1996; Schwartz and Sprouse, 1994, 1996; Vainikka and Young-Scholten, 1996a, 1996b). Because establishing parametric values is a function of the language acquisition mechanism, learners inside the critical period should be more successful than learners outside the critical period in resetting parametric values carried over from the L1 which are inappropriate for the L2.

Parameter setting turns out to be highly problematic, however, as a testing ground for the expected asymmetries. One difficulty is the indeterminacy in the relation between surface patterns and their underlying analysis. Surface patterns can be compatible with two or more competing analyses of the appropriate underlying parametric values. This indeterminacy between surface patterns and the analysis of their parametric values could obscure differences in parameter (re)setting between younger and older speakers in L2 acquisition. It simply may not be clear from the surface manifestation of a given construction in the L2 which setting of a particular parameter is involved (Schwartz and Sprouse, 1996). As an example, examining the same data for the acquisition of order for verb and negative elements in English,

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

Schwartz and Sprouse (1996) and Eubank (1996) derive contradictory interpretations for the relevant parametric value.

There are also empirical reasons to suspect that UG principles and parameters are not the most fertile grounds for testing the hypothesis about a critical period. The literature is certainly mixed regarding the relation between UG and age of acquisition. For example, Johnson and Newport (1991) use the acquisition of subjacency as evidence for a critical period, while Juffs and Harrington (1995), studying a similar population of learners, find no age-related effects in acquiring this principle.

A component of second-language learning that appears more promising as a source of evidence for the predicted asymmetry in L1 effects for younger versus older learners is the acquisition of morphosyntactic categories, including inflectional categories and functional categories such as determiners. These are the categories that were used to demonstrate the existence of a critical period on L2 acquisition in the influential study by Johnson and Newport (1989). These categories reflect aspects of abstract linguistic structure but are clearly visible in surface constructions.

An asymmetry in L1 effects would be expected in circumstances in which the L2 has a morphosyntactic category not found in the L1. Younger learners should readily construct the new L2 category, whereas older learners should have limited success at best. Older learners have neither access to language acquisition mechanisms nor knowledge of the structure, so these categories are presumably difficult to learn. For categories that correspond across the two languages, older learners could profit from the similarity while younger learners may use either analogy with the similar L1 construction, or guidance from language acquisition mechanisms, or both, to master the category. For similar categories, therefore, the gap between older and younger learners would be narrowed, but for reasons that are not necessarily clear. This hypothesis has been explicitly formulated by Kellerman (1995, p. 229).

Finally, a third type of evidence is the relation between the type of L2 competence displayed by younger learners and that of native speakers. Because learning within the critical period should produce native-like competence, young learners should resemble native speakers in specific aspects of their ability. If such factors as testing procedure or specific linguistic categories influence performance of native speakers, they should have the same effect on learners within the critical period. Conversely, learners outside the critical period may demonstrate different influences of these factors. Evidence for this hypothesis would need to compare the effects of testing conditions on the performance of native speakers and second-language learners who are within or beyond the limits of a critical period. The pattern of method variance displayed by native speakers should be the same as the pattern shown by young learners but not necessarily the same as that shown by older learners. Two method features that may reveal such detailed aspects of competence are the mode of testing and the linguistic structure interrogated. Johnson and Newport (1989) showed that age effects on performance were observed for only some of the target structures included in the test items. Further, they found that the same task presented either orally (Johnson and Newport, 1989) or in writing (Johnson, 1992) produced different effects for age of acquisition. Specifically, few of the effects found for the oral task were replicated when the materials were presented in written form. Language competence should not be so fickle. Minimally, the performance of native speakers can be used as a baseline to evaluate the effects of these factors on second-language learners.

To summarise, three types of evidence are needed to reject the null hypothesis that there is no critical period for second-language acquisition. First, there should be a discontinuity in overall achievement that separates learners within the critical period from those outside it. Second, there should be an asymmetry in the effects of the first language on the acquisition of the second; learners inside the critical period should be equally prepared to acquire structures that differ between the two languages and those that correspond. Third, learners inside the critical period should produce the same pattern of results as native speakers in response to testing conditions such as presentation modality.

The study by Johnson and Newport (1989) has been a particularly influential contribution to the field. They demonstrated that native speakers of Chinese and Korean who had arrived in the United States at different ages showed distinct patterns of English proficiency as a function of their age of arrival. Using a grammaticality judgement test, they found that people who arrived before age 6 performed in the same way as native speakers, people who arrived between the ages of 7 and 15 revealed a negative correlation between age and performance, and those who arrived after age 15 exhibited no discernible relation between age and proficiency. Their explanation was that 15 years old marked a boundary that indicated the close of the critical period.

The results of their study are compelling: it appears that there is a clear difference in the success with which learners before and after 15 years old

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

Table 1. *Summary characteristic for participants in the three groups*

Language	Mean age	Age group	N	Mean age at testing	Mean arrival age	Range of arrival age	Mean length of residence	Range of length of residence
Chinese	24 yrs	Younger	15	22 yrs	10 yrs	1–15 yrs	8 yrs	5–18 yrs
		Older	18	27 yrs	23 yrs	17–32 yrs	4 yrs	1–6 yrs
Spanish	25 yrs	Younger	15	24 yrs	10 yrs	3–14 yrs	12 yrs	7–23 yrs
		Older	13	27 yrs	21 yrs	16–41 yrs	6 yrs	2–10 yrs
Native English	28 yrs		38	28 yrs				

learned English. Nonetheless, several aspects of their methodology and interpretation have elicited criticism (Bialystok, 1997; Bialystok and Hakuta, 1994; Kellerman, 1995). The present study builds on the work of Johnson and Newport (1989) by replicating aspects of their original study and extending it in an effort to obtain these three types of evidence. The major additions to the design were the use of contrasting first language groups, first language proficiency tests, and contrasting similarity values for the grammatical structures. The participants were native speakers of either Chinese or Spanish who began learning English as a second language at different ages. Participants were given a proficiency test in their native language to assure adequate competence to claim they were bilingual. A grammaticality test was constructed based on a comparison of Chinese and English. The test sentences were designed so that they contained structures that were either similar in both English and Chinese grammar or different in the two languages. The grammatical features used in constructing the sentences were taken from those investigated by Johnson and Newport (1989), but new sentences were created so that we could better control for the effect of grammatical similarity. Finally, following Johnson and Newport (1989) and Johnson (1992), the test was presented in both an oral and written version.

Method

Participants

There were three groups of participants in the study. All participants were university students or university graduates. The first group consisted of 33 adults who spoke Chinese (Mandarin) as a first language and learned English as a second language. The mean age of participants in this group was 24 years (range 19 to 37 years). Following the study by Johnson and Newport (1989) upon which our design was modelled, the age for dividing participants into older or younger learners was 15 years. Consequently, the first group included 15 younger learners and 18 older learners; age of learning English was considered to be the age at which they arrived in Canada. The second group consisted of 28 adults, with a mean age of 25 years (range 18 to 47 years), whose first language was Spanish. When divided by age of arrival, this produced a group of 15 younger learners and 13 older learners. The third group consisted of 38 native speakers of English. The mean age of this group was 28 years (range 21 to 43 years). Summary information about the age of arrival and length of residence for each group is presented in Table 1.

Task and Procedures

Language questionnaire

All participants completed a Personal Data Sheet. The questions asked about date of birth, country of origin, native language, age of arrival in Canada, language(s) spoken at home in native country, language(s) spoken at home in Canada, age when English instruction began, and TOEFL score (if applicable).

First-language proficiency test

For all participants who were not native speakers of English, proficiency tests in the native language were administered to assure reasonable competence in the native language. If participants who began learning English at a young age did not demonstrate first language competence, it might be argued that English was in fact their first-language and the native language was functionally a second language. Thus they would achieve high scores on the English grammaticality judgement task, but these scores would not necessarily contribute to an investigation of the role of a critical period in second-language acquisition. The Chinese test was in Mandarin. Both the Chinese and English tests had two parts: comprehension-translation and writing.

In the comprehension-translation component of

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

the tests, participants listened to a sentence read on tape in the native language and selected the correct translation from three alternatives. All the alternatives used the same main words. For example, after hearing a Chinese sentence, the following three options were presented:

(a) He goes to the library to study every day.
(b) His sister went to the library to study yesterday.
(c) He doesn't like to go to the library to study.

An example from the Spanish test is:

(a) The doctor told me not to take more than four aspirins.
(b) The doctor told me to take at least four aspirins.
(c) The doctor thought I did not need four aspirins.

Each language test contained 10 sentences in this section.

The second part of the test was different for each of the languages. For the Chinese test, a list of 10 compound words was read on tape and participants were required to write each word in Chinese characters. The ability to do this assured that the subject had more than a casual or conversational knowledge of Chinese. In the Spanish writing test, 10 sentences were read in English and participants were required to produce a Spanish translation.

The comprehension part of the test produced a score out of 10, one point for each sentence. The written section was scored by allotting 2 points for each of the 10 items, allowing participants to obtain part marks. These were used when a written response was not perfect but was close to being correct. A score of 2 indicated that there were no errors at all in the written item. In total, then, the first language test produced a score out of a possible total of 30.

The first language proficiency tests were different for the two languages but were designed by university language instructors and judged to be suitable for intermediate level students. Each language requires slightly different testing procedures, and since strict comparison of the performance across the two languages was not necessary, this did not seem to be a problem.

Grammaticality judgment task

Participants were presented with a set of 160 sentences and asked to decide whether each sentence was correct or contained a grammatical error. The errors were based on one of five structures: plurals, determiners, future tense, present progressive, and collocation restriction. These terms are based on traditional English grammar and are intended as classificatory labels for English construction types that have overt morphological representations. There were 16 examples of each structure, and each appeared in its grammatical and ungrammatical form. All sentences contained between 6 and 9 words and were composed of between 10 and 15 syllables. Examples of sentences from each error type are provided in the appendix.

The 160 sentences were divided into two sets. In Set A, half of the sentences (half from each category) were grammatical and half were ungrammatical. In Set B, the grammaticality of each sentence was reversed. All participants completed the task twice, once with oral presentation of the sentences and once with written. The order in which the oral and written versions were presented was counterbalanced. In addition, for half of the participants, the oral condition was presented in Set A and the written in Set B; for the other half of the participants, this was reversed. Therefore, participants made judgments about 80 items in the oral condition and 80 items in the written condition.

The sentences were developed with an explicit effort to control the relationship between the way in which the structures are formed in English and the way in which the corresponding structures are formed in Chinese. For two of the structures, future tense and present progressive, both languages employ similar means of construction. In the English future tense that was tested, the construction is marked by a free morpheme (the modal auxiliary *will*) to the left of the main verb, with the main verb appearing in its bare stem form. This same pattern occurs in Chinese: the morpheme corresponding to *will* as a marker of future time reference is *hui*. The present progressive construction is also similar in the two languages for morphology and syntax. The English construction involves two morphemes: the auxiliary verb *be*, which occurs to the left of the main verb, and the present participle suffix *-ing*, which occurs on the verb immediately following the auxiliary. Each of the morphemes has a counterpart in Chinese; the counterpart of *be* is the free morpheme *zheng*, and the counterpart of *-ing* the verbal suffix *-zhe*. Further, the morpheme *zheng*, like its counterpart *be*, occurs to the left of the main verb, and *zheng* and the suffix *-zhe* can co-occur in a pattern similar to that for *be* and *-ing*. A difference between the two languages, however, is that the two morphemes *zheng* and *zhe* are not required to co-occur.

For the structures determiner and plural, the two languages differ in their means of construction. Unlike English, Chinese has no system of determiners marking a definite–indefinite contrast and no specific morpheme for indicating plurality.

Finally, the category of collocation restriction does not have a general correspondence between the

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

two languages. The category (corresponding to subcategorization in the study by Johnson and Newport, 1989) refers to restrictions that verbs place on the types of constituents with which they co-occur. For some verbs, the Chinese counterpart shows a similar pattern, but for other verbs, the pattern of the Chinese counterpart differs. Hence, the category of collocation restriction is divided into two subtypes: CR – Same (CRS) and CR – Different (CRD). The 16 sentences for this category were developed so that there were eight for each subtype.

The relationship between Spanish and English is largely the opposite of that between Chinese and English. The future tense, which is similar for Chinese and English, is different for Spanish and English. The Spanish future is marked by an affix on the verb and requires no free-morpheme tense marker analogous to the English modal *will*. The present progressive, also similar for Chinese and English, has two means of construction in Spanish, one of which is similar to that in English. One of the constructions, labelled the simple present, is marked by an affix on the verb, making it different from both English and Chinese. This construction serves a wider range of function associated with the English present progressive. The two categories that are different for Chinese and English are similar for Spanish and English. Like English, Spanish requires a determiner to precede count nouns and indicates plurality by adding a morpheme to the noun. The collocation restrictions are different again for Spanish but not systematically enough to include in the classification of structures as being overall similar or different.

The two presentation modalities were designed to be as comparable as possible. In both cases, participants were seated in front of a Macintosh computer and were given two responses keys, one on each side of the keyboard. In both cases the sequence of information presented on the computer leading up to the subject's response was similar.

For the oral condition, participants wore a set of headphones attached to the computer. A beep was sounded, followed by a sentence read by a female voice. While the sentence was read, a small cross appeared in the centre of the computer screen. Immediately following the sentence, a water drop sound was played and two yellow faces appeared in the bottom corners of the screen. The left face was a "happy face" (☺) signifying a "yes" (or grammatical) response; the right face was a "sad face" (☹) signifying a "no" (or ungrammatical) response. The faces remained on the screen until the subject responded. The computer recorded both the subject's response and the reaction time from the point at which the faces appeared. The reaction time clock was not activated until the water drop sound was heard, so responses attempted before the end of the sentence were not recorded.

The written condition was similar. Each trial began with a beep, but in this case it was followed by the sentence printed in a single line in the middle of the screen. The sentence remained on the screen for 4 seconds, a time judged by pilot testing to be comparable to the time required to listen to the sentence in the oral condition and to read the sentence in this written condition. Following this, the water drop sounded, initiating the reaction time clock, and the two faces appeared. Again, the faces remained on the screen until the subject responded, and both selections and reaction times were recorded.

Half of the participants began with each presentation modality. Each session began with a set of six training items for that condition. Participants were told to respond as quickly as possible but to maintain accuracy. If there were no questions, the task began with the first set of 80 test items. Participants took a break between the two conditions and then completed the other modality, starting again with six new training items.

Results

Language Questionnaire

The countries of origin for the Chinese participants, in decreasing order of frequency, were China, Hong Kong, Singapore, and Taiwan). (One individual from Hong Kong spoke both Cantonese and Mandarin.) The Spanish participants, in decreasing order of frequency, came from El Salvador, Spain, Chile, Argentina, Ecuador, Colombia, Nicaragua, Guatemala, Venezuela, Costa Rica, and Uruguay. The main finding from the language questionnaire is that participants had virtually no experience with English before arriving in Canada.

First-language proficiency

The scores on the L1 proficiency tests by the two groups of bilingual participants are reported in Table 2. These scores are out of a maximum of 30. There was an effect of age, $F\,(1,55) = 14.78$, $p < .0003$, and an interaction of age and group, $F\,(1,55) = 5.15$, $p < .02$. The Chinese native speakers who began learning English at an earlier age did somewhat worse on a test of Chinese than did older learners. There was no reliable difference in Spanish knowledge that resulted from age of acquisition of English.

Another way of considering the effects of profi-

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

Table 2. *Mean score out of 30 and standard deviation on L1 proficiency tests*

Language group	Age of arrival	
	Younger	Older
Chinese	25.0 (2.48)	28.4 (1.71)
Spanish	27.2 (2.90)	28.1 (1.17)

ciency in the first language is to examine the relation between L1 competence and performance in the grammaticality judgement task. For the Chinese group, there was a negative correlation between performance on the Chinese proficiency test and performance in the oral condition of the judgement task, $r = -.63$, $p < .0001$. In other words, higher proficiency in Chinese was associated with lower oral proficiency in English. There was no relation between performance on the Chinese test and scores on the written judgement task. For the Spanish group, there was no significant correlation between Spanish proficiency and judgement scores, although the relation with the oral task was negative, as it was with the Chinese participants, $r = -.21$, n.s.

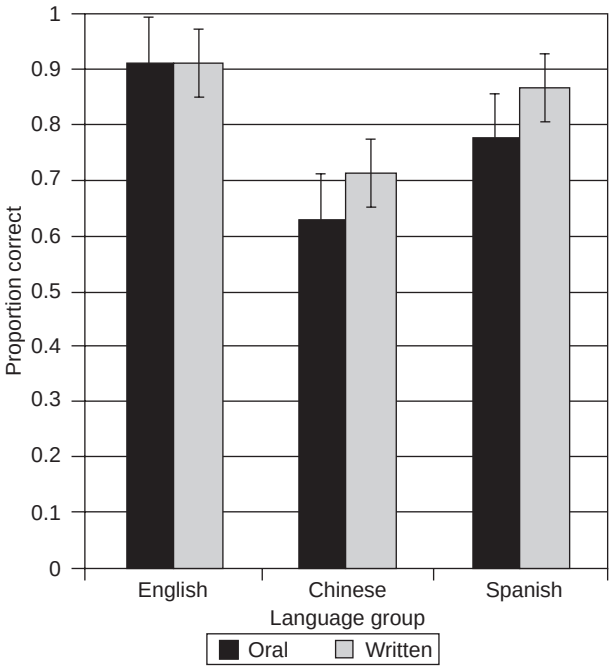


Figure 1. Overall accuracy by group and modality.

Grammaticality judgement

The data set for the grammaticality judgement task includes results for two dependent variables (accuracy and reaction time) for the between-subjects factors group (native English, Chinese, Spanish) and age of arrival (younger, older) and the within-subjects factors grammaticality (grammatical, ungrammatical), modality (oral, written), and structure type (plural, determiner, future, present progressive, and collocation restriction). Because the sentences were created to reflect specific correspondences between English and Chinese, this last category can be interpreted further – as collocation restriction same (CRS) and collocation restriction different (CRD) – for the Chinese group. In addition, the structure type factor can be interpreted to reflect the similarity relations between Chinese and English by considering the structures that are similar (present progressive, future, CRS) or different (plural, determiner, CRD). These are conceptual categories used to interpret the results of the data analysis and are not used to reanalyse or recombine the data. The data are presented separately for each of the two dependent variables, accuracy and reaction time, and individually for each of the three subject groups. The important point is to compare patterns of performance, not absolute levels of performance, so there is no need to combine all participants into a single analysis. There is one analysis, however, that takes the liberty of collapsing across the factors of grammaticality and modality in order to get an overview of differences between groups. Because those two factors are not included in the analysis, the interpretation of the groups analysis is merely suggestive.

Accuracy Data

(a) Overall Performance Across Groups

A 2–way anova for group (3) by modality (2) was conducted to discover the overall patterns of responding across the three groups of participants. The data are plotted in Figure 1. There were main effects of group, $F(2,94) = 58.35$, $p < .0001$, and modality, $F(1,94) = 54.06$, $p < .0001$, and an interaction between them, $F(2,94) = 14.21$, $p < .0001$. Using a Scheff ☺ analysis with an α-level of .01, it was found that the written condition was easier for both learner groups but the two modalities were the same for the native speakers. The Spanish learners did as well as the native speakers in the written items but all three groups were different from each other in the oral items, with English participants scoring the highest and Chinese participants the lowest.

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

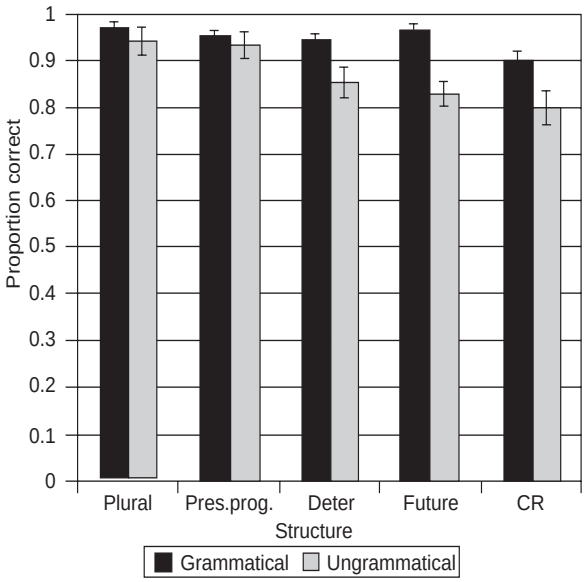


Figure 2. Accuracy by structure and grammaticality for the native speakers.

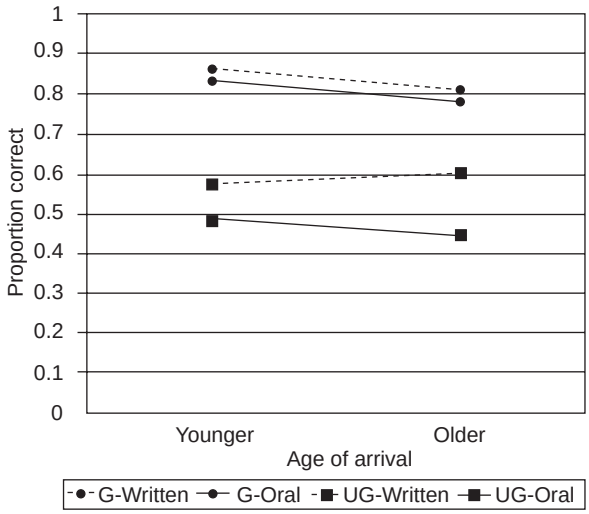


Figure 3. Accuracy by mode and grammaticality for the Chinese group.

(b) English monolinguals

The data from the English native speakers were analysed by means of a 3-way anova for grammaticality (2), modality (2), and structure (5). These data are plotted in Figure 2. There was a main effect for grammaticality, F (1,37) = 16.12, $p < .0003$, grammatical items being easier than ungrammatical ones. There was a main effect of structure, F (5,33) = 15.09, $p < .0001$, as well as an interaction between grammaticality and structure, F (5,33) = 14.97, $p < .0001$. The interaction occurs because all of the variance for the structure effect is produced by the ungrammatical sentences, F (5,33) = 18.87, $p < .0001$. This is logical because participants cannot know by looking at a grammatical sentence how it *could* be made ungrammatical. The absence of a structure effect on these sentences shows that they are not inherently different from each other. Post hoc contrasts using an α-level of .01 showed that the structures in the ungrammatical sentences were organised into two levels of difficulty: the easier were those that violated plural and present progressive rules, the more difficult sentences included violations of the rules governing determiners, future tense, and collocation restrictions. There were no main effects or interaction effects for modality.

(c) Chinese–English bilinguals

A 4-way anova for grammaticality (2), modality (2), structure (6), and age of arrival (2) was conducted on the accuracy data for the Chinese–English bilinguals. There were main effects for grammaticality, F (1,31) = 104.02, $p < .0001$, and modality, F (1,31) = 24.14, $p < .0001$, with both grammatical sentences and written presentations being easier than their counterparts. An interaction of grammaticality and modality, F (1,31) = 5.40, $p < .02$, however, indicates that the difference between oral and written presentation was evident only for the ungrammatical sentences, F (1,31) = 16.95, $p < .0003$. This interaction is shown in Figure 3. The figure also includes the age of arrival even though it was not significant in the analysis.

There was a main effect of structure type, F (5,27) = 6.34, $p < .0005$ and an interaction of grammaticality and structure, F (5,27) = 12.45, $p < .0001$. The effect of structure in the ungrammatical sentences, F (5,27) = 14.96, $p < .0001$, is plotted in Figure 4 (structure is uninterpretable in grammatical sentences). Scheff☺ tests revealed two distinct groups of structures. Those in the easier group were future, present progressive, and CRS; those in the more difficult were plural, determiners, and CRD. These groups correspond to the similarity classifications between Chinese and English. This grouping was confirmed by a supplementary 2-way anova for similarity (2) and age of arrival (2) that showed a reliable effect for similarity, F (1,31) = 20.87, $p < .0001$, and no effect or interaction for age of arrival.

Two kinds of data are relevant for the assessment of the relation between proficiency and age of arrival. First, to establish qualitative differences in overall proficiency as a function of immigrating at a younger

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

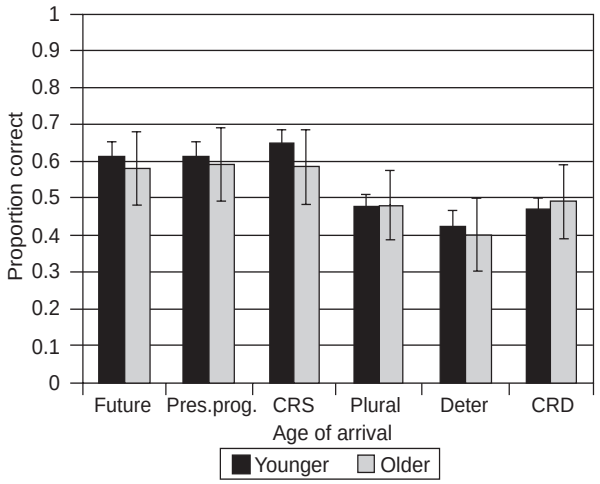


Figure 4. Accuracy by structure in the ungrammatical items for the Chinese group.

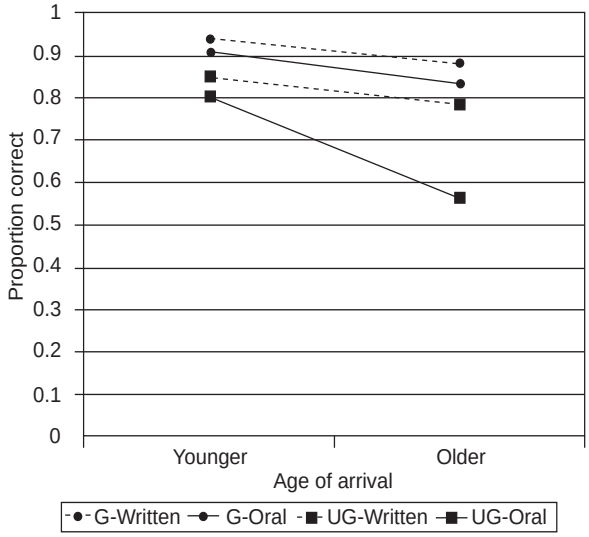


Figure 5. Accuracy by mode, grammaticality, and age for the Spanish group.

or older age, the effect of age of arrival in the anova is examined. There were no main effects or interactions for age of arrival in this analysis. Second, to assess a continuous effect of age of arrival on proficiency, a correlation coefficient was calculated for age of arrival and each of the total oral and written scores. For the oral items, there was a significant relation between age and performance for both the younger, r = -.81, p < .0002, and older, r = -.57, p < .01 learners. For the written items, there was a relation for the younger learners, r = -.71, p < .002, but not the older ones. Using length of residence as a variable, there was a significant relation between length of time in Canada (hence, length of time using English) and performance in the oral items for the younger learners, r = .84, p < .0001, but not the older ones. Similarly, there was a significant relation between length of time in Canada and performance in the written items for the younger learners, r = .78, p < .0007, but not the older ones. Hence, oral proficiency is sensitive to age of arrival throughout the sample, but written proficiency reaches a plateau for the older learners. Nonetheless, there were no overall qualitative differences between younger and older learners.

(d) Spanish–English bilinguals

A 4-way anova for grammaticality (2), modality (2), structure (6), and age of arrival (2) was conducted on the accuracy responses for the Spanish group. There were main effects for grammaticality, F (1,24) = 43.55, p < .0001, and modality, F (1,24) = 54.54, p < .0001, and an interaction between them, F (1,24) = 23.19, p < .0001. This interaction is shown in Figure 5. The difference between oral and written presentation for the grammatical sentences was small but still significant, F (1,24) = 7.60, p < .01. There was a larger effect of modality for the ungrammatical sentences, F (1,24) = 60.29, p < .0001. These effects, grammaticality and modality, enter a three-way interaction with age of arrival, F (1,24) = 12.05, p < .002. The older Spanish learners of English found the ungrammatical sentences disproportionately more difficult than all other items when they were presented orally.

It is worth noting that a comparison of the Chinese (Figure 3) and Spanish (Figure 5) performance reveals that the patterns of response to the grammatical sentences were similar for the two groups (top two lines), even though the Spanish levels were higher. (Scheff☺ test showed this difference to be significant.) In contrast, the two groups handled the ungrammatical sentences very differently, indicating different types of knowledge of English structure. Nonetheless, the written presentation was advantageous for both groups.

Returning to the Spanish learners, there was a main effect of structure, F (5,20) = 7.94, p < .0003, and an interaction of structure and grammaticality, F (5,20) = 7.45, p < .0004. Again, only the ungrammatical sentences varied significantly with structure and only these will be considered, F (5,20) = 19.37, p < .0001. These data are shown in Figure 6. Contrasts showed that the present progressive sentences

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

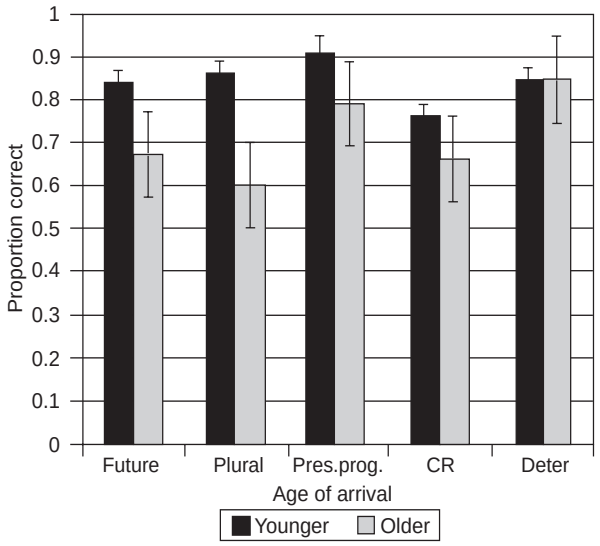


Figure 6. Accuracy by structure in ungrammatical sentences for the Spanish group.

were easier than all other items, none of which differed from each other. There was no interaction with age of arrival, so the patterns were not affected by this factor. There was, however, a main effect of age of arrival, F (1,24) = 8.29, $p < .008$, in which those in the younger group scored higher than those who arrived at an older age.

The effect of age of arrival is again examined in two ways. First, the anova reveals a difference in overall performance between older and younger learners. Second, the correlation between age of arrival and the total oral score produced a significant relation for each of the younger, $r = -.68$, $p < .01$, and older, $r = -.51$, $p < .02$, learners. There was no relation between age of arrival and written performance for either group. Using length of residence, there was a significant relation between length of time in Canada and performance in the oral items for the younger learners, $r = .59$, $p < .03$, but no relation between these variables for the older learners. There was no relation between length of residence and performance on the written items for either group. Therefore, like the Chinese–English bilinguals, these language learners showed a continued sensitivity to age of arrival for the oral task throughout the sample, but unlike the Chinese–English bilinguals, there was an overall qualitative advantage for younger learners. The triple interaction, however, restricts this difference to the ungrammatical sentences in the oral condition.

(e) Age of arrival and overall oral proficiency

In the paper by Johnson and Newport (1989), a convincing part of their argument is presented in a figure that shows the overall performance of the groups as a function of four different ages of arrival. Their categories for age of arrival are 3 to 7 years (N = 7), 8 to 10 years (N = 8), 11 to 15 years (N = 8), and 17 to 39 years (N = 23). The function displayed in the figure they report shows that the first group is equivalent to native speakers, and each subsequent group shows a continuous and significant decrease in performance. Apart from the obvious problem of the unequal Ns in the four groups, an additional problem is that three of the four groups are subdivisions of the younger arrivals, and the fourth group is the intact set of older arrivals. Johnson and Newport (1989) correctly point out in their paper that a critical period would not be supported if proficiency continued to decline beyond the putative critical period. However, their graph does not allow a clear examination of that question, because all of the post-pubescent learners are represented by a single data point.

Following Johnson and Newport (1989), we plotted the overall oral proficiency scores on a single graph using four different ages of arrival. Unlike Johnson and Newport, however, we divided each of the younger and older learners into two subcategories, creating a more even distribution of age of arrival. Because of our distribution of participants, it was impossible for our categories to conform precisely to those used by Johnson and Newport. The categories we used were 1 to 8 years (Chinese N = 4; Spanish N = 6), 9 to 15 years (Chinese N = 11; Spanish N = 9); 16 to 22 years (Chinese N = 11; Spanish N = 10); 23 to 33 years (Chinese N = 7; Spanish N = 3). This function is shown in Figure 7.

Applying the same pairwise series of t-tests to adjacent scores as was done by Johnson and Newport (1989), the results for the Chinese groups were that the native speakers were minimally but significantly different from the youngest learners, t (7.8) = 2.47, $p = .04$, the youngest learners were different from the next group, t (10.9) = 7.69, $p = .0001$; this group was not different from the older group, t (14.9) = .16, $p = .86$, nor were the two older groups different from each other, t (4.7) = 1.34, $p = .24$. For the Spanish group, the differences between native speakers and the youngest learners was not significant, t (21.3) = .89, $p = .37$, between youngest and the next group was significant, t (11.2) = 3.91, $p = .002$, between this and the first of the older learners was not significant, t (13.3) = 1.85, $p = .08$, and between this and the oldest learners was also not significant, t (4.7) = 1.34, $p = .24$.

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

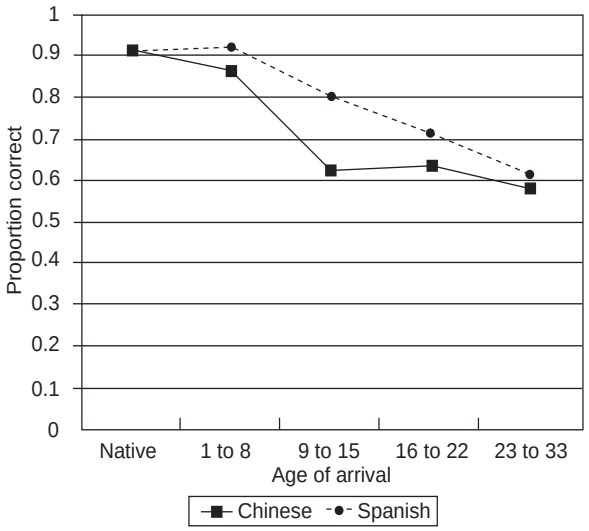


Figure 7. Overall performance on oral task by age of arrival for four groups.

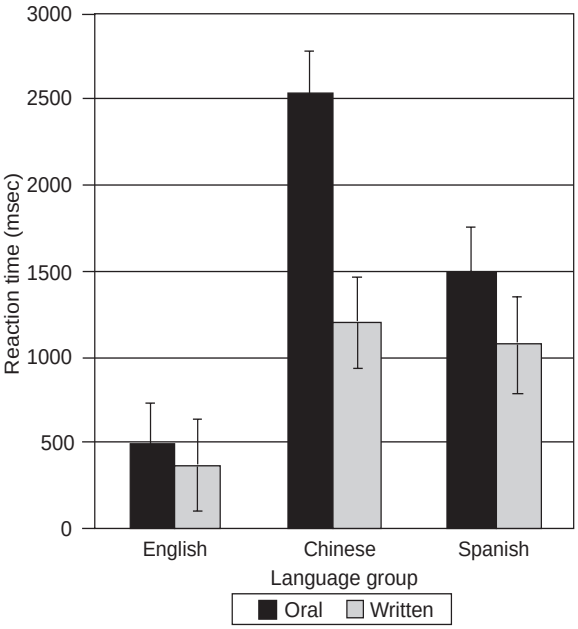


Figure 8. Reaction time by group and modality.

Up until 8 years old, learners in both groups perform with the same proficiency as native speakers (cf. Johnson and Newport, 7 years old). Statistically, both groups show a change in proficiency when age of arrival is greater than 8 years old, although visually the performance of the Spanish group continues to decline throughout the sample. These patterns are consistent with one of two interpretations. First, it may be that there is a critical period that ends at 8 years old, but this age has not previously been suggested as the relevant cut-off point. Second, there is no critical period and other factors explain the changing proficiency. Neither group showed a change in proficiency when age of arrival was about 15 years old, as reported by Johnson and Newport (1989). Nonetheless, this analysis is based on small and unequal numbers in each cell and should not be over-interpreted.

Reaction time data

(a) Overall performance across groups

A 2-way anova for group (3) and modality (2) was conducted using the reaction time to respond as the dependent variable. There were significant effects of group, $F(2,94) = 26.29$, $p < .0001$, modality, $F(1,94) = 44.34$, $p < .0001$, and the interaction between them, $F(2,94) = 15.74$, $p < .0001$. This interaction is plotted in Figure 8. For the written condition, the Chinese and Spanish participants took the same time, and were significantly slower than the native speakers. For the oral condition, all three groups were different from each other. The order of increasing reaction time was English, Spanish, and then Chinese participants. These results parallel those obtained with the accuracy data.

(b) English monolinguals

A 3-way anova for modality (2), grammaticality (2), and type (5) was conducted on the reaction time data for the native speakers. The only effect was modality, $F(1,37) = 35.87$, $p < .0001$, indicating that participants responded to the written items reliably faster than the oral ones.

(c) Chinese–English bilinguals

The reaction times of the Chinese participants were examined by a 4-way anova for modality (2), grammaticality (2), structure (6), and age of arrival (2). There was a main effect of modality, $F(1,31) = 24.41$, $p < .0001$, indicating that the reaction times to the written items were faster than those to the oral ones. There was also a small but significant interaction of grammaticality and age of arrival, $F(1,31) = 4.87$, $p < .04$, which must be interpreted in terms of the three-way interaction between grammaticality, age of arrival, and modality, $F(1,31) = 4.40$, $p < .04$. This interaction is plotted in Figure 9. For the written items, the older learners took slightly longer than the younger learners to respond, but grammaticality was irrelevant. For the oral items, the speed of response by each age of arrival group changed with the grammaticality of the items: The younger learners

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

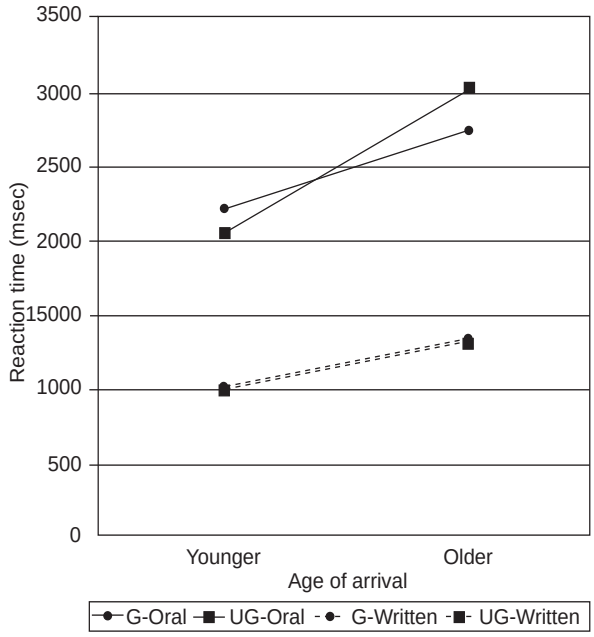


Figure 9. Reaction time by grammaticality, modality, and age for the Chinese group.

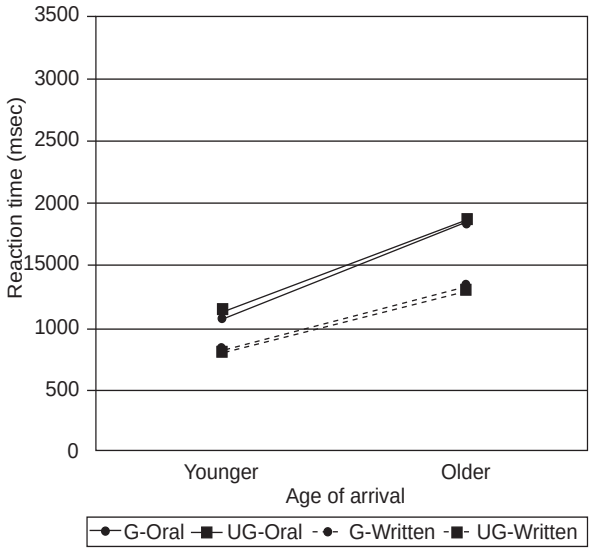


Figure 10. Reaction time by grammaticality, modality, and age for the Spanish group.

responded (slightly) faster to the ungrammatical items while the older learners were faster for the grammatical ones.

(d) Spanish–English bilinguals

The same 4-way anova as used for the Chinese results was applied to the reaction times for the Spanish participants. There were main effects for modality, F (1,27) = 17.2, $p < .0003$, repeating the finding that the written items were faster than the oral ones, and age of arrival, F (1,26) = 7.16, $p < .01$, showing that the younger learners responded more quickly than the older ones. There was a modest effect of structure, F (4,21)=3.15, $p < .03$, because the fastest type, present progressive, was significantly different from the two slowest types, plurals and collocation restrictions. This pattern parallels the results of the accuracy data. These results are displayed in Figure 10.

Discussion

Three types of evidence were identified as necessary to reject the null hypothesis that there is no critical period for second-language acquisition and accept the alternative hypothesis that there is. The first evidence comes from comparing learning outcomes by different learner groups, defined by age of acquisition of the target language, across a putative boundary that corresponds to some reasonable interpretation of a critical period. For the present study, the boundary was adopted from the study by Johnson and Newport (1989). These age-related performance patterns should be evident irrespective of the first language of the learners. Second, if there are effects of specific language structure from the L1, they should emerge primarily for older learners and be most apparent in the acquisition of forms that differ between the two languages. Finally, performance variance attributable to method factors, such as task modality, should mimic that of native speakers. This would be particularly true for younger learners, who had presumably learned the second language in a more "natural" manner and would therefore more closely resemble native speakers. We shall discuss each of these three types of evidence in turn.

Effects of subject groups

Participants in the study from the two language groups were similar in their background characteristics, although the Spanish–English bilinguals have been learning English longer than the Chinese–English bilinguals. Moreover, all the learners demonstrated high levels of competence in their native language, confirming the role of English as a second language, especially for those who arrived at an early age. Nonetheless, the patterns of performance exhibited by the two groups were different from each other.

Overall, the Spanish native speakers were func-

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

tioning at a higher level of English proficiency than were the Chinese speakers. There may be several reasons for this. The first possibility is that the longer length of residence resulted in more advanced language skills for the Spanish group. This difference alone, however, cannot account for the results because of the complexity of the effect of length of residence, which is discussed more fully below. In any case, the addition of an average of four years of length of residence did nothing to improve the scores of the younger Chinese–English bilinguals, but the additional six years in length of residence for the younger Spanish–English bilinguals did produce higher levels of performance. It is unlikely that two more years' residence for the Chinese group would produce a qualitative advance in their proficiency. It is also possible that sampling variance produced a more talented group of language learners for the Spanish group than for the Chinese. More probably, however, the nature of the languages makes it easier for Spanish-speakers to learn English than for Chinese-speakers, given equivalent amounts of time and opportunity. Indeed, the time required for learning a given language as a function of typological similarity between the L1 and L2 is formalised in the Berlitz school programme (Bialystok and Hakuta, 1994).

There were also qualitative differences between the groups in the patterns of problem difficulty. For the Chinese–English bilinguals, age of arrival was a significant factor in determining proficiency only in a three-way interaction with grammaticality and modality for the reaction-time data. Although this interaction did not translate into overall proficiency differences, the older and younger learners were faster to respond to different grammaticality values for the oral items. Specifically, the younger learners were faster for the ungrammatical items but the older learners were faster for the grammatical items. It is possible that this ordering for the younger learners is unreliable. In addition to being only a small difference, it is the only case in the data set in which the grammatical items were not either more accurate or faster than the ungrammatical items. At the very least, it provides little basis for concluding that age of arrival is influential in shaping the pattern of performance by the Chinese native speakers. At the same time, oral proficiency remains correlated with age of arrival throughout the sample, but this relation does not translate into qualitatively different levels of achievement for younger and older language learners.

Age of arrival was consistently influential in determining the performance of the Spanish–English bilingual group, both for qualitative levels of achieved proficiency and an overall relation between age of arrival and proficiency. These effects were reflected in both the accuracy and the reaction time data. Ungrammatical sentences were particularly difficult for the older learners to detect (see for example, Figure 5). A similar result was reported by Birdsong and Molis (1998) who replicated the Johnson and Newport (1989) study with Spanish native speakers and found age of arrival effects at all ages of initial learning. This effect was seen as well in Figure 7. For each increment of age of arrival, there was a decline in performance by the Spanish–English bilinguals, but beyond the age of 9 years, there was no apparent change in the performance of the Chinese–English bilinguals.

Why would age of acquisition have a different role in the performance patterns for the two learner groups? If there were a critical period for learning a second language, one would expect the evidence for it to be broadly based, almost universal. The mechanism for human language learning in childhood is indifferent to which language a child attempts to learn, so efforts at language learning while the natural mechanisms are functional, in other words, during the critical period, should be equally effective in acquiring *any* language. The fact that the younger Spanish learners appeared to be privileged in learning English but the younger Chinese learners were not rules against an interpretation that is based in biology. Instead, the systematic differences between language groups points to the role for more experiential and language-related factors in determining the outcomes of learning, at all ages of second-language acquisition.

The discontinuity of learning outcomes between older and younger learners was identified as an important source of evidence for claiming a critical period. The critical period sets a qualitative boundary on learning, and this boundary should be readily visible by behavioural evidence. For the oral task, there was a significant negative relation between age and proficiency for both older and younger learners in both language groups. For the written task, only the younger Chinese learners revealed a significant negative correlation with performance. Certainly, this would be evidence for a discontinuity in the written task for the Chinese learners, but other corroborating evidence is absent. For example, it would be helpful to the argument if the younger Chinese learners also scored *higher* than the older learners did on the written task, but they did not.

To summarise, there were large differences between the two language groups in their responses to the test items. There were few differences, however, between the older and younger learners. For the Spanish group, age of acquisition was sig-

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

nificant in determining proficiency outcomes, but for the Chinese group it was not. Moreover, the correlation coefficients showed that age continued to influence behaviour in the oral task throughout all the ages studied, even though it did not mark achievement differences between age groups for the Chinese learners. In fact, for the Chinese learners there was some evidence that English proficiency came at the *expense* of Chinese, indicated by the negative correlation between scores on the L1 test and the grammaticality judgement task. Thus, the first type of evidence, namely a discontinuity in performance as a function of a critical period boundary was not found. Moreover, large differences between the two language groups cast further doubt on the interpretation that, even for the younger learners, language acquisition is being controlled by some language-learning mechanism.

Effects of structure

There were differences in participants' ability to respond correctly to the grammaticality of sentences whose errors were caused by violations of different rules. Even the native speakers found that the grammatical structures differed in their detectability, with present progressive and plural violation easier to identify than the others. This pattern can be used as a baseline against which to compare the two learner groups.

The present progressive sentences were significantly easier for all groups, as measured by both accuracy and reaction time data. It is possible that these sentences were intrinsically easier, so the success of the learner groups in detecting these errors should not be interpreted solely as an indication of first-language influence. Although this English structure corresponds to the Chinese form, it corresponds to only one of two possible Spanish equivalents. Nonetheless, these sentences were the easiest of all for the Spanish native speakers. In contrast, the native speakers also found the plural sentences to be particularly easy, but none of the learners did. These correspond to structures in Spanish and not in Chinese, but they were difficult for both groups. Hence, the response patterns of the native speakers point to some differences in inherent difficulty of the sentences but do not explain the performance of the two learner groups.

For both learner groups, there was a significant effect of linguistic structure, but the patterns were different. Evidence for a critical period requires a difference in the manifestation of these effects, with younger learners being less susceptible to similarity patterns than are older learners. Nonetheless, none of the effects of structure interacted with or were influenced by age of learning. Hence, structural effects are an independent factor in determining learning outcomes. Whatever effect structural similarity exerts on learning, it does not change with the age of the learner.

The effect of structure was manifested in two ways. First, the present progressive sentences were simply easier than the others, so irrespective of contrastive differences between the languages, these items produced higher accuracy scores and lower reaction times than the others. Even the native speakers demonstrated this effect. Second, the contrastive relation between the structure in English and in the first language was a significant factor in determining performance for the Chinese speakers but not particularly for the Spanish speakers.

This difference points again to a contrast between the two language groups that bears critically on how the data can be interpreted. For the Chinese group, the similarity between the structures in the two languages accounted for a significant portion of the variance. For the Spanish group, very little impact of structural similarity was found. The design of the study is partly responsible for this asymmetry. The sentences were constructed to create specific similarity patterns between Chinese and English; Spanish was added as a reasonably appropriate language in which these relations were reversed. Because the sentences were not developed with the same sensitivity to Spanish grammar as to Chinese, it is understandable that these structural relations played a smaller role in determining the results.

Another reason may be found in the overall proficiency levels of the two groups. As noted, the Spanish learners were more advanced in English proficiency than were the Chinese. In fact, the older Spanish learners performed at a higher level than even the younger Chinese learners on the written items (.83 for Spanish; .71 for Chinese), although these two groups were the same for the oral sentences (.69 for Spanish; .67 for Chinese). It is possible that the linguistic effects of structural comparison are most evident in the earlier stages of language learning, or at lower levels of proficiency in the language. The Spanish learners in this study may have passed that boundary and had become impervious to details of linguistic structure simply because their English mastery was so advanced. One way of examining this issue would be to repeat the procedure with a set of Spanish native speakers who are at an earlier stage of learning English. Note, however, that even the advanced Spanish speakers did not respond to the structure differences in the same way as native speakers.

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

To summarise, the critical period hypothesis would predict that the effect of the relation between structures in the two languages would be most apparent for the acquisition of the L2 by older learners, and that structures that differ between the languages should be most discriminating of performance. The data showed that there were strong effects of similarity for the Chinese learners, but that these occurred equally for learners of all ages. For Spanish, there were virtually no effects of structural similarity regardless of the age of acquisition. Because of the test design, however, the Spanish data should be interpreted only suggestively. Nonetheless, the pattern of results across the whole sample necessary to support the claim that there is a critical period was not obtained.

Effects of modality

Throughout all the data, there were reliable effects of presentation modality. The same test items presented in an oral or written form and responded to in the same manner led to consistent differences in performance. The written task produced more accurate and more rapid responses than did the oral task for the two learner groups. The native speakers showed no difference at all in accuracy between the two presentation modalities, but they did reply more rapidly to the written items than the oral ones. There may be a built-in difference between the modalities that favours speed to respond to written presentation. Since there were no accuracy differences between the modalities for the native speakers, the reaction time difference has no implications for competence. The important point, however, is that both learner groups differed from the native speakers in this pattern. For both groups of learners at all ages of acquisition, there were reliable differences in accuracy as a function of modality. All the learners found the oral items more difficult than the written items and took longer to decide on grammaticality for these items.

There are two ways in which there could be significant differences in modality but still support the claim for a critical period. The first would be if the modality effects observed for the learners paralleled those obtained from the native speakers. Specifically, there could be reaction time differences in modality but no accuracy differences. The second would be if there were an interaction between age of acquisition and modality, such that the pattern for the younger learners resembled that of the native speakers. Neither of these outcomes was obtained. The modality differences, therefore, present a problem for the conclusion that there is a critical period governing the learning of the second language by this group of language learners.

Effects of length of residence

Differences in performance on the grammaticality judgement task have been discussed in terms of the age of arrival for two groups, specifically, younger learners arriving at 15 years or younger and older learners, at 16 years or older. However, there are also differences between these two groups in the length of time they have been using English. Evidence for the critical period hypothesis requires that the behavioural effects be attributable to the age at which learning began, not the length of time engaged in practising the behaviour. How can we decide whether it is the age of arrival or length of time in the country that has produced different patterns for the two age groups?

Ideally, one would want to conduct a study in which the age of arrival is independent of the length of time spent learning. Further, it would be methodologically desirable for all learners, irrespective of the age at which they began, to have spent precisely the *same* amount of time practising the language, that is, had the same length of residence. This was not the case in the present study, and it is virtually impossible for it to ever be achieved: for university students who are roughly the same age, those who arrived at a younger age have been in the country longer. In the present study, the group of younger learners had been in Canada approximately twice as long as the older learners (see Table 1). In contrast, the study by Johnson and Newport (1989) did equate the two groups for their length of residence in the United States, but this came at a cost. There was a large difference between the two groups in the age of testing, and the difference carried serious implications for the educational experience that each group had with English. In their study, the group of younger learners consisted primarily of undergraduates who had arrived in the United States as children and presumably attended American schools. The older learners were primarily faculty and post-doctoral fellows who had much more diverse learning experience with English and were unlikely to have attended American schools. Hence, the equation of length of residence left a difference in experience that could still easily account for performance differences, especially given the result that there was much greater variance in performance of the older learners than for the younger ones. Being a student in an American school would not only improve one's English but would also reduce the variance in English proficiency for those students who shared this experi-

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

ence. In the present study, the participants were primarily undergraduates with minimal differences between the groups in the age at which testing occurred and marginal differences in their background experience learning English. Consequently, there were greater differences between the groups in the number of years they have been in Canada.

How can the effects of age of arrival and length of residence be disentangled? The present study provides a means of examining these factors individually. There is a strong correlation between age of arrival and length of residence for the younger learners, $r = -.93$, $p < .0001$, but not the older ones, $r = -.13$, n.s. This is logical, because those who arrived as young children went through the same steps as their Canadian peers and arrived at university at approximately the same time. For example, a child arriving at the age of 10 and entering university with the Canadian peer group would be about 22 at the time of testing, having spent 12 years in residence. Those who arrived at older ages underwent more varied experiences and did not necessarily begin university at the same time. For example, someone arriving at the age of 18 may not immediately enter university for a variety of reasons. When they were in university, therefore, the relation between their age and the length of time they have spent in the country would be less systematic. The strong negative correlation between age and length for the younger learners is largely a consequence of having gone through the school system. This difference enables the factors for age and length to be separated. A correlation between age of arrival and performance for younger learners can be attributable either to age or length of residence; a correlation between age of arrival and performance for older learners, however, can only be attributable to age of arrival.

The test is this: if the behavioural results are attributable to a critical period, then there should be a correlation between performance and age of arrival for the younger group but not the older group. This was the pattern reported by Johnson and Newport (1989). If length of residence is accounting for performance, whether or not there is a critical period, then there should be a correlation between performance and length of residence for both the younger and older learners. Finally, if there is no critical period but some other factor is responsible for the familiar child advantage, such as cognitive or educational variables, then there should be a correlation between age of arrival and performance at all ages without any disruption in that relation around a putative critical period. In other words, the older learners, for whom age and length of residence are independent, makes it possible to separate the effects of these factors.

The results showed the last pattern. Age of arrival was related to performance for both the younger and older learners in both language groups but length of residence was related to performance only for the younger learners. These results were more evident for the oral condition than the written one. Three conclusions can be drawn from these results. First, the correlations show that there is an effect of age on performance, but it never ceases to be functional. Individuals who begin learning a second language at an older age are always handicapped with respect to younger learners even for relatively advanced starting ages. Bialystok and Hakuta (1999) report the debilitating effect of age on proficiency in a sample of learners that range in age from about 1 year to 80 years old. This pattern is not consistent with a critical period. Second, the anova results show that there is no mean difference in proficiency, at least for the Chinese learners. Therefore, the ravages of age are possible to overcome and variance from such factors as linguistic similarity overrides some of these effects. Finally, the length of residence did not account for performance. The correlation between length and proficiency for the younger learners was probably an artefact of the correlation between length of residence and age of arrival in this group, and it was age of arrival that was responsible for the relation.

Conclusion

There is no doubt that learners who arrived in Canada at a younger age and began learning English earlier had a better chance of doing well than learners who arrived later. This general trend was found in our data for the Spanish learners and replicates numerous studies that offer empirical support for the widely held view that children are more successful at learning a second language than adults are. The overall performance of the Chinese learners in our study failed to reveal this age-related change in success, but there was a negative correlation between age of acquisition and level of achievement for all the Chinese learners. Even in this case, then, arriving younger was associated with later advantages in English proficiency, although these advantages were more subtle than those observed for the Spanish speakers.

For both of our native language groups, learners who arrived before they were 8 years old performed the same as native speakers on a grammaticality judgement task by the time they were at least university age. This replicates the finding by Johnson and Newport (1989) using a cut-off age of 7. However, this finding is not at all surprising and should not be over-interpreted. Although it is theore-

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

tically possible that it is evidence for a critical period that closes at 8 years old, independent descriptions of the critical period have always placed the putative boundary at an older age, closer to puberty. Children who arrive in a country before the age of formal schooling and receive all their education in the new language undoubtedly have a different experience in learning the language of that country than all other language learners.

In spite of a general trend that relates younger learners with better proficiency, there is no evidence in the data to warrant attributing observed differences in performance to a critical period. All three types of evidence that we identified as indicating that language learning was controlled by a critical period failed to manifest themselves. Instead, we found that language proficiency was governed by such factors as the relation between the two languages, the syntactic structure being tested, and the modality in which the test was presented. These factors point to a view of language learning that is based more on central cognitive learning mechanisms and is sensitive to experiential contingencies. Although it is beyond the scope of this study to elaborate such a theory, elsewhere we have speculated about the reasons why language learning might decline across the lifespan if it were under the control of central cognitive processes (Bialystok and Hakuta, 1999). Yet, even if these cognitive factors had proved significant for older learners but not younger ones, the argument for a critical period could be supported. This was not the case; all learners were influenced by these factors. Hence, we see no reason to reject the null hypothesis that there is no critical period in the acquisition of a second language. If we err, it is on the side of caution.

References

Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. In E. Bates and B. MacWhinney (eds.), *The Crosslinguistic Study of Sentence Processing*, pp. 3–73. Cambridge: Cambridge University Press.

Bialystok, E. (1997). The structure of age: In search of barriers to second-language acquisition. *Second Language Research*, 13, 116–137.

Bialystok, E., & Hakuta, K. (1994). *In other words: The psychology and science of second language acquisition.* New York: Basic Books.

Bialystok, E., & Hakuta, K. (1999). Confounded Age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (ed.), *Second language acquisition and the critical period hypothesis*, pp. 161–181. Mahwah, NJ: Erlbaum.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language*, 68, 23–52.

Birdsong, D., & Molis, M. (1998). *Limits on attainment in second language acquisition: A replication study of Johnson & Newport (1989).* Unpublished manuscript, Department of French and Italian, University of Texas, Austin.

Bongaerts, T., Planken, B., & Schils, E. (1995). Can late starters attain a native accent in a foreign language? A test of the critical period hypothesis. In D. Singleton & Z. Lengyel (eds.), *The age factor in second language acquisition : a critical look at . the critical period hypothesis*, pp. 30–50. Clevedon: Multilingual Matters.

Bornstein, M. H. (1989). Sensitive periods in development: Structural characteristics and causal interpretations. *Psychological Bulletin*, 105, 179–197.

Chomsky, N. (1986). *Knowledge of language: its nature, origins and use.* New York: Praeger.

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

Colombo, J. (1982). The critical period concept: Research, methodological, and theoretical issues. *Psychological Bulletin*, 91, 260–275.

Coppieters, R. (1987). Competence differences between native and near-native speakers. *Language*, 63, 544–573.

Curtiss, S. (1989). The independence and task-specificity of language. In M. H. Bornstein & J. Bruner (eds.), *Interaction in human development*, pp. 105–137. Hillsdale, NJ: Erlbaum.

Danesi, M. (1994). The neuroscientific perspective in second language acquisition research: A critical synopsis, *IRAL*, 32, 201–228.

Elman, J. L. Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Epstein, S. D., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, 19, 677–758.

Eubank, L. (1993/94). On the transfer of parametric values in L2 development. *Language Acquisition*, 3, 182–208.

Eubank, L. (1996). Negation in early German-English Interlanguage: more Valueless Features in the L2 initial state. *Second Language Research*, 12, 73–106.

Gleitman, L. R., & Newport, E. L. (1995). The invention of language by children: Environmental and biological influences on the acquisition of language. In L. R. Gleitman & M. Liberman (eds.), *An invitation to cognitive science: Language, Volume 1*, pp. 1–24. Cambridge, MA: MIT Press.

Harley, B., & Wang, W. (1997). The critical period hypothesis: Where are we now? In A. M. B. de Groot & J. F. Kroll (eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, pp. 19–51. Mahwah, NJ: Erlbaum.

Ioup, G., Boustagui, E., Tigi, M. E., & Mosell, M. (1994). Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition*, 16, 73–98.

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press

Johnson, J. S. (1992). Critical period effects in second language acquisition: The effect of written versus auditory materials in the assessment of grammatical competence. *Language Learning*, 42, 217–248.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99.

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition*, 39, 215–258.

Juffs, A., & Harrington, M. (1995). Parsing effects in L2 sentence processing: Subject and object asymmetries in Wh-extraction. *Studies in Second-Language Acquisition*, 17, 483–516.

Kellerman, E. (1995). Age before beauty: Johnson and Newport revisited. In L. Eubank, L. Selinker, & M. Sharwood Smith (eds.), *The current state of interlanguage: Studies in honor of William E. Rutherford*, pp. 219–231. Amsterdam: Benjamins.

Long, M. (1990). Maturational constraints on language development. *Studies in second language acquisition*, 12, 251–285.

Mayberry, R. I. (1993). First-language acquisition after childhood differs from second-language acquisition: The case of American Sign Language. *Journal of Speech and Hearing Research*, 36, 1258–1270.

MacWhinney, B. (1997). Second Language Acquisition and the Competition Model. In A. de Groot and J. Kroll (eds.), *Tutorials in Bilingualism*, pp. 113–142. Mahwah, NJ: Erlbaum.

Newport, E. L. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Language Sciences*, 10, 147–172.

Newport, E. L. (1991). Contrasting conceptions of the critical period for language. In S. Carey & R. Gelman (eds.), *The epigenesis of mind: Essays on biology and cognition*, pp. 111–130. Hillsdale, NJ: Erlbaum.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, 5, 261–285.

Patkowski, M. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, 30, 449–472.

Patkowski, M. (1994). The critical age hypothesis and interlanguage phonology. In Yavas, M. (ed.), *First and second language phonology*, pp. 205–221. San Diego: Singular Publishing Group.

Pinker, S. (1984). *Language Learnability and Language Development.* Cambridge, MA: Harvard University Press.

Schachter, J. (1990). On the issue of completeness in second language acquisition. *Second Language Research*, 6, 93–124.

Schwartz, B. D., & Sprouse, R. A. (1994). Word order and nominative case in non-native language acquisition: a longitudinal study of (L1 Turkish) German Interlanguage. In T. Hoekstra & B. D. Schwartz (eds.), *Language acquisition studies in generative grammar: Papers in honor of Kenneth Wexler*, pp. 317–68. Philadelphia: John Benjamins.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive stages and the Full Transfer/Full Access model. *Second Language Research*, 12, 40–72.

Singleton, D. (1995). Introduction: A critical look at the critical period hypothesis in second language acquisition research. In D. Singleton & Z. Lengyel (eds.), *The age factor in second language acquisition: a critical look at . the critical period hypothesis*, pp. 1–29. Clevedon: Multilingual Matters.

Snow, C. E., & Hoefnagel-Hohle, M. (1978). The critical period for language acquisition: evidence from second language learning. *Child Development*, 49, 1114–1128.

Vainikka, A., & Young-Scholten, M. (1996a). Gradual development of L2 phrase structure. *Second Language Research*, 12, 7–39.

Vainikka, A. M., & Young-Scholten, M. (1996b). The early stages in adult L2 syntax: additional evidence from Romance speakers. *Second Language Research*, 12, 140–176.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research*, 12, 233–265.

Appendix 1

Examples of Categories and Ungrammatical Items on the Grammaticality Judgement Task

Plural:

The book are too heavy for me to carry.

I bought two bags of rices from the supermarket.

Determiner:

Shirley is a best dancer in the studio.

This is boy who stole my leather wallet!

Future:

Janet pay the bills at 2 o-clock tomorrow afternoon.

Grandma will turning eighty on her next birthday.

Present Progressive:

Sonia is hide underneath the table.

Come and eating your dinner right now!

Collocation Restriction Same:

Mary fainted and collapsed the floor.

Matthew picked the pieces of paper just now.

Collocation Restriction Different:

My parents arrived the airport already.

Jack is studying very hard the exam.

Received February 8, 1999 Revision accepted May 3, 1999

https://doi.org/10.1017/S1366728999000231 Published online by Cambridge University Press