Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-022-02190-1

BRIEF REPORT

Learning words without trying: Daily second language podcasts support word-form learning in adults

Elise Alexander[1] · Stephen C. Van Hedger[1,2] · Laura J. Batterink[1]

Accepted: 12 September 2022

Abstract

Spoken language contains overlapping patterns across different levels, from syllables to words to phrases. The discovery of these structures may be partially supported by statistical learning (SL), the unguided, automatic extraction of regularities from the environment through passive exposure. SL supports word learning in artificial language experiments, but few studies have examined whether it scales up to support natural language learning in adult second language learners. Here, adult English speakers ($n = 70$) listened to daily podcasts in either Italian or English for 2 weeks while going about their normal routines. To measure word knowledge, participants provided familiarity ratings of Italian words and nonwords both before and after the listening period. Critically, compared with English controls, Italian listeners significantly improved in their ability to discriminate Italian words and nonwords. These results suggest that unguided exposure to natural, foreign language speech supports the extraction of relevant word features and the development of nascent word forms. At a theoretical level, these findings indicate that SL may effectively scale up to support real-world language acquisition. These results also have important practical implications, suggesting that adult learners may be able to acquire relevant speech patterns and initial word forms simply by listening to the language. This form of learning can occur without explicit effort, formal instruction or focused study.

Keywords Statistical learning · Implicit learning · Second language learning · Spoken word recognition

Introduction

Imagine that you have travelled to a faraway country and are listening to locals converse in a completely unknown language. Beyond your lack of comprehension of the words themselves, many other aspects of the speech signal may also be unfamiliar. For example, the individual speech sounds may not fall into a phoneme category that you recognize and may be combined into sequences that sound different from your own language. Similarly, the word stress patterns and other rhythmical properties of the language may also sound unfamiliar. Now imagine that you have spent several weeks in this country, frequently overhearing local conversations. At this point, you may start to feel more accustomed to the general speech patterns and rhythms of the language, and you may even be able to recognize a few frequent words, even if you are do not understand their meanings.

The ability to extract linguistic patterns in speech may be supported by a mechanism known as *statistical learning* (SL). SL refers to the process of discovering underlying structure in the environment from repeated exposure to environmental statistics, without external reinforcement, feedback, instruction, or conscious attempts to learn. SL was initially discovered in the context of word segmentation and defined relatively narrowly as the computational process of tracking of syllable patterns to support segmentation (Saffran, Aslin, & Newport, 1996a; Saffran, Newport, & Aslin, 1996b). However, subsequent decades of research have found that SL operates across many other learning contexts, while showing important qualitative differences as a function of sensory modality, type of stimulus material (e.g., linguistic versus nonlinguistic items), type of regularity (e.g., spatial versus temporal; adjacent versus nonadjacent), and other factors (e.g., Arciuli, 2017; Conway & Christiansen, 2005; Frost et al., 2015; Raviv & Arnon, 2018; Siegelman, 2020;

✉ Laura J. Batterink
lbatter@uwo.ca

[1] Department of Psychology, Western Institute for Neuroscience, Western University, London, ON, Canada

[2] Department of Psychology, Huron University College, London, ON, Canada

Published online: 29 September 2022

Siegelman et al., 2017; Siegelman et al., 2018; Siegelman & Frost, 2015). Thus, SL is now taken to refer to a suite of perceptual, cognitive, and linguistic behaviours, rather than to a single computation. In the current study, we consider SL in the context of spoken language learning—specifically, as the ability to acquire word form knowledge through unguided, incidental processes. We operationally define SL as learners' ability to extract characteristic features of spoken words in a novel language, merely through passive, repeated exposure to input, without explicit instruction, feedback, social reinforcement, intention or effort. Importantly, this operational definition of SL contrasts with other, more deliberate forms of learning that may also support the acquisition of initial word forms, such as explicit vocabulary instruction or the intentional application of conscious knowledge (e.g., "when there are two or more of something, add /s/ to the end of a word"; Plante & Gómez, 2018).

Words in continuous speech are cued by overlapping statistical regularities across many levels, representing a rich potential target for SL. While some of these cues are universal, operating in similar ways across different languages, others are specific to a given language and thus must be learned through experience (e.g., Sahni et al., 2010). For example, one type of universal pattern is the sequential structure across neighbouring syllables: regardless of language, syllables that co-occur more frequently, or have higher *transitional probabilities* (TPs), are more likely to belong to the same word (e.g., Swingley, 2005). This type of cue can thus be leveraged by all language learners, regardless of which particular language they are learning. Another universal cue is prosodic utterance and phrasal boundaries, which necessarily align with word boundaries (Sohail & Johnson, 2016). In contrast, other patterns, such as phonotactics and lexical stress, are language specific. For example, in English, words usually follow a strong–weak stress pattern (e.g., KIT–ten), while in other languages it is more common for words follow a weak-strong stress pattern (Cutler & Carter, 1987). It is thought that universal cues, such as TPs and prosodic information, may allow learners to bootstrap language-specific patterns, and together support the identification of words in continuous speech (e.g., Sahni et al., 2010; Sohail & Johnson, 2016; Swingley, 2005; Thiessen & Saffran, 2007). For example, infants (Sahni et al., 2010) and adults (Benitez & Saffran, 2021) can use TPs to discover an overlapping, novel phonetic cue to word boundaries. Similarly, words that are presented in isolation may provide subsequent opportunities for "anchoring," in which the presence of a known word helps to segment adjacent unknown words from fluent speech (Bortfeld et al., 2005; Cunillera et al., 2010). Connectionist models suggest that the presence of overlapping linguistic regularities across multiple levels—far from presenting a complex learning problem that is impossible to solve (Yang, 2004)—may actually be helpful for the learning system (Seidenberg, 1997), reinforcing SL across levels, and supporting further learning (Romberg & Saffran, 2010). SL is thus a potentially powerful mechanism for discovering relevant patterns in natural speech based on mere exposure.

Although SL is present across the life span (e.g., Choi et al., 2020; Palmer et al., 2018; Saffran et al., 1997), the extent to which SL mechanisms actually benefit real-world adult learners of a second language is not yet clear. Adult second language learners typically rely on explicit instruction or intentional study approaches for initial word learning, such as flashcards or word lists (Gu & Johnson, 1996; Rodríguez & Sadowki, 2000; van Hell & Mahn, 1997; Webb et al., 2020). Even when learning occurs outside the classroom or intentional study settings, as in the case of immersion-based learning, adult learners may not necessarily rely on SL for acquiring word forms. For example, by interacting with native speakers, immersion-based learners are afforded opportunities to ask questions, to link specific labels to objects in the environment, to receive feedback on their productions, and to benefit from other forms of intentional and/or reinforced learning. Further, several prominent models in the second language acquisition field have argued that implicit learning play little to no role in adult L2 acquisition (Robinson, 1995; Schmidt, 1990). Thus, it is possible that SL—operationalized here as the learning of relevant patterns and word forms through passive listening to spoken language, divorced from real-world social interactions—results in no discernible learning benefit for adult second language learners. On the other hand, as highlighted above, laboratory experiments highlight the potentially powerful nature of SL in uncovering a range of different types of linguistic patterns simply through passive exposure to language input. The question of whether adult learners might benefit from background, noninteractive exposure to second language speech—for example, through radio, TV, or podcasts—has been informally debated in the second language learning field, but there is limited evidence addressing this question (Frank et al., 2013).

Determining whether SL can effectively scale up to support learning of natural languages is of key interest both theoretically and practically (Erickson & Thiessen, 2015; Frost et al., 2019; Siegelman, 2020). However, much of the evidence for a causal role of SL in word-form learning comes from lab experiments using highly simplified miniature languages, often containing just four or six nonsense words (e.g., Batterink & Paller, 2017; Batterink et al., 2015; Saffran et al., 1997; Saffran, Newport, & Aslin, 1996b; Saffran & Thiessen, 2003; Thiessen & Saffran, 2003, 2007; see Frost et al., 2019, for a review). While several studies have found positive evidence of SL using more complex designs incorporating natural language stimuli (Hay et al., 2011; Kittleson et al., 2010; Pelucchi et al., 2009; Plante et al., 2015), even these studies presented learners with dramatically

fewer words than those found in a complete natural language. Another highly relevant study presented four adult participants with an artificial language made up of 1,000 words over multiple days (Frank et al., 2013). Interestingly, all participants showed positive evidence of segmentation after training, providing promising evidence that SL may scale up to support word-form learning of large-scale lexicons. Nonetheless, the language was still highly artificial, miniature, and monotonized, as well as intentionally devoid of other cues to word boundaries such as prosody. Thus, the "scalability" of SL to fully natural language—particularly with respect to adult second language acquisition—remains an open question, with many researchers calling for more ecologically valid research designs that better resemble the challenges of real-world language acquisition (Erickson & Thiessen, 2015; Frost et al., 2019; Siegelman, 2020).

To shed light on this issue, we tested whether unguided, passive exposure to fully natural L2 input supports initial word-form learning in adult learners. Native English speakers with no prior knowledge of Italian were randomly assigned to listen to either Italian podcasts (L2 exposure group) or English podcasts (control group) for 1 hour a day over a 2-week period. Podcasts were played "in the background" while participants went about their daily activities, such as cooking, exercising, or commuting. We measured participants' Italian word knowledge by asking them to provide familiarity ratings for true words in Italian and nonword foils (Italian-like words that do not truly exist in Italian), both before and after the 2-week period. Our key hypothesis was that unguided exposure to natural Italian speech would allow learners to extract key regularities governing Italian words, as supported by SL mechanisms. Thus, we predicted that the L2 exposure group, but not the control group, should improve in distinguishing true Italian words and nonwords over the 2-week listening period. Such a finding would provide evidence that adult learners can benefit from mere background exposure to natural L2 speech, which may harness SL mechanisms, supporting the discovery of relevant patterns and word forms.

Methods

Participants

A total of 71 young adults (18 male, 53 female) ranging in age from 18–35 years ($M = 21.42$, $SD = 3.14$) completed the experimental protocol. One participant was subsequently excluded as it was discovered at the posttest session that the individual had accidentally listened to half of the podcasts at 1.5 speed, resulting in a total of 70 participants. An additional six individuals completed or partially completed just the first testing session but were not invited to participate in the remainder of the study due to either loss of data resulting from technical issues, or noncompliance with the experimental procedure. All participants were native English speakers with no previous exposure to Italian. In addition, participants did not have significant previous classroom or other experience with any other Romance language, including French, Spanish, Romanian, or Portuguese. Significant classroom experience was defined as having taken more than one Romance language class per year during elementary or secondary school, and/or any postsecondary Romance language courses. All participants reported normal vision and hearing, and no history of learning, hearing, or neurological disorders. All participants completed the protocol sometime between January and July 2021. Participants were not aware of the research hypothesis until completion of the study procedures. All research procedures were approved by Western University's Research Ethics Board.

Participants were randomly assigned to either the control (English listeners; $n = 35$) or experimental (Italian listeners; $n = 35$) group. A sensitivity power analysis conducted using G*Power suggested that an independent-samples t test with 35 participants per group ($N = 70$) would be sensitive to effects of Cohen's $d = 0.60$ with 80% power (alpha = .05, one-tailed). We used a one-tailed test for this analysis as we had specific directional hypotheses. This means the study would not be able to reliably detect effects smaller than Cohen's $d = 0.60$.

Stimuli

L2 exposure stimuli: Podcasts and podcast questionnaires

The L2 exposure stimuli consisted of 14 1-hour Italian podcasts and 14 1-hour English podcasts. The Italian podcasts consisted of select episodes from *Radio Feltrinelli* and *Parole di Storie*. *Radio Feltrinelli* is an Italian news podcast that covers politics and current affairs in Italy, and that includes interviews with various individuals. *Parole di Storie* is a podcast geared towards children, teens, and families, in which the host reads popular stories by classic and contemporary Italian authors. We selected more than one podcast source in order to expose participants to a variety of Italian speakers, narrative styles, and topics. The English podcasts were selected from *Radio Lab* and *Getting Things Done*. *Radio Lab* is a science-focused podcast in which the hosts discuss a variety of science-related questions and stories, as well as related ethical and moral issues. *Getting Things Done* is a motivational podcast that aims to enhance listeners' productivity through various methods and tips, and also features interviews with different individuals who use the methods. All podcasts were available for free download on Apple Podcasts.

For the purposes of the current study, each daily podcast was created by concatenating several podcast episodes from the same show into a single audio file that had a total duration of 1 hour. Each audio file was then further edited by inserting three to six different "secret" English words, each preceded by a chime, at randomly selected time points within the audio. A different set of secret words were used for each of the 14 podcasts within a given language but were the same across the English and Italian podcasts (i.e., the first Italian and English podcast had the same secret words). Participants were required to report on the secret words embedded in each podcast on each day, allowing us to track participant compliance with the daily listening protocol (see Supplementary Materials for more information).

Test stimuli

Exposure + word-detection task The stimuli for this task consisted of two counterbalanced auditorily presented sentence sets (A and B), each containing 300 different spoken Italian sentences. Of the 300 sentences within each set, 250 were designated as "training" sentences. Training sentences collectively contained a total of 10 different "trained" words, each presented a total of 50 times (500 total word presentations). Each training sentence contained between one and three different trained words. All trained words were common trisyllabic Italian nouns that started with a consonant letter (e.g., *galera* [jail], *bambola* [doll], *ragazzo* [boy]). The trained words were embedded within the sentence, never occurring at the beginning or end of a sentence, in order to examine whether participants were able to identify the target words when embedded in continuous speech.

In addition to the 250 training sentences, there were a total of 50 "target" sentences, each containing one of five different "target" words. The target words were also trisyllabic Italian nouns, and also never appeared in the first or last position of a sentence. Each target word was presented 10 times in total. Each target sentence contained only a single target word, and never contained a trained word. Likewise, target words never appeared within a training sentence. Participants responded to the target words during the task (see Procedure).

The 250 training sentences and 50 target sentences were divided into five equal blocks, such that each block contained 50 training sentences and 10 target sentences. A different target word was assigned for each block. Training sentences were randomized across blocks, and block order was randomized for each participant. Within each block, training sentences and target sentences were presented in random order. All training words and target words were unique to each sentence set (A versus B). Individual sentence duration ranged from 2,114 to 9,986 ms (mean = 5,392 ms). The duration of target sentences specifically ranged between 2,114 and 5,954 ms (mean = 3,645 ms). All sentences were separated by a 1-second pause. The onset of a target word and the onset of a new sentence was always greater than 2,000 ms, such that any response considered to be a target "hit" (occurring within 200–2,000 ms of the target word) would not occur during the subsequent sentence.

All sentences were recorded in a male Italian voice using the open access text-to-speech software TTSAutomate, a tool that is designed to produce natural sounding speech from text using a variety of TTS engines. For the present experiment, we used the Google TTS engine, selecting the male Italian (it-IT) voice. The selected voice is designed to be as natural-sounding as possible, with natural articulation, intonation and prosody. As in natural speech, there were no reliable pauses between each word. A native Italian speaker listened to all recorded sentences to confirm that the word pronunciation was acceptable.

Familiarity-rating task The stimuli for this task consisted of two sets (A/B) of 30 trisyllabic test words (60 total words). Each word set contained the 10 trained words from the corresponding sentence set from the word-detection task (i.e., real Italian words that had been previously presented 50 times), 10 "nontrained" Italian words (i.e., real Italian words that had not been previously presented in the word-detection task), and 10 "foil" words (i.e., words that sounded Italian-like but were not real Italian words). Similar to the target Italian words, the nontrained Italian words were also trisyllabic nouns that began with consonants (e.g., *passero* [sparrow], *barchetta* [boat], *tappeto* [carpet]). The foil words were trisyllabic nonsense words that also began with consonants and were formed by recombining common Italian syllables to create Italian-sounding foil words. All foil words were reviewed by a native Italian speaker to ensure they contained syllable combinations that were valid in Italian, while still not being too phonetically similar to any real Italian word. In addition, the initial consonant sounds were closely matched across the three word categories.

As described in the Introduction, transitional probabilities (TPs) between neighbouring syllables represent one possible statistical cue by which learners may discover word forms in spoken language. Thus, to capture TPs as a potential learning cue in our experimental materials, we ensured that the TPs of neighbouring syllables within true Italian words (both trained and nontrained words) were higher than those within the foil items. TPs of test words computed on the basis of the 14 Italian podcasts (see Supplementary Materials for more detail). In addition, trained and nontrained words were well-matched in terms of inherent or baseline linguistic familiarity, as assessed by a separate group of control participants (n = 124; see Supplementary Materials). Finally, for trained and nontrained test words, we also quantified the number of presentations of each word throughout

the 14 Italian podcasts (again, see Supplementary Materials). All words were recorded in TTSAutomate using the male Italian voice. A native Italian speaker listened to all words to confirm that word pronunciation was acceptable.

Procedure

A summary of the procedure is shown in Fig. 1A. The procedure involved two testing sessions separated by a 2-week listening period, during which participants listened to a different one hour podcast each day. Both testing sessions were conducted virtually over Zoom. Participants were asked to use headphones and to keep their video on during the entire testing session.

Pretest session

Upon entering the first virtual testing session with the experimenter, participants confirmed their eligibility for the study and completed the informed consent procedure. Participants were then randomly assigned to either the experimental (Italian listeners) or control (English listeners) group and completed a basic demographic questionnaire.

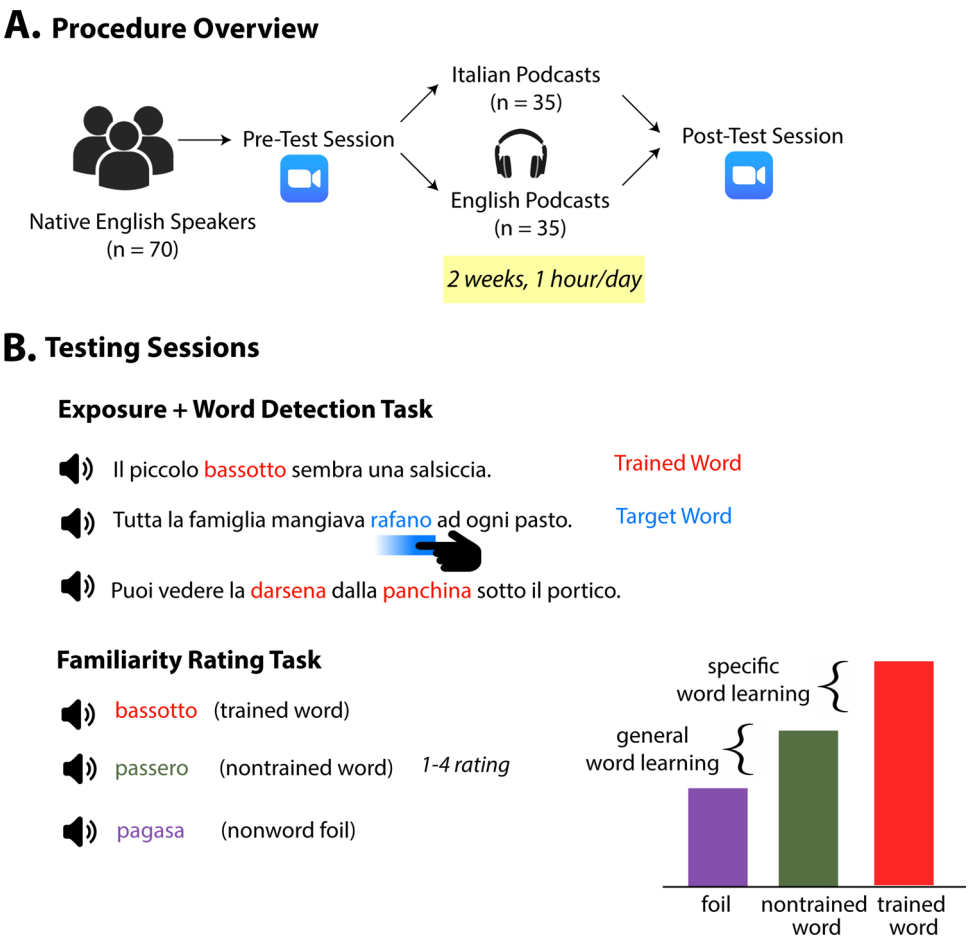

Fig. 1 Overview of experimental procedure and tasks. **A** Participants completed an initial pretest session over Zoom designed to measure Italian word knowledge. They were then randomly assigned to either the L2 exposure group or the control group. Over the next 2 weeks, while going about their daily activities, participants in the L2 exposure group listened to podcasts in Italian, while participants in the control group listened to podcasts in English. Participants then completed final tests of Italian word knowledge. **B** Participants completed two tasks in each testing session. The exposure + word-detection task required participants to respond to specific target words embedded in continuous Italian speech. An additional set of trained words were also presented multiple times throughout the task. The familiarity-rating task required participants to provide familiarity ratings for trained Italian words from the first task, nontrained Italian words (words in Italian that were not presented in the first task) and nonword foils (words that do not exist in Italian). Including both trained and nontrained words in the task design allowed us to examine two separate contrasts and associated hypotheses. Our key contrast of interest was between nontrained words and foil items. An increase in discrimination between these two categories provides a measure of general word-form learning in Italian. The second contrast of interest was between trained and nontrained words, which provides a measure of specific word learning, based on recent word exposures during the exposure task. (Color figure online)

Participants then began the first experimental task, the exposure + word-detection task (see Fig. 1B). This task had two purposes: (1) to expose participants to a set of "trained" words embedded in continuous speech, which they would be tested on later in the familiarity-rating task, and (2) to assess whether participants' ability to detect Italian words in continuous speech was enhanced by L2 listening experience. Given the online testing environment, we also used performance on this task as an objective measure of participant engagement and compliance, with low performance being grounds for exclusion (see Data Analysis section for further details). Participants were first informed that they would be hearing a series of Italian sentences and that they would have to perform a button press when they heard certain words within the sentences. At the beginning of each block, participants were then presented the specific target word for the upcoming block, in both written and spoken form. They were allowed to listen to the target word as many times as they liked before the sentence audio began. Sentences within a block were then presented in succession, separated by a 1-second pause. Optional breaks were provided between each block. No feedback was given to participants when they completed a button press. A static fixation cross was present on a white background as participants completed the task. In total, the task took approximately 35 minutes to complete.

Next, participants completed the familiarity-rating task, in which they were asked to judge how familiar or Italian-like different items were on a 1–4 scale (Fig. 1B). This task represented our main behavioural measure of word learning. One each trial, one of the test 30 items (10 trained Italian words, 10 nontrained Italian words, and 10 nonword foils) was played once. Participants then responded by pressing a number from 1 to 4, with 1 indicating the lowest familiarity with the item and 4 indicating the highest. The precise instructions provided to participants was changed slightly approximately halfway through data collection. The first ~40 participants were asked to "rate how familiar you are with this word on a scale of 1 to 4, with 4 being most familiar." For the final ~30 participants, we decided to make these instructions more precise by asking them to "rate how likely you think this word is a real Italian word on a scale of 1 to 4, with 4 being certain it is a real Italian word." No significant effect of this change was observed, so data from all participants were analyzed as a single group. This task took approximately 5 minutes to complete. Both the exposure + word-detection task and the familiarity-rating task were created in jsPsych (De Leeuw, 2015) and administered using Pavlovia (Ilixa, University of Nottingham).

After completing the familiarity-rating task, the podcast listening procedure was explained to the participants. Participants were informed that they would receive a different 1-hour podcast every day for the next 14 days. They were instructed to listen to the podcast in its entirety as well as complete the corresponding podcast questionnaire on the same day they received it. Participants were told to passively listen to the podcasts through headphones while going about normal activities, such as walking, cooking, and working. Participants were instructed to do their best to listen to the entire podcast in one sitting and to complete the podcast questionnaire immediately after listening. The experimenter remained on Zoom for the duration of the testing session to guide participants through the procedure. Both experimental tasks were created in jsPsych (De Leeuw, 2015) and administered using the online experiment distribution platform Pavlovia (Ilixa, University of Nottingham).

Two-week podcast listening period

At 9 a.m. each day over the next 14 days, participants were emailed a link to their assigned podcast and to the corresponding podcast questionnaire. The email also contained the instructions regarding the podcast listening procedure. Participants in the experimental group were sent Italian podcasts, and participants in the control group were sent English podcasts. The podcasts were stored on Google Drive, with each podcast assigned a unique link. The daily podcast questionnaire was administered via Qualtrics. The questionnaire asked participants to confirm that they had listened to a podcast that day, and then presented them with a list of approximately six words, only some of which had been embedded in the podcast. Participants were asked to select which of the words had been presented in the podcast.

Posttest session

On the day immediately following the 14th podcast listening day (i.e., Day 16), participants completed a second virtual testing session with the experimenter over Zoom. In the second testing session, participants again completed the exposure + word-detection task and the familiarity-rating task, using the counterbalanced sentence set that had not been presented at the pretest. Assignment to Sentence Set A versus B at pretest and posttest was counterbalanced across participants. After completing the experimental tasks, the participants completed an exit questionnaire, which asked them questions regarding their daily listening habits (i.e., the time of day they listened and what they did while listening), subjective reports of attention during podcast listening, their perceived knowledge of Italian, and their perceived difficulty of the experimental tasks. Finally, participants were debriefed about the nature of the experiment and the research question. All participants were compensated with a gift card for Amazon.ca.

Data analysis

Exposure + word-detection task

Responses occurring within 200–2,000 ms of a target word were considered "hits" and included in reaction time analyses. All other responses occurring outside of this window were considered false alarms. As an initial characterization of performance, each participant's mean word-detection rate was computed within each test session by dividing the number of correct responses (hits) by the total number of reaction time words.

We then conducted two analyses on (1) word detection and (2) response time. For word detection, each target word was coded as 1 if successfully detected, and 0 if not detected. We ran a generalized logistic regression model on this dichotomous outcome with group, session, and group × session as fixed factors, and participant as a random intercept. Additional factors were not included as random slopes due to convergence issues. For response time, we conducted a mixed-effect linear regression model with group, session, and group × session as fixed factors, participant as a random intercept, and by-participant random slope for session.

Detection performance as exclusionary criterion Given the online testing environment, participants' word-detection rate was used as an objective measure of their engagement and compliance with the experimental tasks. While the majority of participants performed very well on the task in both testing sessions, a small number of participants performed much more poorly than the group average in one or both sessions. Any participant who performed unusually poorly on this task (defined as an outlier with a word-detection rate below the first quartile by 1.5 interquartile range, on either or both of the two testing sessions) was excluded from subsequent analyses. This resulted in the exclusion of four participants from the exposure group and three participants from the control group, leaving a final sample of 31 participants in the exposure group and 32 participants in the control group

Familiarity-rating task

Linear mixed-effect modelling was used to account for repeated measures. Familiarity ratings (1–4) were measured at the individual trial level for each participant and classified according to the following factors: group (L2 exposure, control), participant, specific trisyllabic item, session (pretest, posttest), and word category (target word, nontrained word, foil). The model consisted of group, session, word category and their factorial interactions as fixed factors, participant intercept and item as random intercepts, and a by-participant random slope for session (pretest, posttest). Due to model convergence issues, a by-participant random slope was not included for word category. As shown in Fig. 1B, we were interested in two specific word category contrasts, each designed to test a separate hypothesis: (1) nontrained words versus foil items and (2) trained versus nontrained words. The first contrast (nontrained words versus foil items) represents the test of our key experimental hypothesis. Improved discrimination between nontrained and foil items over the 2-week period would provide evidence of general word-form learning in Italian, potentially driven by extraction of abstract, characteristic word features through experience with L2 speech. We hypothesized that the L2 exposure group would show greater improvement on this measure than the control group. The second contrast (trained versus nontrained words) provides an assessment of participants' ability to extract and form memories for specific words in continuous L2 speech, which may also potentially be improved with L2 experience. We tested these two key contrasts using treatment (dummy) coding within the main model, with nontrained words set as the reference variable for both contrasts. We used sum coding (−1, 1) for the other two fixed categorical variables in the model (session, group).

As described earlier, TPs between neighbouring syllables represent one possible statistical cue supporting word-form learning, which we captured in the current experimental stimuli. To test whether TPs did indeed serve as a cue to word-form learning in the present experiment, we conducted an additional secondary analysis. This analysis included only nontrained words and foils, as ratings for these items more directly reflect long-term statistical knowledge accrued as a result of L2 listening experience (as opposed to recent, short-term exposure to specific words). Following a similar approach to our original analysis, we tested a mixed effects model consisting of group, session, item TP and their factorial interactions as fixed factors, participant intercept as a random intercept, and a by-participant random slope for session (pretest, posttest). We again used sum coding for the fixed categorical variables in the model (session, group). Finally, as a point of comparison, we conducted the same analysis, but using total number of component syllable occurrences within each word as a predictor (see Supplementary Methods for details on this computation) rather than item TPs.

Results

Exposure + word detection

No significant effects of our experimental manipulation were observed on this task (see Supplementary Materials for complete results).

Familiarity rating task

As expected, trained words were rated as most familiar, followed by nontrained words, with foils rated as least familiar (Fig. 2); word category: $F(2, 57) = 17.1$, $p < .001$; estimated marginal means for trained words = 2.75, $SE = 0.011$; nontrained words = 2.25, $SE = 0.011$; foil items = 2.00, $SE = 0.011$. From the pretest to posttest session, the exposure group's familiarity ratings increased overall while the control group's familiarity ratings decreased, as reflected by a significant Group × Session interaction, $F(1, 61) = 8.56$, $p = .005$. There was a marginally significant Group × Session × Word Category interaction, $F(2, 3557) = 2.44$, $p = .087$, suggesting that from the pre- to the posttest session, the two groups showed a differential change in their ability to discriminate the three word categories. There was no main effect of group, $F(1, 61) = 0.48$, $p = .49$, or session, $F(1, 60) = 0.42$, $p = .52$.

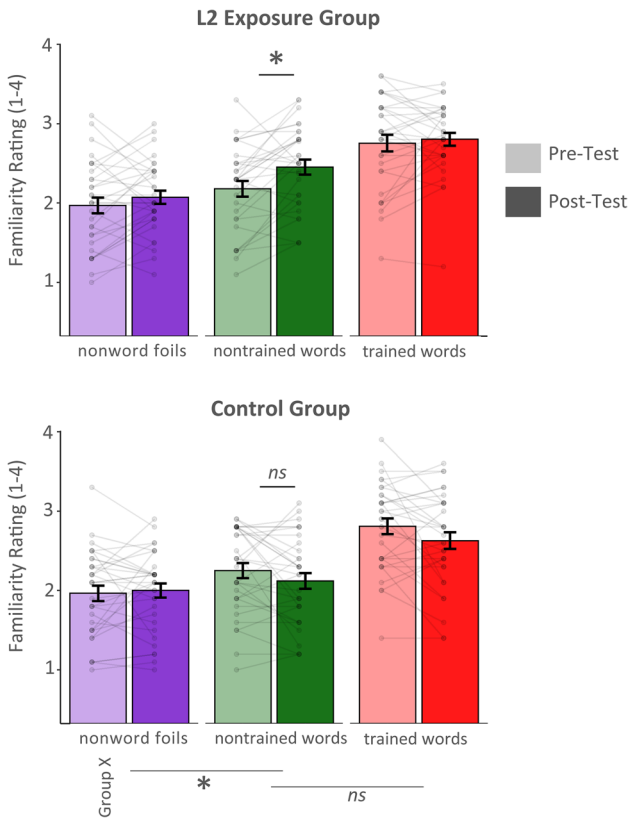

Fig. 2 Results from the familiarity judgment task. As hypothesized, participants in the L2 exposure group showed an improvement in their ability to discriminate between nonwords and foils over the 2-week listening period, whereas participants in the control group showed no such improvement. This was reflected by a significant Group × Session × Word Category interaction for our first contrast of interest. Participants' ability to learn specific, recently presented trained words did not change as a function of training, as reflected by a nonsignificant Group × Session × Word Category interaction for our second contrast of interest

Our first key contrast of interest was between nontrained words and foil items. Critically and as hypothesized, participants in the L2 exposure group showed a stronger improvement in the ability to discriminate between nontrained words and foil items from pre- to posttest, as compared with control participants, Group × Session × Word Category: $t(3556) = 2.20$, $p = .028$; parameter estimate for three-way interaction = 0.069, $SE = 0.031$). Follow-up contrasts indicated that the two groups did not show differences in performance at pretest, prior to the listening period (Session 1: parameter estimate = −0.043, $SE = 0.089$, z ratio = −0.48, $p = .63$). However, by the posttest session, relative to the control group, the L2 group showed significantly stronger discrimination between nontrained words and foils (Session 2: parameter estimate = 0.23, $SE = 0.089$, z ratio = 2.63, $p = .009$). Further, additional contrasts showed that within the L2 group, ratings for foil items did not significantly change from Session 1 to Session 2 (parameter estimate = 0.12, $SE = 0.076$, $p = .12$), whereas ratings for nontrained items significantly increased (estimate = 0.26, $SE = 0.076$, $p = .007$). This result indicates that word learning was primarily driven by increased familiarity for the nontrained words, rather than decreased familiarity for the foil items.

We then examined our second contrast of interest, between trained and nontrained words. Here, there was no significant evidence that the two groups showed differential performance over the two testing sessions, Group × Session × Word Category: $t(3556) = 1.30$, $p = .19$; parameter estimate for three-way interaction = 0.041, $SE = 0.031$.

Improved word discrimination is related to abstract rather than word-specific learning

Next, we ran a secondary analysis to better understand the significant results for the first key contrast (nontrained words versus foils). Specifically, we addressed whether the L2 exposure group's improved ability to discriminate between nontrained words and foils was related to abstract, general knowledge of Italian word features, as opposed to familiarity with specific words that had been presented during the podcasts. As a strong test of generalization learning, we excluded from analysis any item that had appeared even a single time in the L2 podcasts (eight test items excluded out of 40), and then ran the exact same model (with identical contrasts) as before. All our main findings were replicated. As before, we found that compared with control participants, the L2 exposure group showed a stronger improvement in the ability to discriminate between nontrained words and foil items from pre- to posttest, first contrast between nontrained words and foils: Group × Session × Word Category: $t(3067) = 2.08$, $p = .038$; parameter estimate = 0.068, $SE = 0.033$. Again, there was no difference in discrimination performance between the two groups at pretest (Session 1: parameter estimate = 0.047, $SE = 0.092$; z ratio = 0.51, $p = .61$), but a significant group

effect emerged at posttest (Session 2: parameter estimate = 0.23; $SE = 0.093$, z ratio= 2.43, $p = .015$). Also consistent with our main analysis, the L2 group showed no significant change in rating performance for foil items from Session 1 to Session 2 (parameter estimate = 0.12, $SE = 0.077$; z ratio = 1.54, $p = 0.12$), along with a significant boost in familiarity ratings for nontrained items (parameter estimate = 0.29, $SE = 0.082$; z ratio = 3.54, $p < .001$). These results indicate that the improvement in discrimination between nontrained words and foils in the L2 learner group persists even when examining only words that were never presented in the podcasts, and thus cannot be attributed to word-specific knowledge.

Transitional probabilities but not raw syllable occurrences relate to rating performance

In a final analysis, to investigate whether a given word's transitional probability (TP) may provide a potential cue to word-form learning, we examined the relationship between each word's estimated total transitional probability based on exposure to the podcasts and familiarity ratings. There was an overall significant effect of word TP, $F(1, 2389) = 28.6$, $p < .001$, indicating that words with higher TPs were rated as more familiar, across groups and sessions. This result converges with the lower baseline ratings reported for foils by control participants (see Supplementary Materials), reflecting that words with lower TPs may sound inherently less word-like. Interestingly, a significant group x session x TP interaction was also revealed, $F(1, 2368) = 4.11$, $p = .043$, indicating that ratings provided by the L2 group became more strongly correlated with word TP from Session 1 to Session 2, whereas ratings provided by the control group became *less* strongly correlated with TP. Within the L2 exposure group, familiarity ratings did not significantly correlate with TPs at Session 1 (trend = 0.49, $SE = 0.30$), $t(2376) = 1.65$, $p = .098$, but did significantly correlate with TPs at Session 2 (trend = 1.15, $SE = 0.31$), $t(2383) = 3.75$, $p < .001$. Nonetheless, the between-session change in this trend did not reach significance (estimate = 0.66, $SE = 0.43$), $t(2369) = 1.54$, $p = .12$, indicating that L2 participants' ratings were only trending towards being more related to TPs after the 2-week listening period. Within the control group, familiarity ratings correlated significantly with TPs at Session 1 (trend = 1.05, $SE = 0.30$), $t(2377) = 3.52$, $p = .004$, but were not significantly correlated at Session 2 (trend = .50, $SE = 0.29$), $t(2378) = 1.71$, $p = .088$. This between-session decrease was not significant (estimate = 0.55, $SE = 0.42$), $t(2367) = 1.32$, $p = .19$.

Interestingly, as a point of contrast, there was no significant effect of overall syllable frequency of the component syllables in each test word on familiarity ratings, $F(1, 2398) = 0.31$, $p = .57$. Similarly, there was no Group × Session × Syllable Frequency interaction on ratings, $F(1, 2387) = 0.024$, $p = .88$. This comparison analysis indicates that—in contrast to the co-occurrences between neighbouring syllables in the podcasts—participants' ratings were not influenced by the overall number of occurrences of the individual syllables within each test word.

Discussion

We found that unguided exposure to a foreign language facilitates initial word-form learning in adult L2 learners. After 2 weeks of daily exposure to Italian podcasts, English-speaking participants significantly improved in their ability to endorse nontrained Italian words. Control participants who listened to English podcasts showed no such improvement. Our results provide an important proof of concept that adult L2 learners can extract characteristic word features in a novel language just by listening to the target language in the background of day-to-day life, in the absence of formal instruction and intentional study. That we found evidence of learning is especially notable given the relatively brief amount of input, and our use of an explicit behavioural measure of word learning, which may trail earlier, more sensitive neural indicators of learning, such as event-related potentials (e.g., McLaughlin et al., 2004).

Importantly, the improvement in learners' word identification ability was not driven by word-specific knowledge, as learning effects were maintained after excluding any word that had appeared even a single time in the podcasts. Thus, learners' enhanced ability to recognize actual Italian words appears to reflect a form of generalization learning, involving the extraction and generalization of relevant sound patterns to new items (i.e., never-before-encountered words). In principle, learners may have become sensitive to any number of possible characteristic word features, such as syllable TPs, phonetic patterns, and word stress patterns. Although narrowing down precisely which cues were most relevant to learning is beyond the scope of the current study, some evidence supports the role of TPs as one potentially important cue. Relative to control participants, L2 participants' ratings became more strongly related to a given word's TPs over the 2-week listening period and were significantly correlated with word TPs at the second session, but not the first session. These results suggest that L2 participants gained sensitivity to the statistical co-occurrences of syllables present in the podcasts. However, we note that the increase in strength of the TP-rating relationship from Session 1 to Session 2 was only at trend level and did not reach significance. As an interesting point of contrast, the raw frequencies of a given word's individual component syllables in the podcasts were completely unrelated to participants' ratings. Consistent with the broader SL literature (Saffran, Newport, & Aslin, 1996b), these results suggest that distributional statistics represent a more relevant cue for word-form learning than frequency statistics. However, beyond this observation, additional research will be needed to determine which specific

features most strongly drive word segmentation and recognition during early learning stages.

To our knowledge, the current study provides some of the first evidence that unguided SL mechanisms—again, operationalized here as extraction of characteristic word form features through passive listening alone—can scale up to support word-form learning in a fully natural language learning context. As described in the introduction, most past studies of SL have used "toy languages"—highly controlled, miniature systems that are typically devoid of redundant acoustic cues to word boundaries. Advantages to this approach includes tight experimental control over the available statistical information, as well as a simplified learning problem that can be solved in a short period of time. Nonetheless, this approach also comes at the cost of ecological validity, and concerns have been raised as to whether SL can scale up to capture the real-world complexity of natural languages (Erickson & Thiessen, 2015; Frost et al., 2019; Siegelman, 2020). At the same time, it has also been suggested that the presence of overlapping cues to word boundaries in natural language may actually make learning more feasible and could be readily exploited by SL mechanisms (Christiansen et al., 2010). Our results support the idea, providing initial reassurance that SL mechanisms may indeed be powerful enough to extract relevant sound patterns in fully natural speech, thereby supporting initial word-form learning in an unknown language.

These results have important practical implications for adult L2 learners, a group that typically struggles with language learning. The ability to discover individual word forms from a largely continuous stream of sounds is a central and rate-limiting problem for language acquisition, as only after initial word forms are acquired can these individual words serve as input to further language learning. Our findings suggest that beginning L2 learners may be able to leverage passive, unguided listening methods to boost sensitivity to key statistical properties and sound patterns of words in their new language, and thus gain a foothold in their new language more quickly. Given that insufficient exposure to L2 input may present a bottleneck for learning (Marinova-Todd et al., 2000), learners could increase L2 exposure by listening to podcasts, radio, or TV during otherwise unoccupied periods of the day (e.g., commuting, cooking, housework). In turn, this may boost learners' sensitivity to relevant word characteristics in L2 speech, ultimately facilitating word identification. Critically, this type of learning can occur without top-down attention, intention to learn, or conscious effort, and thus represents a relatively low-effort strategy for facilitating learning.

Of course, unguided learning has key limitations, and most aspects of language cannot be acquired simply through passive, context-free exposure to speech. Nonetheless, word segmentation may still bootstrap other aspects of language that depend on explicit or intentional learning, such as vocabulary. For example, previous evidence in infants has shown that learning word forms in continuous speech facilitates subsequent word-meaning mappings (Estes et al., 2007). More generally, familiarity with speech sounds facilitates automatic processing of speech sounds and sound sequences (Bonte et al., 2005; Dehaene-Lambertz, 1997; Huotilainen et al., 2001), and may reduce the computational resources required for word processing (Ylinen et al., 2009). A mere increase in general familiarity with L2 speech patterns could thus potentially facilitate L2 learning at later learning stages by automatizing sensory-level processing and freeing up cognitive resources for other aspects of learning.

Conclusions

Our results suggest that adult learners can extract structure and learn initial word forms from unguided exposure to a novel language, without requiring instruction, conscious effort, and intention to learn. These results provide initial evidence that SL can scale up to support initial word learning in natural language learning conditions. From a practical perceptive, these findings also open up new strategies for adult second language learning.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13423-022-02190-1.

Acknowledgments This work was supported by the National Sciences and Engineering Research Council and by the Social Sciences and Humanities Research Council of Canada. We thank Cecilia Affinita and Giacomo Spinelli for their assistance with Italian language materials.

References

Arciuli, J. (2017). The multi-component nature of statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1711), Article 20160058.

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex, 90*, 31–45.

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language, 83*, 62–78.

Benitez, V. L., & Saffran, J. R. (2021). Two for the price of one: Concurrent learning of words and phonotactic regularities from continuous speech. *PLoS ONE, 16*(6), e0253039.

Bonte, M. L., Mitterer, H., Zellagui, N., Poelmans, H., & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 116*(12), 2765–2774.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16*(4), 298–304.

Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., & Werker, J. F. (2020). Preverbal infants discover statistical word patterns at similar rates as adults: Evidence from neural entrainment. *Psychological Science, 31*(9), 1161–1173.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (2010). Learning to segment speech using multiple cues: A connectionist model.

Language and Cognitive Processes. https://doi.org/10.1080/01690969386528

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 31*(1), 24–39.

Cunillera, T., Càmara, E., Laine, M., & Rodríguez-Fornells, A. (2010). Words as anchors: Known words facilitate statistical learning. *Experimental Psychology, 57*(2), 134–141.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language, 2*(3), 133–142.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavioral Research, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport, 8*(4), 919–924.

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review, 37*, 66–108.

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words?: Statistical segmentation and word learning. *Psychological Science, 18*(3), 254–260.

Frank, M. C., Tenenbaum, J. B., & Gibson, E. (2013). Learning and Long-term retention of large-scale artificial languages. *PLOS ONE, 8*(1), Article e52500.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences, 19*(3), 117–125.

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin, 145*(12), 1128–1153.

Gu, Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning, 46*(4), 643–679.

Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology, 63*(2), 93–106.

Huotilainen, M., Kujala, A., & Alku, P. (2001). Long-term memory traces facilitate short-term memory trace formation in audition in humans. *Neuroscience Letters, 310*(2/3), 133–136.

Kittleson, M. M., Aguilar, J. M., Tokerud, G. L., Plante, E., & Asbjørnsen, A. E. (2010). Implicit language learning: Adults' ability to segment words in Norwegian. *Bilingualism: Language and Cognition, 13*(4), 513–523.

Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly, 34*(1), 9.

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

Palmer, S. D., Hutson, J., & Mattys, S. L. (2018). Statistical learning for speech segmentation: Age-related changes and underlying mechanisms. *Psychology and Aging, 33*(7), 1035–1044.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development, 80*(3), 674–685.

Plante, E., & Gómez, R. L. (2018). Learning without trying: The clinical relevance of statistical learning. *Language, Speech, and Hearing Services in Schools, 49*(3S), 710–722.

Plante, E., Patterson, D., Gómez, R., Almryde, K. R., White, M. G., & Asbjørnsen, A. E. (2015). The nature of the language input affects brain activation during learning from a natural language. *Journal of Neurolinguistics, 36*, 17–34.

Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: Modality-based differences in the effect of age. *Developmental Science, 21*(4), Article e12593.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45*, 283–331.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition: Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(6), 906–914. https://doi.org/10.1002/wcs.78

Rodríguez, M., & Sadowki, M. (2000). Effects of rote, context, keyword, and context/keyword methods on retention of vocabulary in EFL classrooms. *Language Learning, 50*(2), 385–412.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology, 39*(3), 484–494.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science, New Series, 274*(5294), 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*(4), 606–621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental Language Learning: Listening (and Learning) Out of the Corner of Your Ear. *Psychological Science, 8*(2), 101–105.

Sahni, S. D., Seidenberg, M. S., & Saffran, J. R. (2010). Connecting cues: Overlapping regularities support cue discovery in infancy. *Child Development, 81*(3), 727–736.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275*(5306), 1599–1603. https://doi.org/10.1126/science.275.5306.1599

Siegelman, N. (2020). Statistical learning abilities and their relation to language. *Language and Linguistics Compass, 14*(3). https://doi.org/10.1111/lnc3.12365

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language, 81*, 105–120.

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*, 35.

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition, 177*, 198–213.

Sohail, J., & Johnson, E. K. (2016). How transitional probabilities and the edge effect contribute to listeners' phonological bootstrapping success. *Language Learning and Development, 12*(2), 105–115.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50*(1), 86–132.

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*(4), 706–716.

Thiessen, E. D., & Saffran, J. R. (2007). Learning to Learn: Infants' Acquisition of Stress-Based Strategies for Word Segmentation. *Language Learning and Development, 3*(1), 73–100.

van Hell, J. G., & Mahn, A. C. (1997). Keyword Mnemonics Versus Rote Rehearsal: Learning Concrete and Abstract Foreign Words by Experienced and Inexperienced Learners. *Language Learning, 47*(3), 507–546.

Webb, S., Yanagisawa, A., & Uchihara, T. (2020). How effective are intentional vocabulary-learning activities? A meta-analysis. *The Modern Language Journal, 104*(4), 715–738.

Yang, C. D. (2004). Universal Grammar, statistics or both? *Trends in Cognitive Sciences, 8*(10), 451–456. https://doi.org/10.1016/j.tics.2004.08.006

Ylinen, S., Strelnikov, K., Huotilainen, M., & Näätänen, R. (2009). Effects of prosodic familiarity on the automatic processing of words in the human brain. *International Journal of Psychophysiology, 73*(3), 362–368.

Open practices statement The experiment was not preregistered. The data generated and analyzed during the current study are available on the Open Science Framework (https://osf.io/vjbu7/?view_only=2e61b6ee10774181b677e1e9cd0e8ce9).