Implicit language learning: Adults' ability to segment words in Norwegian*

MEGAN M. KITTLESON
University of Arizona
JESSICA M. AGUILAR
University of Arizona
GRY LINE TOKERUD
University of Bergen
ELENA PLANTE
University of Arizona
ARVE E. ASBJØRNSEN
University of Bergen

(Received: April 18, 2008; Revised: December 4, 2009; Accepted: December 21, 2009; First published online 9 March 2010)

Previous language learning research reveals that the statistical properties of the input offer sufficient information to allow listeners to segment words from fluent speech in an artificial language. The current pair of studies uses a natural language to test the ecological validity of these findings and to determine whether a listener's language background influences this process. In Study 1, the "guessibility" of potential test words from the Norwegian language was presented to 22 listeners who were asked to differentiate between true words and nonwords. In Study 2, 22 adults who spoke one of 12 different primary languages learned to segment words from continuous speech in an implicit language learning paradigm. The task consisted of two sessions, approximately three weeks apart, each requiring participants to listen to 7.2 minutes of Norwegian sentences followed by a series of bisyllabic test items presented in isolation. The participants differentially accepted the Norwegian words and Norwegian-like nonwords in both test sessions, demonstrating the capability to segment true words from running speech. The results were consistent across three broadly-defined language groups, despite differences in participants' language background.

Keywords: implicit learning, language, statistical learning, second language acquisition

One of the fundamental questions of language learning is how learners initially segment words from fluent speech. Word segmentation is a challenging problem because continuous speech lacks the acoustic equivalent to the blank spaces that separate words in written text (Cole & Jakimik, 1980). One approach to solving this problem involves the use of statistical cues to segment speech. In 1955, Zellig Harris proposed that words may be identified as clusters of co-occurring sounds (Harris, 1955). In recent years, infants have been shown to utilize these distributional cues when learning language (see an extensive review by Jusczyk, 1999). These studies show that this mechanism of implicit learning emerges early in the period of language acquisition.

Human infants show evidence of using statistical information to identify word units in connected speech. Infants at 7.5 months of age listened longer to repetitions of target words present in a familiarization passage than words not present in the passage (Jusczyk & Aslin, 1995). This result demonstrated the infants' ability to recognize sound patterns heard in fluent speech. Eight-month-old infants were able to use distributional information concerning the predictability of syllable sequences (i.e., their transitional probabilities) to correctly discriminate between words and part-words in an artificial language that contained no pauses or other acoustic cues to word boundaries. Furthermore, they were able to demonstrate learning after only minutes of exposure to the artificial language (Aslin, Saffran & Newport, 1998; Saffran, Aslin & Newport, 1996a).

Adults also appear to use distributional information to identify words. Adults were able to identify words within an unsegmented artificial language based only on the transitional probability between syllables. Their ability to do so was comparable to children who were given the same task (Saffran, Newport, Aslin, Tunick & Barrueco, 1997). Similarly, adults appear to use information characterized by transitional probabilities to learn other aspects of language structure in artificial languages (e.g., Dahan & Brent, 1999; Gómez, 2002; Lany, Gómez & Gerken, 2007; Mintz, 2002; Thompson & Newport, 2007).

Although paradigms that use artificial languages are now quite common, little parallel work with natural language is available. Pelucchi, Hay and Saffran (2009) showed that English-reared infants can segment Italian words, based on their transitional probability in sentence contexts. Another study (McLaughlin, Osterhout & Kim,

* This work was supported in part by NICHD grant 2R01 HD042170 and NIDCD grant R01DC004726.

Address for correspondence:
Elena Plante, Department of Speech, Language & Hearing Sciences, PO Box 210071, The University of Arizona, Tucson, AZ 85721-0071, USA
eplante@u.arizona.edu

2004) suggests that adult learners of a second language begin to build an implicit awareness of word characteristics prior to the time when they can distinguish between actual and invented printed words in the second language. The electrophysiologic responses to real and pseudowords were measured in college students receiving formal instruction in French. The N400 waveform for these adults differed for words taken from their French text compared to French-like pseudowords. The magnitude of this effect correlated with the hours of instruction the students had received, suggesting that this effect was experience-driven. This finding for adult learners parallels the previously reported increases in sensitivity to French spellings by French school children, despite the lack of direct instruction concerning these regularities (Pacton, Perruchet, Fayol & Cleeremans, 2001).

In natural languages, distributional cues co-exist with other cues (e.g., prosodic cues). Jusczyk, Houston and Newsome (1999) found that 7.5-month-old infants acquiring English correctly segmented strong/weak bisyllabic words from natural English passages, a language in which trochaic stress predominates. In addition, infants between 7 and 9 months also showed evidence that they use stress cues preferentially over transitional probability information (Jusczyk et al., 1999). When presented with adjacent syllables that corresponded to a weak–strong–weak pattern, they showed evidence of identifying word-initial syllables on the basis of strong stress (Curtin, Mintz & Christiansen, 2005), even when this segmentation did not reflect an actual word (Jusczyk et al., 1999). For example, when familiarized with passages containing *guitar is*, infants preferred *taris* to *guitar*. However, by 10.5 months infants were able to correctly segment weak/strong words like *guitar* (Jusczyk et al., 1999). Thiessen and Saffran (2003) also examined preference for transitional probability information versus stress cues, but in the context of an artificial language in which only prosodic and statistical cues to word boundaries were present. In contrast to Jusczyk et al. (1999), Thiessen and Saffran's 7-month-old infant participants preferred words that were segmented on the basis of transitional probability information rather than stress cues. In addition, a group of 9-year-old participants showed a preference for stress cues when both stress and distributional cues were available. Like children, adults demonstrate the ability to segment words from an artificial language using transitional probabilities and the adults' performance improves with the addition of prosodic information (Saffran, Newport & Aslin, 1996b). These results suggest that learners can use both distributional and stress cues to help them recognize words in speech, but that their propensity to use one over the other may be age dependent.

It is not the case that learners come to these language learning studies as 'blank slates'. Both infants and adults show effects of prior language experience when learning a new language or language-like information. Infants may use syllable stress as an initial bootstrapping technique to identifying word boundaries (Echols & Newport, 1992). Nine-month-old infants show a listening preference for the prosodic patterns of their native language over those of other languages (Jusczyk, Cutler & Redanz, 1993). Thiessen and Saffran (2003) showed that infants not only preferred the prosodic patterns of their native language, but that this experience overrode the experiment-specific prosodic pattern presented in the familiarization phase of the experiment.

Prior experience also affects the use of other cues to language structure. Lany et al. (2007) found that providing prior experience to a simple structural pattern quickened participants subsequent learning of more complex versions of the basic pattern. These studies all suggest that prior (i.e., native) language experience might influence learning when learners are confronted with a new language. It may be logical to suppose that learners whose native language uses cues to word boundaries similar to those that predominate in a new language may find it easier to segment words in that new language.

Typically, the implicit learning experiments such as those reviewed above use an artificial language that is designed to control for the type and frequency of cues available to the learner. In addition, the artificial languages are often constrained in ways that are intended to make the learning problem simpler to solve in a brief period of time. For example, the phrase length may be very short relative to natural languages and the lexicon limited to very few recurring syllables. Although this approach succeeds in isolating the effect of individual sources of information contained in the stimuli, it greatly reduces the ecological validity of these findings. Natural languages contain a number of other cues which aid in the identification of word boundaries such as prosody, semantics, inflectional morphology, and unstressed function words.

Each of these cues is also present in Norwegian, the experimental language of the present study. Of the cues available to the naïve listener, the dominant trochaic (strong–weak) stress pattern of Norwegian is one that is also common to many other languages (Hyman, 1977). Likewise, the repeated occurrence of a small set of articles and gender markers could also assist the listener in identifying the presence of words, particularly nouns. This is a cue also common to most other natural languages, although the form of these markers is most similar to other Scandinavian languages and, to some extent, to German. In contrast, the phonotactic and allophonic variations are most specific to the Norwegian language relative to other languages. Obviously, the specific sound patterns that constitute Norwegian are largely unique to that language. However, the general phonotactic patterns of Norwegian overlap substantially with other Scandinavian languages and to a lesser extent with those of Germanic languages.

Although no single acoustic cue has been found to be sufficient to support accurate word segmentation, a combination of available cues may improve participants' performance. Conversely, natural languages tend to contain greater variability in terms of phrase length, morphological inflections, syntactic forms, and lexical elements relative to artificial languages. These features may increase the difficulty of the implicit learning task relative to that used in artificial language learning studies. The overall purpose of the current study is to determine if previous research on implicit learning of artificial languages applies to a natural language. In Study 1, we evaluate a set of Norwegian test items to determine whether individuals can distinguish target words from nonword foils without prior exposure to input that contain these test items (Norwegian sentences). In Study 2, we look at the effect of both language background and ambient exposure to Norwegian on adults' ability to segment words from continuous speech in an implicit learning task. In this study, listeners are first familiarized with Norwegian sentences that contain target words, and then are tested using items validated in Study 1.

Study 1

Methods

Participants

The participants were 22 undergraduate students (11 female, 11 male) enrolled at the University of Arizona. They ranged in age from 18 to 28 (mean 26.1 years). All spoke English and the majority had at least some proficiency in a second language (American Sign Language, French, German, Italian, Korean, Japanese, Mandarin, Spanish, Navajo). No participant reported speaking any Scandinavian language. Participants were given a language history survey and were excluded from the study if they indicated that they had a hearing loss, speech impairment, or learning disability. All participants provided informed consent in accordance with institutional review board requirements the University of Arizona. They were paid for their participation.

Materials

Study 1 was designed to evaluate the "guessibility" of the test items to be used in Study 2. These test items were created by clipping the required syllables of the acoustic signal directly from a familiarization stimulus set of Norwegian sentences (described under Study 2). All sentences were digitally recorded by a male native Norwegian speaker using Sound Forge 7.0 (Sony Pictures Digital Inc. 2003). The sentences were spoken with a natural speaking rate and intonation pattern for the structure and content of the sentence.

Three types of test items were created. The first type was correct Norwegian target words (see Table 1). These test items consisted of the bisyllabic real words. The target words occur relatively infrequently in Norwegian. Two types of incorrect items were also created. Since both types of incorrect test items were nonwords, they never occur as words in spoken Norwegian. The first type was a bisyllabic nonword which consisted of one syllable from a target word and one syllable from an adjacent bisyllabic word in one of the familiarization sentences. Half of these nonwords used the first syllable of a correct word and half used the second syllable of a correct word. This type of incorrect test item type will be referred to as PART-TARGET NONWORDS. These two-syllable test items never occur as words in Norwegian. The second type of incorrect item was formed by combining two adjacent monosyllabic words clipped from the familiarization sentences. This type of incorrect test item will be referred to as TWO-WORD COMBINATIONS. These words had a higher average frequency of occurrence in Norwegian (see Table 1) than the CORRECT TARGET WORDS. This is partially a reflection of the fact that some of the single-syllable words in the Two-Word Combinations were functors, which have a relatively high frequency of occurrence in any language. Ten of each test item type were created.

Both sets of nonwords contained some items that had a different prosodic pattern than the real Norwegian words (which all had trochaic stress). Likewise, the phonotactic patterns of the nonwords are necessarily less consistent with those of the natural parent language than are the real Norwegian words. Therefore, it is possible that subjects could distinguish between real and nonwords on those bases alone. If this is the case, then participants would prefer real words to nonword test items, even if they had not had prior exposure to the familiarization sentences from which the test items were obtained.

Procedures

The experiment was administered via laptop computer to one individual at a time. All individuals were tested in a quiet room at the University of Arizona. The instructions, familiarization stimuli, and test items were all presented using E-Prime Version 1.0 software (Copyright 2002 Psychology Software Tools, Inc.). Keyboard responses to the test items were recorded by E-Prime. The test items were presented in random order and participants were asked to indicate, by button press, if each item was a real Norwegian word or not.

Results

The acceptance rate for each of the test items was examined on an item-by-item basis. Although there was a

https://doi.org/10.1017/S1366728910000039 Published online by Cambridge University Press

Table 1. *Stimulus characteristics of test items for Study 2.*

	Frequency of occurrence per 1000 words in Norwegian[1]		Frequency of occurrence in familiarization sentences (542 words)	Forward transitional probability[2]	Two-way syllable dependency (r_{φ})[3]
CORRECT TARGET WORDS*					
lev'-er	0.07		14	1.0	1.0
mell'-om	0.97		12	.750	.798
kikk'-ert	0.002		12	1.0	1.0
van'-drer	0.008		12	1.0	1.0
hø'-re	0.09		12	1.0	1.0
brik'-ke	0.003		12	1.0	1.0
rø'-de	0.06		12	.462	.469
ba'-nan	0.0006		12	.750	.864
grøn'-ne	0.03		12	1.0	1.0
AVERAGE	**0.14**		**12.2**	**.884**	**.903**
PART-TARGET NONWORDS*					
de-van'	0		2	.077	.111
ne-pær'	0		2	.091	.299
som-lev'	0		2	.125	.122
bor'-mell	0		2	.200	.149
de'-ja	0		2	.091	.166
ke'-kan	0		2	.111	.090
mell'-bri	0		2	.111	.139
drer'-man	0		2	.167	.174
kert'-kås	0		2	.167	.406
AVERAGE	**0**		**2**	**.127**	**.184**
TWO-WORD COMBINATIONS*					
	1st word	2nd word			
du-skal'	1.73	3.18	2	.166	.284
ha-med'	1.67	11.28	2	.125	.168
bra'-som	0.28	15.77	2	.500	.0001
vi-vil'	4.22	4.12	2	.111	.139
jeg-har'	3.93	10.57	2	.500	.350
oss-en'	0.78	14.84	2	.500	.102
to'-fram	1.85	0.42	2	1.0	1.0
vil-ha'	4.12	1.67	2	.200	.131
når-du'	1.70	1.73	2	.333	.230
AVERAGE	**2.25**	**7.06**	**2**	**.382**	**.267**

*Words are divided into syllables and the stressed syllable is indicated by the ' symbol. Underlined portions of Part-Target words are syllables contained in the Correct Target words.

[1]Frequency in 1000 words. Based on a corpus of 18.5 million words from three Norwegian newspapers (http://www.tekstlab.uio.no/norsk/frekvensordlister/bokmaal/aviser_og_lover/aviser.frek.html).

[2]$Py|x =$ (freqency of syllable x followed by syllable y)/(frequency of x followed by anything else)

[3]r_{φ} reflects the overall association between two syllables (forward and backwards transitional probability) relative to all other possible combination of syllables.

https://doi.org/10.1017/S1366728910000039 Published online by Cambridge University Press

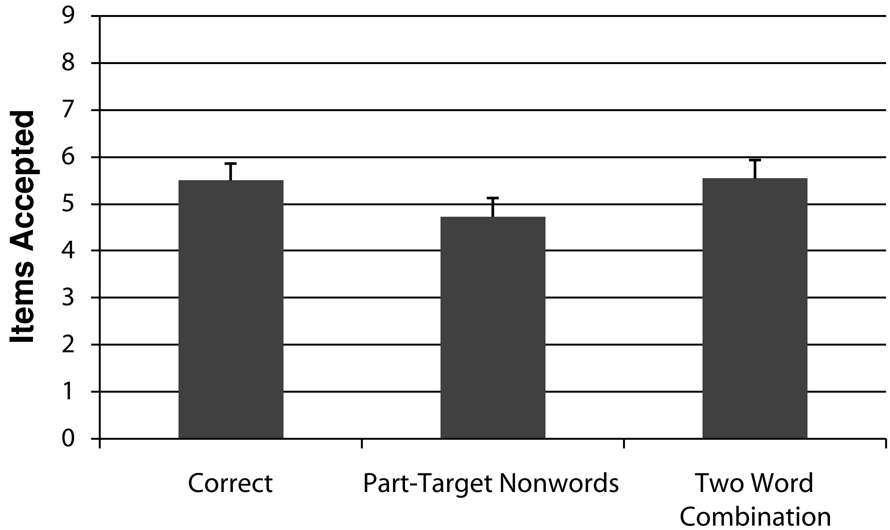

Figure 1. Mean rate of accepted test items as a function of item type when participants were not first familiarized with items embedded in Norwegian sentences. Data shown for the nine-item test set.

range of acceptance rates for individual items, one correct Norwegian word appeared to be an outlier in terms of their overall acceptance rates ("reise", probability of acceptance = .900). The probability of this word was outside the remaining range for all other Correct Target words and nonwords and was more than 2 standard deviations from the average acceptance rate of .515 (SD = 1.68).

Because of the probability that this one particular correct items might be highly guessable, the decision was made to exclude it from the analysis of the participant responses in Study 2. To achieve an equal number of test items across the three item types, for purposes of within-subject statistical analysis, the most guessable words from the remaining incorrect item classes were also selected for exclusion from subsequent analysis.

The performance characteristics of the test words analyzed in Study 2 is displayed in Figure 1. The rate of acceptance of the correct words (mean = 5.5 items accepted per person), Part-Word targets (mean 4.7) and Two-Word Combinations (5.5) did not differ significantly from each other.

Study 2

Methods

Participants

Twenty-two adults (13 female, 9 male) who were foreign students either attending the University of Bergen or working in the University Hospital participated in this study. The participants' ages ranged from 20 to 54 years old (mean age 30 years). The participants originated from 13 different countries and spoke 12 different primary languages. Many participants spoke multiple languages. Twenty-one participants were right-handed and one participant was left-handed. The amount of time the participants had previously spent in Norway ranged from two days to 28 months. Participants were excluded from the study if they rated themselves higher than a 2 on a self-rating scale that reflected how much Norwegian they understood. A "1" on the scale indicated they understood "nothing" and a "10" indicated their understanding was "equal to their primary language". Participants were given a language history survey and were excluded from the study if they indicated that they had a hearing loss, speech impairment, or learning disability.

Participants were recruited for the study with flyers posted in an international University dorm, the Department of Biological and Medical Psychology, the University hospital at the University of Bergen, and public areas around the University. Individuals who spoke Norwegian dialects, Swedish, Danish, Faroese and Icelandic were excluded from the study because these languages are similar to Norwegian. Twenty-nine participants were recruited for the initial testing phase of the study, but only 22 were included in the final data set. The other seven participants were excluded because five of the participants did not return for the second of two experimental sessions used in this study, one participant's session was disrupted and then discontinued, and one participant's data was lost due to an experiment error.

Based on the self-reports of primary language, the participants were divided into three language groups: Germanic (English, German), Indo-European which excluded Germanic languages (Punjabi, Polish, Portuguese, French, Spanish, Bulgarian), and Other Languages not related to the prior groups (Chinese, Arabic, Kiswahili, Amharian). There were eight participants in the Germanic group, nine participants in the Indo-European, and 11 participants in the Other Languages group. Note that there was no significant difference in the participants ratings of their own level of fluency (either a 1 or a 2 on the 10-point rating scale, $\chi^2 = 3.23$, $p > .05$). Germanic is the language group most similar to Norwegian, then the Indo-European group, and finally the Other Languages. The groups were used to look at the impact of similarity between the participants' primary languages and Norwegian on their ability to learn Norwegian.

All participants provided informed consent in accordance with institutional review board requirements of the experimenters' home universities. Participants were paid for their participation.

Materials

The stimuli for the experiment were presented in Norwegian. An initial set of familiarization stimuli consisted of 60 sentences with a total of 374 words. The test items from Study 1 were heard embedded in these sentences. All but one Correct Target word occurred 12 times embedded within the familiarization sentences; one word appeared 14 times (see Table 1 above).

Repeating each Correct Target word within the familiarization set had two consequences that can each increase learning. The repetitions increased the frequency of occurrence of the individual syllables that made up the Correct Target words and also increased the number of times that the syllables that composed the Correct Target words co-occurred relative to the syllables in the remainder of the familiarization set. The degree of syllable co-occurrence can be expressed in various ways. One is by calculating the forward transitional probability, the metric most frequently manipulated in artificial learning studies. This metric ($Py|x$) expresses the likelihood that when a participant heard the first syllable of a test item within the familiarization sentences, this syllable was followed directly by the second syllable of that test. A second metric (r_φ) expresses the overall degree to which the syllables that compose the test items co-occurred within the corpus of words participants heard. This metric accounts for all possible syllable combination in the corpus, not just the transitional probability of the syllables in the test words. There is some evidence that this metric, in comparison to transitional probabilities, may more accurately reflect the information that learners use to identify phonological units (Perruchet & Peereman, 2004). The values for both metrics were higher on average for the Correct Target words relative to the two types of nonword test items.

The Part-Target Nonword items were designed to control for the possibility that the higher frequency of occurrence alone could produce a general sense of recognition and cause participants to accept these items. For Part-Target Nonwords, one of the two syllables that made up these test items was heard as often as the syllables in the Correct Target words. However, the bisyllable combinations that constituted the Part-Target Nonwords co-occurred only twice within the familiarization period. Therefore, the co-occurrence of the syllable combination constituting Part-Target Nonwords was low relative to the Correct Target words. Finally, the two words that composed the Two-Word Combination test items were heard as a combination just twice during familiarization. Therefore this combination of syllables had a low frequency of occurrence. However, their transitional probability and r_φ values fell between those of the Correct Target and Part-Target Nonwords. This was a logical consequence of the higher co-occurrence rates for Correct Target words relative to other words in the sentences. In comparison to Correct Target words, other content words appeared either just once or very infrequently within the stimulus sentences. Therefore, the low frequency of occurrence tended to elevate the co-occurrence rate of these words with their neighbors.

The familiarization stimuli were read by a male native Norwegian speaker and recorded with Sound Forge 7.0 (Sony Pictures Digital Inc. 2003). The recording was a total of 3.61 minutes and was clipped into 12 blocks, each lasting approximately 18 seconds and containing five sentences. The full familiarization set was played twice over the course of the experiment, totaling 7.2 minutes. Therefore each of the sentences was heard twice during the familiarization period. The duration of this period was based on prior pilot testing that showed that a shorter duration resulted in poor task performance.

The initial familiarization period was followed immediately by a test phase. The test phase consisted of the 30 bisyllabic test items used in Study 1. The test phase contained no novel items in that all of the syllables used in the test sessions co-occurred in the familiarization stimulus set. The 30 test items were randomly ordered and then the auditory file containing the test items was clipped into three blocks of 10 items each. Each of the blocks contained each of the three test item types. The participants received the three blocks in a fixed order and as an uninterrupted set of 30 test items in the test phase. These blocks were created to test for the possibility that exposure to the test items, particularly the nonword test items, might erode the mental representation of Correct Target words over the test period. Therefore, the use of three blocks allowed us to test for changes in overall accuracy across the three blocks.

Apparatus

The experiment was conducted either in an enclosed room at an international dorm ($n = 18$) or in a sound booth ($n = 4$) at the University of Bergen. The choice between testing locations was determined by whichever offered more convenience to the participants. The experiment

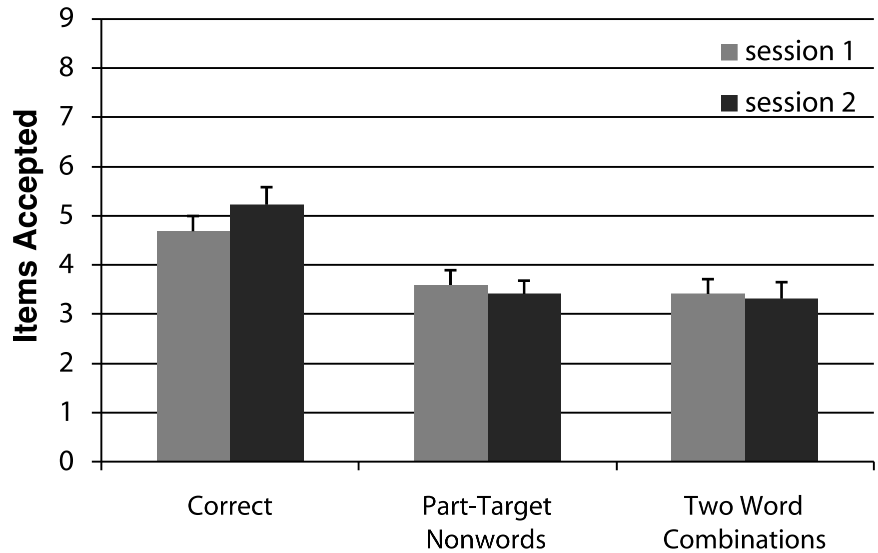

Figure 2. Mean rate of accepted test items as a function of item type when participants were first familiarized with items embedded in Norwegian sentences. Participants were tested over two sessions.

was administered via laptop computer to one individual at a time. The instructions, familiarization stimuli, and test items were all presented using E-Prime Version 1.0 software (Copyright 2002 Psychology Software Tools, Inc.). Keyboard responses to the test items were recorded by E-Prime.

Procedures

The instructions for the first experiment were presented both in print and in auditory modalities in English, a language spoken by all participants. The participants were instructed that they would first hear several sentences in Norwegian and then they would hear a series of words. They were informed that their task was to indicate whether the word was a correct Norwegian word by pressing the "yes" key or an incorrect Norwegian word by pressing the "no" key. The participants were not told that their task would be to determine where words began and ended within the stimulus set. During the familiarization phase, participants advanced the experiment at their own pace by pressing a button after each familiarization block. Following this period of familiarization, reminder instructions about the participants' task were presented both in print and in auditory modalities immediately before the testing block began. The designation (left- or right-hand) of the yes/no buttons was randomized across participants. No practice trials were given prior to the first test item.

The experiment consisted of two sessions. The second session was held an average three weeks after the first test session with a range between five and 38 days. The experimental content was the same for both test sessions. The presentation order of the familiarization set remained the same and the presentation order for the test items were randomized as discussed above. The same participants participated in both testing sessions. The purpose of the second test session was to test the effect that ambient exposure familiarization to Norwegian had on our participants' learning.

Results

There was no evidence that subjects' responses changed significantly across the three blocks of test items. Therefore, responses across all three blocks were combined for further statistical analysis. Recall that there were 10 responses to each item type Correct Target words, Part-Target Nonwords and Two-Word Combinations. However, given the outcome of Study 1 (above), we analyzed only the responses to the nine items of each type shown in Table 1. All statistics were based on the participants' rate of acceptance of these items. Therefore "word learning" is demonstrated by a high acceptance rate for actual Norwegian words accompanied by a low rate of acceptance of the two types of nonwords.

The mean rate of accepted items for the three item types for each of the two testing times are presented in Figure 2. One-tailed t-tests were used to determine whether the performance on these items deviated in the expected direction from chance. The mean acceptance rate for the correct items was 4.68 and 5.23 for the first and second experimental sessions, respectively. Although the first of these acceptance rates was not significantly above chance ($t(1,21) = 0.58$, $p = .28$, $d = 0.12$), by the second session acceptance rates for correct items were above chance ($t(1,21) = 2.06$, $p = .026$, $d = 0.44$). Acceptance

https://doi.org/10.1017/S1366728910000039 Published online by Cambridge University Press

rates of all incorrect items were significantly below chance at both test sessions. The mean acceptance rate for the incorrect Part-Target items was 3.59 in the first session (t(1,21) = 3.04, p = .003, d = 0.79) and 3.41 on second testing (t(1,21) = 4.06, p = .0003, d = 0.87). The mean score of the incorrect Two-Word Combination items was 3.41 and 3.32 on the first session (t(1,21) = 3.65, p = .0006, d = 0.78) and the second test session (t(1,21) = 3.56, p = .0006, d = 0.76).

The relative pattern of responses across test item types was formally analyzed with a repeated measures ANOVA, with time (test session 1 vs. test session 2) and item type (correct, Part-Target incorrect items, Two-Word Combination incorrect item types) as within-participant factors. The results indicated that there was a significant effect of item type (F(2,38) = 4.48, p = .018, $\eta_p^2 = 0.19$) on the participants' rate of acceptance. In addition, there was a significant interaction effect for item type × time (F(2,38) = 8.76, p = .0007, $\eta_p^2 = 0.32$). No other main effects or interactions were statistically significant.

Planned comparisons concerning acceptance rates of correct vs. each of the incorrect item types were conducted within each test session. One-tailed tests were used because learning would be reflected by a higher rate of correct accepts than incorrect accepts. There was a significant difference in the rate of acceptance for Correct Target items and Part-Target Nonwords during the first test session (t(1,21) = 2.26, p = .016, d = 0.75) and the second test session (t(1,21) = 3.49, p = .001, d = 0.87). A similar pattern was found for the difference in acceptance rates for Correct Target items and Two-Word Combination items during session 1 (t(1,21) = 2.70, p = .005, d = 1.10) and session 2 (t(1,21) = 3.19, p = .002, d = 1.15). The difference in acceptance rates for correct items from session 1 to session 2 was marginal (t(1,21) = 1.57, p = .05, d = 0.33).

As Table 1 indicates, the incorrect test items contained exemplars that either conformed or diverged from the trochaic (strong–weak) stress pattern that is typical of Norwegian words. Prosodic cues might used by listeners to determine where words occur within an acoustic stream. We sought to determine whether participants were sensitive to word-level stress cues within the incorrect test items. This was investigated by recombining the incorrect items by syllable stress pattern. This gave us two new groups of incorrect test items that reflected strong–weak (SW) and weak–trong (WS) stress patterns. There were 8 test items for each of the two sessions that represented a SW stress pattern for a total of 16 SW items. There were 10 test items for each of the two sessions that presented a WS stress pattern for a total of 20 WS items. Because the two conditions did not have an equal number of items, they were converted to a proportion by dividing by the total number of test items in each of the respective groups. The statistics were calculated according to how often the participants accepted a test item as correct. Because only the incorrect item types were evaluated, a low score indicates high accuracy. The acceptance rate for both words were similar (37% for SW and 39% for WS). Both these rates were significantly below chance (50%) (iambic stress: t(1,21) = 5.10, p = .00005, one-tailed, d = 1.08; trochaic stress t(1,21) = 4.24, p = .0004, one-tailed, d = 0.92). The difference in response to words of each stress pattern was also compared in the subjects' respective language groups (Germanic, Indo-European without Germanic, and other Language groups). None of the three language groups were significantly different from chance for words of either stress pattern. This means that the difference in participants' performance between test conditions was not dependent on the differences between their language groups.

The participants' primary language did not have a significant effect on their performance either between item types or test sessions in the correct, incorrect Part-Target, and incorrect Two-Word Combination item types. The ANOVA main effect for group did not reach statistical significance (F(2,19) = 3.08, p = .069, η_p^2 = 0.24). Likewise the interactions of test session and group (F(2,19) = 0.15, p = .86, $\eta_p^2 = 0.02$) or test session, item type and group (F(4,38) = 0.43, p = .79, $\eta_p^2 = 0.04$) were not significant. Therefore, the participants' language group, which was based on their primary language background, did not significantly influence performance.

The possibility that other demographic variables might have influenced the participants' performance was also examined. Specifically, the number of languages the participants spoke, the amount of time they had spent in Norway, their perceived fluency in Norwegian, and the ordinal scale value reflecting the relationship among the participant's primary languages in terms of their distance from the Norwegian language was explored. A Pearson product moment correlation between time in Norway at the time of study and performance failed to produce any significant correlations for Correct Target words (r(22) = −.22), Part-Target Nonwords (r(22) = .25) or Two-Word Combinations (r(22) = .27). A point-biserial correlation indicated that the participant's perceived fluency did not correlate with performance on any of the item types (Correct Target r(22) = −.37; Part-Target Nonwords r(22) = −.02; Two-Word Combinations r(22) = .05). In addition, a Spearman Rank Order correlation failed to demonstrate a significant relation between performance and the participants' primary languages (Correct Target rho(22) = .04; Part-Target Nonwords rho(22) = .09; Two-Word Combinations rho(22) = .23). The only statistically significant correlation (Pearson product moment) that occurred was between the number of languages spoken and acceptance of Two-Word Combination items (r(22) = −.55). However, this result was not corroborated by similar correlations between the other items types across test sessions or within test sessions. Therefore, there is little support for interpreting this finding as a true relationship rather than a chance product of multiple comparisons.

The lack of a significant influence from these demographic variables supports the experiment's validity because it reinforces the interpretation that the participants' performance was due to their ability to use information inherent to the Norwegian stimuli rather than being associated with individual differences in native language background and ambient experience.

Discussion

The results indicate that adults can differentiate between actual Norwegian words and nonwords after a brief period of familiarization to spoken sentences that contained the target words. This ability was due primarily to the participants' ability to correctly reject nonwords, whereas their ability to positively identify real words was at chance levels after the first session and just above chance after the second. In other words, participants were more sure of when a test item was NOT a word than of when it truly WAS a word. This pattern has also been seen for children in the initial stages of word learning (Alt, Creusere & Plante, 2004; Alt & Plante, 2006). This pattern, along with the modest effect sizes, suggests that our adult learners were still at an initial level of word learning after the brief period of exposure they received. In contrast, the group of adults in Study 1, who did not hear the items embedded in a language context, did not differentiate between correct and incorrect words.

When participants were divided into their respective language groups (Germanic, Indo-European without Germanic, and Other Languages) no significant effects were found among groups either within test item types or across testing sessions. Additionally, demographic factors like the participants' primary language, the number of languages they spoke, or the amount of time they had previously spent in Norway and their perceived fluency in Norwegian did not have a significant effect on their performance.

The performance of our participants is consistent with the finding of Saffran et al. (1996b). Adults in that study differentiated between target words and nonwords or part-words. Those adults accepted the correct items more often than they accepted the incorrect items. This is also a skill infants demonstrate. Saffran et al. (1996a) showed that infants can segment continuous speech with only two minutes of familiarization. Aslin et al. (1998) also demonstrated that eight-month-old infants could discriminate words from part-words based on the syllable sequences' transitional probabilities. These findings demonstrate that humans are sensitive to the distributional properties of the input from a young age. The literature shows that both adults and infants are successful at segmenting words from fluent speech in an artificial language.

The findings in the current study support the idea that adults are also capable of segmenting words from fluent speech in a natural language that contains multiple cues to word structure. However, it is important to note that the real language context used in the present study presented a considerably more complex set of stimuli to the listener than is typically used in artificial language studies. Artificial language studies typically use a very limited syllable set which repeats frequently; our stimuli used sentences comprised of a large number of different syllables. Therefore, our stimuli were much more varied than those typically encountered in artificial learning contexts.

One consequence of using a natural instead of an artificial language is that a number of cues were present in the stimuli in addition to the distributional information. However, direct examination of these failed to support a significant use of cues other than distributional information by the participants. Possible additional cues included the frequency of occurrence of syllables in the input. In order to manipulate distributional cues in natural language, it was necessary to also manipulate the frequency with which the syllables that comprise real words co-occurred. This was done by raising the frequency of these words relative to others in the input. This was an expected limitation of using a natural language. To compensate for this reality, we composed nonword test items that contained a syllable from the frequently-presented test words. Therefore, if listeners were simply influenced by frequency of presentation, they should have been attracted by the Part-Target Nonwords. However, not only were these test items accepted significantly less often than the correct Norwegian words, they were selected at rates similar to the Two-Word Combination test items, for which each syllable had a low frequency of occurrence in the input. This suggests that factors other than syllable frequency played a more important role than syllable frequency in participants' responses. This is consistent with a recent study of infants that demonstrated that word frequency alone was not the basis of word segmentation (Pelucchi et al., 2009).

A second potential cue was syllable stress. Stress patterns can be either eliminated or controlled in artificial languages but co-occur with distributional cues in natural languages. There is evidence that listeners are sensitive to stress patterns when performing word segmentation tasks. Saffran et al. (1996b) conducted a word segmentation experiment both with and without prosodic cues (specifically syllable lengthening). The study demonstrated that in all three test conditions (no prosodic cue, initial-syllable lengthening, and final-syllable lengthening) participants performed better than chance. However, in the final lengthening condition, participants performed significantly better than in either the no prosodic-cue or the initial lengthening conditions. This result indicates that listeners were capable of learning words using information concerning transitional probabilities alone (as in the no prosodic-cue condition), but listeners are further aided by the addition of some types of prosodic cues. Furthermore, it was reasonable to believe that those exposed to linguistically similar languages, like English and German (trochaic stress), would still be aided by the similarities to Norwegian in terms of word stress. Such effects have been demonstrated in studies of infants who prefer the prosodic patterns of their native language (Echols, Crowhurst & Childers 1997; Jusczyk et al., 1999). Therefore, we considered the possibility that participants would be predisposed to rely on familiar stress cues to identify words.

In the present study, the impact of prosodic cues in the current study was explored by investigating the influence of syllable stress on participants' performance. We were unable to find evidence that participants were using this particular prosodic cue in isolation to aid performance. Instead, they were as likely to accept incorrect items that were either consistent or inconsistent with the dominant trochaic pattern of Norwegian. This outcome is consistent with recent studies that show that adults can be resistant to using syllable-level stress cues in an artificial language learning context. In two studies that required listeners to generalize rules of stress assignment in an artificial language, adults only showed learning when word lengths were unusually long (Guest, Dell & Cole, 2000) or when the intensity of stressed items were artificially enhanced (Bahl, Plante & Gerken, 2009). In contrast, infants were readily able to learn stress assignment without such additional manipulations (Gerken, 2004). These studies, as a whole, suggest that adults may not rely on the types of intensity variation that differentiates trochaic and iambic stress patterns as their default mechanism for learning. However, the convergence of phonological and stress cues can aid both adults and infants in segmenting words (Curtain et al., 2005; Saffran et al., 1996b). In the present study, the two syllables that made up each of the incorrect nonwords in this study were acoustically identical to those heard during familiarization. However, the relative stress on a syllable as it was perceived in a Target word was sometimes different than when it was paired with alternate syllables for incorrect test items. This may have facilitated the differential response to Target and incorrect item types (cf. Curtain et al., 2005).

There is reason to believe that attention to stress cues may not be a primary strategy for adult listeners. First, not all languages provide reliable stress cues to word boundaries. For example, only 211 out of the 444 languages surveyed in a study by Hyman (1977) had reliable stress on either the word-initial syllable or the word-final syllable. Because stress is not a reliable cue in all languages, and word learning is necessary to all languages, relying on syllable stress cannot be the only strategy to word segmentation. Moreover, prior knowledge of word boundaries might facilitate recognition of stress cues. Identifying syllables as belonging to individual words would assist in appreciating whether stressed elements were cuing the beginning, middle or end of a word (Saffran et al., 1996b). Therefore, it may be that the learners in the present study had not yet reached a stage in their ability to segment Norwegian words when stress cues could be functional for them.

Figure 2 suggests only a modest improvement in the ability to differentiate between correct and incorrect test items from session 1 to session 2. This pattern was confirmed statistically with a significant item $\times$ time interaction effect, although the overall effect was small. This marginal improvement could be due to the ambient exposure to Norwegian that participants received as a byproduct of living in Norway during the weeks between the first and second testing sessions. However, there was no correlation between the amount of time participants spent in Norway prior to the first testing session and their initial performance. This result suggests that extra ambient exposure to Norwegian may not have provided a performance advantage for our participants. There was no correlation between the length of time in Norway and performance on this task. It is possible that we selected some individuals whose natural propensity for learning languages was not strong, leading them to report little or no knowledge of Norwegian even when having resided in the country for some months. However, it should also be noted that the frequency and quality of ambient exposure participants received varied from person to person. A limitation of the naturalistic approach used in this study is that the amount of exposure a participant received (based on their length of stay Norway) did not necessarily predict the frequency and quality of ambient exposure they actually received. Moreover, the density of target word presentations within the study may be relatively high relative to densities that happen under many conditions when listeners are exposed to ambient language. However, it is also possible that specific forms of input (i.e., input designed to highlight the co-occurrence of particular syllables) is necessary for individuals to segment specific words from running speech. Future research is necessary to resolve this issue.

Past research has demonstrated the sensitivity of both adults and infants to word boundaries in a stream of continuous speech within the context of an artificial language. The adults' performance in the current study extends the validity of these findings to natural language conditions. Importantly, the results demonstrate that it is not necessary for the input to the listener to be highly constrained in terms of the phonology, syllable length, or the artificial isolation of distributional cues relative to other potential cues in order for word segmentation to occur. Overall, participants representing different native language backgrounds appeared to be performing similarly on this task, despite differences based on their respective language groups, length of time in Norway, and other demographic factors. This suggests that regardless of such differences in background, participants may be using a common strategy to segment words from continuous speech.

References

Alt, M., Creusere, M., & Plante, E., (2004). Semantic features in fast-mapping performance of preschoolers with specific language impairment versus preschoolers with normal language. *Journal of Speech, Language, and Hearing Research*, 47, 407–420.

Alt, M., & Plante, E. (2006). Factors that influence lexical and semantic fast-mapping of young children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 49, 941–954.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9, 321–324.

Bahl, M., Plante, E., & Gerken, L. A. (2009). Processing prosodic structure by adults with language-based learning disability. *Journal of Communication Disorders*, 42, 313–323.

Cole, R. A., & Jakimik, J. (1980). Segmenting speech into words. *Journal of the Acoustical Society of America*, 64, 1323–1332.

Curtin, S., Mintz, T. H., & Christiansen, M. H. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition*, 96, 233–262.

Dahan, D., & Brent, M. R. (1999). On the discovery of novel wordlike units from utterances: An artificial-language study with implications for native-language acquisition. *Journal of Experimental Psychology: General*, 128, 165–185.

Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language*, 36, 202–225.

Echols, C. H., & Newport, E. L. (1992). The role of stress and position in determining first words. *Language Acquisition*, 2, 189–220.

Gerken, L. A. (2004). Nine-month-olds extract structural principles required for natural language. *Cognition*, 93, B89–B96.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Pscyhological Science*, 13, 431–436.

Guest, D. J., Dell, G. S., & Cole, J. S. (2000). Violable constraints in language production: Testing the transitivity assumption of Optimality Theory. *Journal of Memory and Language*, 42, 272–299.

Harris, Z. S. (1955). From phoneme to morpheme. *Language*, 31, 190–222.

Hyman, L. M. (1977). On the nature of linguistic stress. In L. M. Hyman (ed.), *Studies in stress and accent* (Southern California Occasional Papers in Linguistics 4), pp. 37–82. Los Angeles, CA: University of Southern California.

Jusczyk, P. W. (1999). How infants begin to extract words from fluent speech. *Trends in Cognitive Science*, 3, 323–328.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology*, 29, 1–23.

Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development*, 64, 675–687.

Jusczyk, P. W., Houston, D., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39, 159–207.

Lany, J., Gómez, R. L., & Gerken, L. A. (2007). The role of prior experience in language acquisition. *Cognitive Science*, 31, 481–507.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, 7, 703–704.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition*, 30, 678–686.

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Pscyhology: General*, 130, 401–426.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, 80, 674–685.

Perruchet, P., & Peereman, R. (2004). The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics*, 17 (2), 97–119.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606–621.

Saffran, J., Newport, E. L., Aslin, R. N., Tunick, R., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, 8 (2), 101–105.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, 39, 706–716.

Thompson, S., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, 3 (1), 1–42.