A Non-critical Period for Second-language Learning

James Emil Flege
University of Alabama at Birmingham

Abstract
Early learners usually enjoy greater success in second-language (L2) learning than Late learners do. This is often interpreted to mean that the capacity for L2 learning diminishes after the close of a critical period. However the seeming limits on Late learners' success in learning an L2 following immigration, even after years of regular L2 use in the host country, may not be the unwanted consequence of normal neurocognitive maturation. It may instead arise from differences in the quantity and quality of input that Early and Late learners typically receive. This hypothesis was supported by the research reviewed in this chapter for both L2 speech learning and some aspects of L2 morphosyntax learning, leading to the proposal that long-term success in L2 learning is determined probabilistically by a *non-critical period* defined by age-related variation in L2 input.

1. Introduction

Eric Lenneberg (1967) laid out a nativist account of second-language (L2) acquisition that continues to influence research. He provided convincing evidence that native-language (L1) acquisition has a strong biological component and that, to be completely successful, the L1 must be learned before the close of a *critical period*. Lenneberg then extended his critical period (CP) hypothesis for L1 acquisition to the learning of L2 speech based on a simple observation, namely that a foreign accent (FA) is usually evident in the speech of those who began learning their L2 after puberty (1967, p. 176).

As Lenneberg (1967) showed, normal neurological development tends to follow a fixed schedule across individuals. If the capacity for

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), ***A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn*** (pp. 501-541). Dept. of English, School of Communication & Culture, Aarhus University.

learning L2 speech diminishes at a certain point, the effect of a reduced capacity for learning should be evident at roughly the same developmental state – and by extension, roughly the same chronological age of first exposure – for everyone who learns an L2. Although Lenneberg (1967) did not say so explicitly, he probably assumed that immigrants to a predominantly L2 speaking country receive abundant input from native speakers who provide a correct model of how the L2 should be pronounced. If that were true, then the observation of detectable foreign accents in immigrants who arrived in the host country after puberty might reasonably be interpreted as evidence for a diminished capacity for L2 speech learning.

This chapter will examine FA research in the light of the CP hypothesis. As a starting point, I question the tacit assumption that all immigrants receive abundant and adequate native-speaker input (see also Moyers, 2009). In fact, we know relatively little about the input that immigrants receive. This is because, as noted by Piske and Young-Scholten (2008, pp. 12-13), "only recently [has] input begun to receive consideration". Indeed, these authors acknowledge that we simply "do not know how much input second language learners actually get [nor] how much exposure a learner requires".

Researchers have tended to overlook input as a potential cause of differences between individual L2 learners, potentially leading to the misattribution of inter-subject differences. For example, a difference between two learners who immigrated to a predominantly L2 speaking country at the same age and lived in the host country for the same length of time might be ascribed to *individual differences* of unknown etiology or to differences in language learning aptitude. However if more were known about the quantity and quality of L2 input the two hypothetical individuals had received the difference between them might simply be a manifestation of differences in the input that they had received.

Few researchers would agree that an understanding of how input varies across individuals and groups is crucial to an understanding of age-related differences in L2 learning. This is due, at least in part, to the mistaken view that L2 input can be adequately assessed by length of residence (LOR) in a predominantly L2-speaking country. LOR is often used in L2 research because it can be readily obtained from language background questionnaires- It is, unfortunately, an imprecise measure and sometimes misleading index of the quantity of L2 input immigrants have received. This is because not all immigrants begin using their L2 immediately (e.g., Flege, Munro & MacKay, 1995a, Table I) nor use their L2 on a regular basis (Moyer, 2009, p. 162). The results of Flege and Liu (2001) suggested

that LOR provides a valid index of quantity of L2 input only for immigrants who have had both the opportunity and the need to use their L2 on a regular basis. These two crucial conditions for success in L2 learning are more likely to exist for Early learners than for Late learners (Moyers, 2009, pp. 360-363).

Another problem is that the LOR variable provides no information regarding quality of L2 input. Immigrants can hardly avoid trying to use their L2 when speaking to monolingual speakers of the target L2 after arriving in the host country. It may be just as difficult for them to avoid using the L2 in *linguistically mixed company*, that is, in conversations involving a monolingual L2 native speaker and one or more fellow immigrants from the same L1 background. In such situations the foreign-accented L2 speech to which immigrants are exposed is likely to provide an incorrect model of the target L2, one that reinforces the learners' natural tendency to adapt L2 speech to their existing L1 phonetic/phonological system.

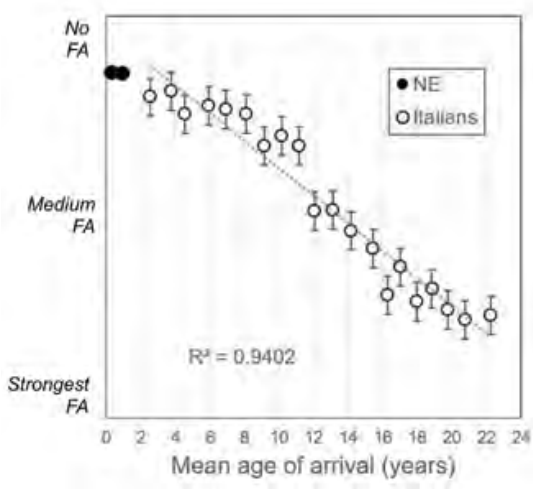

Figure 1. Mean FA ratings obtained for 20 groups of Italians and two groups of native English speakers (+/- 1 Sem)

This chapter is organized as follows. Section 2 considers speech learning, first through an examination of FA and then segmental production and perception accuracy. The focus of Section 3 is the learning of L2 morphosyntax. Section 4 directly compares the learning of L2 speech and morphosyntax. Finally, based on the preceding synthesis, Section 5 proposes that a non-critical period exists for L2 speech and morphosyntax learning and that Early vs Late differences derive primarily from age-related differences in the quantity and quality of input received.

2. L2 speech learning

2.1 AOA conditions input

Flege, Munro and MacKay (1995a) examined English sentences spoken by 240 Italian adults who immigrated to Canada between the ages of 2-23 years and had lived in Ottawa, ON for decades. The Italians recruited for this study had not received formal classroom instruction in English before im-

migrating to Canada, and all continued to speak Italian, especially at home, with close friends, and at church-related events. A delayed repetition task was used to elicit the sentences, which were later presented along with sentences spoken by native English (NE) speakers to listeners who were native speakers of Canadian English. The listeners rated the randomly presented sentences for overall degree of perceived FA using a continuous scale ranging from 1 (*strongest foreign accent*) to 256 (*no foreign accent*).[1]

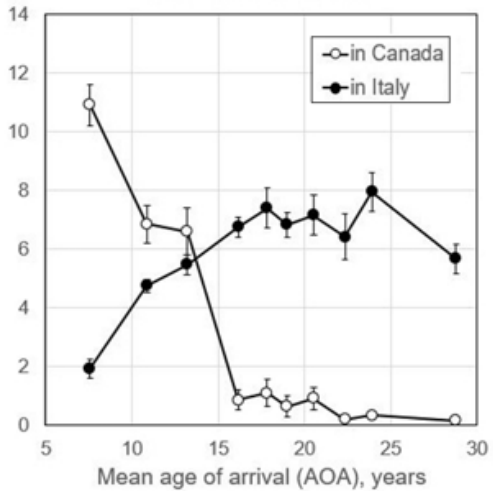

Figure 2. Mean years of formal education for 10 groups of 19 each.

Fig. 1 shows the mean FA ratings obtained for 20 groups of 12 Italians each differing in age of arrival (AOA) in Canada. The groups' mean AOA values accounted for 94% of the variance in the mean FA ratings, and a small increase in strength of FA is evident at an AOA of 11.7 years. Many would consider these findings to be convincing proof that a CP exists for L2 speech learning and that the capacity for L2 learning diminishes after about the age of 12 years as the result of normal neurological maturation.

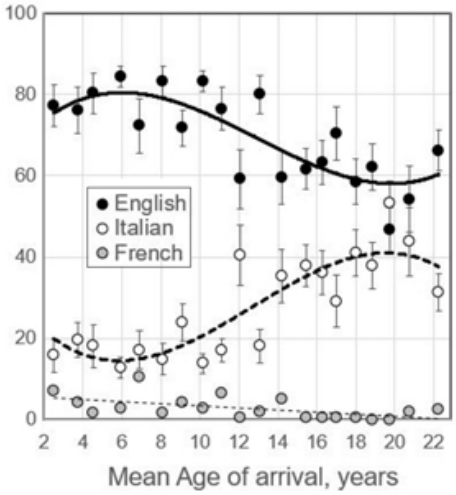

Figure 3. Mean self-estimated language use for 20 groups of 12 Italians each differing in AOA.

Both conclusions may be unwarranted. First, as far as I know, age of exposure to an L2 has never been directly linked to state of neurological development (see Flege, 1987; Hartshorne, Tenenbaum & Pinker et al., 2018, p. 274). Second, there is another, potentially better way to interpret these

[1] The five sentences examined were randomly presented four times each. The first set of ratings were treated as practice and so discarded. The mean rating computed for each participant was thus based on 150 ratings. The 240 Italians were originally assigned to 10 AOA-defined groups of 24 each rather than the 20 AOA-defined groups shown here.

data. The potency of AOA as a predictor of strength of FA may be due to its association with variation in input rather than with state of neurological and/or cognitive maturation at the time of first exposure to an L2. Immigrants' AOA largely conditions their later experience with the L2 (Stevens, 1999, p. 556) and so immigrants' success in learning the target L2 may be the result of differences in the input they receive.

Fig. 2 shows the mean number of years of formal education that AOA-defined groups of Italian immigrants had received both in Italy and later, after arriving in Canada.[2] Those who arrived in Canada before the age of 15 years obtained a substantial amount of formal education in English-speaking Canadian schools whereas most who arrived after that age received little if any formal education in Canada.

A long period of education in Canadian schools was unlikely to have directly affected the phonetic variables of interest here. However it was likely to have impacted the quantity and quality of English language input that the Italian immigrants to Canada later received over the course of their lives (Stevens 1999, p. 563). Specifically, the Italians who were enrolled at a local school soon after arriving in Canada learned English from their NE teachers and their NE classmates, with whom they often developed lifelong friendships and sometimes married. However most of those who arrived in Canada after the age 15, and so did not begin attending school on a full-time basis, lacked an important opportunity for establishing a strong social network in the English-speaking community.

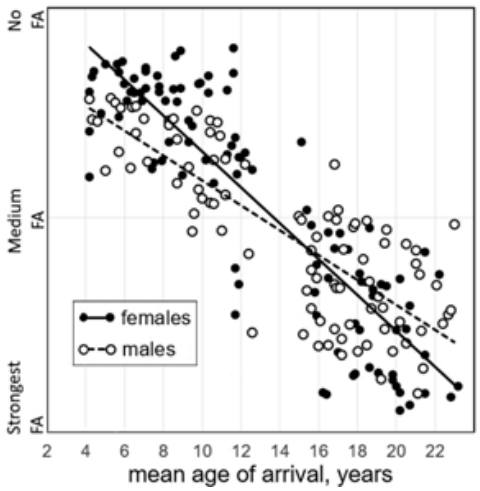

Figure 4. Foreign accent ratings obtained for Italian males and females differing in age of arrival in Canada.

LOR continues to be used in L2 research, but self-estimated L2 use provides a better index of quantity of input, especially if used in combination with LOR. Fig. 3 shows the Italians' mean self-estimates of percentage use

[2] These data were drawn from an unpublished study of aging that examined participants drawn from the same population in Ottawa ON as those tested by Flege et al. (1995a).

of English and Italian. The earlier the Italians arrived in Canada, the more they tended to use English and the less they used Italian.[3] The fact that most (182 or 76%) of the Italians tested by Flege et al. (1995a) reported using English more than Italian is hardly surprising because the Italian-speaking community in Ottawa is relatively small (< 20,000).

Percentage use estimates like these are limited in two ways. They are based on self-report and, perhaps more importantly, they tell us nothing about the quality of L2 input. It is plausible to think that the more frequently the Italians spoke their native language, the more often they were involved in linguistically mixed conversations in which they were exposed to Italian-accented English. The influence of foreign-accented input can be inferred from the gender effects reported by Flege et al. (1995a).

Fig. 4 shows the mean FA ratings obtained for 96 Early learners (AOA = 4.2-12.6) and 96 Late learners (AOA = 15.0-23.2) differing in gender. As expected (e.g., Geary, 1998, p. 263 ff), Early females had a significantly better pronunciation of English than Early males did. The expected female advantage was not evident for Late learners, however. Late females' pronunciation of English was in fact significantly worse than that of Late males ($p<.05$ by Mann-Whitney *U* tests).

This gender effect can be attributed to quality of input. As already mentioned, Italians who arrived early in life learned English at school from NE teachers and classmates. Males who arrived after the age of 15 years typically worked outside the home and learned English from both NE speakers and fellow Italian immigrants (the proportion is unknown). Female Late learners, on the other hand, usually stayed at home in their first few years in Canada. Their first model of English was likely to have been the foreign-accented English spoken by male relatives.

This reconstruction of the Italians' earliest phase of L2 learning was supported by an analysis of rate of learning. A variable called "Time needed to learn English" was derived by subtracting the Italians' age of arrival from their estimates of the age at which they were first able to speak English "comfortably" (see Flege et al., 1995a, Table 1). The times needed by Early males and females to reach this important milestone in L2 acquisition (M=11.8 vs 10.2 months) did not differ significantly (Mann-Whitney $U(1)=.966$, $p>.05$). However Late females needed a year longer to speak English comfortably according to self-report than Late males did

[3] The low frequency of French use for some participants who arrived in Canada before the age of 15 reflects the fact that French is taught as a foreign language in English-speaking Canadian schools.

(*M*=28.6 vs 16.4 months; Mann-Whitney *U*(1)=-2.94, *p*<.05). The Late males' relatively rapid learning of English was likely the result of an opportunity and need to use English at work, motivations that Late females may have lacked.

This inference was supported by a factor analysis of language background questionnaire data. Flege et al. (1995a) carried out separate Principal Components Analyses of responses obtained for males and females. The factors identified for the two genders were then used as predictors of FA in step-wise multiple regression analyses. Some factors identified for both males and females accounted for a significant amount of variance in the FA ratings. Other factors, however, were unique to one gender. For males but not females, variables designated *Languages used at work*, *Strength of concern for pronunciation*, and *Instrumental motivation* accounted for a significant amount of variance in the FA ratings. For females but not males, factors named *Overall language use* and *Language loyalty* were significant predictors of FA.

2.2 Problems for the CP hypothesis

For some researchers, the observation that AOA accounts for more variance than any other variable that can be derived from a language background questionnaire provides strong support for the existence of a CP. However, this form of evidence does not in itself prove that a CP exists nor that the basis for a CP – should one exist – is L2 learners' state of neurological maturation at the time of first exposure to the L2.

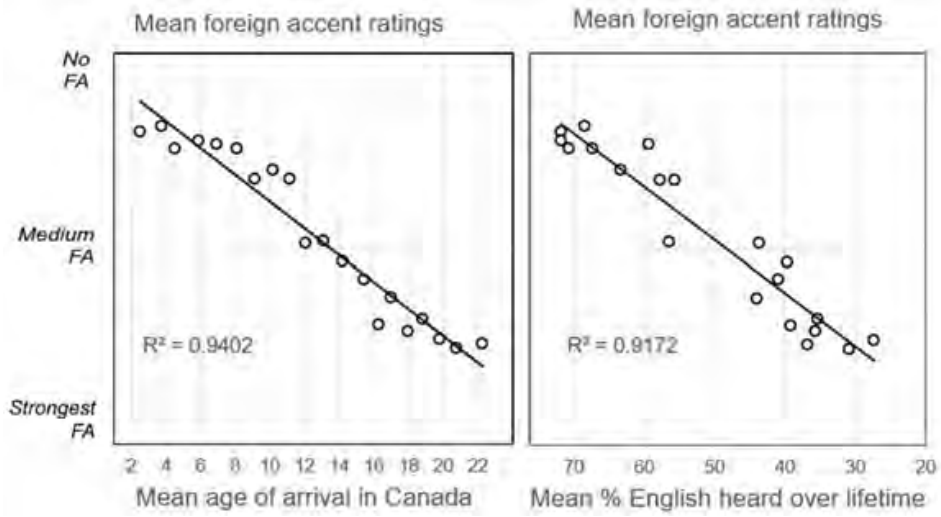

Figure 5. The relation between strength of foreign accent and age of arrival (left) and a measure of input (right) for 20 groups of 12 Italians each.

A substantial amount of variance in FA can also be accounted for without reference to AOA. I derived an input variable called "Percent English Heard in Life" from the Italians' age at the time of testing, length of residence in Canada, and self-estimated percentage use of English. As seen in Fig. 5, this derived input variable captured nearly as much variance in the FA ratings as AOA did (92% vs 94%).

The CP hypothesis formulated by Lenneberg (1967) predicts the presence of FA for the Italians who arrived after puberty but not the presence of FA in individuals who arrived in Canada as young children. Asher and Garcia (1969) was one of the first studies to show that even children may speak their L2 with a FA. The authors recorded 71 Cuban children living in the San Francisco area. None were judged by NE-speaking listeners to speak English without a FA. However more children who had lived in the US for 5-8 years were judged to have a "near native" pronunciation of English than those who had lived in the US for just 1-4 years (51% vs 15%). This suggested that the Cuban children were making progress over time in the pronunciation of English but left open the question of whether the children would eventually manage to speak English without a detectable FA.

The Italian adults tested by Flege et al. (1995a) had lived in Ottawa far longer than the Cuban children had lived in San Francisco. Some of the Italian Early learners – adults who had begun learning English as young children – obtained FA ratings that were more than 2 SDs below the mean rating obtained for NE speakers. The aim of Flege, Frieda and Nozawa (1997), therefore, was to determine if these Early learners spoke English with a detectable FA.

Flege et al. (1997) re-examined sentences spoken by 40 Early learners drawn from the Flege et al. (1995a) study. These Early learners arrived in Canada at a mean age of 5.8 years and had lived there for an average of 34 years. Most Early leaners tested by Flege et al. (1995a) reported using English more than Italian, but seven of the Italian Early learners examined by Flege et al. (1997) reported using Italian more than

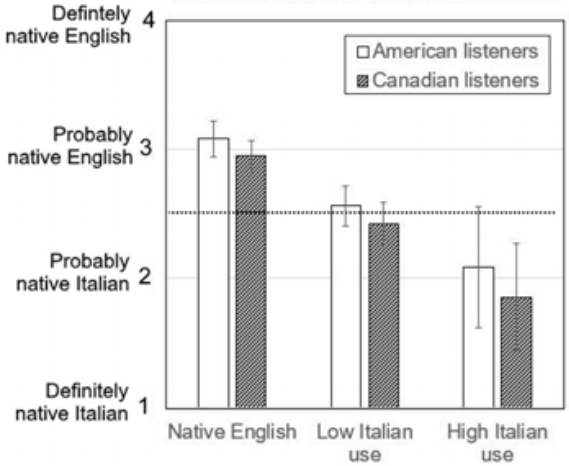

Figure 6. Mean ratings of English sentences spoken by native speakers of English and groups of Early learners differing in language use.

English. The Early learners examined Flege et al. (1997) were assigned to subgroups of 20 each according to percentage use of Italian (*M*=3% vs 36%). The two groups did not differ significantly in AOA (*M*=5.6 vs 5.9 years) or the time that elapsed between arrival in Canada and first exposure to English (4 vs 3 months).

NE-speaking listeners from both Canada and the US classified sentences spoken by the Early learners and 20 NE speakers as having been (1) *definitely* spoken by a native Italian (NI) speaker, (2) *probably* spoken by a NI speaker, (3) *probably* spoken by a NE speaker, or (4) *definitely* spoken by a NE speaker. These classifications have been treated in Fig. 6 as a 4-point rating scale. An ANOVA examining the ratings in Fig. 6 yielded a significant main effect of Group, but not a significant main effect of Listener group (Canadians from Ontario vs Americans from Alabama) nor a significant interaction.

A Tukey test revealed that all between-group differences were significant ($p<.05$). This means that both groups of Early learners spoke English with a FA, and that the strength of the Early learners' foreign accents depended on language use. The foreign accents were stronger for the Early learners who used English seldom/Italian often than for the Early learners who used English often/Italian seldom. Neither finding is compatible with the CP hypothesis.

Flege et al. (1997) carried out additional analyses to verify these theoretically important findings. The ratings just considered were based on 216 judgments obtained for each of the 60 participants (12 listeners x 2 listener groups x 3 sentences x 3 replicate judgments). The ratings obtained from the Canadian and American listener groups can be considered replicate experiments. To determine if the results would generalize to other listeners, *Listener-based* analyses were also carried out. In these, a mean rating was obtained for each of the 12 Canadian and 12 American listeners by averaging over the responses obtained for the 20 talkers in each group. The same findings were obtained again.

Significant differences between the two Early learner groups were also obtained using a non-parametric, bias-free measure of sensitivity to the presence of foreign accent (A-prime). The A-prime scores were based on the number of *hits* (i.e., "Definitely" and "Probably native Italian" responses for sentences spoken by the Early learners) and *false alarms* (the same responses given to sentences spoken by NE speakers). Both the Canadian and American listeners were significantly more sensitive to the presence of FA in sentences spoken by Early learners who used English

seldom/Italian often than for the Early learners who used English often/ Italian seldom ($p<.05$).

I carried out additional analyses for this chapter to better understand the presence of FA in the Italian Early learners. This involved selecting the 13 (of 40) Early learners who had the best pronunciation of English (the relatively *Good pronouncers*) and the 12 who had the worst pronunciation of English (the *Poor pronouncers*). English sentences spoken by the Poor pronouncers were usually judged to have been "definitely" spoken by a native speaker of Italian whereas the Good pronouncers' sentences were usually classified as "probably" having been spoken by a NE speaker.

The ratings obtained for the 25 Early learners just described are shown in Fig. 7 in correspondence with the ratings obtained for the same sentences by Flege et al. (1995a). The two sets of ratings were strongly correlated, $r=.94$, despite a difference in the scaling procedures used in the two studies (a continuous vs 4-point scale) and even though the ratings obtained by Flege et al. (1997) occupied just a small portion of the range of perceptibly different strengths of FA. The crucial point to note, however, is that the two groups of Early learners were judged independently to differ in strength of FA by three groups of NE-speaking listeners.

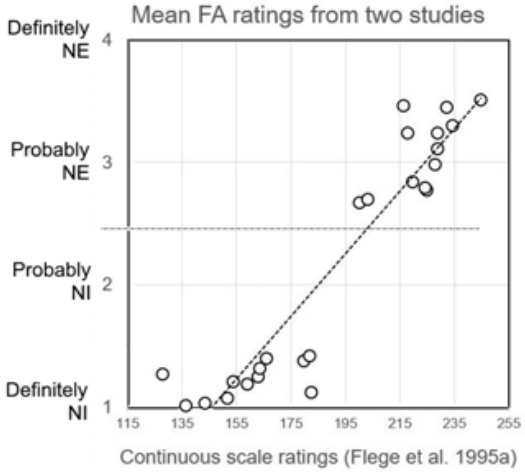

Figure 7. The mean foreign accent ratings obtained for 25 native Italian Early learners of English using two procedures for evaluating strength of foreign accent.

I next examined the 25 Early learners' responses to 7-point rating scales on a language background questionnaire. Participants were asked, in separate items, to self-rate their accuracy in pronouncing English and Italian. The two groups' self-rated pronunciation of English did not differ significantly, but the Poor pronouncers of English judged their Italian pronunciation to be significantly better than the Poor pronouncers did ($p<.05$ by Kruskal-Wallis test). As well, the Good pronouncers of English reported using English significantly more often, and Italian significantly less often than the Poor pronouncers did ($p<.05$ by Kruskal-Wallis tests). These results support the conclusion that differences in language use measurably affected the Early learners' pronunciation of English after decades of predominantly English use.

Another problem for the CP hypothesis is that foreign accents grow increasingly strong after the supposed closure of the CP. It is widely assumed that the ill effects of beginning to learn an L2 after the closure of a biologically based CP will be much the same for all *post-critical period* learners because the biological changes that trigger the close of the CP occur at roughly the same chronological age for all normally developing individuals (see, e.g., Lenneberg, 1967, Johnson & Newport, 1987; Birdsong, 2013).

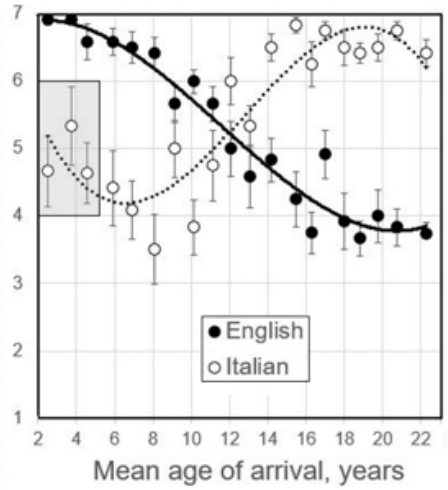

Figure 8. Mean self ratings of English and Italian pronunciation by Italians differing in AOA.

Flege & MacKay (2011) tested this prediction by extending upward the AOA range examined in previous research. The participants in this study were three groups of 20 Italians each who had arrived in Canada at mean ages of 10, 18 and 26 years as well as an age-matched group of NE speakers. The ratings obtained for the NE speakers and the three AOA-defined groups of Italian immigrants all differed significantly from one another ($p<.05$).

The CP hypothesis proposed by Lenneberg (1967) correctly predicted a stronger FA for members of the AOA-18 than the AOA-10 group. However it did not predict the significantly stronger foreign accents evident for members of the AOA-26 group than the AOA-18 group. Johnson and Newport (1989), in apparent agreement with Lenneberg, observed that "there are not many important maturational differences between … the brain of a 17-year old and the brain of a 27-year old" (p. 79). Nor did the CP hypothesis predict the significantly lower ratings obtained for the pre-critical period members of the AOA-10 group than for the NE speakers.

2.3. Experienced L2 learners are bilinguals

The CP hypothesis presented by Lenneberg (1967) , and later variants of this hypothesis, failed to consider a basic aspect of L2 learning. Immigrants who learn an L2 through immersion become bilinguals possessing two partially over-lapping phonetic subsystems which they usually cannot turn on or off either instantly or completely. Learning an L2 influences the native language (see Hopp & Schmid, 2013). According to the Speech

Learning Model (Flege, 1995) the phonetic elements comprising the L1 and L2 phonetic subsystems of a bilingual necessarily interact because they occupy the same phonetic/phonological space.

The 240 Italians tested by Flege et al. (1995a) self-rated their ability to pronounce English and Italian using 7-point rating scales. Fig. 8 shows the mean ratings obtained for 20 AOA-defined subgroups of 12 each. Most Italians who arrived in Canada before the age of 12 years reported having a better pronunciation of English than Italian whereas the reverse held true for those who arrived later in life. The self-ratings obtained for the 36 Italians with AOA values ranging from 1.7-5.3 years (grey box) may seem anomalous but are probably due to the fact that these participants were kept at home when they first arrived in Canada and so heard mostly or only Italian until school age.

The data in Fig. 8 supported the "common space" hypothesis but at the time of publication were judged to be of limited value because they were based on self-report. To further test the common space hypothesis, therefore, Yeni-Komshian, Flege and Liu (2000) recruited six male and six female Koreans at each age of arrival in the US ranging from 2 to 22 years. The 240 Koreans will be described in greater detail later when we consider the learning of English morphosyntax. For now, I simply note that the Koreans' AOA values correlated as expected with their chronological ages, $r(238)=.51$, self-estimates of English use, $r=-.50$, self-estimates of Korean use, $r=.51$, years of US residence, $r=-.42$, and years of US education, $r=-.92$. All these variables, in turn, were inter-correlated ($p<.05$).

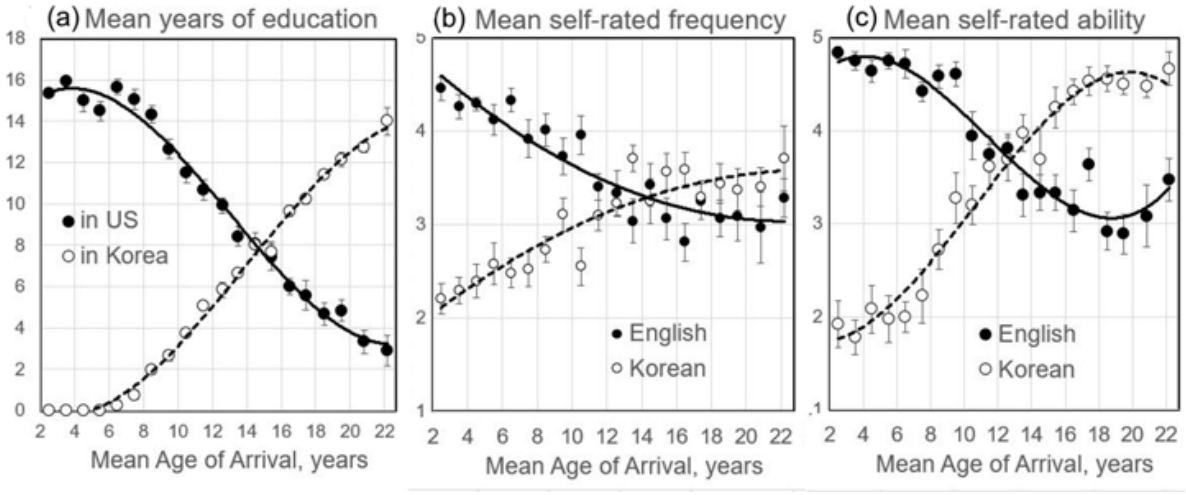

Figure 9. Relation between AOA and (a) years of formal education, (b) self-estimated frequency of use, and (c)self-estimated proficiency in Korean and English for groups of 12 immigrants each.

Fig. 9 shows the relation between the Koreans' mean AOA values and variables thought likely to influence their learning of English. Third-order polynomials have been fit to the mean values in all three panels to visually organize the data. Fig. 9(a) shows that an inverse relation existed between years of formal education obtained in Korean and American schools. Fig. 9(b) shows that Koreans who arrived in the US before the age of 12 years generally used English more than Korean whereas the reverse usually held true for Koreans who arrived after the age of 12 years. Finally, Fig.9(c) shows that most Early learners judged themselves to be more proficient in English (5-point rating scales regarding pronunciation, reading/writing ability, and grammatical knowledge) than in Korean whereas the reverse usually held true for Late learners.

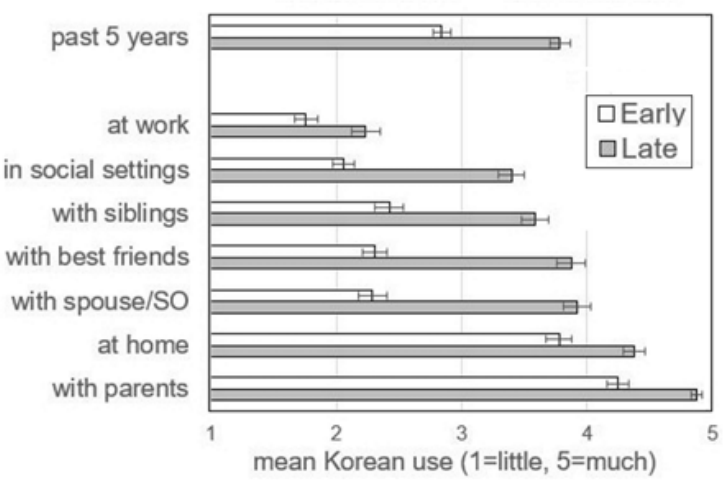

Figure 10. Mean ratings of Korean use in seven contexts by Early and Late learners.

As observed earlier for Italian immigrants to Canada, the Koreans' use of their two languages depended importantly on context. Fig. 10 shows the mean ratings of frequency of Korean use obtained using scales that ranged from 1 (*very little*) to 5 (*very much*). The Koreans indicated using Korean far more frequently with their parents and when at home than while at work. We might think of these contexts as representing *obligatory contexts* in which the use of Korean or English was the norm. The remaining four contexts shown in Fig. 10, however, were intermediate in value to those observed in the home and work contexts. The ratings obtained for 120 Late and 120 Early learners in these *optional contexts* differed substantially (*M*=2.34 vs 3.51,

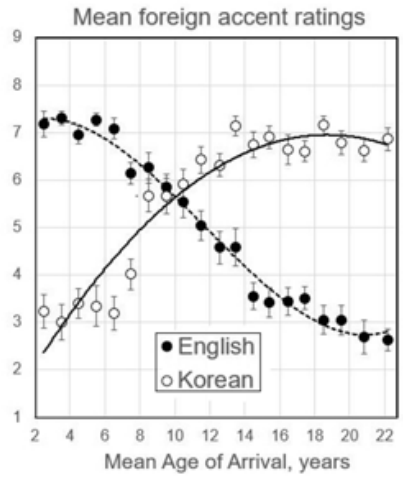

Figure 11. Mean foreign accent ratings obtained for Korean and English sentences.

$F(1,238)=134.5$, $p<.05$). Variation in language use in optional contexts like these was probably due to social factors rather than to the Koreans' state of neurocognitive maturation when they immigrated to the US.

Yeni-Komshian et al. (2000) had English and Korean listeners rate the 240 Koreans' pronunciation of Korean and English sentences for degree of FA using a 9-point scale. Fig. 11 reveals a crisscross pattern much like the one seen earlier for the Italians' self-ratings of L1 and L2 pronunciation ability. Higher ratings, indicating the presence of milder foreign accents, were obtained for Early learners' productions of English than Korean sentences whereas the opposite held true for the Late learners. This finding supported the hypothesis (Flege, 1995) that bilinguals' L1 and L2 phonetic systems interact in a common space.

2.4 Segmental production and perception

Lenneberg's (1967) CP hypothesis focused on global pronunciation of the L2, that is, the presence or absence of foreign accents. It was soon extended to phonetic research examining other aspects of L2 speech, namely the production and perception of L2 vowels and consonants. Some of that work will be presented here.

Later work in Ottawa with Italian immigrants employed an orthogonal design in which participants were selected of the basis of both AOA and language use. In one project 36 Early and 36 Late learners differing in AOA (M=7.5 vs 20.0 years) were subdivided into subgroups of 18 each based on self-reported Italian use (Low Italian use M=7-10%, High Italian use M=43-53%). Given that the Italians' use of English was inversely related to their use of Italian, I will refer to the groups differing in language use as the *English often/Italian seldom* and the *English seldom/ Italian often* groups.

The 72 Italians took part in experiments examining overall degree of perceived FA (Piske, MacKay, & Flege, 2001), the identification of word-initial and word-final consonants (MacKay, Meador, & Flege, 2001), the production of English vowels (Piske, Flege, MacKay, & Meador, 2002), and the perception of English vowels (Flege & MacKay, 2004). In all four studies the Italians who used English often/Italian seldom obtained significantly higher scores than those who used English seldom/Italian often. This held true for both Early and Late learners. When taken together, the results of this research demonstrated the importance of language use but did not, of course, rule out a possible role of neurological maturation at the time the Italians immigrated to Canada.

One of the perception experiments carried out by Flege and MacKay (2004) focused on the vowels in English words like *beat* and *bit* (/i/, /ɪ/). These English vowels were of special interest because Italian has an /i/-quality vowel but not an /ɪ/-quality vowel. When Italians are first exposed to English they tend to hear both English vowels as being instances of their Italian /i/ category. However one experiment by Flege & MacKay (2004) suggested that although it is initially difficult for Italians to perceptually distinguish English /i/ from /ɪ/, doing so is eventually possible because of a perceived difference in the relation of the two English vowels to Italian /i/.[4]

Flege and MacKay (2004) examined the Italian immigrants' ability to detect errors in the production of English /i/ and /ɪ/. The stimuli were short phrases edited from the spontaneous conversations of Italian immigrants from an earlier study. The stimulus set included (a) correct productions of the vowel /i/ in phrases like "*speak the*", (b) correct productions of /ɪ/ in phrases like "*my kids they*", (c) incorrect productions of the target vowel /i/ as [ɪ], as in the phrase "*and reading*", and (d) incorrect productions of the target vowel /ɪ/ as [i], as in the phrase "*very difficult*".

The test phrases were presented auditorily and visually on a computer screen. An asterisk replaced the target vowel in the written phrases (e.g., sp*k the; very d*fficult) to localize the target vowel of interest in each phrase. The task on each trial was to decide if the target vowel had been produced correctly or incorrectly. An unbiased measure of sensitivity, A', was computed based on hits (correct detections of errors) and false alarms (classifications of correct productions as incorrect).

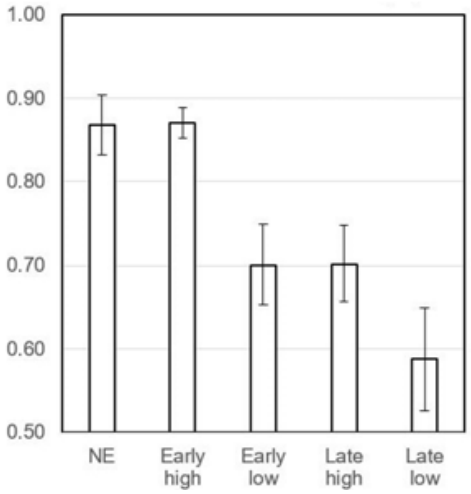

Figure 12. The mean perceptual sensitivity of NE speakers and four groups of Italians differing in AOA (Early vs Late) and language use (High vs Low) to vowel production errors.

[4] Young Italian adults who had little conversational experience in English classified English /ɪ/ tokens as Italian /i/ less often than they classified English /i/ as Italian /i/ (*M*=65% vs 87%). Moreover, the English /ɪ/ tokens were rated as being poorer instances of Italian /i/ than the English /ɪ/ tokens were (*M*=2.9/5 vs 4.2/5).

The mean A' scores obtained for the four Italian groups are shown in Fig. 12. Three Italian groups (Early-low use of English, Late-high use of English, Late-low use of English) showed significantly less perceptual sensitivity to vowel production errors than the NE speakers did ($p<.01$). The only Italian group that did not differ from the NE group was Early-high, the group consisting of Early learners who used English often/Italian seldom. These findings are problematic for the CP hypothesis in two ways. First, a group of Early learners (Early-low) differed significantly from the NE speakers. Second, Early and Late learner groups (Early-low, Late-high) obtained virtually identical scores.

Language use turned out to be a slightly better predictor of the error detection scores than AOA did[5] and captured slightly more variance than AOA (η^2=.162 vs .158). However the small advantage of Language use over AOA does not provide convincing evidence that input is more important than age of first exposure to an L2. Language use and AOA together accounted for only 32% of the variance in the error detection scores. This means that some other dimension(s) contributed importantly to the error detection scores.

The important dimension that was missing, I suspect, was the quality of English input that the Italian immigrants had received during their many years of Canadian residence. The more often the Italians used English, the more often they were likely to have taken part in linguistically mixed conversations involving a NE speaker and one or more other Italian immigrants. If so, they may have been more likely than Italians who used English seldom to hear English vowels spoken with an Italian accent, and so to have developed inaccurate perceptual representations for English vowels.

3. Learning L2 morphosyntax

In this section we turn our attention to the learning of English morphosyntax.

---

[5] AOA showed a stronger correlation to the error detection scores than either % English use or LOR examined individually, $r(61)$=-.402 vs .340, .230). However the strength of correlation between the detection scores and *Years of English use* (LOR multiplied by % English) and AOA did not differ, $r(61)$=.412 vs -.402. The correlation between % English use and the error detection score correlation remained significant when AOA was partialled out, $r(60)$=.26, $p<.05$, but the AOA-detection correlation became non-significant when % English was partialled out, $r(60)$=-.24, $p>.05$.

3.1. Maturational limits

Johnson and Newport (1989) extended Lenneberg's (1967) CP hypothesis from L2 speech learning to the learning of L2 morphosyntax. These authors (henceforth "J&N") recruited a total of 46 Korean and Chinese adults who had arrived in the US between the ages of 3-39 years. To ensure that all participants had attained their ultimate level of performance in English, all were required to have lived in the US for at least five years, at least three years of which had to be continuous years of US residence prior to testing at the University of Illinois.

J&N developed a 276-item grammaticality judgement test (GJT) to evaluate knowledge of the "most basic aspects of English sentence structure" (1989, p. 72). Their test consisted of grammatical and ungrammatical versions of sentences such as *Last night the old lady died/*die in her sleep*. The task of the adult native and non-native participants was to judge the randomly presented sentences as grammatical or ungrammatical. The authors described their test as "minimally demanding" because 6 to 7-year old NE-speaking children obtained "virtually perfect" scores on it (1989, p.70).

The percent correct scores obtained for the Korean and Chinese participants generally decreased as AOA increased. The scores obtained for seven immigrants having AOAs in the 3 to 7-year range did not differ significantly from the scores obtained for 23 NE speakers (M=97.6% vs 97.4%, p>.05 by two-sample t-test). However the scores obtained for participants having AOA values in the 8 to 10-year range (n=7, M=92.8%), in the 11 to 15-year range (n=8, M=85.5%) and in the 17 to 39-year range (n=23, M=76.2%) were all significantly lower than the NE speakers' scores (p<.05). J&N also noted the existence of a significant correlation between AOA and test scores for participants having AOA values in the 3 to 15-year range but not in the 17 to 39-year range (r=-.87 vs -.16).

These findings led J&N to conclude that a critical (or sensitive) period exists for the learning of L2 morphosyntax. The period in question, which ranged from AOA values of 7 to 15 years, represented a period of rapid decreases in the GJT scores flanked by AOA ranges in which the GJT scores varied little or not at all. The authors hypothesized that the mechanism underlying the CP was either the reduction of an unidentified "language [learning] faculty" or a general change in "cognitive abilities involved in language learning", itself a consequence of normal neurological and/or cognitive maturation (p. 61).

The conclusions drawn by J&N regarding both the mechanism underlying the hypothesized CP as well as its offset have been contested. Hakuta and Bialystok (1994) noted, for example, that when the CP offset was shifted from an AOA of 15 to 20 years, a significant correlation was found to exist between AOA and the scores obtained for both the newly defined pre- and post-critical period learners. Hartshorne et al. (2018, p. 271) observed that J&N would almost certainly have observed a significant correlation between test scores and the AOA values of Later arrivals had they simply recruited more participants. This observation applies equally to the apparent existence of an *optimal age*, that is, the absence of a significant difference between NE speakers and the seven nonnatives having AOA values in the 3 to 7-year range.

3.2. Korean immigrants in the US

3.2.1 Participants

The aim of Flege et al. (1999) was to replicate and extend the classic Johnson and Newport (1987) study. These authors tested 295 Korean adults in 1992-1996 at the University of Maryland (UMD). All were tested individually by a bilingual Korean-English research assistant in a single 1.5-hour session. Most participants were current or former UMD students or faculty members. All were able to speak English as shown by their ability to repeat English sentences fluently following a filled delay. Twenty-six Koreans had to be excluded from the study because they could not fluently repeat Korean sentences following a delay and so might not have been bilinguals. Another 29 Koreans were excluded for one or more of the following reasons: not speaking the Seoul dialect of Korean, speaking a language other than English and Korean; having lived in a country other than Korea and the US.

The experimental protocol specified a minimum of 10 years of residence in the US. When it became evident after three years of recruitment that the 10-year minimum would make it impossible to fill all 20 one-year AOA bins with 6 males and 6 females, the minimum was reduced to 8 consecutive years of US residence. Two Koreans retained for the study were married to a NE speaker but even they reported using Korean in the home. The ages of the 240 Koreans retained for the study differed little (*M*=26 years, *range*=17-46) from the ages of the 24 NE speakers who formed the comparison group (*M*=27 years, *range*=20-45). Statistical analyses presented by Flege et al. (1999) focused on groups formed by combining adjacent 1-year

AOA groups, thereby creating 10 AOA-defined groups of 24 each. The mean AOA values of these groups of 24 each ranged from 3-22 years.[6]

3.2.2 Grammaticality Judgment Test

The 18-min test used by Flege et al. (1999) was derived from the GJT developed earlier by Johnson and Newport (1987). Sentences from the original test that probed knowledge of Auxiliaries, Word order and the Present progressive were eliminated because they had served little to distinguish Early from Late learners . By adding to the length of the test, these sentences may have contributed to errors due to inattention or fatigue. Sixteen new items testing lexically specified subject/object raising were added, however, yielding a GJT test that consisted of 144 sentences, half grammatical and half ungrammatical.

The sentences comprising the new GJT were recorded at a constant moderate rate by a single native speaker of English (JEF) who took care to articulate word-final stops (e.g., the /d/ in *died*) and fricatives (e.g., the /s/ in *paints*). A short break occurred between the two halves of the GJT. The grammatical and ungrammatical versions of each sentence pair were presented in separate halves of the test. The written presentation of each sentence was accompanied by a single aural presentation of the same sentence, with a fixed 4.0 sec interval between sentences. Participants were required to respond "Yes" (grammatical) or "No" (ungrammatical) to each sentence before moving on to the next sentence.

3.2.3 Results

As in the study by Johnson and Newport (1987), the GJT scores obtained for Korean immigrants by Flege et al. (1999) generally decreased as AOA increased. Korean groups having mean AOA values in the 7 to 22-year range, but not those having mean AOA values in the 3 to 5-year AOA range, obtained significantly lower GJT scores than the NE speakers did

---

[6] Flege et al. (1999) largely avoided a confound between AOA and years of formal education. The Koreans' highest educational attainment averaged 15.7 years (*SE*=.11, *range*=12-21 years). The effect of Group (10 levels) on education was significant, $F(9,230)=2.1$ $p<.05$. However just one pair-wise difference between the ten groups reached significance, that between Koreans having AOA values of 5 and 19 years (M=14.8 vs 16.5, $p<.05$ by Tukey test). This difference arose because ten members of the AOA-5 group had completed less than 3 years of college because of their young age (17-20 years) but this held true for just three members of the somewhat older AOA-19 group.

(Bonferroni-corrected $p<.05$). A Gumpertz-Makeham growth function was fit to the percent correct GJT scores obtained for the 240 Koreans to visually organize the data. Inspection of this function suggested that most Koreans having AOA values in the 2 to 6-year AOA range obtained scores resembling those of the NE speakers. Scores for Koreans in the 7 to 15-year AOA range decreased in a near linear fashion as AOA increased whereas the scores for Koreans having AOA values in the 16 to 22-year range continued to decrease as AOA increased, but at a slower rate.

Flege et al. (1999) submitted the Koreans' responses to 39 questionnaire items to a Principal Components Analysis (PCA) to identify factors underlying the AOA effects just described. Factor scores derived from the PCA were then used as predictors of the morphosyntax scores in a step-wise multiple regression analysis. A factor named *Age of L2 learning* accounted for far more variance in the GJT scores than a factor named *Length of residence* did (67.7% vs 3.6%). This might be taken as support for the traditional view of AOA effects, namely that Early vs Late differences are due to the loss of capacity to learn an L2 following closure of a CP and that variation in input contributes little if at all to Early vs Late differences.

There are two reasons to question this interpretation. First, consider the regression analysis used to identify factors underlying age-related variation in the GJT scores. The factor accounting for most (67.7%) of the variance was named Age of L2 learning because the questionnaire variable having the highest loading on it (.912) was AOA. However, five other variables also had high loadings on the Age of learning factor: Years of education in the US (-.856), Use of Korean with spouse (.786), Use of Korean with close friends (.729), Use of Korean at social gatherings (.737) and Use of English at social gatherings (-.712). The multi-collinearity lurking beneath the surface of the Age of L2 learning factor made it impossible to directly evaluate the role of AOA in differentiating Early from Late learners of English.

Second, the influence of AOA was not the same for all sentences examined. AOA correlated significantly with all nine sentence types in the GJT used by Flege et al. (1999). However the strength of correlation ranged from a low of $r(238)=-.44$, for sentence pairs examining Third-person singular (e.g., *Every Friday our neighbor* <u>*washes/*wash*</u> *her car*), to a high of $r(238)=-.74$, for pairs examining Determiners (e.g., *The boy is helping the man build* <u>*a/*Ø*</u> *house*). Johnson and Newport (1989, Fig. 3) also noted substantial differences across sentence types for Korean and Chinese and immigrants.

The analysis of sentence types provided little general understanding of the learning of L2 morphosyntax. This is because the grammatical structures that proved difficult for the Koreans might not prove difficult, or else manifest a different degree of learning difficulty, for speakers of other languages. To provide a more general understanding of L2 morphosyntax learning, therefore, Flege et al. (1999) calculated two morphosyntax subscores that will be referred to here as the *Rule-based* and *Lexicon-based* scores. The two scores represented a functional rather than syntactically motivated grouping of sentences from the GJT since the items upon which the scores were based were drawn from multiple sentence types (see Appendices 1 & 2 in Flege et al., 1999). Both scores were based on responses to 44 sentences (half grammatical, half ungrammatical) identified through Principal Components Analyses.

The Rule-based scores probed knowledge of regular, productive and generalizable rules of the surface morphology of English. All involved case or number assignment on nouns, or person, or tense markers on verbs. For example: **A/The* *boys are going to the zoo this Saturday*; *The man *paints/painted his house yesterday*, **Them/They worked on the project all night.* The Lexicon-based sentences, on the other hand, probed irregular and ungeneralizable aspects of English morphosyntax involving the proper assignment of particles or prepositions with verbs, or knowledge of idiosyncratic features of English verbs. For example: *The farmers were *hoping/hoping for rain*; *The little boys *laughed/laughed at the clown*. All ungrammatical Lexicon-based sentences could be made grammatical by replacing the verb (e.g., changing "lets" to "permits" in **The man lets his son to watch TV*). However the ungrammatical Rule-based sentences could not be corrected in this way.

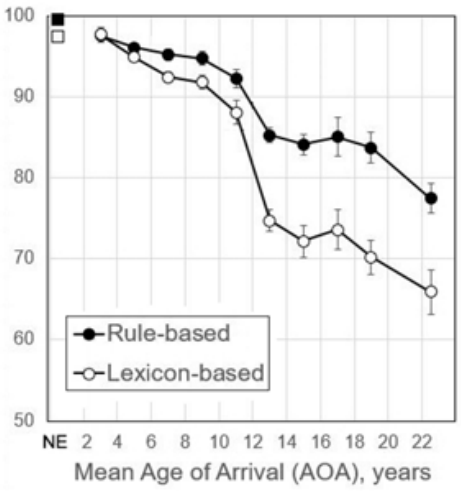

Figure 13. Mean Rule-based and Lexicon-based scores obtained for native English (NE) speakers and 10 groups of Korean differing in age of arrival (AOA).

The Lexicon-based and Rule-based scores were likely to draw on different forms of memory: declarative vs procedural. Also, the Lexicon-based scores were likely to be more susceptible to variation in input than the Rule-based scores (Pinker & Prince, 1991; Prasada & Pinker, 1993;

Beck, 1997; see also Birdsong & Flege, 2001). That being the case, we expected to observe larger Early vs Late differences for Lexicon-based than Rule-based scores. This expectation was confirmed as can be seen in Fig. 13. The effect of AOA was greater for the Lexicon-based than Rule-based scores. AOA showed a significantly stronger correlation with the Lexicon-based than Rule-based scores (r=-.71 vs -.58. $X(1)$=12.3 p<.05) and captured more variance (η^2=.513 vs .318).

Differences between the NE group and 10 AOA-defined groups of 24 Koreans each were evaluated using one-sample t-tests. All 10 Korean groups obtained significantly lower Rule-based scores than the NE speakers did (t- values ranging from -3.1 to -10.1, Bonferroni-corrected p<.05). However just eight Korean groups – those having mean AOA values in the 7 to 22-year range – obtained significantly lower Lexicon-based scores than the NE speakers (t- values ranging from -3.5 to -13.3, Bonferroni-corrected p<.05). Koreans having mean AOA values of 3 and 5 years did not obtain significantly lower Lexicon-based scores than the NE speakers (t=0.4 and -2.9, respectively, Bonferroni corrected p>.05).

The absence of significant differences in Lexicon-based scores between NE speakers and Koreans with AOAs of 3 and 5 years was probably not the result of a ceiling effect for the NE speakers. There was less variance, to be sure, in the NE speakers' Rule-based scores (M=99.5% correct, SE=.19, *range*=97.7 to 100.0) than in their Lexicon-based scores (M=97.3, SE=.90, *range*=81.8 to 100.0). Of the 24 NE speakers tested, 19 were at ceiling from the Rule-based scores as compared to just 12 for the Lexicon-based scores. Had a more difficult test been administered more NE speakers would have been off ceiling for the Rule-based scores; but even with 19 of 24 at ceiling, significant native vs nonnative differences were detected. A more likely explanation for the absence of a difference between the NE and two Korean groups was a small effect size. If so, native vs nonnative differences for the Lexicon-based scores would probably have been obtained for NE speakers and Koreans with mean AOAs of 3 and 5 years if a larger number of Koreans had been tested.

Two-sample t-tests were used to evaluate differences between the Rule-based and Lexicon-based scores. The Koreans having mean AOA values in the 13 to 22-year range obtained significantly lower Lexicon-based than Rule-based scores (t-values ranging from 4.6 to 8.9, df=23, Bonferroni-corrected p<.05) whereas the Lexicon-based and Rule-based scores obtained for Koreans having mean AOA values in the 3 to 11-year range did not differ significantly (t-values ranging from -0.2 to 2.2, df=23, p>.05).

Why did an AOA of 12 years mark the point of demarcation between the Rule-based and Lexicon-based scores? In the popular mind – and that of many researchers as well – the notion of a critical period is associated with the age of 12 years. As seen in Fig. 9(c) an AOA of 12 years was the point of demarcation between Koreans who usually judged themselves to more proficient in English than Korean from those who usually reported the opposite. The age of 12 years is also relevant in another way. It is the age at which mandatory instruction in English begins in Korean schools. Johnson and Newport (1987, p. 69) raised the issue of whether immigrants' *age of exposure* to an L2 should be defined as the age at which they began to study the L2 at school in their home country or age of emigration to a predominantly L2-speaking country.

To evaluate the influence of school study before immigration, I examined the results obtained for Koreans in 16 of the original 20 one-year AOA bins of 12 each. Participants who arrived in the US before the age of 4 years were excluded from this analysis because of uncertainty as to whether they only (or mostly) heard Korean until school age after arriving in the US. The Koreans who arrived in the US at the age of 12 years were also excluded because of uncertainty as to whether they had already begun to study English at school before departing for the US. Finally, to create balanced groups of Early and Late learners, Koreans who arrived in the US after the age of 20 years were also excluded.

This procedure yielded groups of 96 Early and 96 Late learners having AOA values that both spanned a seven-year range: 4 to 11 years for the Early learner and 13 to 20 years for the Late learners. As expected, significantly higher scores were obtained for Early than Late learners (M=93.1% vs 77.2%, $F(1,190)=179.5$, $p<.05$) and for the Rule-based than Lexicon-based scores (M=88.9% vs 81.5%, $F(1,190)=158.3$, $p<.05$). A significant interaction was obtained ($F(1,190)=62.8$, $p<.05$) because the Rule-based and Lexicon-based scores differed more for Late learners (83.2% vs 71.1%) than for Early learners (94.5% vs 91.8%).

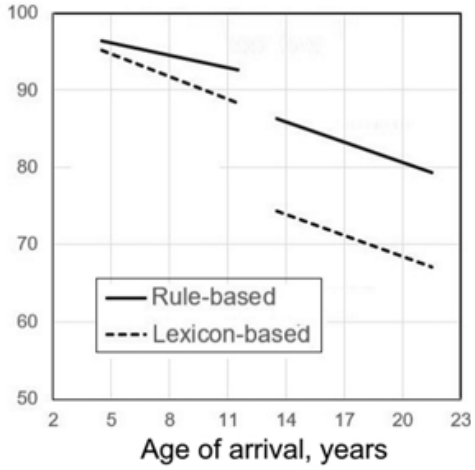

Figure 14. Best fitting functions relating AOA to scores obtained for Early and Late learners.

The best fitting linear functions for the two GJT scores are shown in Fig. 14 for the Early and Late learners. The slopes of all four functions differed significantly from zero ($t(94)$=-2.35 to -4.29, Bonferroni-corrected p<.05) indicating that the GJT scores decreased as AOA increased over a 7-year range for both groups. The Early and Late learners' slopes did not differ significantly from one another for either the Rule-based or Lexicon based scores (p>.05). If exposure to English at school in Korea really mattered, we would have expected steeper slopes for the Early than for the Late learners, whose age of first exposure to English at school in Korea was a constant 12 years.

There is no guarantee, of course, that the downward trends seen in Fig. 14 were due to 7-year increases in AOA, nor that the abrupt decrements in performance at an AOA of 12 years were due to the closure of a CP at that age. As seen earlier in Fig. 9, AOA was correlated with variables that might have affected the Rule-based and Lexicon-based scores: Years of US education, Korean use, and English use. As well, Years of US education and frequency of English use correlated with one another, r(238)=.53, p<.05. As years of education in American schools increased, the likelihood that the Korean participants had received explicit instruction on some of the grammatical structures probed by the GJT was also likely to have increased. Perhaps more importantly, Years of education in the US was probably related to the number of enduring relationships the Korean immigrants established with NE speakers. This may have resulted in greater exposure to correct models of English morphosyntax and more English input overall.

Bahrick et al., (1994) noted that AOA and Years of education is commonly confounded in studies of immigrant populations. This is important in that Years of education exerts a strong effect on immigrants' self-reported proficiency in the L2 (Hakuta et al., 2003). Flege et al. (1999) used a subgroup matching task to circumvent the AOA-

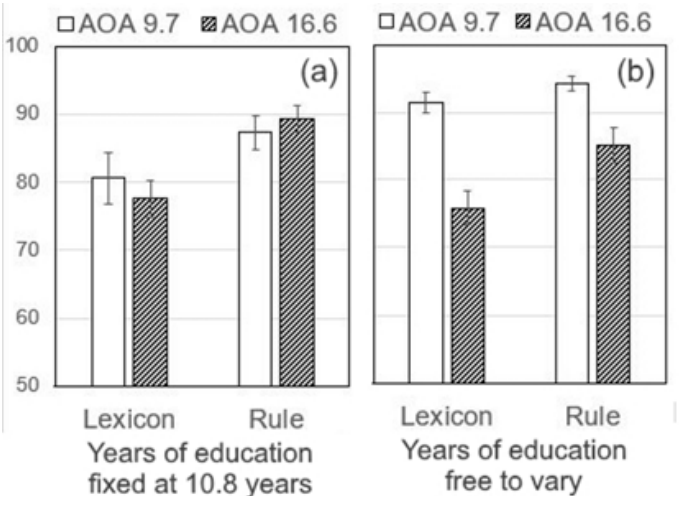

Figure 15. Mean Lexicon-based and Rule-based scores for subgroups of 20 Koreans each who did or did not differ in years of education in the U.S.

Years of education confound. Two subgroups of 20 Koreans each were identified from the larger sample of 240 participants. Members of the two groups had non-overlapping AOA values (*M*=9.7 vs 16.6 years) but were matched for Years of education in the US (*M*=10.8 years for both).

The GJT scores obtained by Flege e al. (1999) for the *matched* subgroups are shown in Fig. 15(a). As expected, the Rule-based scores were significantly higher than the Lexicon-based scores (*M*=89.4% vs 79.1%, $F(1,38)=31.3$, $p<.05$). Importantly, the difference between the Early and Late learners (*M*=84.0% vs 83.5%) was non-significant for both GJT scores ($p>.05$). The absence of a significant difference between the education-matched groups of Koreans having mean AOA values of 9.7 vs 16.5 years undermines the view that Early vs Late differences are due to age-related differences in state of neurocognitive maturation at the time L2 learning begins.

The absence of an AOA effect for matched subgroups might be attributed to a lack of statistical power. Flege et al. (1999) therefore carried out a control analysis comparing two new groups of 20 Early and 20 Late learners. These *unmatched* subgroups had the same mean AOA values as the those in the matched subgroup analysis (9.7 and 16.6 years). However participants in the two new groups of Early and Late learners differed significantly in terms of Years of education in the US (*M*=14.4 vs 8.0, $p<.05$) and the ratio of English/Korean use (*M*=1.6 vs 0.9, $p<.05$). because they were selected without regard to any variable other than AOA.[7]

As shown in Fig. 15(b), the unmatched Early and Late learners differed significantly as is usually the case ($F(1,38)=24.4$, $p<.05$). A two-way interaction was obtained, $F(1,38)=7.91$, $p<.05$, because the difference between Rule-based and Lexicon-based scores was significant for the Late learners ($p<.01$) but not the Early learners ($p>.05$). This finding suggests that the lack of a significant difference between Early and Late learners in the matched analysis was not due to a lack of statistical power.

As seen earlier in Fig. 9(b), AOA was also confounded with language use in the sample of Korean immigrants tested by Flege et al. (1999). Accordingly, a similar matched subgroup analysis was undertaken to evaluate the effect of AOA when language use was controlled. This analysis compared subgroups of 20 Koreans each who were matched for AOA but differed in terms of language use. Participants selected for the "High-ratio" subgroup had English/Korean use ratios greater than

[7] The small LOR difference between members of the two unmatched groups (*M* = 14.0 vs 12.1) was non-significant ($p > .05$).

1.4 (*M*=2.07, *range*=1.43-3.24) whereas those selected for the "Low-ratio" subgroup all had ratios less than 1.0 (*M*=0.73, *range*=0.50-0.99).[8] Both subgroups had a mean AOA of 13.0 years (*range*=5-22 years). The matched subgroups did not differ significantly in either age (*M*=25.4 vs 27.1 years) nor Years of education in the US (*M*=11.0 vs 11.6 years).

Selecting participants for the High-ratio and Low-ratio groups required the identification of Koreans who did not show the usual effect of AOA on language use. As seen in Fig. 9(b), most Late learners in the original sample of 240 used English slightly less often than Korean, yielding English/Korean ratios that were less than 1.0. On the other hand, most Early learners used English more than Korean, yielding ratios that were greater than 1.0. These statistical regularities did not hold true for all individuals, however. For example, a Late learner might have had an English/Korean use ratio greater than 1.0 if s/he used English exclusively at work and had mostly NE-speaking friends. An Early learner might have had an English/Korean ratio less than 1.0 if s/he needed to use Korean at work and was married to a Korean Late learner who wanted to speak only Korean at home.

Fig. 16 shows the scores obtained for the High-ratio and Low ratio subgroups. An ANOVA examining these scores yielded a significant interaction, $F(1,38)=4.49$, $p<.05$, for two reasons. First, the Rule-based and Lexicon-based scores differed significantly for the Low-ratio group (*M*= 87.5% vs 79.7%, $p<.05$) but not for the High-ratio group (*M*=92.2% vs 87.5%, $p>.05$). More importantly, the Koreans who used English more than Korean (i.e., High-ratio group members) obtained significantly higher Lexicon-based scores than those who used English less than Korean (i.e., Low-ratio group members, *M*=87.5% vs 79.7%, $p<.05$). However the High-ratio and Low-ratio groups did not differ significantly for the Rule-based scores (*M*=92.2% s 91.3%, $p>.05$). This find-

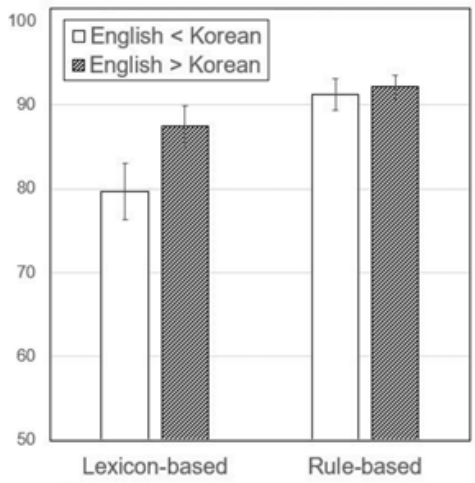

Figure 16. Mean percent correct scores obtained for Koreans who differed in language use but were matched for AOA.

[8] The ratio of self-reported use of Korean (9 questions) and English (7 questions) were used here to select participants. However Flege et al. (1999) used the mean frequency of Korean use.

ing suggests that a relatively large quantity of English input facilitates the learning of morphosyntactic properties of English, but only those properties that must be learned on a word-by-word basis (i.e., the Lexicon-based scores).

It is important to note, however, that both the High-ratio and Low-ratio groups obtained significantly lower Rule-based and Lexicon-based scores than the NE speakers did ($p<.05$ by Bonferroni-corrected one-sample t-tests). This should probably not be interpreted to mean that complete learning is impossible, however. This is because we cannot be sure that the Koreans had obtained adequate English input.

First, we have no information regarding how much of the English input the Koreans received consisted of correct productions of English sentences by NE speakers. If the Koreans often heard ungrammatical English sentences produced by other Koreans they would not, of course, be expected to respond just like monolingual native speakers of English. This assumes, of course, that the NE speakers who formed the comparison group had heard only well-formed utterances produced by fellow NE speakers.

Second, the quantity of input the Korean immigrants had received may have been insufficient for complete learning to have occurred. I calculated percent English use values from the ratings presented earlier. Members of the High-ratio group used English more frequently than members of the Low-ratio group did (M=66.7% vs 41.8%) and had also lived somewhat longer in the US than members of the Low-ratio group had (M=14.5 vs 12.7 years). Multiplying the percentage use estimates by LOR provides an estimate of Years of full-time English input. By this measure, members of the High-ratio group had received substantially more English input than members of the Low-ratio group had (M=9.7 vs 5.4 years, $p<.05$).

However, 9.7 years of full-time English input distributed over 15 years of residence may not have been enough for the Korean immigrants to have learned some aspects of English morphosyntax completely. Even though members of the NE comparison group had an average age of 27 years, only half of them were at ceiling for the Lexicon-based scores. Learning some things takes time. Indeed, some of the 90 native English adults tested by Dąbrowska (2018) performed at a chance level for certain English grammatical constructions even though they had a mean age of 38 years (*range*=17-65) and had performed well on other constructions. Importantly, Dąbrowska (2018) found that some aspects of grammatical knowledge depended on amount of formal education and print exposure.

This finding points to the importance of input, even for monolingual adult native speakers of English.

Hartshorne et al. (2018) proposed that NE speakers may need as much as 30 years of full-time input to completely learn English grammar. One member of the High-ratio group had in fact lived in the US for 30 years. However this outlier had received only the equivalent of 20.3 Years of full-time English input because he used English only part of the time. His Lexicon-based score was 91% correct. One wonders if this participant's learning of English morphosyntax had stopped irreversibly, or if it would continue improving slowly over time until reaching a native-like level of attainment once 30 years of full-time English input had been received.

4. Speech vs morphosyntax learning

The CP hypothesis has been applied to the learning of both L2 speech and morphosyntax. Investigators have tended to treat L2 pronunciation and morphosyntax as separate entities that require different kinds of learning and different explanations for age-related effects. Some think that a CP closes sooner for the learning of L2 speech than morphosyntax (e.g., Long, 1990) and others that a CP exists only for L2 speech learning (e.g., Scovel, 1988; Bahrick et al., 1994). The seemingly greater difficulty in pronouncing an L2 without a FA than in obtaining morphosyntax scores equaling those of L2 native speakers has been attributed to differences in the extent to which learners relate structures found in the L1 and L2 (e.g., MacWhinney, 1992) or the neuromotor component involved in speech production (Zatorre, 1989) and possibly speech perception (Best, 1995).

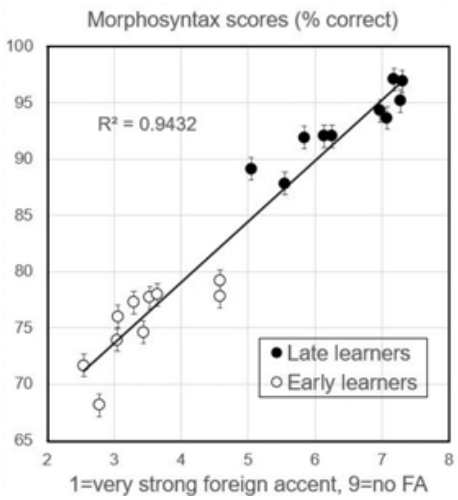

Figure 17. Mean foreign accent ratings and morphosyntax scores for 20 groups of 12 Korean each who differed in AOA.

Oyama (1973) and Patkowski (1980, 1990) reported somewhat stronger correlations between AOA and L2 pronunciation than between AOA and L2 morphosyntax scores. Not surprisingly, Flege et al. (1999) observed a stronger correlation between AOA and overall degree of FA than between AOA and GJT scores, $r(238)=-.85$ vs $-.75$, $X(1)=19.9$, $p<.05$.

I carried out an *F*-test comparing the Early and Late learners tested by Flege et al. (1999). AOA accounted for slightly more variance in FA ratings than GJT scores did (η^2=.623 vs .548). Fig. 17 shows that the FA ratings and GJT scores obtained for 20 AOA-defined Korean groups correlated strongly with one another ($r(18)$=.97, p<.05). The FA ratings and GJT scores also correlated significantly when the scores obtained for 120 individual Early learners and 120 individual Late learners were examined ($r(118)$=.62 vs .58, p<.05).

The similar correlations obtained here for learning in the two linguistic domains, both for Early and for Late learners, suggest that an important commonality exits for the learning of L2 speech and morphosyntax. That underlying commonality may be input. Differences in the quantity and quality of input that Early and Late learners receive exert a strong effect on the learning of L2 speech and at least some aspects of L2 morphosyntax learning.

5. A non-critical period for L2 learning

5.1 LOR and ultimate attainment

The critical period (CP) hypothesis proposed by Lenneberg (1967) for L2 speech learning derived from the observation that people who learn an L2 after puberty usually speak it with a FA. Lenneberg probably assumed that a detectable FA remains evident in the speech of post-pubescent learners even after they have received abundant native-speaker input for many years. Had this not been so, Lenneberg's observation regarding foreign accent would not have evoked widespread interest and launched a new research subdiscipline.

The unspoken assumption I just attributed to Lenneberg (1967) later surfaced in the L2 acquisition literature under names such as *ultimate attainment*, *end state* and *asymptotic learning* (Birdsong, 2013). Researchers soon became attentive to the potential role of L2 experience. For example, to ensure ultimate attainment in English morphosyntax learning, DeKeyser (2000) required that all 57 of the Hungarian immigrants he tested in the Pittsburgh area have lived in the US for at least 10 years. The GJT scores obtained for Hungarians who arrived in the US before vs after the age of 15 years showed little overlap. This led DeKeyser to conclude that Early vs Late differences derive from changes in cognitive processing. More specifically, DeKeyser concluded (2000, p. 518) that "somewhere between the ages of 6-7 and 16-17, everybody loses the mental equipment required for [learning] the abstract patterns underlying a human language".

DeKeyser also concluded (2000, p. 518) that input differences "are not a good explanation for age effects" on L2 morphosyntax learning. This conclusion was based on the complete absence of correlation between the Hungarians' lengths of residence in the US and their GJT scores. It seemed that once the Hungarians had reached what DeKeyser thought was their ultimate attainment in English after 10 years of US residence further increases in LOR no longer resulted in much if any further improvement.

The conclusion drawn by DeKeyser (2000) regarding the inefficacy of LOR beyond 10 years of residence in a predominantly L2-speaking country was supported by two findings. For the 240 Koreans tested by Flege et al. (1999), GJT scores showed a modest correlation with LOR, $r(238)=.39$, $p<.05$. However the strength of correlation between test scores and LOR decreased to $r(181)=.23$ ($p<.05$) for the subset of Koreans who had lived in the US for more than 10 years, and to a non-significant $r(151)=0.12$ for Koreans who had lived in the US for more than 12 years. In an analysis of US census data, Stevens (1999) found that immigrants' self- reported proficiency in English increased rapidly as a function of LOR until about 10 years of residence in the US, but that additional increments in LOR beyond 10 years were associated with little further improvement in self-reported English-language proficiency.

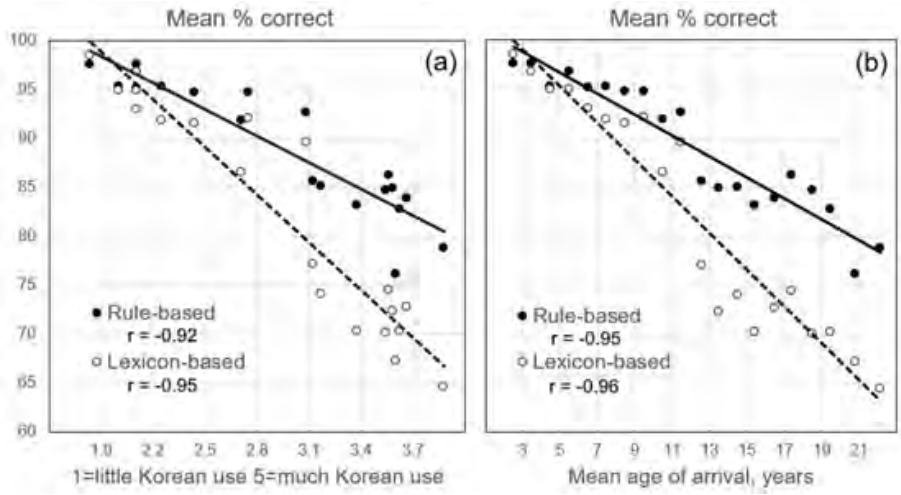

Figure 18. Mean morphosyntax scores obtained for 20 groups of 12 Koreans each as a function of self-estimated Korean use (left) and AOA (right).

The conclusion drawn by DeKeyser (2000) regarding the scant role of input in L2 morphosyntax learning, however, is questionable. This is because LOR does not provide an adequate index of quantity of L2 input (e.g., Flege & Liu, 2001) and offers no insight at all regarding the quality

of L2 input that immigrants receive. DeKeyser (2000) did not attempt to determine how much, with whom, in what social contexts, and for what purpose(s) his native Hungarian participants used English. Had this been done he might well have drawn a different conclusion regarding the role of input.

Fig. 18(a) shows the Rule-based and Lexicon-based scores obtained by Flege et al. (1999) for 20 AOA-defined groups of Koreans as a function of the groups' average estimated frequencies of Korean use in four language-optional contexts (see Fig. 10).[9] The same GJT scores are shown in Fig. 18(b) as a function of AOA.

The correlations between the GJT scores and the two predictor variables (frequency of Korean use, AOA) were both very strong. Somewhat weaker but still highly significant correlations between the GJT scores and the two predictor variables were obtained when the scores for all 240 Koreans were examined rather than mean values obtained for groups of 12 Koreans each. In these analyses the correlations between Rule-based and Lexicon-based scores and AOA were $r(238)$=-.60 and -.74, $p<.05$, while the correlations between the test scores and self-reported frequency of Korean use were $r(238)$ =-.48 and -.60, $p<.05$.

The correlations with AOA obtained for individual participants were significantly stronger than the correlations with the Korean use estimates. This held true for both GJT scores ($p<.05$). The difference in strengths of correlation was probably due to measurement precision. Language use estimates obtained using rating scales on a language background questionnaire are inherently noisier than AOA, an objective measure that is correlated with other variables likely to influence GJT scores. In fact, when the effect of Years of education in US schools was partialled out, the only correlation that remained significant was that between the Lexicon-based morphosyntax scores and Korean use, $r(237)$=-.21, Bonferroni corrected $p<.05$.

The results of this analysis suggest that there is no real justification, other than tradition, for concluding that AOA offers a better explanation of Early vs Late differences than variation in language use does. Indeed, it is plausible to hypothesize that variation in input is the single most important

[9] The bilinguals tested by Flege et al. (1999) spoke only English and Korean, which means that English use was the inverse of Korean use. In the analysis presented here I used Korean rather than English use estimates because the Koreans were not asked to estimate English use in the four context-optional contexts which, in retrospect, appear to have been the most indicative.

predictor of Early vs Late differences. AOA is regarded as a stand-in for other potentially causative variables such as state of neurocognitive maturation at the time of first exposure to an L2. Frequency of language use, on the other hand, relates directly to the input that is essential for the learning of L2 speech and at least some aspects of L2 morphosyntax.

5.2 Large sample studies

Most investigations of L2 learning by immigrants (e.g., Flege et al., 1999; DeKeyser, 2000) have examined relatively small numbers of participants. Here I consider two studies that examined far larger numbers of participants than is typical for L2 research.

Hartshorne et al. (2018) obtained responses to a 10-min grammar quiz from individuals who had learned English as a foreign or second language. These authors obtained quiz scores via the internet from 45,067 *immersion learners,* respondents who had spent at least 90% of their lives in an English-speaking country once having arrived there; from 266,701 *non-immersion* learners, who had spent at most 10% of their lives – but never more than one year – in an English-speaking country since first exposure to English; and from 246,497 native speakers of English from around the world. Some non-immersion learners reported having first been exposed to English at school and to have received subsequent input by watching TV programs and movies in English (2018, p. 266) although the sources of input and the context of first exposure to English were not systematically examined.

Hartshorne et al. (2018) obtained substantially higher grammar quiz scores for immersion than non-immersion learners of English (see Fig. 6). This supports the view that performance in an L2 depends on the amount of input received, assuming of course that the immersion learners had indeed received more English-language input than the non-immersion respondents had.

The aim of one analysis of special interest was to estimate the age of first exposure to an L2 beyond which "mastery … [of L2 grammar] is no longer attainable" (Hartshorne et al., 2018, p. 270). This analysis examined scores obtained from 25% of the original immersion group respondents and 11% of the non-immersion respondents. To be included in this analysis respondents had to be less than 70 years of age and to have had at least 30 years of experience in English. The first criterion was meant to obviate the influence of cognitive losses due to normal aging. The second criterion was meant to ensure ultimate attainment in English proficiency. The

variable *Years of experience* indicated the difference, in years, between the respondents' age at test and the age or arrival in a predominantly English-speaking country (immersion learners) or the age of first exposure (non-immersion learners).

Hartshorne et al. (2018) found that the immersion learners showed "little decline in ultimate attainment [in English] until an age of first exposure of 12 years" whereas the non-immersion learners showed no decline until the age of 9 years and a "sharp decline" thereafter (p. 270). The authors concluded that to reach a "native-like level of proficiency in English [grammar]" the slow process of L2 learning must "start by 10-12 years of age" (p. 270). An early start is needed, according to the authors, because progress must begin well before a "sharp drop in learning rate [occurs] at about 17-18 years of age" (p. 270).

The conclusion that a CP for L2 learning closes at around 17-18 years of age might be questioned for several reasons. First, the test used by Hartshorne et al. (2018) to evaluate knowledge of English morphosyntax was an exclusively written instrument. It is safe to assume that all respondents to the internet grammar quiz could read English but not that all of them could speak English. Stevens (1999) found that the proportion of immigrants to the US who were unable to speak English increased steadily between the ages of 5 and 60 years.

Second, the Hartshorne et al. (2018) findings for immigrants (i.e., immersion learners) differed substantially from the findings obtained in another study examining a large number of immigrants. Hakuta et al. (2003) analyzed data obtained in the 1990 US Census for 2.02 million immigrants from a Spanish language background and 0.32 million immigrants from a Chinese language background. Respondents to the US census had been asked a series of questions regarding their proficiency in English. Their responses were used to construct a 4-point English proficiency scale for each respondent; these scores were then modelled. The functions obtained by Hakuta et al. (2003) showed no evidence of a discontinuity at the age of 12 or 15 years nor a discontinuity at any other age of immigration. Instead, self-rated proficiency in English was observed to decline gradually from ages of arrival that ranged from 5 to 60 years for both linguistic groups. The authors attributed this slow decline to normal cognitive aging over the lifespan.

Hartshorne et al. (2018, p. 269) questioned the validity of data obtained from the US census because these data were based on 4-point rating scales. However Hakuta et al. (2003, p. 32) cited validation research

which yielded moderate correlations of r=.52 and .54 with the English proficiency ratings they analyzed. Moreover, the findings of Flege et al. (1999) indicated that immigrants have a realistic understanding of their own L2 competence. The 240 Koreans rated their own proficiency in English using 5-point rating scales. Their self-ratings of English pronunciation correlated strongly with native English listeners' ratings of their pronunciation of English, and the Koreans' self-ratings of English grammatical knowledge correlated with the scores they obtained on the144-item GJT described earlier (r=.64, p<.05).

The two large-sample studies just cited led to very different patterns of data, and so to different interpretations of Early vs Late differences. The between-study differences may have arisen, at least in part, from sampling procedures. A general problem for L2 acquisition research is that participants are not randomly selected, leading to interpretive difficulties. For example, studies in which relatively well-educated participants are recruited on or near a university campus (e.g., Johnson & Newport, 1987; Flege et al., 1999) may not generalize to the population of persons around the world who learn English as a second or foreign language.

A potentially more serious problem arises for the analysis of samples that are systematically biased or skewed. For example, the census data analyzed by Hakuta et al. (2003) did not include respondents who reported no longer using their L1 while living in the US. This eliminated respondents who were likely to have lived longer than average in the US, and so to have achieved greater than average proficiency in English than respondents who continued to use their L1 and whose proficiency ratings were available for analysis (Stevens, 2004, p. 215). The possibility exists, therefore, that selection bias contributed to the absence of support for a CP in the Hakuta et al. (2003) study.

One wonders, in the same vein, if selection bias contributed to the strong support for a CP obtained by Hartshorne et al. (2018) . The individuals who provided data for this innovative internet study were not, in fact, selected. They volunteered to participate after having learned about the grammar quiz, usually via social media. The authors selected a subset of respondents for the analysis mentioned earlier after having excluded respondents who provided obviously spurious data. The respondents retained for the analysis were nevertheless likely to have had a greater than average interest in language and language learning than the population of humans around the world who learn English as a foreign or second language. The influence of this selection bias, if indeed one existed, is unknown.

5.3 The crucial role of input

The conclusion by Hartshorne et al. (2018) that a critical period for the learning of L2 morphosyntax closes at 17-18 years of age is difficult to evaluate in the absence of information regarding input. The authors did not ask respondents to indicate how often, with whom, in what contexts, or for what purposes(s) they used English. Use of the internet to obtain L2 acquisition data is an important step forward and is likely to become common. Such research will need to provide information regarding input, however, to yield interpretable results.

The need for input data is evident from inspection of individual data obtained by Hartshorne et al. (2018), which can be downloaded from https://osf.io/pyb8s/wiki/home/. Respondents who were speakers of the same L1 and reported having begun to learn English at the same age sometimes responded differently when asked to indicate their primary language(s). Many indicated having just one primary language, their native language. However others indicated that both their native language and English were primary languages, and others still indicated that English was their only primary language at the time of test.

Self-reported primary language(s) was not used in statistical modeling by Hartshorne et al. (2018), but one naturally wonders what accounted for such variation. The respondents who did/did not report English to be a primary language may have differed in one or more important ways, for example: living on a long-term basis with a native speaker(s) of English, being required to use English at work, or having the desire or need to use English for important social, recreational or religious activities.

The belief that a critical period (CP) exists for L2 learning has motivated a large amount of L2 research. Very little of this research, however, has aimed to determine if a CP actually exists. Most researchers simply assume the existence of a CP because of the ubiquitous presence of Early vs Late differences and simply seek to determine when and how rapidly the CP closes. Another aim of CP-inspired research has been to identify the underlying cause(s) of a loss or diminution of learning capacity that is hypothesized to occur after closure of the CP. This research, unfortunately, has met with scant success. To take an early example: Lenneberg (1967) thought that a CP for L2 speech learning closes when hemispheric specialization for language functions reaches completion. This proposed mechanism was soon discarded when evidence contradicting it emerged (e.g., Krashen, 1973).

Hartshorne et al. (2018, p. 263) defined the critical period for L2 learning as a "theory-neutral descriptor of diminished achievement, whatever its cause". These authors culled seven possible causes for diminished achievement from the L2 acquisition literature. It has been hypothesized that, in comparison to Late learners, Early learners may: (1) have greater neural plasticity; (2) have the opportunity to learn the L2 over a longer period of time; (3) have less cognitive processing ability, which prevents them from being "distracted by irrelevant information"; (4) experience less "interference" from previously learned L1 structures; (5) show a "greater willingness to experiment and make errors"; (6) feel a "greater desire" to conform to peers; or (7) be more likely to immerse themselves in a "community of native speakers" (2018, p. 263).

It is noteworthy that four of the seven items on this list of potential causes of diminished achievement (viz., 2, 5, 6, 7) relate in some way to input. For example, someone who has a relatively strong desire to speak an L2 like his/her peers presumably spends time with those peers, which raises the question of "Who are they? Are they native or non-native speakers of the L2?" In a similar vein, one might ask "Who are the L2 learners who decide to immerse themselves in a community of L2 native speakers?" According to Stevens (1999, p. 574) learning an L2 at any age "requires exposure to the language, motivation, and opportunities to practice receptive and active skills. In short, language learning requires communicative and social interactions." Moyers (2009, p. 161) argues that age of first exposure to an L2 "leaves much to be desired as an explanation for what is a very complex endeavor – one that is, by its nature, grounded in a social framework".

6. Conclusions

Evidence reviewed in this chapter indicated that the quantity and quality of L2 input that learners receive influence the long-term learning of L2 speech and at least some aspects of L2 morphosyntax learning. The age at which immigrants are first exposed to an L2 conditions the quantity and quality of L2 input they are likely to receive over the course of their lives in the host country. Early arrivals usually receive more and better L2 input than do immigrants who arrive later in life. This, in my opinion, is the primary basis for the Early vs Late differences that have been widely reported in the literature examining second-language acquisition.

My conclusion may surprise some readers given that input has tended to be ignored or downplayed in L2 research (Flege, 2009). The absence of attention to input seems to have derived, at least in part, from the mistaken belief that LOR provides an adequate measure of input. Another explanation is that many researchers believe that obtaining adequate measures of input is impossible. However the technology needed to obtain precise measures of the quantity and the quality of input now exists (Flege & Wayland, 2019). For those who are unwilling or unable to use this admittedly expensive and time-consuming approach, a simple paper and pencil instrument might be used to good advantage to assess L2 input.

As illustrated in Fig. 19, participants could be asked to indicate two percentages of language use, one for their L1 and the other for their L2, in a wide range of specific social contexts. When taken together, the responses would serve to define the participants' everyday language use patterns. (The two percentages would necessarily sum to 100% in research with bilinguals.) In the final four items of such an instrument, participants would be asked to indicate their overall percentage use of the L1 and L2 in their first, third, and fifth years of residence in the host country, and also their overall language use in the year preceding the test. The purpose of these last four items is to evaluate possible changes over time in L2 use.

Figure 19. Illustration of an instrument for assessing language use in social contexts. Responses for the L1 and L2 must total 100%.

In summary, the pattern of data reviewed in this chapter leads me to conclude that long-term success in L2 learning is not limited by the closure of a critical period resulting in either the loss or the diminution of L2 learning capacity. Instead, degree of long-term achievement in L2 learning is determined probabilistically by a *non-critical period* that is defined primarily by age-related variation in the L2 input that learners normally receive owing to their differing motivations to learn the L2 and exposure to the L2 in differing social contexts.

Input is crucial for the learning of L2 speech and some aspects of L2 morphosyntax, just as it is for other socially defined human activities. Input is unlikely, of course, to be the only factor that influences long-term success in L2 learning. Other factors, for example, language learning aptitude (e.g., Abrahamsson & Hyltenstam, 2008), might also be found to influence long-term success in L2 learning. However, the role of these other factors cannot be understood via research designs that fail to control for variation in input.

The hypothesis that the closure of a critical period for L2 learning limits long-term success in L2 learning (Lenneberg, 1967) has inspired a large amount of research examining both L2 speech and morphosyntax. However the widespread appeal of the CP hypothesis has also impeded progress in L2 research focusing on age-related effects by encouraging investigators to ignore or downplay the crucial role of input. The time has come for the establishment of a new paradigm based on research designs favoring the evaluation of input and other factors that influence long-term success in L2 learning. The participants in future research with immigrants, in particular, should be selected on the basis of the input they have received at the time of test rather than on the basis of their presumptive stage of neurocognitive maturation years earlier at the time of immigration.

Acknowledgments

This research was supported by grants from the National Institute of Deafness and Other Communicative Disorders to the University of Alabama at Birmingham. I thank Ocke-Schwen Bohn, Anders Højen, Thorsten Piske, Anja Steinlen, and Tullia Trevisan for comments on an earlier version.

References

Abrahamsson, J., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*(4), 481-509.

Asher, J., & Garcia, R. (1969). The optimal age to learn a foreign language. *The Modern Language Journal*, *55*(5), 334-341.

Bahrick, H., Hall, L., Goggin, J., Bahrick, L., & Berger, S. (1994). Fifty years of language maintenance and language dominance in bilingual Hispanic immigrants. *Journal of Experimental Psychology: General*, *123*(3), 264-283.

Beck, M.-L. (1997). Regular verbs, past tense and frequency. Tracking down a potential source of NS/NNS competence differences. *Second Language Research*, *13*, 93-115.

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-206) Timonium, MD: York Press.

Birdsong, D. (2013). Age and the end state of second language acquisition. In W. Ritchie & T. Bhatia (Eds.) *The new handbook of second language acquisition (*pp. 401-424). Bingly, England: Emerald Group Publishing.

Birdsong, D., & Birdsong, D. (2001). Regular-irregular dissociations in L2 acquisition of English morphology. BUCLD 25: *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 123-132). Boston, MA: Cascadilla Press.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499-533.

Flege, J. (1987). A critical period for L2 learning to pronounce foreign languages? *Applied Linguistics*, *8*, 162-177.

Flege, J. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229-273). Timonium, MD: York Press.

Flege, J. (2009). Give input a chance! In Piske & Young-Scholten (Eds.) *Input Matters in SLA* (pp. 175-190). Bristol: Multilingual Matters.

Flege, J. (2018). It's input that matters most, not age. *Bilingualism: Language and Cognition*. https://doi.org/10.1017/S136672891800010X, Published online 15 May 2018

Flege, J., Frieda, E., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*, 169-186.

Flege, J., Frieda, E., Walley, A., & Randazza, L. (1998). Lexical factors and segmental accuracy in second language speech production. *Studies in Second Language Acquisition*, *20*, 155-187.

Flege, J., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*, 527-552.

Flege, J., & MacKay, I. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*, 1-34.

Flege, J., & MacKay, I. (2011). What accounts for "age" effects on overall degree of foreign accent? In M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in the acquisition of second language speech. New Sounds 2010*, Vol. 2 (pp. 65-82). Bern: Peter Lang.

Flege. J., & Munro, M. (1994). The word unit in second language speech production and perception. *Studies in Second Language Acquisition*, *16*, 381-411.

Flege, J., Munro, M., & MacKay, I. (1995a). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*, 3125-3134.

Flege, J., Munro, M., & MacKay, I. (1995b). Effects of age of second-language learning on the production of English consonants. *Speech Communication*, *16*, 1-26.

Flege, J., & R. Wayland (2019). The role of input in native Spanish Late learners' production and perception of English phonetic segments. *Journal of Second Language Studies*, *2*(1). In press.

Geary, D. (1998). *Male, female. The evolution of human sex differences*. Washington, D.C.: American Psychological Association.

Hakuta, K., Bialystok, E. & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, *14*(1), 31-38.

Hartshorne, J., Tenenbaum, J., & Pinker, S. (2018) A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263-277.

Hopp, H., & Schmid, M. (2013). Perceived foreign accent in first language attrition and second language acquisition. *Applied Psycholinguistics*, *34*, 361-394.

Krashen, S. (1973). Lateralization, language learning, and the critical period: Some new evidence. *Language Learning 321*(3), 63-74.

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

Long, M., (1990) Maturational constraints on language development. *Studies in Second Language Acquisition*, *12*, 251-285.

MacWhinney, B. (1992). Transfer and competition in second language learning. In R. Harris (Ed.) *Cognitive processing in bilinguals* (pp. 371-390). Amsterdam: North Holland.

Moyer, A. (2009). Input as a critical means to an end: Quantity and quality of experience in L2 phonological attainment. In T. Piske & M. Young-Scholten, M. (Eds.), *Input matters in SLA* (pp. 159-174). Bristol: Multilingual Matters.

Oyama, S. (1973). *A sensitive period for the acquisition of a second language*. Unpublished Harvard University Ph.D. thesis.

MacKay, I., Meador, D., & Flege, J. (2001). The identification of English consonants by native speakers of Italian. *Phonetica*, *58*, 103-125.

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

Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics*, *11*, 73-89,

Pinker, S., & Prince, A. (1991). Regular and irregular morphology and the psychological status of rules of grammar. In Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on the Grammar of Event Structure, pp. 230-251. DOI: http://dx.doi.org/10.3765/bls.v17i0.1624.

Prasada, S., & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. Language and Cognitive Processes, *8*(1) , 1-56.

Piske, T., Flege, J., MacKay, I., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica*, *59*, 49-71.

Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*, 191-215.

Piske, T., & Young-Scholten, M. (2009). *Input matters in SLA*. Bristol: Multilingual Matters.

Scovel, T. (1988). *A Time to Speak: A Psycholinguistic inquiry into the critical period for human speech*. New York: Newbury House Publishers.

Stevens, G. (1999). Age at immigration and second language proficiency among foreign-born adults. *Language in Society*, *28*, 555-578.

Stevens, G. (2004). Using census data to test the critical-period hypothesis for second-language acquisition. *Psychological Science*, *15*(3), 215-216.

Yeni-Komshian, G., Flege, J., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition*, *3*(2), 131-149.

Zatorre, R. (1989). On the representation of multiple languages in the brain: Old problems and new directions. *Brain & Language*, *36*, 127-147.