Title	Proficient foreign-language users show faster symbol processing
Sub Title	
Author	尾島, 司郎(Ojima, Shiro) 永井, 敦(Nagai, Atsushi) 田谷, 文彦(Taya, Fumihiko) 大津, 由紀雄(Otsu, Yukio) 渡辺, 茂(Watanabe, Shigeru)
Publisher	Centre for Advanced Research on Logic and Sensibility The Global Centers of Excellence Program, Keio University
Publication year	2012
Jtitle	CARLS series of advanced study of logic and sensibility Vol.5, (2011.) ,p.163- 180
JaLC DOI	
Abstract	
Notes	I. Study of Logic and Sensibility Part 3 : Cognition and Language
Genre	Research Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO12002001-20120224-0163

16 Proficient Foreign-language Users Show Faster Symbol Processing

Shiro Ojima[1], Atsushi Nagai[2],
Fumihiko Taya[1], Yukio Otsu[3],
Shigeru Watanabe[4]

[1] Centre for Advanced Research on Logic and Sensibility (CARLS), Keio University
[2] Graduate School of Human Relations, Keio University
[3] Institute of Cultural and Linguistic Studies, Keio University
[4] Department of Psychology, Keio University

Behavioral evidence suggests that higher foreign-language (FL) proficiency is associated with higher meta-linguistic awareness about the mother tongue (first language or L1). Or simply put, people who have higher FL proficiency know their L1 better. Evidence for this view rests on participants' conscious judgments on language materials. The present study shows that differences associated with FL proficiency can be seen in participants' brain activity that precedes their conscious judgments. During the experiment, Japanese adults with either high or low English (FL) proficiency silently read sentences in Japanese (L1) displayed on a monitor. The participants' event-related brain potentials (ERPs) time-locked to the presentation of the critical words indicated that the individual sub-stages of sentence processing in the L1 occurred earlier in the participants with higher FL proficiency. The same participants were also tested in the processing of arithmetic, in order to see if the effects of FL proficiency extend to other types of symbol processing. Higher FL proficiency was again found to be associated with faster processing. These data suggest that higher FL proficiency may be correlated not only with higher meta-linguistic L1 awareness but also with faster symbol processing commonly involved in L1 processing and arithmetic processing.

I. Introduction

Not everyone has a high command of a foreign language (FL). Some people have high FL proficiency while others don't. It has been of great scientific and educational interest whether people with high FL proficiency and those with low proficiency differ in aspects other than FL proficiency itself. The existence of two languages in one child's mind is sometimes considered to have negative effects on the child's mother tongue (L1), but this view is refuted by many empirical studies, which as a whole showed that FL learning and bilingual acquisition have positive effects on children's L1 (Armstrong & Rogers, 1997; Cunningham & Graham, 2000; Garfinkel & Tabor, 1991; Taylor & Lafayette, 2010; Thomas, Collier, & Abbott, 1993; Yelland, Pollard, & Mercuri, 1993). Positive relations between FL learning and the learner's L1 have also been found in teenagers (Cooper, 1987; Kecskes & Papp, 2000; Masciantonio, 1977).

For this to be possible, there must be something common across the FL and the L1 in one's mind. This possibility has been conceptualized as the dual-iceberg representation of bilingual proficiency (Cummins, 1984, 1991). An iceberg floating on the ocean has a massive volume of ice hidden beneath the sea surface. The iceberg may have two or more separate tips or peaks above the sea surface, but they are connected under the sea surface. The FL and the L1 may look separate and different superficially, but there may be common underlying proficiency (CUP) that is invisible on the surface. Through this CUP commonly shared by the FL and the L1, the effects of learning in one language are considered to get transferred to the other. It is also possible that there is genetically determined individual variation in CUP. If so, people with superior CUP will be more likely to attain high proficiency in both the L1 and the FL, than people with inferior CUP who are given the same language-learning experience.

The present study compares two groups of adults who differ greatly in their proficiency of a foreign language, using not only conventional behavioral measures but also recent measures of brain activity. The CUP, other related hypotheses consistent with it, and the supporting evidence for these views, all focus on language speakers' conscious behavior, often involving academic language skills and meta-linguistic awareness. By measuring brain activity online, we test the possibility that people with different FL proficiency may also differ in some basic stages of L1 processing that occur before conscious behavior is produced. To record brain activity that occurs during language processing, we used the event-related brain potential (ERP) technique. ERPs are electrical reflections of ongoing neural activity that is triggered by some sensory or cognitive event. ERPs have the potential to visualize neural activity related to language processing even without assigning the participant an explicit task (Bentin, Kutas, & Hillyard, 1995). Below we analyze ERPs obtained from native Japanese speakers while they were silently reading Japanese sentences.

The second possibility that we try to test is that people with different FL proficiency may differ in other kinds of symbol processing such as the processing of arithmetic. If differences in FL proficiency are associated with differences in basic aspects of L1 processing, a natural next question would be whether such differences exist only in the domain of language or extend to other kinds of cognition that involves symbols. Hence we obtained ERP data from the same participants while they were silently viewing arithmetic equations.

II. Methods

1. Participants

Two groups of healthy, right-handed, native Japanese-speaking adults participated after providing written consent. The two groups primarily differed in English proficiency as measured by TOEIC (Test of English for International Communication), which tests receptive English proficiency in both listening and reading. The high English proficiency group (n = 22, 10 men, mean age 31.0 years old, age range 24–36) had a TOEIC score of 850 or higher, whereas the low proficiency group (n = 20, 9 men, mean age 31.5, age range 23–35) had a score of 350 or lower (the score range of TOEIC is 5 to 995). The handedness quotients (Oldfield, 1971) were similar between the groups (high: .98, low: .97, where 1 = completely right handed and –1 = completely left handed). All the participants were born to native Japanese-speaking parents, had grown up in Japan, and started English learning at or slightly before 12 years of age (high: mean 11.8 years of age, min 10, low:

mean 11.9, min 10). Through a detailed questionnaire, we estimated total hours of English learning that each participant had done since their first contact with English. As expected, the high English proficiency group had learned English for far longer hours (average 4,380 hours, median 3,210, range 1,740–12,300) than had the low proficiency group (average 1,416 hours, median 1,260, range 540–2,970).

Information about the participants' families was also obtained by the questionnaire. The two groups were compared by Mann-Whitney's U-test in the relevant variables. First, they did not differ in the father's and mother's level of English as a foreign language (ps > .9), rated by the participant on a 5-point scale (1 = novice or no knowledge, 2 = intermediate level, 3 = advanced level, 4 = near-native, 5 = native). Nor did they differ in the income level of the family in which the participant had grown up, subjectively rated on a 9-point scale (1 = lowest, 5 = average of Japan, 9 = highest, p =.349), and in the degree of economic satisfaction of the family, again subjectively rated on a 9-point scale (1 = maximally dissatisfactory, 9 = maximally satisfactory, $p = .220$). However, both the father's and mother's education level (such as high-school graduates, college graduates, university graduates, etc.) were higher for the participants in the high English proficiency group than for those in the low proficiency group ($p = .047$ for father's education level, and $p = .013$ for mother's). The participants themselves were all junior-college or university graduates, and we compared the two groups in the levels of the universities (or colleges) they had graduated from. In Japan, each academic department in each university is associated with a particular *deviation score*, which is an index of the difficulty of entrance and thus the level of the university (more precisely, a particular department of a particular university). With deviation scores, it is possible to compare university levels. In our samples, the high English proficiency group showed significantly higher university levels than did the low proficiency group (mean 61.6 vs. 48.9, $t(40) = 6.127$, $p < .001$, two-tailed). This is an inevitable difference, because English is a major part of entrance exams of Japanese universities (one of the three core subjects beside Japanese and mathematics), and students with higher English proficiency can enter higher-levels of universities with other things being equal. In statistical comparisons of the two participant groups in various behavioral and neural variables to be presented below, we used university level (quantified as deviation score) as a covariate where necessary, to see if the group differences are truly due to the difference in English proficiency or a byproduct of the difference in the university level.

The study was approved by the ethics committee of the Faculty of Letters at Keio University.

2. Behavioral test batteries

2.1. English grammar

Knowledge of basic aspects of English grammar (syntax and morphology) was inquired by means of a grammaticality judgment task on written English sentences. This was done as a complement to TOEIC, which focuses on more advanced aspects of English usage. We compiled 100 test items (50 grammatical, 50 ungrammatical) from example sentences described in two previously published papers, one on agrammatic aphasia (Linebarger, Schwartz, & Saffran, 1983) and the other on second language acquisition (Johnson & Newport, 1989), excluding duplicates between the two studies. Test sentences were displayed one by one on a monitor. Upon reading each sentence, the participant made a grammaticality judgment and indicated their judgment by pressing one of the two active buttons on a response pad. Half of the participants pressed the right button for grammatical sentences and the left button for ungrammatical sentences. For the other half of the participants, the sides of the buttons swapped. The sentence remained on the monitor until a button press, and the participant did this task on their own pace. The accuracies of the participants' responses were analyzed statistically by means of an analysis of variance (ANOVA) with a between-subject factor Group (high vs. low English proficiency) and a within-subject factor Sentence Type (grammatical vs. ungrammatical). A response was considered accurate either if a grammatical sentence was judged as grammatical or if an ungrammatical sentence was judged as ungrammatical.

2.2. Japanese grammar

In a procedure identical to the English grammaticality test, we tested the participants' meta-linguistic awareness of the Japanese grammar. All the participants were adult native speakers of Japanese and were considered to share the core knowledge of the Japanese grammar. However, individual variation is likely to exist among native speakers in meta-linguistic awareness about the grammar of one's L1. To create a Japanese grammaticality judgment test, we consulted linguistic papers and books on the Japanese grammar and collected samples of ungrammatical Japanese sentences used for the construction of linguistic theories of the Japanese grammar. We discarded ungrammatical sentences that can be easily judged as ungrammatical by any native speaker of Japanese, retained only sentences that required difficult meta-linguistic judgments, and created 86 test items consisting of 43 grammatical and 43 ungrammatical Japanese sentences. The accuracies of the participants' responses were statistically analyzed in an identical manner to the English grammar test.

2.3. Japanese ambiguity

An ambiguity judgment task is another major task of meta-linguistic nature (Keil, 1980), alongside a grammaticality judgment task. We used a modified version of an ambiguity test developed by the second author (A.N.) for his Master's thesis. This pencil-and-paper test consists of 33 Japanese sentences, each of which is ambiguous between two interpretations. The participant was told that each sentence had two meanings and was asked to write down the two meanings if they could detect them. The numbers of correctly detected and explained sentences were analyzed statistically as the dependent variable in a *t*-test (two-tailed).

2.4. Verbal working memory

As an index of verbal working memory, we measured the participants' reading spans in Japanese. For the knowledge of language to be used for comprehension and production, some cognitive resources are necessary, which at least include verbal working memory and (perhaps domain-general) attention. To probe verbal working memory, we used the Japanese version (Osaka & Osaka, 1994) of the so-called reading span test originally developed for English (Daneman & Carpenter, 1980). Using the reading span as the dependent variable, the two English proficiency groups were compared by a *t*-test (two-tailed).

2.5. Non-verbal IQ

We measured the participants' non-verbal intelligence quotients (IQs) by Cattell Culture-Free Test (Cattell, 1961). This multiple-choice test is entirely based on figures, and the questions in the test can be answered without using any verbal knowledge. The IQs of the two English proficiency groups were analyzed statistically by a *t*-test (two-tailed). As will be clear below, the groups differed significantly in this variable. This raises the possibility that the statistically significant group differences in other variables may be due to the difference in non-verbal IQ rather than to the difference in English proficiency. Hence we used non-verbal IQ as a covariate for group com-

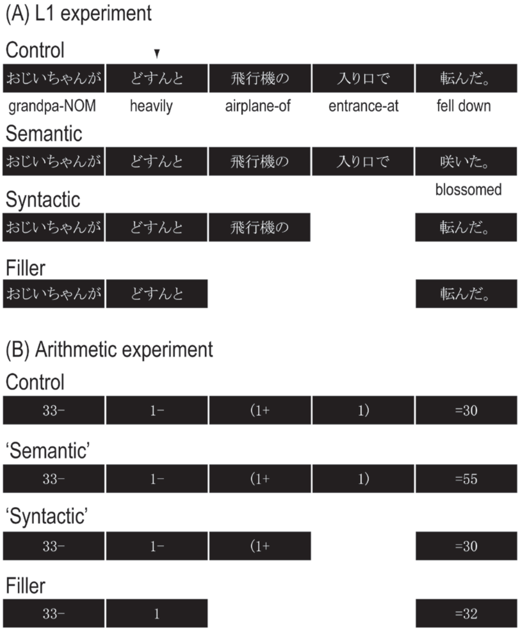

Figure 1. Stimuli in the ERP experiments

Examples of stimuli are shown for each of the four conditions (control, semantic, syntactic, and filler) in each ERP experiment (L1 and arithmetic). In each condition, there were short and long versions of stimuli. Shown here are long versions. The corresponding short versions did not have the second stimuli, indicated by a triangle in the figure. English translations for the examples in the L1 experiment would be: (control) 'My grandpa fell down heavily at the entrance of the airplane', (semantic) 'My grandpa blossomed heavily at the entrance of the airplane', (syntactic) 'My grandpa fell down heavily of the airplane', and (filler) 'My grandpa fell down heavily'.

parisons in other variables, where necessary.

3. ERP experiments

We recorded ERPs to study participants' brain activity that precedes their conscious judgments. Two ERP experiments were run successively in one recording session. The first experiment tested neural activity related to the processing of the mother tongue (L1), while the second studied the processing of arithmetic. The two experiments differed in the stimuli but were identical in other aspects unless otherwise stated.

3.1. Stimuli

In the L1 experiment, 160 sentences in Japanese were used as stimuli (Figure 1). All the sentences were fairly simple; they consisted of easy, frequent words, were in the active voice, and in the canonical word order, without embedding or subordination. Each participant read 40 tokens for each of the four types of the sentences. The control sentences were normal, grammatical sentences. The semantically anomalous sentences contained a semantic violation at the sentence-final verb; that is, the verb was not congruous with the preceding semantic context. The third type of sentences contained a syntactic violation at the sentence-final verb. The violation was created by deleting a necessary constituent from the position immediately preceding the verb. Essentially this is a kind of phrase-structure violation often used to study syntactic processing in ERP studies (Friederici, Pfeifer, & Hahne, 1993; Neville, Nicol, Barss, Forster, & Garrett, 1991). The fourth type was grammatical filler sentences, which were not subjected to subsequent analyses.

In the arithmetic experiment, 160 arithmetic problems were used as stimuli (Figure 1). The participant viewed 40 tokens for each of the four types of problems. The control stimuli were correct arithmetic problems consisting of addition and subtraction. The second type of arithmetic problems contained a calculation error; that is, the answers at the end of the equations were mistaken. This is analogous to sematic violations in the L1 experiment, in that calculation errors are violations of arithmetic facts stored in the brain as part of the conceptual memory, as semantic violations are violations of pragmatic facts that constitute part of the conceptual memory. Hereafter we describe both these types of violations as *semantic* to indicate the underlying commonalities at abstract levels. The third type of arithmetic problems violated a rule of the formation of arithmetic problems. In arithmetic equations, brackets are used in pairs as in “(1+1)”; if a first bracket is introduced, a second bracket must appear to close the first bracket. We created arithmetic equations violating this rule, by deleting a second bracket (as well as one operand preceding it). This is analogous to syntactic violations in the L1 experiment, where a rule of phrase structure is violated by deleting a necessary constituent. Hereafter we describe both types of violations as *syntactic*. In addition to the three types of arithmetic equations above, we used as fillers correct arithmetic equations that were shorter than the control stimuli. The filler items were not the target of analyses.

3.2. Procedure

The stimuli were visually presented on a 60-Hz CRT monitor (Diamondtron M2 RDF223G, Mitsubishi). The participant performed 12 practice trials and received feedback about their behavioral responses in each experiment. To start a trial, the participant pressed either one of the two active buttons on a response pad, while the start cue was being displayed. Then a white-line rectangle appeared against a black background. This rectangle remained on the display until the prompt for a button press appeared. The participant was asked to hold blinks and not to move while the rectangle was on. Stimuli were presented in the center of the rectangle on a chunk basis. In the L1 experiment, one chunk consisted of one content word and one grammatical particle or morpheme. In ERP studies of a European language such as English, it is common to present stimuli word by word, treating content words and function words equally. In Japanese, it looks rather odd if content words and grammatical items are presented separately. To equate the mode of stimulus presentation across the L1 and arithmetic experiment, the arithmetic problems were also presented in chunks of an operand(s) and an operator(s). In both experiments, each chunk was presented in white for 500 ms against a black background, with an inter-stimulus interval of 300 ms. The participant was instructed to judge whether the sentence or arithmetic problem was correct or incorrect, by pressing one of the buttons. Half of the participants were asked to press the left button for correct stimuli and the right button for incorrect stimuli. For the other half of the participants, we swapped the sides of the buttons.

3.3. Recording

Electroencephalograms (EEGs) were digitized at a sampling rate of 500 Hz by a Neurofax amplifier (EEG 9100, Nihon Kohden, Tokyo). The participant wore an elastic cap (EasyCap, Germany), which had 23 Ag/AgCl sintered electrodes at the following scalp locations: Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC6, T7, C3, Cz, C4, T8, CP5, CP6, P7, P3, Pz, P4, P8, O1, and O2. The impedances at the scalp electrodes were below 5 kilo ohm. An additional electrode was placed at the AFz as the ground. The left earlobe was used as the online reference. Horizontal electrooculograms (EOGs) were recorded through two electrodes placed besides the outer canthi. Vertical EOGs were recorded through an electrode placed below the left eye and one placed above the left eyebrow. A .08–120 Hz bandpass filter was applied online. Offline, periods of amplifier blocking were first detected, to prevent them from distorting the subsequent data processing and analyses. The EEGs were then re-referenced to the average of the left and right earlobe, and were lowpass-filtered at 30 Hz. EEGs of 1,400-ms length including pre-stimulus 200 ms were segmented. EEG segments were averaged for each condition for each scalp location. The trials on which the participant responded incorrectly were excluded from averaging. The trials such that the participant made a blink during the period of stimulus display (i.e., post-stimulus 500-ms period) were also rejected, as were trials with movement artifacts and amplifier blocking. To reduce blink artifacts that occurred outside the period of stimulus display, contributions from the vertical EOG channel were regressed out of the EEG segments (Semlitsch, Anderer, Schuster, & Presslich, 1986). All the participants analyzed here had a minimum of 20 accepted trials per condition.

3.4. Analysis

We measured the latencies of ERP components to compare the high and low English proficiency group in the speed of L1 and arithmetic processing. ERP latencies were measured in two ways. Measurement of peak latency, or more precisely, *local* peak latency (Luck, 2005), was applied to the early negative ERP components which peaked in a time range of 200–500 ms in both semantic and syntactic condition and in both L1 and arithmetic experiment. Late positive ERP components appeared after the early negative ERPs but lacked uniquely identifiable peaks in most cases. Hence for the late positive ERPs, we measured *fractional area latencies* using 50% fractions (Luck, 2005). A 50% area latency was measured by first calculating the positive area in a specified time window and then searching for the time point which divided the area into two equal fractions. A major advantage of fractional area latency is its resistance to high-frequency noise; it can be measured for a component that does not have one unique peak.

For statistical analyses of peak latency and 50% area latency, we obtained averages among electrodes C3, Cz, C4, CP5, CP6, P7, P3, Pz, P4, and P8 for each participant. If latencies could not be measured precisely at some electrodes, those electrodes were excluded from averaging. This happened in a few cases, either because early negative ERP components lacked unique peaks, or because there were no positive areas in a late time window to allow

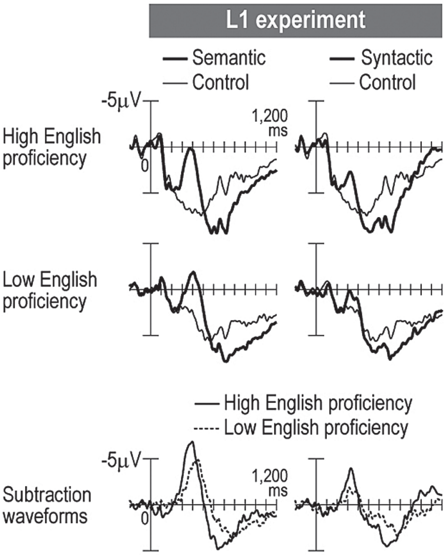

Figure 2. ERP waveforms in the L1 experiment

ERP waveforms obtained in the mother tongue (i.e., Japanese) experiment at a representative electrode (Pz) are shown. Negative is plotted upwards. The upper two rows compare semantically anomalous (left column) or syntactically anomalous sentences (right column) against control sentences, separately for the high (first row) and low English proficiency group (second row). The bottom row compares the high and low group directly in the subtraction waveforms obtained by subtracting waveforms for control sentences from those for anomalous sentences.

measurement of area latency. Degrees of freedom varied among statistical analyses, because there were participants (at most 2) who could not contribute latencies for the above reason.

It should be noted here that the present paper analyzes only latency measures. ERP data can be analyzed with amplitude measures as well. The results presented here should be taken as preliminary.

III. Results and discussion

1. Behavior

In the English grammar test, the high English proficiency group responded far more accurately overall than did the low proficiency group ($F(1, 40) = 103.3$, $p < .001$). A significant Group x Sentence Type interaction was also found ($F(1, 40) = 9.910$, $p = .003$), because the low English proficiency group was worse at rejecting ungrammatical English sentences than at accepting grammatical sentences, whereas the high proficiency group judged both types of sentences equally well. In fact, the rate of rejection of ungrammatical English sentences by the low English proficiency group was not statistically different from chance level ($t(19) = 1.539$, $p = .14$).

In the Japanese grammar test, the high English proficiency group responded significantly more accurately overall than did the low proficiency group ($F(1, 40) = 21.26$, $p < .001$). The interaction between Group and Sentence Type only approached significance ($F(1, 40) = 3.695$, $p = .062$). The rejection of ungrammatical Japanese sentences was significantly more difficult than was the acceptance of grammatical Japanese sentences ($F(1, 40) = 30.86$, $p < .001$).

In the Japanese ambiguity test, the high English proficiency group detected more ambiguities than did the low proficiency group ($t(40) = 5.480$, $p < .001$).

Verbal working memory in Japanese did not differ significantly between the two groups ($t(40) = 1.867$, $p = .069$). The reading spans of both groups were around the previously reported average of Japanese university students tested with the same materials (Osaka & Osaka, 1994).

Finally, higher non-verbal IQ was shown by the high English proficiency group ($t(40) = 5.236$, $p < .001$). However, the low proficiency group was not a group of low-IQ people; their IQs were significantly higher than the average of healthy adults (108.35 vs. 100, $t(19) = 3.333$, $p = .003$).

We reanalyzed the statistically significant group differences by means of analyses of covariance (ANCOVA) using non-verbal IQ and university level (expressed as deviation score) as covariates. After regressing out the effects of these two variables, the accuracies of the English grammar differed significantly between the two groups ($F(1, 38) = 37.01$, $p < .001$). So did the accuracies of the Japanese grammar ($F(1, 38) = 9.658$, $p = .004$). In both tests, the high English proficiency group was more accurate. The effects of non-verbal IQ and university level were not statistically significant as covariates in either analysis (ps > .4). We obtained a different result for the Japanese ambiguity test. When non-verbal IQ and university level were used as covariates, the group difference did not reach significance ($F(1, 38) = 2.592$, $p = .116$), whereas the covariates themselves approached significance (non-verbal IQ, $F(1, 38) = 3.909$, $p = .055$; university level, $F(1, 38) = 3.446$, $p = .071$). We reanalyzed the non-verbal IQs using university level as the sole covariate. Even if university level was taken into account, the high English proficiency group had higher non-verbal IQs ($F(1, 39) = 7.812$, $p = .008$).

To summarize the results of behavioral tests, we found statistically significant group differences in favor of the high English proficiency group in most cases. The high English proficiency group showed higher meta-linguistic awareness about the L1 in a grammaticality judgment task, in accordance with numerous previous studies which have reported positive relations between FL learning and L1 skills involving awareness. Effects of FL proficiency were less clear in another meta-linguistic task conducted in the L1, an ambiguity judgment task, in which the two groups significantly differed when no covariates were used, but not when non-verbal IQ and university level were analyzed as covariates. Perhaps the detection of ambiguities at sentence levels relies on domain-general cognition more strongly than does the detection of ungrammaticalities, in that the former requires the participant to deliberately think about the stimulus sentences that are quite normal in themselves, whereas the latter depends on a feeling of abnormality that arises, to some degree, automatically when the participant processes the stimulus sentences. The two groups of participants did not differ in verbal working memory in the L1, which seems to suggest either that FL learning does not have beneficial effects on the development of verbal working

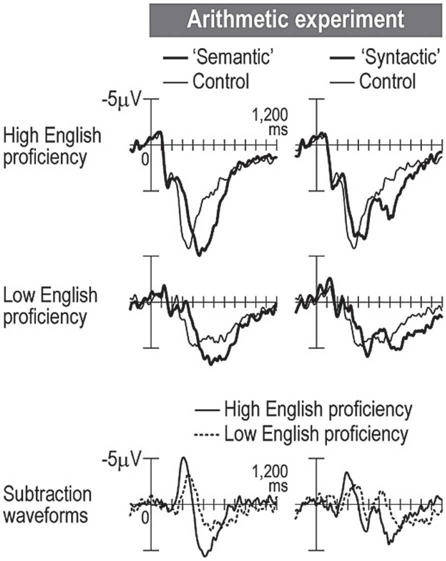

Figure 3. ERP waveforms in the arithmetic experiment
ERP waveforms obtained in the arithmetic experiment are shown in an identical manner to those in the L1 experiment shown in Figure 2.

memory or that large verbal working memory is not always an advantage in acquiring high FL proficiency. Non-verbal IQ was higher in the high English proficiency group, suggesting that higher FL proficiency may be associated with higher capacity for the processing of non-verbal information.

2. ERPs

2.1. Mother tongue

ERP waveforms obtained for the control sentences and two types of anomalous sentences (i.e., semantic and syntactic) are shown in Figure 2. Subtraction waveforms were created by subtracting the waveforms for the control sentences from those for the anomalous sentences. The resulting [anomalous - control] subtraction waveforms showed a negative-then-positive biphasic pattern of ERP responses in both semantic and syntactic condition. In the semantic condition, the earlier negative ERP can be identified as the N400 from its polarity, latency, and scalp distribution (Kutas & Hillyard, 1980).

The later positive ERP following the N400 is commonly called the Late Positive Component (LPC) (Juottonen, Revonsuo, & Lang, 1996). The high English proficiency group showed significantly earlier N400 peak latencies than did the low proficiency group ($F(1, 40) = 32.76$, $p < .001$). This group difference remained significant after regressing out the effects of the two covariates, non-verbal IQ and university level ($F(1, 38) = 14.91$, $p < .001$; *p*s for the covariates > .2). The 50% area latency of the subtraction LPC was also earlier in the high English proficiency group ($F(1, 40) = 12.93$, $p = .001$). This group difference was significant when the covariates were taken into account ($F(1, 38) = 5.633$, $p = .023$; *p*s for the covariates > .7).

In the syntactic condition, the groups significantly differed in the peak latency of the early negative ERP component obtained in the [anomalous - normal] subtraction waveforms ($F(1, 38) = 4.545$, $p = .040$), but this group difference was not significant when non-verbal IQ and university level were included as covariates ($F(1, 36) < 1$; *p*s for the covariates > .2). The groups differed in the 50% area latency of the following positive ERP component ($F(1, 39) = 9.157$, $p = .004$). This difference remained significant when the two covariates were analyzed together ($F(1, 37) = 7.037$, $p = .012$; *p*s for the covariates > .3).

2.2. Arithmetic

The features of the [anomalous - control] subtraction waveforms obtained in the arithmetic experiment were roughly similar to those in the L1 experiment (Figure 3). The subtraction waveforms in the semantic condition showed a clear N400-like early negative ERP component, which is sometimes called the arithmetic N400 (Niedeggen, Rösler, & Jost, 1999). The peak latency of the arithmetic N400 in the subtraction waveforms was significantly earlier in the high English proficiency group than in the low proficiency group ($F(1, 40) = 17.77$, $p < .001$). The group difference remained significant after removing the effects of non-verbal IQ and university level ($F(1, 38) = 6.828$, $p = .013$; *p*s for the covariates > .6). The subtraction waveforms in the semantic condition contained an LPC-like positive ERP component. Its 50% area latency was earlier in the high English proficiency group ($F(1, 40) = 15.96$, $p < .001$). This was true when the covariates were included in the statistical model ($F(1, 38) = 9.870$, $p = .003$; *p*s for the covariates > .25).

In the syntactic condition of the arithmetic experiment, we obtained results that were different from those in the semantic condition. The peak latency of the early negative ERP component in the [anomalous - normal] subtraction waveforms was earlier in the high English proficiency group ($F(1, 39) = 7.987$, $p = .007$), but this difference between the groups disappeared when non-verbal IQ and university level were included in the analysis as covariates ($F(1, 37) < 1$; *p*s for the covariates > .07). We obtained a similar pattern of results for the 50% area latency of the late positivity. The high English proficiency group showed a significantly earlier latency initially ($F(1, 38) = 10.261$, $p = .003$), but the group difference was not significant when the two covariates were used ($F(1, 36) = 1.420$, $p = .241$; *p*s for the covariates > .13).

To summarize, the individual sub-stages of L1 and arithmetic processing as indexed by early negative ERPs and late positive ERPs were earlier in the high English proficiency group than in the low proficiency group. Hence higher FL proficiency may be correlated with faster symbol processing commonly involved in L1 processing and arithmetic processing. While we have clear evidence for this correlation, the direction of causality is not clear. One possibility is that FL learning improves the common neural bases of language and arithmetic. A second possibility is that genetically given individual variation in the functioning of those neural bases leads to individual variation in the final abilities of language (both L1 and FL) and arithmetic. Both these possibilities may be correct. The superiority of the high proficiency group was more consistently found in sematic processing than in syntactic processing in both L1 and arithmetic experiment. That is, the group differences remained statistically significant in semantic processing even if covariates were used, while this was not always the case in syntactic processing. This could be because there are more overlaps among the L1, the FL, and arithmetic in the neural bases of semantics than in those of syntax.

References

Armstrong, P. W., & Rogers, J. D. (1997). Basic skills revisited: The effects of foreign language instruction on reading, math, and language arts. *Learning Languages, 2*(3), 20–31.

Bentin, S., Kutas, M., & Hillyard, S. A. (1995). Semantic Processing and Memory for Attended and Unattended Words in Dichotic-Listening: Behavioral and Electrophysiological Evidence. *Journal of Experimental Psychology: Human Perception and Performance, 21*(1), 54–67.

Cattell, R. B. (1961). *Culture-free intelligence test, scale3*. Champaign, IL: Institute for Personality and Ability Testing.

Cooper, T. C. (1987). Foreign Language Study and SAT-Verbal Scores. *The Modern Language Journal, 71*(4), 381–387.

Cummins, J. (1984). Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students. In C. Rivera (Ed.), *Language Proficiency and Academic Achievement* (pp. 2–19). Clevedon: Multilingual Matters.

Cummins, J. (1991). Language Development and Academic Learning. In L. Malavé & G. Duquette (Eds.), *Language, culture and cognition: a collection of studies in first and second language acquisition* (pp. 161–175). Clevedon: Multilingual Matters.

Cunningham, T. H., & Graham, C. R. (2000). Increasing Native English Vocabulary Recognition Through Spanish Immersion: Cognate Transfer from Foreign to First Language. *Journal of Educational Psychology, 92*(1), 37–49.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450–466.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-Related Brain Potentials during Natural Speech Processing - Effects of Semantic, Morphological and Syntactic Violations. *Cognitive Brain Research, 1*(3), 183–192.

Garfinkel, A., & Tabor, K. E. (1991). Elementary School Foreign Languages and English Reading Achievement: A New View of the Relationship. *Foreign Language Annals, 24*, 375–382.

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

Juottonen, K., Revonsuo, A., & Lang, H. K. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cognitive Brain Research, 4*(2), 99–107.

Kecskes, I., & Papp, T. (2000). *Foreign Language and Mother Tongue*. Mahwah, NJ: Lawrence Erlbaum Associates.

Keil, F. (1980). Development of the ability to perceive ambiguities: Evidence for the task specificity of a linguistic skill. *Journal of Psycholinguistic Research, 9*(3), 219–230.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207*(4427), 203–205.

Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition, 13*, 361–392.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Masciantonio, R. (1977). Tangible benefits of the study of Latin: A review of research. *Foreign Language Annals, 10*, 375–382.

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically Based Sentence Processing Classes - Evidence from Event-Related Brain

Potentials. *Journal of Cognitive Neuroscience, 3*(2), 151–165.

Niedeggen, M., Rösler, F., & Jost, K. (1999). Processing of incongruous mental calculation problems: evidence for an arithmetic N400 effect. *Psychophysiology, 36*(3), 307–324.

Oldfield, R. C. (1971). Assessment and Analysis of Handedness: the Edinburgh Inventory. *Neuropsychologia, 9*(1), 97–113.

Osaka, M., & Osaka, N. (1994). Working memory capacity related to reading: measurement with the Japanese version of reading span test [Article in Japanese]. *Shinrigaku Kenkyu, 65*(5), 339–345.

Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology, 23*(6), 695–703.

Taylor, C., & Lafayette, R. (2010). Academic Achievement Through FLES: A Case for Promoting Greater Access to Foreign Language Study Among Young Learners. *The Modern Language Journal, 94*(1), 22–42.

Thomas, W. P., Collier, V. P., & Abbott, M. (1993). Academic Achievement through Japanese, Spanish, or French: The First Two Years of Partial Immersion. *The Modern Language Journal, 77*(2), 170–179.

Yelland, G. W., Pollard, J., & Mercuri, A. (1993). The metalinguistic benefits of limited contact with a second language. *Applied Psycholinguistics, 14*, 423–444.