Neural correlates of second-language word learning: minimal instruction produces rapid change

Judith McLaughlin, Lee Osterhout & Albert Kim

Adult second-language (L2) learning is often claimed to be slow and laborious compared to native language (L1) acquisition, but little is known about the rate of L2 word learning. Here we report that adult second-language learners' brain activity, as measured by event-related potentials (ERPs), discriminated between L2 words and L2 'pseudowords' (word-like letter strings) after just 14 h of classroom instruction. This occurred even while the learners performed at chance levels when making overt L2 word-nonword judgments, indicating that the early acquisition of some aspects of a new language may be overlooked by current behavioral assessments.

ERPs, measured from the scalp, provide a nearly continuous sampling of the brain's electrical activity[1]. We focused on learning-related changes to the N400 component, a negative wave that peaks at 400 ms after the visual presentation of a word[2]. The N400 is sensitive to lexical status (that is, whether or not a letter string is a word)[3] and word meaning[2,4]. For native speakers of a given language, the N400 amplitude is largest for pronounceable, orthographically legal nonwords (hereafter, 'pseudowords'); it is intermediate for words preceded by a semantically unrelated context; and it is smallest for words preceded by a semantically related context[5]. Our goal was to determine how much L2 exposure is needed before a learner's brain activity reflects the lexical status and meaning of L2 words.

Our participants included a group of university students who were enrolled in an introductory French course but had not had formal instruction in or significant exposure to French before the study ('learners'), and a control group who had never received any French instruction or significant exposure to French ('nonlearners'). All participants reported at least 1 year of instruction in another foreign language. We longitudinally obtained ERPs and behavioral responses from both groups in three separate sessions (for the learners at session 1: mean 14 h of instruction, range 5–28 h; session 2: mean 63 h, range 59–67 h; session 3: mean 138 h, range 126–150 h). Five learners left the French course and the experiment before the end of the 9-month instructional period. All participants were included in single-session analyses, but only those who participated fully were included in multi-session comparisons.

Stimuli were two lists of 112 'prime-target' pairs of letter strings. Each list contained 40 semantically related (*chien-chat*) word pairs, 40 unrelated (*maison-soif*) word pairs and 32 word-pseudoword (*mot-nasier*) pairs. Prime and target words (nouns, adjectives and verbs) were selected from the assigned textbook and roughly matched in first appearance and distribution in the text. Target words were identical across lists but counterbalanced across prime type. Pseudowords were derived by taking words from the text and replacing one or two of the central letters. Each participant saw one list per session. Across sessions, lists were alternated for each participant, with the list in session 1 repeated for session 3.

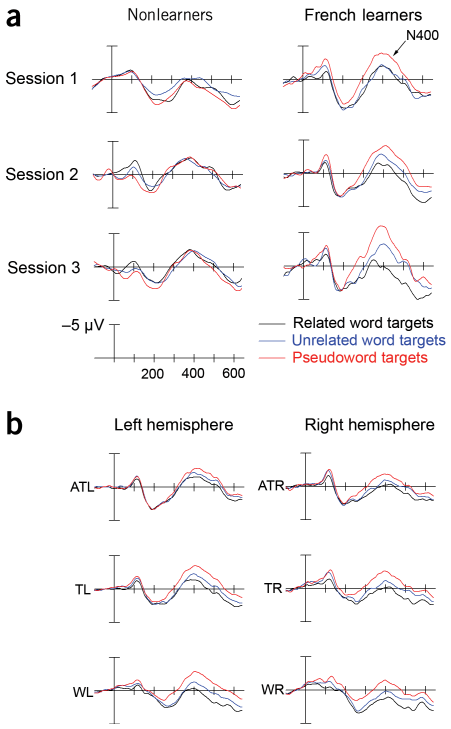

Figure 1 Event-related potentials to target stimuli. (**a**) ERPs to word and pseudoword targets during the three testing sessions, for the nonlearners ($n = 8$; mean age: 27.6 years) and French learners ($n = 18$, 16 and 13 for sessions 1, 2 and 3, respectively; mean age, 21.3 years). Informed consent was obtained from all participants. Data acquired over the central midline site (Cz) are shown. The vertical calibration bar indicates target onset. Each tick mark represents 100 ms. (**b**) Learners' ERPs to targets, averaged over sessions. ATL/R, anterior temporal left/right; TL/R, temporal left/right; WL/R, Wernicke's area left/right. Trial sequence: fixation cross (500 ms); blank screen (500 ms); prime (400 ms); blank screen (400 ms); target (400 ms); blank screen (400 ms); response prompt. Electroencephalographic activity was sampled at 200 Hz from 13 scalp sites (three midline, five lateral pairs; 0.01–100 Hz bandpass; 3 dB cut-off; left mastoid reference). Trials contaminated by artifacts (17%) were excluded. N400 amplitude was quantified as mean voltage within a 300–500 ms window, relative to a baseline of mean voltage from 100 ms before to 50 ms after stimulus onset. We used a repeated-measures ANOVA with the Greenhouse-Geisser correction.

Participants made a lexical decision (word/nonword) for each prime-target pair, concurrent with ERP recording. Sensitivity in the lexical decision task was assessed using a d' measure ($d' = 0$ indicates no sensitivity; $d' = 4$ indicates near-perfect sensitivity[6]; **Table 1**). For the nonlearners in all three testing sessions, as well as for learners during session 1, d' was near zero. Learners showed moderate increases in sensitivity during sessions 2 and 3.

We did not observe any N400 amplitude modulations for the no-instruction participants ($F < 1$; **Fig. 1a**). For the learners, beginning in

Department of Psychology, University of Washington, Guthrie Hall, Box 351525, Seattle, Washington 98195-1525, USA. Correspondence should be addressed to J.M. (giuditta@u.washington.edu).

Published online 13 June 2004; doi:10.1038/nn1264

Table 1 Proportion of words and pseudowords identified as a word in the lexical decision task, and the *d′* measure of sensitivity

	Words			
	Related	Unrelated	Pseudowords	d′
Nonlearners				
Session 1	0.57	0.57	0.60	–0.1
Session 2	0.56	0.54	0.56	0.0
Session 3	0.56	0.55	0.54	0.0
French learners				
Session 1	0.61	0.62	0.63	0.0
Session 2	0.70	0.65	0.48	0.5
Session 3	0.74	0.71	0.44	0.8

We calculated *d′* using a formula provided by Miller[8]: $d' = z(\text{h}) - z(\text{fa})$, where d' represents the likelihood of a real target being recognized, *z* (h) represents the *z*-score that corresponds to the proportion of real words that were identified as words (hits), and *z* (fa) represents the *z*-score that corresponds to the proportion of nonwords that were identified as words (false alarms).

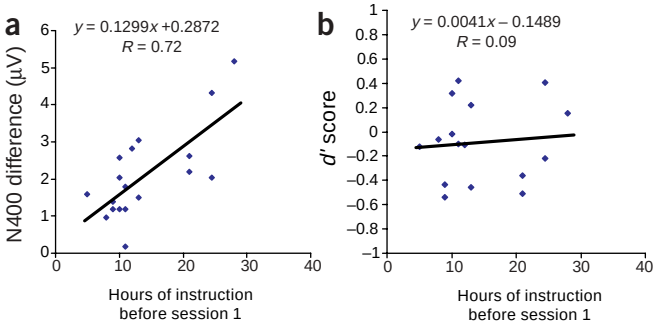

Figure 2 Session 1 N400 amplitude word/pseudoword differences and *d′* scores regressed onto hours of instruction before session 1. (**a**) N400 amplitude difference between words and pseudowords. (**b**) *d′* scores. Two subjects were excluded from the *d′* analysis due to technical problems.

session 1 and continuing across sessions, pseudowords elicited larger N400s than did related or unrelated words (pseudowords vs. related words: session 1: midline electrodes , $F_{1,17} = 15.69$, $P = 0.001$; lateral electrodes, $F_{1,17} = 14.38$, $P < 0.01$; session 2: midline, $F_{1,15} = 39.56$, $P < 0.001$; lateral, $F_{1,15} = 19.19$, $P < 0.001$; session 3: midline, $F_{1,12} = 49.64$, $P < 0.001$; lateral, $F_{1,12} = 45.91$, $P < 0.0001$; pseudowords vs. unrelated words: session 1: midline, $F_{1,17} = 19.52$, $P < 0.001$; lateral, $F_{1,17} = 4.06$, $P = 0.05$; session 2: midline, $F_{1,15} = 7.71$, $P < 0.02$; lateral, $F_{1,15} = 4.39$, $P = 0.05$; session 3: midline, $F_{1,12} = 6.60$, $P < 0.03$; lateral, $F_{1,12} = 6.59$, $P < 0.03$). This word-pseudoword difference increased across the three sessions ($F_{4,48} = 3.82$, $P < 0.05$). Effects of word meaning, manifested as smaller-amplitude N400s to words preceded by related versus unrelated words, were observed in sessions 2 and 3 (session 2: midline, $F_{1,15} = 5.35$, $P < 0.05$; session 3: midline, $F_{1,12} = 8.94$, $P < 0.02$). This effect also increased in amplitude across sessions ($F_{2,24} = 4.15$, $P < 0.03$). By session 3, learners' ERP responses were qualitatively similar to analogous native language responses. N400 effects were evenly distributed over midline sites (target type: $F_{2,92} = 32.74$, $P < 0.0001$; target type × electrode: $F < 1$) and posteriorly distributed over lateral sites (target type: $F_{2,92} = 29.82$, $P < 0.0001$; target type × electrode: $F_{8,368} = 3.22$, $P = 0.001$; **Fig. 1b**).

The most striking finding was the presence of a word-pseudoword N400 effect after approximately 14 h of L2 instruction. To determine whether this effect was a function of L2 exposure, we regressed the N400 differences and *d′* scores of session 1 onto the number of instructional hours each learner experienced before session 1. N400 differences were correlated with hours of instruction ($r = 0.72$, $P < 0.001$; **Fig. 2a**), but *d′* scores were not ($r = 0.09$, $P > 0.3$; **Fig. 2b**). N400 differences in session 1 were also correlated with the proportion ($r = 0.64$, $P < 0.001$) and frequency ($r = 0.67$, $P < 0.001$) of target words in the text material assigned to each learner before session 1.

People rapidly extract co-occurrence statistics for letter and sound combinations within a language[7]. Perhaps the word-pseudoword N400 difference was due to smaller mean grapheme co-occurrence values for the pseudowords than for words. We tested this hypothesis by computing bigram, trigram and quadragram frequencies using a corpus of all French words appearing in the first four chapters of the learners' French text. The target words and pseudowords did not differ in bigram or trigram frequency ($P > 0.3$), but they did differ in quadragram frequency ($t_{128} = 1.76$, $P < 0.04$). Quadragram frequency, however, was itself highly correlated with target word frequency ($r = 0.99$). Thus, our results probably reflect whole-word rather than grapheme co-occurrence frequency.

Our results show that adult language learners rapidly accrue information about L2 words (initially about word form, and then about word meaning). The earliest evidence of learning, even before changes in overt lexicality judgments, is seen in ERPs to L2 words and pseudowords. This early learning might involve elemental aspects of linguistic knowledge (*e.g.*, knowledge about word forms) that serve as prerequisites to linguistic competence. Our results indicate that adult L2 learning is not uniformly slow and laborious; some aspects of the language are acquired with remarkable speed. Our results also suggest that some behavioral assessments of L2 learning might underestimate what has been learned. ERPs might more accurately reflect implicit learning and continuous change in knowledge than do explicit, categorical judgments. The method used here could be extended to examine the effects of L1-L2 similarity, instructional methods and learners' age on L2 acquisition. Scalp recordings of the brain's electrical activity might thus represent a uniquely sensitive means for investigating the developing linguistic competence of adult second-language learners.

ACKNOWLEDGMENTS

This work was supported by National Institute on Deafness and Other Communication Disorders grants R01DC01947, P30DC04661 and F32DC005756.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 16 January; accepted 29 April 2004
Published online at http://www.nature.com/natureneuroscience/

1. Osterhout, L., McLaughlin, J. & Bersick, M. *Trends Cogn. Sci.* **1**, 203–209 (1997).
2. Kutas, M. & Van Petten, C. in *Handbook of Psycholinguistics* (ed. Gernsbacher, M.) 83–143 (Academic, New York, 1994).
3. Bentin, S. *Brain Lang.* **31**, 308–327 (1987).
4. Kutas, M. & Hillyard, S.A. *Science* **207**, 203–205 (1980).
5. Chwilla, D.J., Brown, C.M. & Hagoort, P. *Psychophysiology* **32**, 274–285 (1995).
6. Green, R.M. & Swets, J. *Signal Detection Theory and Psychophysics* (Krieger, Huntington, New York, 1974).
7. Saffran, J.R., Aslin, R.N. & Newport, E.L. *Science* **274**, 1926–1928 (1996).
8. Miller, J. *Percept. Psychophys.* **58**, 65–72 (1996).