KNOWLEDGE OF NATIVELIKE SELECTIONS IN A L2

The Influence of Exposure, Memory, Age of Onset, and Motivation in Foreign Language and Immersion Settings

Pauline Foster, Cylcia Bolibaugh, and Agnieszka Kotula
St. Mary’s University College

It is well established that part of native speaker competence resides in knowledge of conventionalized word combinations, or nativelike selections (NLSs). This article reports an investigation into the receptive NLS knowledge of second language (L2) users of English in both the United Kingdom and Poland and the influence of a variety of independent variables on this knowledge. Results indicate that only an early start (< 12 years old) in an immersion setting guarantees nativelikeness. Long exposure in late starters brings moderate gains in both settings but not to nativelike levels; positive feelings toward the L2 and motivation to interact in it bear little to no relationship with NLS; phonological short-term memory (pSTM) is the only predictor of NLS ability in immersion late starters, with no effect found in a foreign language setting. Our results suggest that NLS is subject to age effects and that, for late starters, a good pSTM and L2 immersion are necessary for the acquisition of this dimension of L2 knowledge.

This study was funded by a grant from the UK Economic and Social Research Council, reference RES-000-22-2645.
Correspondence concerning this article should be addressed to Pauline Foster, St. Mary’s University College, School of Arts and Humanities, Waldegrave Road, Twickenham, Middx TW1 4SX, U.K. E-mail: pauline.foster@smuc.ac.uk
In their seminal article, Pawley and Syder (1983) challenged the prevailing generative account of language by observing that, although novel and grammatical sequences of words are part of native speaker capacity, they are not a regular part of native speaker behavior. They argued that, poets aside, people are conservative rather than original in forming utterances, spurning novelty in favor of well-worn combinations of words used by everyone else in their speech community. In the following 30 years, extensive work in corpus analysis and phraseology has confirmed beyond room for doubt that native language performance is characterized by communally preferred ways of saying things and is built not so much word by word as it is through an interconnected network of commonly associated words (see, e.g., Altenberg, 1990; Ellis, Simpson-Vlach, & Maynard, 2008; Erman & Warren, 2000; Forsberg, 2008, 2010; Howarth, 1998; McCarthy, 1990; Nattinger & DeCarrico, 1992; Sinclair, 1991; Van Lancker-Sidtis & Rallon, 2004; Warren, 2005; Weinert, 1995; Wray, 2008; Wray & Perkins, 2000). Pawley and Syder used the term *nativelike selection* (NLS) to describe a speech community’s favored collocations, compound nouns, phrasal verbs, colligations, sentence stems, prepositional phrases, metaphors, sayings, quotations, exclamations, and the like. Given that a native speaker of English is estimated to know more than 20,000 word families that consist of base words with their various inflectional forms and derivational spin-offs (Nation & Waring, 1997), the number of NLSs of these must run into hundreds of thousands.

To take a famous example of the persistent generative position, Pinker (1994) states, “Virtually every sentence that a person utters or understands is a brand-new combination of words appearing for the first time in the history of the universe” (p. 22). Whether or not this is true for whole sentences is not the point at issue. A sentence, which may be novel when viewed as a whole, is overwhelmingly likely to be constructed from smaller parts that are not novel at all. This is easily shown. At the time of writing this article, the four-word combination *for the first time* returned 10.72 billion hits in a Google search. The eight-word combination *for the first time in the history of* returned 196 million hits. *Virtually every* got 10.1 million hits, and there were 3.9 million hits for *brand new combination of*. By contrast, a paraphrased version of Pinker’s sentence, “Almost any word ordering which a person may select for the forming of a sentence is nearly certainly something for which human history provides no perfect match,” was demonstrably less novel. Searches for the three phrases *almost any word ordering*, *for the forming of a sentence*, and *history provides no perfect match* returned no hits at all. But the collocations *human history* (6.03 million), *almost any* (33.3 million), and *perfect match* (21.0 million) were abundant. In the pursuit of total novelty, it takes some effort to come up with a sentence from which commonly collocated words are entirely absent.
To reject the view of human language as infinitely innovative is not, of course, to advocate the idea that it consists merely of memorized phrases. Dual-system accounts of language (e.g., Sinclair, 1991; Van Lancker Sidtis, 2004) accommodate two distinct processes, one operating on a set of grammatical rules and a lexicon of individual words to generate novel combinations and the other operating on partially or fully fixed sequences of words stored in memory. Both processes are drawn on in the encoding and decoding of sentences, although as Ellis (1996) has put it, language processing is most efficient when it is “as deep as necessary but as shallow as possible” (p. 116). The original quotation from Pinker and its clunky paraphrase can both be processed by a speaker of English, but the original offers an economy of effort precisely because so much of it is not novel at all; it is easier to locate combinations in long-term memory than to compute them word by word (Ellis et al., 2008). From a dual-system perspective, Pinker has used both his grammatical knowledge and his memory to bolt together a brand-new sentence from some very old parts.

The notion of shared word sequences arising in a speech community fits with usage-based or emergentist accounts of first language (L1) acquisition (e.g., Bates, 1994; Beckner et al., 2009; Bybee, 2001; Ellis, 1996, 2002, 2003, 2008; Goldberg, 2006; Hopper, 1988; Langacker, 2000; MacWhinney, 2006; Tomasello, 2003). In these accounts, speakers base their linguistic behavior on their experience of interacting with one another. Their sociopragmatic norms, their pronunciation, and their colligational and collocational choices all arise from previous experience of interaction. As young children become members of the speech community, they take on these norms through repeated and lengthy exposure to language samples in their normal daily lives. Their grammatical competence is seen as emerging through the implicit induction of the structural regularities in the samples. Ellis (2003) describes a process whereby abstract hierarchical constructions (phrase structure grammar) emerge from the acquisition of formulas and low scope patterns. He writes:

The learning of abstract constructions . . . begins with chunking and committing formulae to memory. . . . Once a collection of like examples is available in long-term memory, there is scope for implicit processes of analysis of their shared features and for the development of more abstract summary schema, in the same way as prototypes emerge as the central tendency of other cognitive categories. (p. 69)

Children’s typical overgeneralizations of grammatical morphemes and word meanings (e.g., how did you unsqueezed it, I beed a good typewriter, let’s get brooming, these flowers are sneezing me; Stilwell-Peccei, 2006) are evidence for how they have perceived regularities in the
language they encounter before they have had enough experience to mark the restrictions of their use. Knowledge rooted thus in experience results in the child acquiring not just the meaning of a word, its pronunciation, its grammatical role, and its morphology but also a detailed and networked memory of its commonly associated grammatical and semantic structures, or what Cook (1997) has described as the accretion of its individual idiosyncrasies (see Hoey, 2005, for a similar argument). It is this accretion of knowledge about how words are conventionally sequenced that explains synchronic and diachronic dialect variation across communities and why native speakers today could, and yet do not, produce grammatical but nonnativelike selections such as She kicked the ball strongly.²

VOCABULARY SELECTION IN A SECOND LANGUAGE

Although there is good evidence that child L1 learners acquire both system knowledge (grammar) and knowledge of use (NLS) at the same time, there is less research on the extent to which adult second language (L2) learners do the same thing, although Ellis (2003, pp. 72–74) has argued that both L1 and L2 knowledge could be described essentially as arising from the perception and learning of sequences of words. Research interest into L2 vocabulary acquisition (e.g., Bardel, Gudmundson, & Lindqvist, 2012; Heinrich & Schoonen, 2008; Laufer & Nation, 1995; Meara, 1998) has tended to look at the size, richness, or sophistication of the L2 lexicon in terms of individual words and not at how they work in combinations. Studies focusing on learners’ written production of NLSs (most often operationalized as collocations) have found that learners systematically deviate from native norms through over- or underuse (e.g., Granger, 1998; Nesselhauf, 2005) and are particularly susceptible to cross-linguistic influence (e.g., Laufer & Waldman, 2011). Hoey (2005) notes that L2 learners in a classroom are not in an environment rich enough for them to be able to build nativelike networks of L2 word associations, and, so, what he terms L1 primings are likely to be activated, unhelpfully, by L2 words.

In light of this, it is interesting to review research that has looked at the effect of learning environment on the use of largely fixed formulas in a L2. Möhle and Raupach (1983) and Regan (1998) both found that study abroad learners were more fluent and sounded more natural when compared to groups of learners who were studying outside the L2 community. They attributed this, in part at least, to the study abroad learners’ greater use of formulaic sequences. Marriott (1995) and Siegal (1995) also showed greater use of formulaic sequences in learners who were studying abroad. Towell, Hawkins, and Bazergui (1996) were unusual in adopting a within-subjects design to compare the L2 French of English
learners before and after they had spent a year in France. The authors reported that, on their return from France, learners were able to use a greater number of formulaic sequences of words and consequently were more fluent. These studies show that lexical organization of the formulaic kind is an area of benefit for learners living within the target language (TL) community and suggest that context can significantly influence acquisition of L2 lexical knowledge.

This is a potentially fruitful area to explore more broadly by looking beyond fixed or partially fixed formulas. Foster and Tavakoli (2009) compared the performance of three groups of participants who each told two of four narratives from cartoon picture prompts. One group consisted of native speakers of English, a second consisted of intermediate-level adult learners of English based in Tehran, and a third group was comprised of intermediate-level adult learners of English based in London. Measures of accuracy and complexity showed no difference between the performances of the learners in London and Tehran. On a measure of lexical variety, VocD (Malvern & Richards, 2002), an ANOVA revealed no significant differences between the native speakers and the learners in London but did reveal a significant difference between the learners in London and the learners in Tehran (on all four narratives: \( p < .001, \eta^2 = .212–.322 \)). A further qualitative analysis (Foster, 2009) showed that in many instances the learners in London selected word combinations (e.g., *they stopped off for a coffee*; *she went for a swim*) that were the same as those selected by the native speakers to describe the same narrative events, whereas the learners in Teheran selected grammatical but non-nativelike word combinations (e.g., *they drank a coffee*; *she swam*). It was reasoned that these NLSs had been acquired from the rich L2 environment that the learners in London were exposed to and that they were accessed as fully or partially fixed items, which contributed to the superior fluency in the task performance of the London-based learners. As there was no detectable difference in the syntactic accuracy of performance between the London-based learners and the Tehran-based learners, these findings indicate that, in the context of a wider L2 environment, idiomaticity was not developing in step with grammaticality.

A common feature of these studies is that they looked at L2 output. They did not investigate what knowledge a L2 speaker might possess but not draw on productively. Just as everyone has a much larger passive than active vocabulary of individual items, they also know more word combinations than they display in use. This is inevitable given that how people choose to say things reflects how they wish to be identified (e.g., in terms of age, gender, status, education, etc.), whereas what they hear reflects the identities of other people, which will not necessarily overlap with their own identities (Wray & Perkins, 2000). However, whereas the learners who appear to be idiomatic and fluent in a L2 may have an extensive passive knowledge of NLSs, they could
instead be successfully masking a rather limited knowledge of this sort by relying on a relatively small number of NLSs that are frequently called into use, their own “islands of reliability” (Dechert, 1983, p. 184) in L2 production (see also Granger, 1998). To investigate the extent to which a learner has acquired passive knowledge of NLSs in a L2 is therefore an interesting step for research to take, as would be exploring what variables might be involved in this dimension of L2 knowledge.

POSSIBLE INFLUENCES ON THE ACQUISITION OF NATIVELIKE SELECTIONS IN A L2

As previously noted, language users are presumed to have a vast store of knowledge of how words in their L1 most naturally combine, and this is acquired, like the bulk of their L1 knowledge, incidentally through a lifetime of interactions. If such knowledge is built also in a L2, it is likely that this is mainly through exposure rather than through explicit teaching or study. Whereas some idiomatic sequences, especially of the fully fixed variety, can obviously be learned through conscious effort (e.g., Gray’s [1999] textbook), this could only represent a tiny proportion of the quantity needed to endow nativelikeness to L2 attainment. Because of the logical problem of the sheer number of sequences and combinatorial restrictions that determine NLS, it is not plausible that a L2 learner could set out deliberately to acquire them. After all, it was not until the advent of computer analysis of large language corpora that a broad picture of the combinatorial relationships between words was made available to explicit scrutiny (Sinclair, 1991).

Thus, in investigating the development of NLS in a L2 lexicon, three areas of influence can be suggested that merit exploration. First and most simply, the sheer quantity of exposure to the TL (Robinson & Ellis, 2008; Schumann, 1978) is likely to be important and will in part depend on learners’ levels of engagement with the L2 community and their motivation to become part of it. People who enthusiastically embrace a L2 are likely to cultivate personal and professional relationships within the TL community and to conduct these entirely in the L2. They are also likely to have further contact with the language through newspapers, TV, film, Internet, and social media. Their daily interactions with their environment are thus mediated primarily and by choice through the L2, giving them, over many years, a counterpart to the massive exposure that a child uses to develop L1 knowledge. Bardovi-Harlig and Bastos (2011) investigated the development of lexical sequences in L2 learners during extended periods of residence in the L2 community and found (a) that not all L2 users avail themselves of the contact opportunities available to them and (b) that measures of exposure such as length of residence mask a variety of experiences. Dörnyei, Durow, and Zahran (2004) found
that success or failure to acquire a target set of lexical sequences seemed to depend on learners’ determination to “break out of the ‘international ghetto’ they [found] themselves in” (p. 105). Thus, the exposure relationship is crucial but more complex than it appears at first sight. Interaction with members of the TL community seems to be key, but L2 users within the TL community can be more or less willing or able to exploit the opportunities that exist.

Second, given that learners appear to be variable in attainment of collocations even in a L2 immersion situation (Dörnyei et al., 2004), a second factor could be memory (Baddeley, 2000; Gathercole, 2006; Gathercole, Willis, Emslie, & Baddeley, 1992). Following Ellis’s claim that “much of language acquisition is in fact sequence learning . . . determined by individual differences in learners’ ability to remember simple verbal strings in order” (1996, p. 91), adult L2 learners with a longer memory span should be able to hold L2 phonological information (such as an interlocutor’s side of a conversation) long enough in short-term memory for a memory trace to be established and, after several encounters of the same thing, for it to be consigned to the long-term memory store. This, in turn, yields a database for the abstraction of regularities in collocation and colligation, such as two or more words frequently co-occurring, or a word and a particular syntactic structure never co-occurring. There is a large body of evidence linking phonological short-term memory (pSTM), as described by the phonological loop architecture of Baddeley and colleagues’ working memory model (Baddeley & Hitch, 1974), to the acquisition of word forms in vocabulary development. The relationship has been documented in children acquiring their L1 (e.g., Gathercole et al., 1992), in children learning a L2 (e.g., Service & Kohonen, 1995), and in adults learning a L2 (e.g., Papagno & Vallar, 1992). For recent reviews relating pSTM functioning to SLA more generally, see French and O’Brien (2008), Hummel (2009), and Kormos (2013). This evidence has led a number of researchers to call for the addition of measures of pSTM to aptitude batteries (e.g., DeKeyser & Koeth, 2011; Dörnyei & Skehan, 2003; Robinson, 2005). In addition to its influence on single-item vocabulary learning, pSTM has been found to influence the learning of L2 multiword units (Ellis & Sinclair, 1996) and collocations (Skrzypek, 2009). A measure of pSTM might therefore illuminate which individuals are more apt to detect nonnativelike word combinations in a L2 (Ellis & Schmidt, 1997).

A third factor to influence the development of NLSs could be age of L2 onset (AoO). Age effects on different dimensions of L2 attainment have been widely reported. For example, in the development of a native accent, Flege, Munro, and MacKay (1995) and Munro, Flege, and MacKay (1996) suggest that exposure to a L2 before age 7 is necessary. For nativelike attainment in a L2 grammar, DeKeyser (2000), Johnson and Newport (1989), and Patkowski (1980) indicate that exposure by puberty or
thereabout is necessary, with DeKeyser’s study implicating a subcomponent of aptitude—verbal-analytical ability (as measured by the Modern Language Aptitude Test, part IV)—in the attainment of high, although not fully nativelike, proficiency in learners who had an age of first exposure beyond 7. More recently, the data gathered by Abrahamsson and Hyltenstam (2008) suggest an earlier age limit. Their participants were L2 speakers of Swedish of such high proficiency that they were assumed by native Swedes to be fellow compatriots. But given a much more taxing grammaticality judgment test than that used by Johnson and Newport or DeKeyser, the L2 speakers scored lower than a native speaker baseline even though their AoO was in early childhood. In the case of L2 vocabulary, two studies (Spadaro, 2013, and Lee, 1998) have found evidence of age effects on measures of lexical patterning. Spadaro (2013), looking at collocational knowledge, compared native speakers with three groups of nonnative speakers who had different AoOs: < 6 years, 7–12 years, and > 13 years. Her results showed that even highly advanced L2 users with AoOs greater than 6 years performed below native speaker levels. Lee’s (1998) data also showed that collocational ability was related to AoO. She divided her L2 users into the same AoO bands as Spadaro and found a strong negative correlation between AoO and collocational test scores, with only the < 6 age group achieving native speaker levels (see Long, 2007, pp. 50–55, for a full discussion of these studies). With AoO thus framed as a pervasive influence on ultimate L2 attainment in grammar, phonology, and collocations, it is interesting to explore how this might influence the development of NLS ability more widely.

THE RESEARCH STUDY

Research Aims

The study reported here investigated the extent of L2 learners’ memory store of NLSs and how this might be related to the following six independent variables: (a) engagement with the L2 community, (b) motivation to acquire a high level of L2 attainment, (c) AoO, (d) length of exposure (LoE) to the L2, (e) pSTM, and (f) context of acquisition (inside or outside the L2 community).

Participants

The participants were 79 adult Polish-born speakers of English; 39 were residents of West London in the United Kingdom, and 40 were residents...
of Szczecin in Poland. As determined by a face-to-face interview in both languages, all were comfortably bilingual in Polish and English and were frequent (i.e., daily) users of spoken and written English at a level of B2/C1 on the Common European Framework (CEF). To avoid criticism such as that applied to studies that involved L2 users with only 5 years of exposure (e.g., Johnson & Newport, 1989), the participants in our study had a minimum of 12 years of exposure to English, and, for many, their exposure spanned several decades. According to DeKeyser (2012), there is empirical evidence that “learners asymptote after ten years for just about any area of language, except vocabulary” (p. 456). It is an open question whether NLS ability develops with continued exposure, and so it was important to include in our study individuals with many years of exposure. Although it is possible that the participants residing in Poland were exposed to nonnative English in their daily lives, it is equally possible that this happened also to the participants residing in the United Kingdom, who were all part of the large Polish-speaking community in London. In the United Kingdom cohort (NNS UK), the range of LoE was 12–67 years ($M = 37$); for the Poland cohort (NNS PL), it was 12–35 years ($M = 17$). The AoO for the NNS UK cohort was 1–35 years ($M = 19$). For the NNS PL cohort, AoO was 5–30 years ($M = 12$). In the NNS UK cohort, the age range at testing was 19–79 years ($M = 56$) and 19–59 years ($M = 30$) in Poland. Twenty adult monolingual native speakers of English (i.e., people born in the United Kingdom) aged between their late 20s and late 50s were also recruited to give baseline scores. They were all residing in West London. We recruited speakers for whom English was their only language, learned from birth, so as to avoid any possibility that the lexicon of another language might act as a confound.

**Research Instruments**

Four instruments were used in this study. To explore the nature of the L2 participants’ daily usage of English, their attitudes to British life and culture, and their motivation to be highly competent users of English, a questionnaire was devised, following Freed, Dewey, Segalowitz, and Halter (2004). This asked about participants’ level of contact with English speakers, their level of engagement with various types of English-language media, and how important it was to them to be accepted as full members of an English-speaking community. The questionnaire is included in Appendix A. Answers were recorded on Likert scales to give quantitative data.

To measure pSTM, a serial recall task developed by Pisoni and colleagues was adapted in Superlab software (Conway, Baurnschmidt,
This pSTM task is the result of a long-standing research program to index phonological processing (e.g., pSTM, sequence learning, etc.) in a population that cannot be easily assessed with traditional instruments (e.g., deaf children with and without cochlear implants; see Karpicke & Pisoni, 2004, for the development and validation of these measures). During the task, participants listened to a recorded voice in Polish saying sequences of four color names (i.e., red, green, blue, and yellow). The sequences were four to eight items in length. After listening to each sequence, participants saw a question mark prompt on a computer screen, followed by a screen on which the four colors were presented on a four-square grid. They then repeated the sequence aloud while using a computer mouse to click the colors on the grid in the order they had heard them. There were three practice sequences and then 62 test sequences. Superlab recorded their mouse clicks. Because this measure was also devised to index pattern sensitivity (not reported here), the first 32 sequences and half of the last 30 sequences were generated by a finite state grammar. A participant’s serial recall span was measured as the weighted total (e.g., a correctly repeated sequence five items in length was awarded 5 points) of those 15 sequences, unconstrained by any underlying pattern.

To measure ability to detect nonnativelike selection in English, a test was devised that consisted of two stories (193 and 130 words in length) that were based on strip cartoon prompts concerning a lost football and a walk around town. The full texts are provided in Appendix B. Authentic nonnative speaker transcripts were used to generate these stories, with the grammar corrected but the nonnative word combinations left in. To ensure the construct validity of this test, it was extensively piloted and adjusted to remove any genre-specific material or uncommon vocabulary. All the words used were compared to the word frequency lists that are based on the British National Corpus (in Cobb’s, n.d., Compleat Lexical Tutor). The two texts returned, respectively, 99.3% and 96.4% of vocabulary in the 3,000 most common words in the British National Corpus and could therefore be considered well within the capacity for daily users of English at the CEF B2/C1 level of proficiency. Furthermore, at the time of testing, all the participants were invited to seek clarification of any words in the texts that they did not know; none asked for clarification. Extensive native speaker testing of the texts coupled with searches in the British National Corpus identified 24 nonnative selections in collocation (imagine an idea, get success), derivational morphology (to gunfight), and colligation (the ball came up by floating, reply by a shrug). Participants were assured the stories contained no grammatical, punctuation, or spelling errors. They were asked to read the stories and underline anything they found odd in expression. They scored 1 point for each correctly identified nonnative selection.
Following DeKeyser (2000), we did not score incorrect answers on this test. A final instrument used in this study was a grammaticality judgment test, reported elsewhere (Foster, Bolibaugh, & Kotula, 2013).

**Procedure**

The native English speakers were asked to do the NLS test. This was to determine, rather than assume, baseline native speaker performance. The Polish-born participants did all the tests. After an initial contact, participants were visited at a convenient place, usually their own homes. The pSTM test was administered first, as piloting had shown it to be the most tiring. The other two tests were administered in the following order: NLS test, grammaticality judgment test (reported elsewhere), and questionnaire. Rest and refreshment breaks were given between tasks. A payment of £30 or 150 PLN (Polish złoty) was made to participants for their time. All data were coded, and 10% were checked by a second rater. There was more than 95% agreement on all measures. These data were then put into a SPSS file for analysis.

**RESULTS**

This study investigated the effects of six independent variables on L2 users’ ability to detect NLSs. Following a summary of results for our dependent measure, this section presents the results for five of the independent variables for NNS UK and NNS PL participants separately; these are (a) degree of engagement with the TL community, (b) motivation to be like a native speaker, (c) AoO, (d) LoE, and (e) pSTM. The sixth independent variable—context of acquisition as inside or outside the TL community—will be considered simultaneously with the first five. We conclude with two sequential regression models, one for NNS UK participants and one for NNS PL participants, to estimate the joint contributions of the independent variables.

**NLS Test**

Native speakers \((N = 20)\) detected between 17 and 24 of the total 28 nonnativelike selections in the text. The mean score was 20.50 \((SD = 1.79)\). Nonnative speakers in the United Kingdom \((N = 39)\) detected between 0 and 24, with a mean of 14.35 \((SD = 6.80)\), whereas nonnative speakers in Poland \((N = 40)\) detected between 2 and 19, with a mean of 10.25 \((SD = 5.40)\).
Results from the Kolmogorov-Smirnov test indicated that distributions for each group were normal.

To see whether differences in the means were significant, we conducted a one-way ANOVA with NLS as the dependent variable and group as the independent variable. Results revealed a main effect of group, $F(2, 62.97) = 65.46$, $p < .001$. Because Levene’s test indicated that variances between groups were significantly different, Welch’s $F$-ratio is reported here. Planned contrasts demonstrated that the native speakers’ NLS ability was significantly better than either of the Polish L1 groups, $t(91.80) = -10.31$, $p < .001$ (one-tailed). It was also found that the NNS UK group’s NLS ability was significantly better than the NNS PL group, $t(74.23) = -2.99$, $p < .001$ (one-tailed).

However, it seemed possible that the differences in sample characteristics (i.e., age, AoO, LoE) might account for the differences in NLS ability between the NNS PL and NNS UK cohorts. A second ANOVA was therefore run between the two nonnative speaker groups with AoO and LoE as covariates. Age at testing was not included, as it is the sum of a participant’s AoO and LoE. There was still a significant effect of group (NNS PL or NNS UK) on NLS ability after controlling for the effects of both AoO and LoE, $F(1, 76) = 5.37$, $p < .05$, $r = .25$. We therefore report the results for the remaining independent variables separately for each group.

### Independent Variables

**Engagement with the L2 and Motivation for a High Level of L2 Attainment.** Table 1 reports the means, standard deviations, and correlations with NLS of the questionnaire items that investigate engagement with the L2 and motivation to acquire a high level of L2 attainment. Five items index engagement (Questions 1–5), whereas five other items index motivation to acquire a high level (Questions 6–10).

The results of the NNS UK group’s questionnaires were compared to their NLS scores through Pearson correlations. Only two moderately significant correlations were found: hours spent reading and watching television in English, $r = .328$, $p = .041$, and number of days per week using English online, $r = .391$, $p = .014$. Otherwise, no relationship was detected between NLS test scores and questions regarding liking the English language, admiring the culture and art, wishing to be like a native speaker, proportion of social interactions in English, proportion of daily life conducted in English, or language in which they felt more comfortable. Given that all participants who had an AoO of 12 years or earlier were performing at native speaker levels in the NLS test, the data were divided to include only those participants who had come
## Table 1. Mean scores and standard deviations for the questionnaire and Pearson correlations with NLS

<table>
<thead>
<tr>
<th>Question</th>
<th>Descriptive statistics</th>
<th>Correlations with NLS scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NNS UK</td>
<td>NNS PL</td>
</tr>
<tr>
<td>Daily use of English</td>
<td>2.87</td>
<td>1.21</td>
</tr>
<tr>
<td>Number of friends who use English</td>
<td>2.97</td>
<td>.96</td>
</tr>
<tr>
<td>Books, magazines, TV—days per week</td>
<td>7.44</td>
<td>1.02</td>
</tr>
<tr>
<td>Books, magazines, TV—number of hours</td>
<td>3.28</td>
<td>1.52</td>
</tr>
<tr>
<td>Internet—days per week</td>
<td>5.10</td>
<td>3.05</td>
</tr>
<tr>
<td>Internet—number of hours</td>
<td>2.10</td>
<td>1.35</td>
</tr>
<tr>
<td>More comfortable language</td>
<td>2.13</td>
<td>.92</td>
</tr>
<tr>
<td>Speaking to communicate with NSs</td>
<td>4.89</td>
<td>.31</td>
</tr>
<tr>
<td>Speaking as well as NSs</td>
<td>4.36</td>
<td>.96</td>
</tr>
<tr>
<td>Degree of liking of English language</td>
<td>4.46</td>
<td>.94</td>
</tr>
<tr>
<td>Culture and art</td>
<td>4.26</td>
<td>.82</td>
</tr>
<tr>
<td>Being similar to NSs</td>
<td>3.23</td>
<td>1.33</td>
</tr>
</tbody>
</table>

Note. NNSs = native speakers.

* p < .05.

** p < .01.
to the United Kingdom after this point, and the correlations were run again. This showed no significant correlations for any part of the questionnaire except for the weekly use of English online, $r = .667, p = .001$. It is intriguing to note that this result has been replicated in a follow-up study (Bolibaugh & Foster, 2013).

For the NNS PL participants, the only significant correlations of the NLS test scores and answers to the questionnaire were number of days per week on which they use English for reading and watching television and films, $r = .467, p = .002$, and weekly Internet use, $r = .367, p = .02$. Otherwise, measures such as daily use of English, motivation to learn, number of acquaintances who speak English, liking the culture, and wanting to be like a native speaker were again shown to bear no relationship to the development of an ability to detect nonnative selections. When the data were selected to include only participants with an AoO of 13 years or later, no correlation was found between NLS test scores and any part of the questionnaire.

**Age of Onset.** For the NNS UK participants ($N = 39$), the effect of AoO on NLS is illustrated by the scatterplot in Figure 1(a). This shows the native speaker baseline scores, ranging from 17 to 24 out of a possible 24. The regression line between AoO and NLS for native speakers has an intercept at the mean, 20.50, and zero slope. Nonnative speakers whose AoO was 12 or earlier scored within this range. With two exceptions, the nonnative speakers whose AoO was 13 or later scored below this range (and in most cases well below it). The two exceptions have an AoO of 17 and 21 and scored in the lower end of the native speaker range. A Pearson correlation analysis shows a large and significant negative relationship between AoO and NLS scores, $r = .748, p < .001$, which accounts for 56% of the variance ($R^2 = .56, p < .001$), a large effect.

For the NNS PL participants, the pattern is rather different. The scatterplot in Figure 1(b) shows that a few of them have scored within the bottom of the native speaker range, and all but two of these were exposed to English before the age of 10. Nevertheless, the scores generally appear to have a normal distribution, with no one scoring in the higher native speaker range. A Pearson correlation shows a weak negative relationship that does not reach significance, $r = -.277, p = .084$, which accounts for a nonsignificant 8% of the variance, $R^2 = .08, p = .084$.

In light of the fact that all NNS UK participants whose AoO was 12 or earlier scored within native speaker range, we divided each participant group into early AoO ($\leq 12$ years old) and late AoO ($> 12$ years old) starters and reexamined the relationship between AoO and NLS. Figure 1(c) shows the separate regression lines for early ($N = 14$) and late ($N = 25$) starters in the United Kingdom. For early starters, there is no relationship between AoO and NLS, $r = -.004, p = .989$. For late starters, there appears to be a negative relationship between AoO and
Figure 1. Scatterplots between NLS and age of onset (AoO) in (a) native speakers (NSs) and NNS UK, (b) native speakers and NNS PL, (c) NNS UK grouped by early and late starters, and (d) NNS PL grouped by early and late starters. In plots (a) and (b), native speaker range is marked by a horizontal line at 17, and division between early and late starters by a vertical line at 12.

NLS, but this relationship is not significant, $r = -.366, p = .072$. For the NNS PL participants, early starters ($N = 20$) appear to show a negative effect of increasing AoO on NLS, but this relationship is not significant, $r = -.210, p = .374$. The correlation in late starters ($N = 20$) is similarly nonsignificant, $r = -.02, p = .926$. 
**Length of Exposure.** For the NNS UK participants, the effect of LoE (i.e., length of their residence in the United Kingdom) on NLS in the entire group regardless of AoO is illustrated by the scatterplot in Figure 2(a). A Pearson correlation analysis shows a moderate positive relationship overall between NLS test scores and LoE, \( r = .457, p = .003 \), representing 21% of the variance, \( R^2 = .21, p = .003 \), a medium effect. When the effects of exposure are considered separately for early and late starters in Figure 2(c), however, the effect disappears. Figure 2(c) shows that the native speaker range of 17–24 is achieved here by a cluster of early starters with 55–65 years of exposure to English and by a smaller number of early starters with 15–30 years of exposure. The association between NLS and LoE for this group \((N = 14)\) is very slight and nonsignificant, \( r = .122, p = .677 \). For the late starters \((N = 25)\), who did not score within

![Figure 2](https://www.cambridge.org/core/images/figure2.png)

**Figure 2.** Scatterplots between NLS and length of exposure (LoE) in (a) NNS UK, (b) NNS PL, (c) NNS UK grouped by early and late starters, and (d) NNS PL grouped by early and late starters.
the native speaker range, there is no clear pattern relating NLS test scores to LoE, \( r = .051, p = .810 \), and there are two participants who, despite 60 years or so of residence in the United Kingdom, have NLS test scores well below the native speaker range. There is also one with 39 years of residence who scored zero.

As shown in Figure 2(b), a few NNS PL participants have scores in the low end of the native speaker range. There is a moderate positive relationship between NLS test scores and LoE, revealed in a Pearson correlation analysis, \( r = .407, p = .009 \), which suggests that NLS ability does improve with LoE to the language, even though participants do not achieve more than the lower end of native speaker levels. In the NNS PL group, exposure accounts for 17% of the variance in NLS scores, \( R^2 = .17, p = .009 \), a medium effect. Looking at the effects of exposure on the early (\( N = 20 \)) and late (\( N = 20 \)) starters separately in Figure 2(d), however, it can be seen that this association is due to the strong effects of exposure in late starters only. In early starters, there is no linear relationship between exposure and NLS, \( r = .193, p = .414 \), whereas in late starters, there is a strong positive correlation, \( r = .689, p = .001 \), which accounts for 46% of the variance, \( R^2 = .457, p = .001 \), a large effect.

**Phonological Short-Term Memory (pSTM).** Participants in the United Kingdom (\( N = 38 \)) scored between 0 and 40 points on the pSTM measure with a mean of 17.84 (\( SD = 10.58 \)). Participants in Poland (\( N = 40 \)) scored between 6 and 53 with a mean of 26.58 (\( SD = 11.32 \)). Figures 3(a) and (b) show the relationship between NLS and pSTM in the full NNS UK and NNS PL groups regardless of AoO. Pearson correlations show no effect of pSTM in either the NNS UK participants, \( r = .235, p = .17 \), or the NNS PL participants, \( r = -.027, p = .87 \).

Figure 3(c) shows the separate regression lines for early- and late-starting NNS UK participants, and it can be seen that pSTM affects NLS differently depending on AoO. Early starters (\( N = 14 \)) still show no effect of pSTM, \( r = .131, p = .655 \), but in late starters, there is a moderate to strong positive correlation, \( r = .535, p = .007 \), which accounts for 29% of the variance, \( R^2 = .287, p = .007 \), a large effect. The two late starters who scored within the native speaker range are shown with filled-in squares. It is notable that they have pSTM scores 1.4 and 1.6 standard deviations above the mean. For the NNS PL participants, there is no evidence of a similar pattern. Figure 3(d) shows no influence of pSTM in either early starters, \( r = .001, p = .996 \), or late starters, \( r = -.132, p = .579 \).

Given these differing effects for pSTM depending on AoO and context and the idea that the effects of exposure would themselves be mediated by pSTM, we wanted to see how much variance exposure and pSTM would account for once AoO had been taken into account. The final step in our analysis was, therefore, to run separate hierarchical (sequential)
regression models for NNS UK and NNS PL participants with NLS as the outcome measure and AoO, LoE, and pSTM as the three predictors.

A hierarchical regression model (Table 2) was fitted with AoO entered first, LoE second, and pSTM last. In the final model, the three predictors together account for 65% of the variance in NLS, $R^2 = .648$, $F(3, 33) = 21.49$, $p < .001$, with AoO predicting the largest part of the variance (57%). Length of exposure and pSTM together explain 8% additional variance in NLS ability, and their standardized regression coefficients ($\beta$) reveal a moderate and significant effect of each.

The three-step model clarifies the influence of exposure in the NNS UK group as a whole. In Step 2, adding LoE to the model with AoO results in very little additional variance explained (1%), and neither the change...
in the multiple correlation squared ($R^2$) nor the regression coefficient is significant. Adding pSTM as a predictor in Step 3, however, not only improves the general fit of the model by significantly increasing the multiple correlation squared ($R^2$) but also increases the size and significance of LoE’s coefficient. The same effect can be seen in the increase in size of pSTM’s standardized regression coefficient relative to its bivariate correlation, which is nonsignificant, $r = .235$, $p = .075$. This relationship of reciprocal suppression (Lutz, 1983; Tzelgov & Henik, 1991) indicates that pSTM accounts for a portion of the “noise,” or unsystematic variance, in the relationship between exposure and NLS scores, and that the converse is also true.

A second hierarchical regression model was fitted for participants residing in Poland (Table 3). Again, AoO was entered first, LoE second, and pSTM last. In the final step, the three predictors together account for 32% of the variance in NLS, $R^2 = .323, F(3, 36) = 5.72, p = .003$. This model does not appear to be an improvement on Step 2, however: pSTM’s coefficient does not reach significance in Step 3 nor does the change in $R^2$.

Step 2, with AoO and LoE as predictors, therefore appears to be the better model.

In Step 2, the two predictors together account for the same amount of variance (32%) as Step 3, $R^2 = .317, F(2, 37) = 8.59, p = .001$. Age of onset is moderately negatively associated with NLS ability, and LoE is moderately positively associated. The inclusion of LoE in the model increases the size and significance of AoO’s coefficient. This indicates that some

Table 2. Results of hierarchical regression analysis for variables predicting NLS in the NNS UK group ($N = 38$)

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$SE$ $B$</th>
<th>$\beta$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1$^a$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>24.10</td>
<td>1.53</td>
<td>-.77</td>
<td>.000</td>
</tr>
<tr>
<td>AoO</td>
<td>-0.51</td>
<td>0.07</td>
<td>.00</td>
<td>.000</td>
</tr>
<tr>
<td>Step 2$^b$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>21.76</td>
<td>2.82</td>
<td>-.70</td>
<td>.000</td>
</tr>
<tr>
<td>AoO</td>
<td>-0.47</td>
<td>0.08</td>
<td>.12</td>
<td>.329</td>
</tr>
<tr>
<td>LoE</td>
<td>0.04</td>
<td>0.05</td>
<td>.35</td>
<td>.020</td>
</tr>
<tr>
<td>Step 3$^c$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>12.79</td>
<td>4.40</td>
<td>.33</td>
<td>.015</td>
</tr>
<tr>
<td>AoO</td>
<td>-0.37</td>
<td>0.09</td>
<td>-.55</td>
<td>.000</td>
</tr>
<tr>
<td>LoE</td>
<td>0.13</td>
<td>0.05</td>
<td>.35</td>
<td>.020</td>
</tr>
<tr>
<td>pSTM</td>
<td>0.21</td>
<td>0.08</td>
<td>.33</td>
<td>.015</td>
</tr>
</tbody>
</table>

$^a R^2 = .57$ for Step 1 ($p < .001$).
$^b \Delta R^2 = .01$ for Step 2 ($p = .329$).
$^c \Delta R^2 = .07$ for Step 3 ($p = .015$).
of the unsystematic variance in the relationship between AoO and NLS is due to LoE, and the inclusion of LoE in the regression model allows for a more accurate estimation of the relationship between AoO and NLS. Our earlier observation that pSTM appears to be unrelated to NLS in the NNS PL participants is confirmed in Step 3, which considers the relationship after both AoO and LoE have been accounted for.

Comparing the regression models for NNS UK and NNS PL participants reveals several interesting findings. First, far more of the variance in the NNS PL participants remains unexplained (68%) than in the NNS UK participants (35%), and this is mostly due to the large influence of AoO in the NNS UK group. Second, exposure accounts for a much greater portion of the variance in the NNS PL participants (24%) than in the NNS UK participants (8%, when considered jointly with pSTM). Finally, these exposure effects in NNS UK participants are only revealed when pSTM is taken into account; no such relationship is evident in the NNS PL participants.

**Summary of Independent Variables.** Here we summarize the results for the impact of each of the independent variables on NLS ability.

1. Motivation and engagement: The questionnaire showed that for neither cohort were positive feelings about the target language and culture related to NLS scores. In terms of engagement with the English language, there was one significant relationship, that of Internet use. This relationship disappears when an analysis is run on data from only the late starters.

---

**Table 3.** Results of hierarchical regression analysis for variables predicting NLS in the NNS PL group (N = 40)

<table>
<thead>
<tr>
<th>Step</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>13.55</td>
<td>2.04</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>AoO</td>
<td>-.27</td>
<td>0.15</td>
<td>-.28</td>
<td>.084</td>
</tr>
<tr>
<td>Step 2b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>5.82</td>
<td>2.78</td>
<td>.043</td>
<td></td>
</tr>
<tr>
<td>AoO</td>
<td>-.39</td>
<td>0.14</td>
<td>-.40</td>
<td>.007</td>
</tr>
<tr>
<td>LoE</td>
<td>0.55</td>
<td>0.15</td>
<td>.51</td>
<td>.001</td>
</tr>
<tr>
<td>Step 3c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>6.80</td>
<td>3.32</td>
<td>.048</td>
<td></td>
</tr>
<tr>
<td>AoO</td>
<td>-.40</td>
<td>0.14</td>
<td>-.41</td>
<td>.007</td>
</tr>
<tr>
<td>LoE</td>
<td>0.56</td>
<td>0.16</td>
<td>.51</td>
<td>.001</td>
</tr>
<tr>
<td>pSTM</td>
<td>-.04</td>
<td>0.07</td>
<td>-.08</td>
<td>.581</td>
</tr>
</tbody>
</table>

---

\(a \quad R^2 = .08 \text{ for Step 1 (} p = .084)\).

\(b \quad \Delta R^2 = .24 \text{ for Step 2 (} p = .001)\).

\(c \quad \Delta R^2 = .01 \text{ for Step 3 (} p = .584)\).
2. Age of onset: An early AoO (< 13) serves as a guarantee of nativelikeness in an immersion context. In a foreign language learning context, two participants with AoOs of 14 and 15 reached native speaker levels. Otherwise, an early start was necessary but not sufficient. Only a very few of those participants who started early reach nativelike levels of attainment.

3. Length of exposure: Longer exposure brings moderate improvement to NLS for NNSs living within the TL community but is not sufficient to bring participants to nativelike levels without the benefit of an early start. In a foreign language learning context, exposure accounts for the largest portion of the variance in NLS scores, but this effect was only found in those participants who started after the age of 12. Those who started earlier showed no effect of exposure on NLS.

4. Phonological short-term memory: For late starters within the TL community, pSTM is the only significant predictor of NLS ability. There is also evidence of a relationship of reciprocal suppression between pSTM and LoE, indicating that each accounts for unsystematic variance in the other’s relationship to NLS. For participants living outside the TL community, pSTM has no relationship with NLS, regardless of AoO or LoE.

5. Context: NLS ability is significantly affected by type of exposure (foreign language or immersion), with the NNS UK cohort outperforming the NNS PL cohort. Nonetheless, even after a minimum of 12 years of immersion, and in many cases considerably longer, the NNS UK cohort is still significantly different from the native speaker mean.

**DISCUSSION**

Our results show that AoO and context of learning exert the strongest and most predictable influence on the ability of a learner to detect non-native selections in a text. In our regression models, AoO accounts for 57% of the variance in NLS scores in NNS UK participants—a number in line with effects found for AoO in studies investigating ultimate attainment in morphosyntax in immigrants (DeKeyser, 2012)—in comparison with 8% of the variance in NNS PL participants. The scatterplots in Figure 1 show clearly different regression lines for early and late starters and are compatible with other research into age effects on language acquisition, particularly in an immersion context (e.g., Abrahamsson, 2012; Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000). A question that obviously arises here is this: Can the NNS UK NLS test data be described as showing a critical period (CP) effect (e.g., Bialystok, 2002; Birdsong, 2005, 2006; DeKeyser, 2000; Lenneberg, 1967; Singleton, 2005)? For the most part, the discussions about the evidence of a CP in language acquisition revolve around the shape of the line when plotting a measure of language performance against AoO, with a CP entailing a flat start, a sudden downward turn at a certain age, and then a flattish decline afterward (Birdsong, 2005, p. 113). A more general relationship between age and learning (i.e., not a CP) would show a step-by-step decline in
performance as AoO advances. Our results seem more consistent with a CP than a general age effect in that the discontinuity in NLS ability between the younger and older starters is so marked; the regression lines are more or less flat for the younger starters and steeply decline for the older starters. Eubank and Gregg (1999) have posited different CPs for different aspects of language proficiency, such as phonology and morphosyntax, and our results suggest that a CP for the ability to detect grammatical but nonnativelike word combinations could be added to these.

This might be considered in the light of Bley-Vroman’s fundamental difference hypothesis (Bley-Vroman, 1989), which posits that children use a domain-specific implicit learning mechanism for language, whereas adults rely instead on a more general explicit problem-solving ability. DeKeyser (2000) showed that the small set of late L2 learners in his study who scored within the native speaker range on a grammaticality judgment test also had an unusually good verbal aptitude, defined by him as “explicit, analytic problem-solving capacities” (p. 518). Our study investigated an aspect of L2 knowledge that we have argued cannot plausibly be explained this way; it seems likely that the knowledge of many thousands of nativelike word combinations used by a speech community is part of what Abrahamsson (2012) has termed “acquired, nonverbalized linguistic intuition” (p. 197) and is gained incidentally through experience rather than deliberate and explicit analysis. For the late United Kingdom-based learners in our study, an ability to build these intuitions was mediated through their span of pSTM, in tandem with massive exposure to the TL. As we previously noted, only two of the late learners in the United Kingdom scored inside the native speaker range (Figure 3[c]), and they had pSTM scores around 1.5 standard deviations above the mean. Apart from these two exceptions, the results of the NLS test could not distinguish between native speaker participants and the nonnative participants who came to the United Kingdom at or before the age of 12. For these, we conclude that AoO had more impact than motivation or engagement or simple amount of exposure. For the rest of the late learners, even several decades of daily use of English in the United Kingdom did not compensate for starting after the age of 12. For the participants residing in Poland, only six achieved NLS scores (just) within the native speaker range, and four of these had AoOs under 10 years. The two late starters who scored just within the native speaker range (17 and 18 out of 24) had ages of onset of 14 and 15, respectively. We are not able to account for why their NLS scores were higher than the rest of the NNS PL late starters (several of whom have NLS scores of under 5 out of 24), but it is perhaps noteworthy that when these two participants began to study English, in 1974 and 1981, studying English was highly unusual for any schoolchild in Poland, so they are rather exceptional in this regard.
Additionally, the results of this study suggest that the effects of memory on the ability to distinguish nativelike from nonnativelike selections also depend both on type of exposure and on AoO. That memory effects are only seen in an immersion context supports usage-based accounts’ predictions that NLS ability requires massive input from the target speech community. That memory is the single best predictor of NLS in those United Kingdom-based learners with a later (> age 12) start provides support for the notion that some adult onset L2 learners can incrementally accumulate knowledge about the lexical patterning of language in much the same way as L1 learners are posited to (Ellis, 2003, p.74), whereas the suppression effects found between memory and amount of exposure lend additional support to chunking accounts based on the repetition of phonological sequences in short-term memory. Nonetheless, superior pSTM in combination with many years of exposure to the TL community enabled only two late-starting learners to score within native speaker range.

As a further explanation for why knowledge of NLS does not fully develop in the majority of late learners, we can consider Hoey’s (2005) work on lexical priming. He argues that when a word is learned in a L2, initially at least, it will activate the related collocational and syntagmatic relationships of its L1 counterpart, and these are not likely to be nativelike to the L2. A L2 word, even if it has an exact semantic equivalent to something in the L1, will not share the same network of relationships with other L2 words. The learner may assume it does and may produce a nonnative selection as a result. The learner may also find it difficult to distinguish what has been genuinely encountered in the L2 (NLS) from what might be transferred from the L1 (nonnativelike selection). Also, given the perceived tendency for adult learners to focus on word-sized units in the L2 (Fitzpatrick & Wray, 2006, p. 54) they might not pay sufficient attention to the other words or grammatical constructions that co-occur with it. Accordingly, adult learners are likely to analyze a L2 sequence into its constituent parts and, in doing so, lose sight of valuable information about how the individual words work together. Consequently, the nonnativelike primings that attach themselves to L2 vocabulary items are less likely to be supplanted by newer, nativelike primings. In contrast, children acquiring two languages have relatively less developed primings in whatever language came first, and, as a consequence, there is less interference from that quarter when they encounter a new language. Given that child learners do not “atomize” a sequence of words, simply because they have not developed the explicit learning mechanisms necessary to do this, their more holistic implicit learning mechanisms capture collocational and colligational information, and their memories retain the trace. The learner of English with an early AoO will possibly reject the ball came up by floating as a NLS because she will never have encountered a verb of motion relegated to an adverbial phrase such as this, whereas she will have
encountered them as main verbs on innumerable occasions. The learner with a later AoO might atomize the sentence and find it well formed, which of course it is, without recourse to knowledge of how nonnativelike such a colligation might be.

Bringing empirical evidence to bear on this idea, Silverberg and Samuel (2004) investigated the lexical primings of early and late L2 learners and concluded that early L2 learners are able to develop conceptual and lexical representations at the same time, whereas late L2 learners appear to map their newly formed L2 word representations onto those ready made in their L1. This supports the notion that for late learners L1 word associations are carried over into their L2, impeding the development of nativelike L2 lexicon. Similarly, Satterfield (2008) shows how a number of studies into the neural underpinnings of bilingual representations of word meanings all suggest very different architectures for the lexicons of late and early L2 learners (Kim, Relkin, Lee, & Hirsch, 1997; Weber-Fox & Neville, 1999; Wuillemin & Richardson, 1994). Early learners (i.e., before age 7) are able to map L2 words onto the same conceptual spaces used for L1 words, whereas older L2 learners do not do this. It is possible that the small number of our NNS PL starters with ages of onset between 7 and 15 who did well on the NLS test were able to map L2 word representations onto L1 conceptual space.

These differences between early and late L2 lexical acquisition also serve to explain why pSTM acts as an aptitude for L2 NLS ability in late-onset learners. In Ellis’s (1997) account of the development of collocational knowledge in a L1, associations between word forms (chunks) are created through repeated exposure and are thus mediated by pSTM. The chunking process is largely implicit and is also influenced by the semantic or functional properties of a particular utterance: Sequences that are reliably associated with an important meaning or function are more likely to become associated in long-term memory. Thus, there are separable processes for the development of form and meaning, and, in a L1, the simultaneous development of these lexical (formal) and conceptual (semantic) levels means they are mutually influential. This is not the case for late L2 learners, who are particularly dependent on lexical learning given their preexisting conceptual representations. Phonological short-term memory operates at the lexical level, mediating the development of these formal representations.

This research study explored a dimension of L2 attainment—receptive knowledge of NLS—that has hitherto received scant attention, and our findings produced evidence that, similar to phonology and morphosyntax, receptive knowledge of NLS is sensitive to age effects that are consistent with a CP. This study has produced interesting results despite a number of limitations, and it suggests a few useful directions for future investigation. For example, Figure 2(b) shows that for the early starters in a foreign language context, there is no association between LoE and NLS scores,
but for late starters, there is a very strong one. So although starting early is a necessary but not sufficient condition for NLS ability, it nullifies the benefit of LoE. In contrast, starting late (excluding the two exceptional cases mentioned previously) means NLS will not reach native speaker levels but will improve with more exposure. This is an intriguing finding, and it will be interesting to see if it is replicable.

The NLS test worked well enough, but, in spite of extensive piloting, the native speaker range of 17–24 was wider than we had anticipated. When dealing with notions of oddness, there are never going to be clear-cut boundaries, and informants are always going to have individual levels of tolerance; thus, better categorization of the nativelike equivalents of the nonnative speaker selections might have resulted in a narrower native speaker range. The test of pSTM succeeded in accounting for why some adult learners can do much better than others at detecting nonnative selections. Further tests, such as a nonword repetition task, might give a better idea of the importance of item memory in developing these intuitions. Additionally, what Stevens (2006) has called the age-length-onset problem prevented us from including age as a covariate in our analyses. In light of evidence that short-term memory decreases as a function of age (Multhaup, Balota, & Cowan, 1996), another possible interpretation of our results is that pSTM is serving as a proxy for (life-stage or cognitive) age-related variance in our analyses. Given that we were interested in exploring the possible effects of very long lengths of exposure to the L2, this confound was inevitable. But, addressing these limitations, we conducted a second study in which we physically controlled multicollinearity between age at testing on the one hand and biographical and cognitive variables on the other. All the participants were under the age of 50, all had an AoO greater than 13, and all had resided within the United Kingdom for a minimum of 10 years. Findings (reported in Bolibaugh & Foster, 2013) have replicated a clear effect of pSTM on NLS ability and lend weight to the idea that the full complexity of the architectural structure of a L2 lexicon rests on the incidental learning mechanisms of early life working in a rich language environment.

Received 3 November 2011
Accepted 17 October 2012
Final Version Received 7 February 2013

NOTES

1. The hit counts presented here are the results of Google searches performed by the authors on August 30, 2013.
2. Although performance can be disrupted by brain damage and can occasionally fail due to extreme fatigue or intoxication, this does not change the essential point here that, under normal conditions, healthy individuals produce language that is nativelike.
3. One United Kingdom participant was removed from the analysis due to highly unusual home circumstances. She had married a fellow Pole and lived with her mother, grandmother, and two children, who all spoke only Polish to one another, making her AoO hard to determine.

4. Although the two groups differed statistically on predictor variables, an ANCOVA showed that there was a significant effect for context of learning even when these differences had been accounted for.

5. The CEF was established by the Council of Europe to standardize the levels of language exams. Level C1 describes effective operational proficiency, in terms of appropriateness, sensitivity, and the capacity to deal with unfamiliar topics. Level B2 is just below this.

6. Twelve years of exposure may still be insufficient for ultimate attainment, particularly in a nonimmersion setting, but this minimum nevertheless improves considerably on earlier studies.

7. A reviewer noted that a monolingual native speaker standard might not be considered appropriate in some areas of SLA research, but here it is very important to be sure that the baseline against which the L2 speakers’ performance is measured is not compromised by the influence of any other language.

8. Both the correct rejection of nonnativelike selections and the incorrect rejection of NLSs are of interest, but, for the purpose of the analysis presented here, only those word combinations that we had validated and piloted as nonnativelike were taken into account.

9. One participant’s data recorder did not activate properly.

10. This confound of course does not crop up in age effect studies for L2 morphosyntax or accent, as these can safely be said to be asymptotic after about 10 years.

REFERENCES


**APPENDIX A**

**QUESTIONNAIRE FOR UK PARTICIPANTS (ADAPTED WHERE NECESSARY FOR USE WITH THE PL PARTICIPANTS)**

1. Approximately what percentage of your daily language use is English?
   (i) less than 25%, (ii) 26%–50%, (iii) about half, (iv) 76%–90%, (v) more than 90%

2. Approximately how many of your friends and acquaintances normally speak English with you?
   (i) almost none, (ii) a few, (iii) about half, (iv) most, (v) nearly all
3a. Typically, how many days per week do you read English language 
newspapers, magazines, or novels, or watch English language movies, 
television, or videos?
0 1 2 3 4 5 6 7
3b. Typically, how many hours per day?
0–1 1–2 2–3 3–4 4–5 more than 5
4a. Typically, how many days per week do you use English online?
0 1 2 3 4 5 6 7
4b. Typically, how many hours per day?
0–1 1–2 2–3 3–4 4–5 more than 5
5. What language do you feel more comfortable speaking in?
1) Polish, 2) English, 3) equally comfortable in either
6. How important is it to you to speak English in order to interact with 
native speakers of English?
5 = very, 4 = quite, 3 = so-so, 2 = not very, 1 = not at all
7. How important is it to you to speak English as well as a native 
speaker?
5 = very, 4 = quite, 3 = so-so, 2 = not very, 1 = not at all
8. How much do you like the English language? (Circle your answer.)
5 = very much, 4 = quite a lot, 3 = so-so, 2 = not really, 1 = not at all
9. How much do you like the culture and art of native English 
speakers?
5 = very much, 4 = quite a lot, 3 = so-so, 2 = not really, 1 = not at all
10. How much would you like to become similar to native speakers of 
English?
5 = very much, 4 = quite a lot, 3 = so-so, 2 = not really, 1 = not at all

APPENDIX B

NATIVELIKE SELECTION TEST

Nonnative selections are numbered and in bold.

A young man was 1strolling his way in the street, listening to music 
on a Walkman. He had the volume up so loud, he couldn’t hear anything 
going on around him. This was a pity, as a lot of unusual things 2missed 
his notice.

For example, just after he left home, a car 3was crashed by another in 
a terrible accident, but it happened 4at his back and he saw 5nothing 
about it.

A little while later, a robber smashed a shop window and 6stole away 
some jewellery, but the young man didn’t hear a thing.
Later, in another street, two robbers of a bank were gunfighting with the police, but still the young man heard nothing, not even the shots of guns.

Finally, he went into a park to read his newspaper, still with his Walkman on loud. An escapee tiger crossed close behind him, but his ears were too full and he paid no attention.

Back home, as he was taking the coat off, his wife asked if he had enjoyed an interesting walk. He replied by a shrug, because for him there had been nothing of the extraordinary.

During a hot day, a group of boys were playing football in the park. Suddenly, one of them kicked the ball so strongly that it went sailing over the trees. When the boys eventually found it, it had gone down a deep hole in the ground.

The boys tried many efforts to reach the ball, but did not get success because the hole was so deep. When they saw a dangerous snake nearby they thought they should probably forget their play and go home.

But one of the boys was cleverer than the rest and imagined a good idea.

He ran off, and soon came back with a big bucket of water.

He poured it into the hole fully. The ball came up by floating and they got it back.