English Teaching & Learning (2024) 48:23–47
https://doi.org/10.1007/s42321-022-00129-2

ORIGINAL PAPER

# Resolving Referential Ambiguity for Nouns and Verbs Across Situations by Foreign Language Learners

外語學習者如何運用跨情境訊息解決名詞與動詞的指意模糊問題

**Chieh-Fang Hu**[1]

Received: 14 March 2022 / Revised: 9 October 2022 / Accepted: 16 October 2022 /
Published online: 7 November 2022


## Abstract

The study examined the process used by foreign language learners in resolving referential ambiguity across situations. Two hundred learners of English as a foreign language took two cross-situational word learning tasks, one on nouns and the other on verbs. In each trial, participants heard a novel word, observed two referents (animals or dynamic events), and selected one referent under uncertainty. Following zero or two intervening trials, they were tested by one of three probes, which preserved the word-referent mapping previously selected, switched the referent previously selected, or switched the form previously heard. Learners performed above chance on all the probes, indicating that they tracked the unselected but potential word-referent mappings across situations for later learning. The performance pattern was similar for nouns and verbs, though verb learning was slightly impaired when the referent was switched over intervening trials. Moreover, verb learning was associated with phonological short-term memory, but not with statistical learning and English vocabulary. Noun learning was associated with none of the learner variables. These results suggest that foreign language learners track all the word-referent co-occurrences across situations in resolving referential ambiguity, though success of tracking can be constrained by the complexity of the co-occurrences and individual learners' memory capacities.

摘要

本研究探討外語學習者如何運用跨情境訊息解決指意不明的情形。200位英語為外語的學習者接受兩項跨情境詞彙學習任務，一為名詞學習，另一為動詞學習。參與者每次聽到一個新的英語新詞，看見兩個可能的指涉項目（動物

✉ Chieh-Fang Hu
cfhu@utaipei.edu.tw

[1] Department of English Instruction, University of Taipei, Taipei 10001, Taiwan

或動態事件），並於指涉不明的情況下選擇一個指涉。在零個或兩個干擾項後，其學習情形透過三種偵測項來檢視。三個偵測項分別為同原選偵測項、非原選偵測項、換詞偵測項。學習者在三種偵測項的表現都高於隨選機率，顯示學習者能跨越情境追蹤原先沒有選擇的指涉項目，以作為後續學習所需。學習者在名詞和動詞的整體表現相當類似，只有動詞在經過幾個干擾項後，在非原選偵測項上的表現後略低於名詞。動詞學習表現與語音短期記憶相關，但與統計學習能力、英語詞彙能力無關。名詞學習與上面三項學習者因素都無關。研究結果顯示在新詞指意不明的情形下，外語學習者跨越情境追蹤所有新詞與指涉的共現可能性，找出新詞的可能指意。至於跨情境追蹤能否成功則可能因新詞與指涉共現的複雜度以及學習者的記憶能力而有所限制。

**Keywords** Foreign language learning · Cross-situational word learning · Individual differences · Memory · Statistical learning · Fast mapping

關鍵詞 外語學習 • 跨情境詞彙學習 • 個別差異 • 記憶 • 統計學習 • 快速配對

## Introduction

The study investigated a fundamental question unresolved in foreign language (FL) word learning: How do FL learners resolve word-referent mappings across referentially ambiguous situations? Do they track one mapping for one word at a time or do they track multiple potential mappings across situations? Is resolving referential ambiguity across individually ambiguous situations too overwhelming that they just bypass the ambiguous data and show no evidence of learning? These questions are rarely discussed in FL word learning. It is generally assumed that word learning in a FL classroom is direct and thus unambiguous. Even when there is ambiguity, FL learners can quickly form a "fast map" between a word and its referent based on some lexical assumptions, such as mutual exclusivity (Markman & Wachtel, 1988). For example, hearing a novel word *snood* upon seeing a book and a ring of knitted material, child FL learners tend to infer that *snood* refers to the knitted material rather than to the book, for which they already have a name (Hu, 2012, 2017). With the mutual exclusivity assumption, referential ambiguity can be easily resolved at a single time of learning.

Nevertheless, given the growing popularity of content and language integrated learning in many Asian and European countries (Coyle et al., 2010), ambiguity is not uncommon in FL classrooms. In content-rich FL classrooms, FL learners may hear a novel word in the presence of a number of name-unknown referents with little explicit instruction about its intended meaning. Referential ambiguity cannot be resolved by the classic mutual exclusivity assumption—a preference for establishing one-to-one word-referent mappings. Nonetheless, learners can resolve referential ambiguity if they track word-referent co-occurrences across individually ambiguous situations and across time, known as cross-situational word learning (CSWL). There has been ample evidence supporting CSWL on both adults and children learning L1

(e.g., Chen et al., 2017; Fitneva & Christiansen, 2017; Vlach & DeBrock, 2017; Yu & Smith, 2007; Yurovsky & Frank, 2015; Yurovsky et al., 2014). However, cross-situational word learning is generally outside the scope of FL literature even though in content-rich language classes or in language-rich content classes, FL learners also encounter similar situations where referential ambiguity cannot be resolved in a single situation, for example, when watching a video of environmental issues, discussing a gender issue, or engaged in learning an academic subject. FL literature has typically focused on how learners infer a word's meaning from contextual clues in a single text or situation (e.g., Hamada, 2014; Nassaji, 2003; Van Zeeland, 2014) rather than across texts or situations. This study addressed this gap in FL literature by examining the process used by FL learners to resolve lexical ambiguity over a number of individually ambiguous situations, along with learner variables that may contribute to their CSWL performance.

### The Two Models: Associative Learning and Hypothesis Testing

In a CSWL paradigm, learners hear one or multiple novel words and observe multiple name-unknown objects. They are not able to infer which word refers to which object at that single time by the mutual exclusivity assumption, within that single situation. However, across situations, a certain word consistently co-occurs with a particular object. The consistency of word-object pairing provides information about the intended meaning of the novel word (Yu & Smith, 2007). In this paradigm, the referent of a novel word is determinable across situations, though not in a single situation (i.e., a learning trial).

The mechanisms underlying CSWL have been under considerable debate. At issue of the debate is whether and how learners track word-referent co-occurrences across situations. There are two general models of learning mechanisms: associative learning and hypothesis testing. The major difference between the simplest versions of the two models is whether learning passes through a state of partial knowledge (Yurovsky et al., 2014). In CSWL, partial knowledge refers to the sensitivity to all the word-referent co-occurrences of the input. It does not refer to individual word knowledge along the partial-precise continuum (shades of meaning) or the depth of vocabulary knowledge often discussed in L2 vocabulary acquisition (Henriksen, 1999; Schwanenflugel et al., 1997). According to the associative-learning model, learners track all possible co-occurrences between words and referents across situations in time and use cross-situational probabilistic information to find the underlying word-referent mappings (Roembke & McMurray, 2016; Scott & Fisher, 2012; Vlach & DeBrock, 2017; Yu & Smith, 2007). In each learning situation, multiple word-referent associations are sampled, computed, weighted, and carried from one situation to another. Evidence supporting the model generally shows that learners demonstrate partial knowledge about the mappings that previously conflict with their prediction (Yurovsky et al., 2014).

In contrast, according to the hypothesis-testing model, learners proceed CSWL by conjecturing one meaning for one word at a time and track that hypothesized mapping across situations (for adults, see Medina et al., 2011; Trueswell et al., 2013; for children, see Aravind et al., 2018; Woodard et al., 2016). If the hypothesized mapping is

verified in the next encounter, the mapping is maintained. If it is not verified, the mapping is reset from the scratch. Learners do not encode all the co-occurrences between words and referents. No partial information is maintained for alternative mappings. Evidence supporting the model generally shows that if learners initially conjecture a wrong mapping, they rarely remember the alternative referents and can hardly recover from the wrong mapping (e.g., Medina et al., 2011; Woodard et al., 2016).

How do FL learners discover a word's mapping in ambiguous contexts? Among the few studies directly examining CSWL in FL learners, Hu (2017) found that 8-year-old Mandarin-speaking learners were able to map two novel English words (e.g., *wedge* and *snood*) in the presence of two novel referents under uncertainty across multiple learning trials. This study measured global learning outcomes at the end. It is not clear what mechanism subserved CSWL in FL learners. Both the associative-learning and the hypothesis-testing models could predict above-chance performance at the end of the learning trials. The participants might have completed the task by adopting the hypothesis-testing model in CSWL as one mapping at a time is not only manageable but also the typical way of word learning in instructional settings, where FL learners usually rote-rehearse one-to-one mapping between a word and a picture or a translation equivalent. It was also possible that the participants adopted the associative-learning model as tracking probabilistic information is a fundamental learning mechanism independently of one's linguistic learning experience or the language to learn (Poepsel & Weiss, 2016). To understand the mechanism underlying CSWL in FL, the present study tracked whether FL learners formed partial knowledge during CSWL instead of measuring CSWL performance at the end.

### Individual Learner Variables

A second focus of the research is to explore the learner variables contributing to CSWL performance in FL. FL learners differ in their cognitive and processing strengths that they can draw on in learning another language. FL research has shown that individual learner variables can play a more important role in FL learning than the input factor in some situations (e.g., Brooks et al., 2017; Hu & Maechtle, 2021; Sunderman & Kroll, 2009). The first learner variable that may constrain CSWL is the learners' vocabulary capacity of the FL. Theoretically, learners' vocabulary level reflects, in part, their ability to learn words and thus should be associated with CSWL (Scott & Fisher, 2012; Vlach & Debrock, 2017).

Phonological short-term memory (STM) is another learner variable that may constrain CSWL (Roembke & McMurray, 2021; Vlach & DeBrock, 2017; Vlach & Sandhofer, 2014). This should be especially true for CSWL under the associative-learning model, by which learners encode, retain, and retrieve highly uncertain information across moments in time. In contrast, phonological STM should play a minimal role in CSWL if learners propose and verify only one hypothesis at a time, without forming partial knowledge for all alternative mappings. As we are interested in the memory constraints in CSWL, we chose to measure phonological STM through forward and backward digit span rather than nonword repetition. First, nonword repetition can be confounded with nonword learning in the present study.

Second, nonword repetition is a test of phonological short-term memory, complicated by vocabulary knowledge (Gathercole, 1995). And we had a separate test of vocabulary.

The third learner variable examined in the present study is statistical learning. Statistical learning refers to the ability to use statistical information to learn the structure of one's environment. In language learning, it is broadly defined as the sensitivity to the statistical properties (e.g., repeated patterns of regularities and transitional probabilities) embedded in a continuous stream of stimuli (Kidd et al., 2018). Statistical learning is intimately associated with, for example, sentence comprehension (Kidd et al., 2016) and morphosyntactic learning (Brooks et al., 2017; Kidd, 2012; Misyak & Christiansen, 2012). Tasks of statistical learning usually present participants with structured material with no instruction or feedback (Aslin, 2017; Marcus et al., 1999, 2007). For example, one paradigm of statistical learning requires participants to detect and generalize a pattern (e.g., ABB) recurring in a long, continuous string of auditory stimuli (e.g., gatitilinana) without explicit instruction about what information to attend to (Marcus et al., 1999). Statistical learning can be a basic ability associated with CSWL, as both require sensitivity to the statistical pattern embedded in the input.

### The Overview

The present study sought to examine the mechanism underlying CSWL in FL learners, along with the individual learner variables that may contribute to CSWL. Instead of focusing on CSWL performance at the end, the study is designed to track learners' responses to the selected as well as the unselected referent in a $1 \times 2$ CSWL task, involving one word and two referents. The two CSWL models make the same prediction about the selected referent but diverge in their predictions about the unselected referent. The associative-learning model predicts that learners form partial knowledge about the unselected (but still possible) referent, track the co-occurrences of the whole data set, and thus can recover from initial mis-mappings. The hypothesis-testing model predicts that learners encode one mapping at a time, show little knowledge of the unselected referent, and thus cannot recover from initial mis-mappings.

There were two CSWL tasks, noun learning and verb learning. Most CSWL studies focused on nouns (but see Scott & Fisher, 2012). How verbs are learned across situations is massively underrepresented in literature. Compared to nouns, verb meanings are relatively difficult to acquire in a single learning situation (Childers et al., 2002; Imai et al., 2008; Monaghan et al., 2015). Nouns are constant across time and space; verbs are not. Hearing a verb across situations is necessary to gather consistent information about a verb's meaning, known as a "packaging problem" (Gleitman & Gleitman, 1992). For example, English verbs can lexically specify different perceptual elements as part of their meanings: manner (e.g., *stir*, *shake*, *beat*), path (e.g., *exit*, *enter*, *cross*, *follow*), instrument (e.g., *whip*, *club*, *hammer*), result (e.g., *break*, *slice*, *shatter*), or instrument together with result (e.g., *cut*, *slice*). It is the consistency across situations, along with variability, that provides clues to the perceptual elements of an event that are packaged and lexicalized in a verb

(Childers, 2011; Childers et al., 2017). Thus, discovering referential meaning for a verb is more cross-situational and more complex than for a noun (Childers, 2011). It should be interesting to see whether FL learners handle CSWL following different models for nouns and verbs.

Q1: Do FL learners resolve referential ambiguity via CSWL in an associative-learning or a hypothesis-testing manner?
Q2: Do FL learners proceed with CSWL in a similar way for nouns and verbs?
Q3: What learner variables contribute to FL learners' CSWL performance?

## Methods

### Participants

Two hundred college students, recruited from 19 departments of a university in Taipei, participated in the study. Their English scores in the General Scholastic Ability Test range from 4 to 15. In Taiwan, the score for each academic subject in the test is scaled from 0 to 15. About 78% of the participants fell between 10 and 14. Aside from receiving formal English education in school, over 50% of the participants had studied English in private language programs during primary school (67%), junior high school (59%), or senior high school (55%). Only 2% reported that they frequently used English in interpersonal communication. The participants were randomly divided into two groups, each for one interval condition in the two word-learning tasks.

### Noun Learning

#### Stimuli and Design

The study adopted a 3 (probe) × 2 (interval) design, with probe as a within-participants factor and interval as a between-participants factor. One group of the participants learned the words presented in consecutive trials, without intervening trials. The other group learned the words with presentations intervened by two trials. With the two intervals, we could examine how far the partial knowledge, if any, was carried to support subsequent word learning.

The participants' CSWL performance on each word was examined by one of the three probes, using a same-switch procedure similar to that used in Woodard et al. (2016) and Yurovsky and Frank (2015). The three probes were Same, Form-switch (F-switch), and Referent-switch (R-switch). On the Same probe, participants heard the same word and saw a referent previously selected. On the R-switch probe, participants heard the same word but saw a referent previously seen but not selected. One the F-switch probe, participants heard a different word but saw a referent previously selected. The probe trials were designed to understand about the information

the participant encoded and carried from previous trials, based on the consistency (or inconsistency) of the word-referent co-occurrences across the observations.

Participants were exposed to 18 blocks of experimental trials, with each block for one target word. In each trial, the participants heard a word and observed two referents. Their task was to select from the two referents the one they thought was the referent of the word they heard. Word stimuli were 18 bisyllabic nonce words that were phonotactically probable in English, six examined by each of the three probes (Same, R-switch, F-switch). The nonce words were chosen from the English Lexicon Project nonword database (http://elexicon.wustl.edu). Examples of nonce words were *borale*, *ferpent*, and *jatal*. Another six bisyllabic nonce words were selected and used for the F-switch probe trials. Appendix A shows the nonce words for the noun learning task.

The referents for selection were 54 photographs of weird animals/insects (e.g., rhipiceridae, dumbo octopus, lip batfish, Brazilian treehopper, star nose mole, axolotl, kakapo), partitioned into 18 sets of three. Two animals served as the potential referents for a novel word. The third animal served as the exposed foil in the familiarization trial. The images of the animals/insects were strategically selected so that they could not be readily labeled with a general name. For example, the image of a kakapo was the domed head of the kakapo without revealing its wings, claws, or tail so that it could not be readily associated with a generic label (i.e., bird). The images selected were those which were name-unknown to four English majors in a pilot screening test.

### Trials

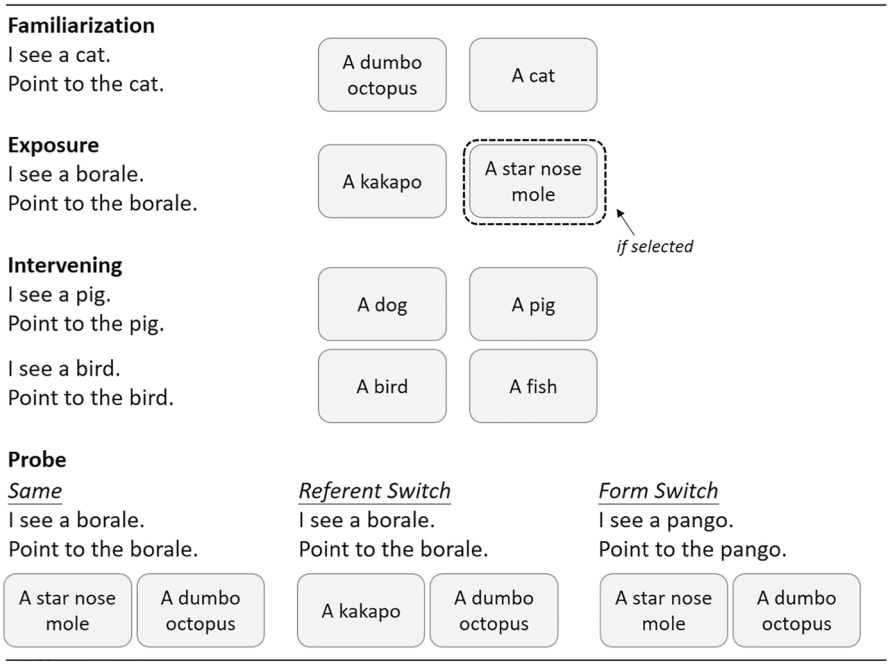
Each word was presented in a block of trials, consisting of one familiarization trial, one exposure trial, zero or two intervening trials, and one probe trial, in that order. There were 18 blocks. Please see Table 1 for examples of the trial sequence in a block.

**Familiarization** In the familiarization trial, the participant observed two pictured referents, one with a known name and the other without (e.g., a cat and a dumbo octopus) and heard a familiar word, for example, "I see a cat. Point to the cat." The novel referent served as a potential referent for the subsequent trial. The purpose of the familiarization trial was to control the participants' familiarity with the novel referents used for the probe trial.

**Exposure** In the exposure trial, the participant saw two novel referents (e.g., a star nose mole and a kakapo) and heard a novel word, for example, "I see a borale. Point to the borale." The participant selected the one thought to be the referent of the novel word. This trial was ambiguous as to which picture went with the novel word. No feedback was given to the choice made by the participant.

**Intervening** Following the exposure trials were the intervening trials. The number of the intervening trials (zero or two) was a between-participants factor. The

**Table 1** Sample materials and sequence of trials for CSWL of nouns in a block




Note. The referent position (left or right) for the target was randomized.

intervening trials involved only familiar words (e.g., *dog*, *cat*, *pig*, *fish*). The intervening trials simulated FL word learning in real-life situations. FL learners did not always hear novel words consecutively. They heard novel words, along with familiar words, with time passing between repetitions.

**Probe** The probe trials were designed to understand about the information the participant encoded and carried from the exposure trial. Although the novel word was ambiguous in the exposure trial, its referent became clear in the probe trial, due to the consistency (or inconsistency) of the word-referent co-occurrences between the exposure and the probe trials. Each word was examined by one of the three probes, Same, R-switch, and F-switch, each probe for six target words. On the Same and the R-switch probe, the participant heard the same novel word presented in the exposure trial, for example, "I see a borale. Point to the borale." On the F-switch probe, the participant heard a second novel word, different from the one in the exposure trial, for example, "I see a pango. Point to the pango." On the Same probe, the target referent was the one the participant had previously selected during exposure; the foil was the name-unknown referent during familiarization. On the R-switch probe, the target referent was the one the participant did not select during exposure; the foil was the name-unknown referent presented during familiarization. On the F-switch probe, the target referent was the name-unknown referent presented during familiarization; the foil was the one previously selected by the participant during exposure. The

two referents for each probe were equally familiar and displayed randomly on left or right sides of the screen. The order of the probes was also randomized with the stipulation that probes of the same type were not administered in more than two successive blocks.

### Procedure

Each participant took two practice trials, followed by 18 blocks of experimental trials (familiarization, exposure, intervening, and probe), with each block for one target word. All experimental trials were administered using a prerecorded interactive video format, which allowed for consistency of length and presentation and controlled for unintentional social cues or attentional biases. No feedback about the correctness of referent selection was given. The participants were told that they went night hunting. They would learn to recognize the animals they were going to hunt in the field. Some of the animals were unfamiliar. They should point to the one they believed to be the correct referent of each word they heard.

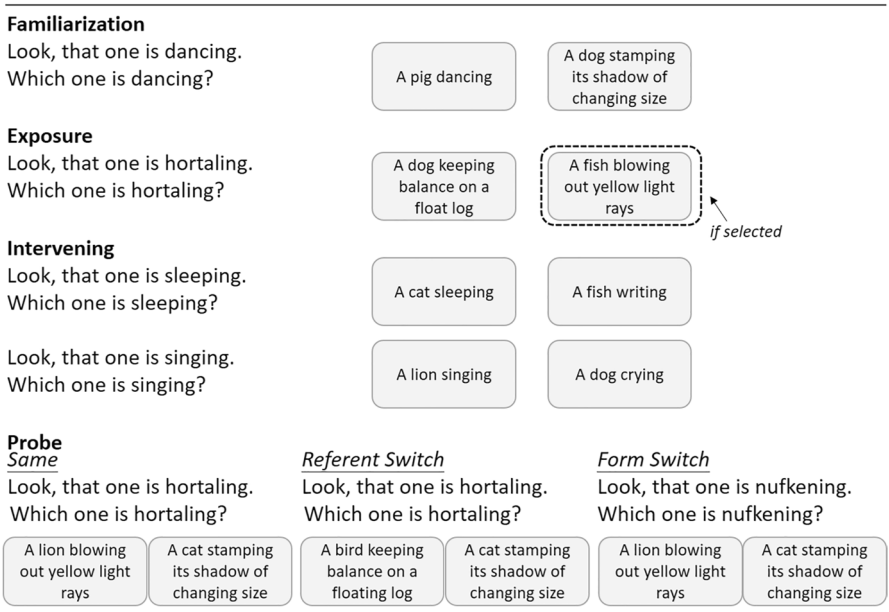
### Verb Learning Task

The design of the verb learning task was similar to that of the noun learning task, except that the 18 nonce words were presented in a verb frame and the referents were dynamic events. Each novel word was accompanied with a unique set of three dynamic referents. Two events served as the potential referents for the verb. The third event served as the exposed foil in the familiarization trial. Each of the events depicted a novel, animated actions played by familiar animals, such as a rabbit stamping on its shadow which changed its size with each stamp, a fish blowing out yellow light rays, and a dog finding balance on a floating log. Please see Table 2 for the sample stimuli for a block of verb learning. Appendix B shows the nonce words for the verb learning task.

As in the noun learning task, each word was presented in a block of experimental trials: familiarization, exposure, intervening, and probe. Participants heard a word and observed two events. The word presented in an intransitive verb frame: "*That one is* ____*ing*." The subject was "that one" so that the utterance could refer to either of the scenes on the screen. The duration of the two events was kept identical on the screen. In each block, the performers of the same action were different. As shown in Table 2, the performers involved in the two events of blowing out yellow light rays were a fish in the exposure trial and a lion in the probe trial. The participant was told that they were to host an animal show. They would learn how to introduce the animal show.

As in the noun learning task, the participants were assigned to two interval groups, consecutive and intervening. Those assigned to the consecutive group in the noun learning task were assigned to the intervening group in the verb learning task, and vice versa.

**Table 2** Sample materials and sequence of trials for CSWL of verbs in a block




Note. The referent position (left or right) for the target was randomized.

## Learner Factors

### English Vocabulary

Each participant completed a standardized test of receptive vocabulary, Peabody Picture Vocabulary Test-III (Dunn & Dunn, 1997). The participant selected from four pictures the one that best represented the word said by the test giver. Following the standardized procedure, test administration terminated after an error criterion was reached. The score was the number of items answered correctly.

### Phonological STM

**Forward Digit Span** It is a widely used test of the basic storage capacity of phonological STM. The participant verbatim repeated a sequence of synthetic spoken digits. The number of digits increased by one until the participant consecutively failed two trials of the same length. The score was the maximum number of digits an individual recalled accurately.

**Backward Digit Span** The test is typically considered as a simple test of the processing component of phonological STM because it requires manipulation of the items that are stored in short-term memory (Miller et al., 2009). It was identical to the forward test only that the participant repeated the digits in a reverse order, for example,

saying 3, 9, 4 as 4, 9, 3. Span was scored as the longest string length of the digits repeated accurately.

### Statistical Learning

The test consists of a familiarization phase and a surprise test phase. In the familiarization phase, the participant heard 16 tri-syllabic nonce words in a continuous stream as in Marcus et al. (1999). The nonce words followed either the ABB or the AAB pattern. The A syllables were *lay*, *wee*, *jee*, and *day*; the B syllables were *dee*, *jay*, *lee*, and *way*. The nonce words (e.g., *laydeedee*) were concatenated together with no acoustic markers of boundaries, using synthesized speech software. Immediately after the familiarization phase was the test phase. The participant heard 13 tri-syllabic nonce words constructed with novel syllables. Five of the novel sequences followed the ABB pattern (e.g., *bakoko*) and the other five AAB (e.g., *babako*). Three fillers followed the ABA pattern (e.g., *bakoba*) with no adjacent repetition of the syllables. For each sound sequence, the participant indicated whether it sounded like a code from the alien language just heard. The design was counterbalanced: the ABB test sequences were novel to the participants familiarized with the AAB sequences, and vice versa. The score was the proportion of the total number of correct responses (out of 13).

### General Procedure

The participants were individually administered the tasks in a quiet room on campus. They were initially surveyed about their language learning background. They then took the two word-learning tasks. One word-learning task was administered first, and the other last. Between the two word-learning tasks were the three measures of learner variables. The order of the two word-learning tasks was counterbalanced. For the first word-learning task, the participants were randomly assigned to one of the two interval groups, consecutive or intervening. The consecutive interval group in the first word-learning task took the second word-learning task in an intervening condition and vice versa. Thus, with respect to the two word-learning tasks, there were four order of task administration: noun consecutive – verb intervening, noun intervening – verb consecutive, verb consecutive – noun intervening, and verb intervening – noun consecutive.

## Results

### Preliminary Analyses

Recall that the participants were randomly assigned to two groups. The two groups took the two CSWL tasks in different interval conditions. Preliminary analyses indicate that the two groups did not differ in any of the learner variables, $t(198) = 0.14$, $p > 0.05$ for forward digit span, $t(198) = 0.94$, $p > 0.05$ for backward digit span,

$t(198) = 0.12$, $p > 0.05$ for statistical learning, and $t(198) = 0.55$, $p > 0.05$ for English vocabulary.

### Noun Learning

The participants' responses were coded for accuracy and transformed into proportional scores. One point was awarded if the participant selected the referent previously selected on the Same probe, the referent not previously selected on the R-switch probe, and the familiarized referent on the F-switch probe. Figure 1 displays the overall mean performance for the three probes in noun learning. A 2×3 analysis of variance (ANOVA) was performed on the noun learning scores with interval (consecutive vs intervening) as a between-participants factor and probe as within-participants factor (Same, R-switch, F-switch). The assumption of homogeneity of variances was met based on Levene's test, $F(1, 198) = 2.9$, $p > 0.05$. The assumption of normal distribution was not met. However, we chose not to transform the data given that the sample sizes were sufficiently large and the distributions were unimodal and similar in shape (e.g., both were negatively skewed). The results of ANOVA revealed that the main effects were both significant, $F(1, 198) = 33.6$, $p < 0.001$, partial $\eta^2 = 0.15$ for interval, and $F(2, 396) = 14.9$, $p < 0.001$, partial $\eta^2 = 0.07$ for probe. The interaction between interval and probe was not significant, $F(2, 396) = 0.3$, $p > 0.05$. Post hoc analyses (with Bonferroni correction) revealed that the participants performed better on the Same probes than on the R-switch probes, $t(199) = 4.2$, $p < 0.001$, or on the F-switch probes, $t(199) = 5.3$, $p < 0.001$. Performance on the two switch probes did not differ, $t(199) = 1.2$, $p > 0.05$.

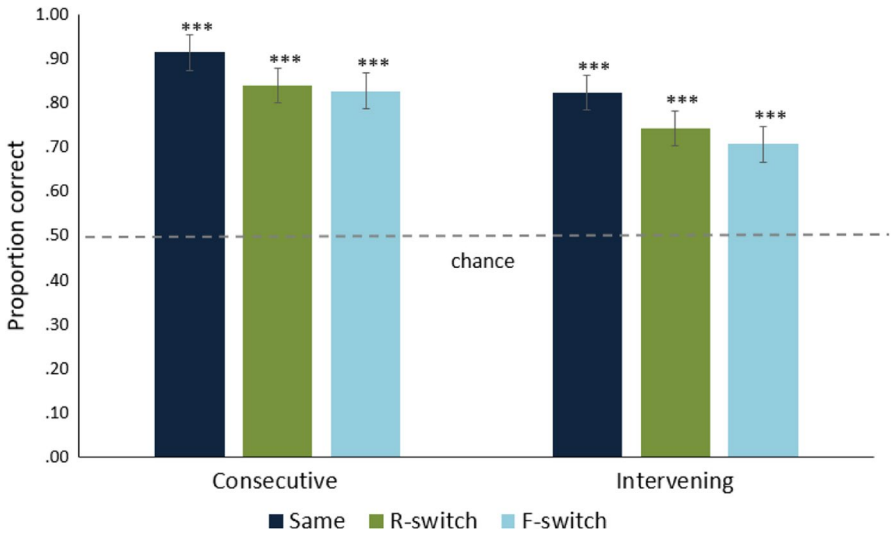



**Fig. 1** Mean proportion correct for the three probes administered to the two interval groups in noun learning. Error bars show 95% confidence intervals

To understand whether the participants aggregated and retrieved information across observations, their performance was compared to the chance level 0.50. The results revealed that mean performance was all above chance: for the consecutive interval group, $t(100) = 20.0$, $p < 0.001$ for the Same probes, $t(100) = 16.5$, $p < 0.001$ for the R-switch probes, and $t(100) = 16.5$, $p < 0.001$ for the F-switch probe; for the intervening interval group, $t(98) = 16.8$, $p < 0.001$ for the Same probes, $t(98) = 12.0$, $p < 0.001$ for the R-switch probes, and $t(98) = 17.0$, $p < 0.001$ for the F-switch probes. These results are more in accordance with the associative-learning model than the hypothesis-testing model.

One may argue that the data averaged over the 18 blocks of learning might not fully reveal the mechanism used for learning. As learning proceeded, the participants might learn that they had to reverse their referent selection in later trials and thus make efforts to encode the alternative referent on the scene. In other words, it is likely that the adult FL learners initially employed the hypothesis-testing model but shifted to the associative-learning model when they noticed that their initial hypothesis was not always correct. To assess this possibility, participants' scores on the very first of the three probes (Same, R-switch, and F-switch) were compared to chance. For the consecutive interval group, mean performance on the first probes was all above chance, $t(100) = 17.0$, $p < 0.001$ for the first Same probe, $t(100) = 9.4$, $p < 0.001$ for the first R-switch probe, and $t(98) = 5.9$, $p < 0.001$ for the first F-switch probe. The intervening interval group showed similar results, with mean performance all above chance for the first Same probe, $t(98) = 6.6$, $p < 0.001$, the first F-switch probe, $t(98) = 7.7$, $p < 0.001$, and the first F-switch probe, $t(98) = 2.8$, $p < 0.01$. Please see Fig. 2 for the performance on each of the first probe trials. There is no indication that the participants employed

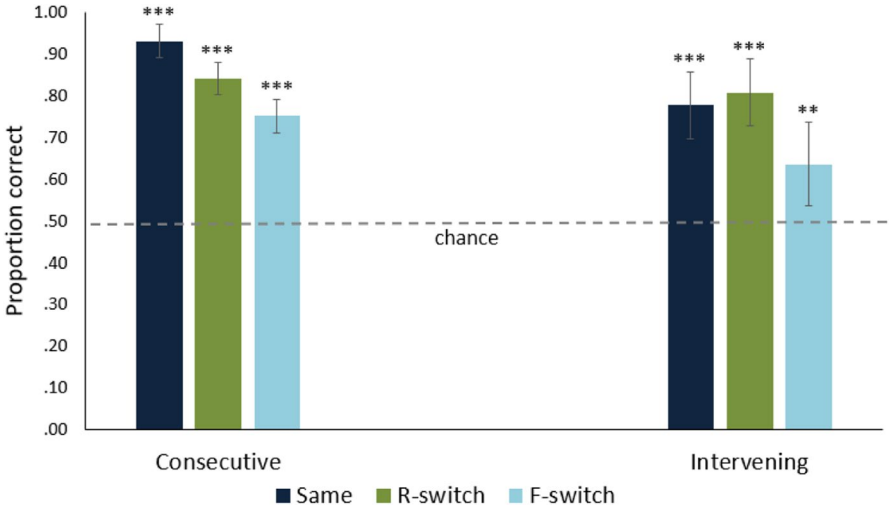



**Fig. 2** Mean proportion correct for the first of each probe in noun learning. Note. Error bars show 95% confidence intervals

the hypothesis-testing model in the noun learning task. They retained the possible referent of their guess and tracked the referent they did not choose.

### Verb Learning

Figure 3 displays the overall mean performance for the three probes administered to the two interval groups in verb learning. Performance on verb learning was subject to a 2×3 ANOVA, with interval as a between-participants factor (consecutive vs intervening) and probe as a within-participant factor (Same, R-switch, F-switch). The result of the Levene's test indicated that the assumption of homogeneity of variance was met $F(1, 198) = 3.1$, $p > 0.05$. The assumption of normal distribution was not met. We did not transform the data as the sample sizes were large and the distributions were unimodal and similar in shape (e.g., both were negatively skewed). The results of ANOVA revealed significant main effects of interval, $F(1, 198) = 23.7$, $p < 0.001$, partial $\eta^2 = 0.11$, and probe, $F(2, 396) = 29.4$, $p < 0.001$, partial $\eta^2 = 0.13$. The interaction between learning condition and probe was not significant $F(2, 396) = 1.6$, $p > 0.05$. Post hoc analyses revealed that the participants performed better on the Same probes than on the R-switch probe, $t(199) = 7.9$, $p < 0.001$, or on the F-switch probe, $t(199) = 5.5$, $p < 0.001$. Verb learning on the two switch probes did not differ, $t(199) = 2.0$, $p = 0.046$ with an adjusted $p$ value 0.017 (Bonferroni correction). These results were similar to the results obtained for noun learning.

The results of chance level analyses on the scores averaged over the 18 blocks revealed that participants' verb-learning performance was all above chance: for the consecutive interval group, $t(98) = 28.4$, $p < 0.001$ for the Same probes, $t(98) = 11.4$, $p < 0.001$ for the R-switch probes, and $t(98) = 14.7$, $p < 0.001$ for the F-switch probe; for the intervening interval group, $t(100) = 15.0$, $p < 0.001$ for the Same probes,

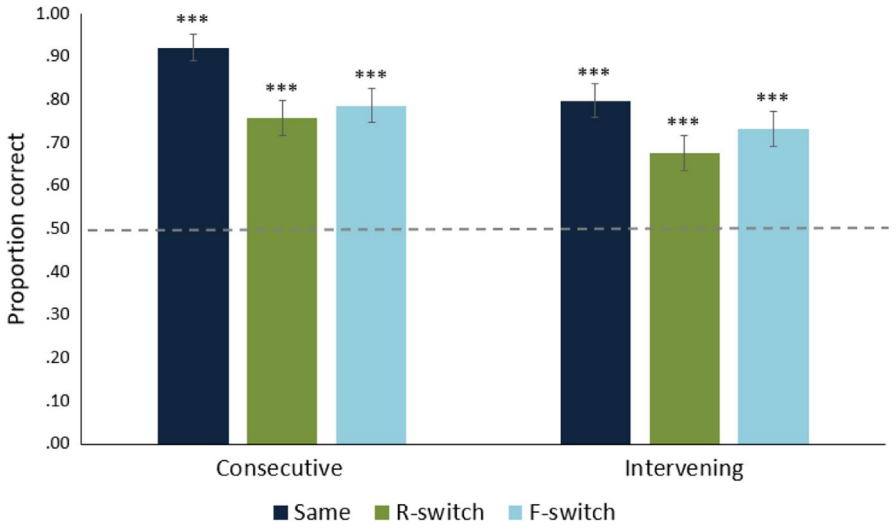



**Fig. 3** Mean proportion correct for the three probes administered to the two interval groups in verb learning. Error bars show 95% confidence intervals

$t(100) = 7.8$, $p < 0.001$ for the R-switch probes, and $t(100) = 11.5$, $p < 0.001$ for the F-switch probes.

Figure 4 displays mean performance for the first probes administered to the two interval groups in verb learning. The results of chance level analyses on the very first of each probe revealed that participants in the consecutive interval group performed above chance on all the first probes, $t(98) = 23.2$, $p < 0.001$ for the Same probe, $t(98) = 4.0$, $p < 0.001$ for the R-switch probe, and $t(98) = 5.6$, $p < 0.001$ for the F-switch probe. Participants in the intervening interval group performed above chance for the first Same probe, $t(110) = 6.2$, $p < 0.001$, and for the first F-switch probe, $t(110) = 4.4$, $p < 0.001$, but not for the first R-switch probe, $t(110) = 1.7$, $p = 0.09$. These results indicate that the participants faced challenge in tracking all the co-occurrences in their first encounter with a verb.

### Noun and Verb

The next question is whether verb learning is more difficult than noun learning. As shown by Figs. 1 and 3, the performance patterns for noun and verb learning are similar, except that verb learning performance was relatively low on R-switch probes. A 2×3 ANOVA was conducted on the performance of the two tasks, with word type (noun vs verb) as a between-participants factor and probe as a within-participants factor. The assumption of homogeneity of variances was met for the consecutive interval condition, Levene's $F(1, 198) = 0.01$, $p > 0.05$ and for the intervening interval condition, Levene's $F(1, 198) = 0.00$, $p > 0.05$. For the consecutive interval condition, there was a significant effect of probe, $F(2, 396) = 22.6$, $p < 0.001$, partial $\eta^2 = 0.10$. The effect of word type was also significant, $F(1, 198) = 5.7$, $p = 0.02$, partial $\eta^2 = 0.03$. The interaction effect was approaching the significant level, $F(2,$

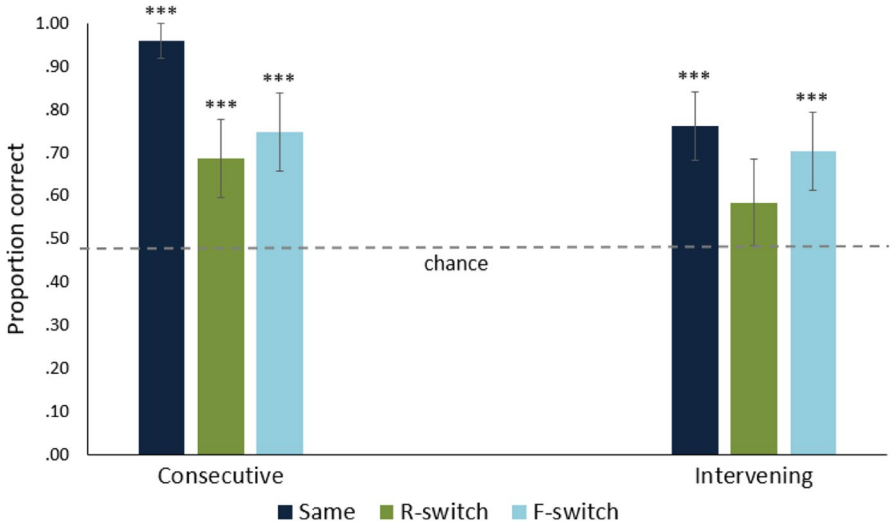



**Fig. 4** Proportion correct for the first of each probe in verb learning. Note. Error bars show 95% confidence intervals

396) = 2.5, $p = 0.08$. For the intervening interval condition, there was a significant effect of probe, $F(2, 396) = 17.7$, $p < 0.001$, partial $\eta^2 = 0.08$. The effect of word type was not significant, $F(1, 198) = 1.5$, $p > 0.05$. The interaction effect was marginally significant, $F(2, 396) = 3.0$, $p = 0.052$, partial $\eta^2 = 0.02$. The marginal interaction effect was due to the relatively lower performance in verb learning than in noun learning on the R-switch probes, $t(198) = 2.7$, $p < 0.01$ for the consecutive intervals and $t(198) = 2.2$, $p = 0.03$ for the intervening intervals.

## Learner Variables

Table 3 presents a correlation matrix along with means and standard deviations of the variables. The correlations above the diagonal are the zero-order correlations, and the correlations below the diagonal are partial correlations controlling for group differences. Performances in the two word-learning tasks were not correlated probably due to the design of the present study. In the present study, those assigned to the consecutive interval group in one learning task were assigned to the intervening interval group in the other. With respect to individual learner variables, none of them was correlated with noun learning. However, the two phonological STM measures were moderately correlated with verb learning, $r = 0.18$ for forward digit span and $r = 0.27$ for backward digit span. To examine the relative contribution of the two measures, a hierarchical regression was conducted with verb learning scores as the dependent variable. Interval group was entered on the first step. Forward and backward digit span scores were entered on the second step with stepwise selection. After partialling out the difference due to interval groups, backward digit span predicted 6% of unique variance in CSWL of verbs. Forward digit span was not a significant predictor in the regression model.

Notably in statistical learning, a large proportion of the participants performed around the chance level. These data points were the noise when individual performance was considered in the correlational analysis (Siegelman et al., 2017). To complement the correlational analysis, we used the binomial distribution to determine the minimal number of successful trials needed to present significantly above-chance learning at the individual level. According to the binomial distribution, the

**Table 3** Means, standard deviations, and correlations for word learning and learner variables ($N = 200$)

| | Mean | SD | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| 1. Noun learning | .81 | .14 | | − .02 | .07 | .08 | .05 | − .05 |
| 2. Verb learning | .78 | .13 | .12 | | .18* | .27*** | .06 | .04 |
| 3. Forward digit span | 9.70 | 1.5 | .08 | .18* | | .43*** | .10 | .07 |
| 4. Backward digit span | 7.81 | 2.01 | .12 | .27*** | .43*** | | .13 | .13 |
| 5. Statistical learning | .61 | .13 | .03 | .08 | .10 | .13 | | − .22** |
| 6. English vocabulary | 83.64 | 35.43 | − .03 | .02 | .07 | .13 | − .21** | |

*$p < .05$; **$p < .01$; ***$p < .001$

The upper diagonal represents zero-order correlations; the lower diagonal represents partial correlations controlling intervals (consecutive vs intervening)

probability of 9 hits out of 13 with random guessing at the individual level is less than 0.05. We thus classified those who succeeded in 9 or more than 9 trials as statistical learners ($N=68$), and those below as non-statistical learners ($N=132$). A multivariate analysis of variance (MANOVA) was conducted on the word learning scores from the Same, R-switch, F-switch probes for each of the word learning tasks. Results of MANOVA revealed no differences between the two groups in noun learning, $F(3, 196)=0.1$, $p>0.05$, or in verb learning, $F(3, 196)=1.2$, $p>0.05$, further confirming the null relationship between statistical learning and word learning performance in the correlational analyses. As a final check, we ran correlational analyses between statistical learning and word learning scores for the 68 participants dubbed as statistical learners. The correlations were not significant, $r=0.05$, $p>0.05$ for statistical learning and noun learning and $r=-0.09$, $p>0.05$ for statistical learning and verb learning.

## Discussion

Overall, the results of the present study extended our understanding of CSWL by FL learners in previous studies (Hu, 2017). First, adult FL learners used cross-situational information to home in the meaning of an ambiguous word. They carried the partial information about word-referent co-occurrences to subsequent learning even when they were asked to propose a meaning for the word under uncertainty. Second, the performance pattern was similar in noun and verb learning though tracking multiple event referents for a verb with intervening intervals was more challenging than tracking multiple object referents. Third, the three learner variables, English vocabulary, phonological STM, and statistical learning, have limited predictive value of CSWL performance in adult FL learners. Only phonological STM was moderately correlated with verb learning performance.

The design of the study was unique in several aspects. First, the participants were asked to select one referent under uncertainty during exposure. In a typical CSWL paradigm, participants select referents after multiple observations of the word-referent co-occurrences (e.g., Yu & Smith, 2007; Yurovsky et al., 2014). The current paradigm forced the participants to assign one meaning each time they heard a word (Woodard et al., 2016; Yurovsky & Frank, 2015). In a sense, this task favored the hypothesis-testing mode; however, it reflected the communicative demands in naturalistic settings, where listeners rapidly select one meaning for an ambiguous word rather than maintain multiple meanings until some deterministic cues appear (Swinney, 1979). The paradigm used in the present study allowed for an examination of whether word learning still proceeded in a state of partial knowledge even when a meaning was tentatively assigned to a word.

Second, the target referents for the novel words were not pre-determined by the experimenter. They were determined by the selection each participant made during the exposure trial and were tested by various probes in a subsequent trial. On the Same probe, the target referent was the one the participant previously selected; on the R-switch probe, the target referent was the one the participant did not previously select; on the F-switch probe, the target referent was not the one the participant

previously selected. Referential ambiguity was resolvable at the moment when a probe was given if the participants tracked all the word-referent co-occurrences across the observations. The R-switch probe was used to examine whether learners tracked the alternative referent and recovered from their incorrect mapping; the F-switch probe enabled further examination of whether the initial selection made by the participants was a real word-referent mapping process, as opposed to a simple referent selection process.

### Associative Learning Versus Hypothesis Testing

Consistent with previous studies on L1 learners (e.g., Medina et al., 2011; Yurovsky & Frank, 2015), word learning performance shown by the FL learners was best on the Same probe than the other two switch probes. The proportion correct was 90% for both noun and verb learning on the Same probe when there were no intervening trials. Although the proportion correct dropped significantly when the exposure and the probe trials were intervened by two familiar word trials, the proportion correct was still over 80% for both noun and verb learning. These results indicate that the participants were reliably engaged in the learning process, albeit starting with a random guess.

Performance on the F-switch probes provides some insight into the robustness of CSWL in adult FL learners. The results on F-switch probes indicated that the random guess made by the participants was a conjecture about the word's referential meaning rather than simply a random selection of an animal or a scene. The participants registered not only the information about the animal or the scene they selected but also the mapping between the word and the referent. They carried the mapping information across trials and used that information to disambiguate another novel word, rejecting a second novel word for a referent previously assigned to a word, over at least 70% of the time. Adult FL learners seemed to proceed the CSWL task in a conventional mutual exclusivity manner (Markman & Wachtel, 1988), but with some extensions. Recall that the participants were asked to select a referent for a novel word under uncertainty during the exposure trial. Subsequently, at the moment of hearing the second novel word from the F-switch probe, no feedback was yet given to the referential meaning of the first novel word. Neither of the referents in display had an ascertained name, be it the one that had been selected or the one that appeared in the familiarization trial. Nevertheless, the participants still performed in a mutual exclusivity manner, rejecting to apply the second word to something with a conjectured but unascertained meaning. The non-random selection on the F-switch probes indicated the robustness of the effect in the first random selection upon subsequent learning via CSWL. The inference-like performance can be understood in both the hypothesis-testing and the associative-learning frameworks (Yu & Smith, 2012). The hypothesis-testing model predicts a selection of a referent that has not been hypothesized to be linked to another word; the associative-learning model predicts a selection of a referent with relatively weak association with another word so as to "optimize the whole data set" (Yu & Smith, 2012).

The critical data to the disentanglement of the two learning models were from the R-switch probes. As indicated earlier, when an R-switch probe was given, neither of the referents was the one that the participant selected during exposure. The only referent that previously co-occurred with the target word was the one that the participant did not select. If the participants followed the hypothesis-testing model, maintained one single conjecture at a time and disregarded the alternative referent, they should not be able to recover from their incorrect hypothesis and select the referent that previously co-occurred with the target word (Trueswell et al., 2013; Woodard et al., 2016). On the other hand, if they followed the associative-learning model and tracked the co-occurrence probabilities for the whole data set, they should be able to recover from their incorrect guess. Different from the results from the study on children learning L1 novel names (Woodard et al., 2016; see also Trueswell et al., 2013; Medina et al., 2011), adult FL learners in the present study performed better than expected by chance on the R-switch probes in noun as well as in verb learning. The proportion correct was 84% and 74% in noun learning and 76% and 67% in verb learning for the two interval groups, similar to their performance on the F-switch probes. These results indicated that the FL learners carried information about the co-occurrence structure from one observation to another. In addition to the chosen referent, they also encoded some partial information about the unchosen one. And this was the case despite that the participants were asked to make an immediate decision about a word's meaning under uncertainty the first time when they heard the word.

The results of the present study support the associative-learning model of CSWL in FL learners engaged in one-to-two mapping. The participants aggregated information about word-referent co-occurrences across observations to guide subsequent learning not only in noun learning but also in verb learning, as demonstrated by their above-chance performance on the two switch probes. Their CSWL was robust. The information they encoded during exposure enabled them to recover from their mistakes (as indicated by their performance on the R-switch probes) and to further resolve referential ambiguity of another novel word (as indicated by their performance on the F-switch probes).

### Noun Versus Verb

The FL learners performed rather similarly in the CSWL tasks for nouns and verbs. Both involved accumulation of partial knowledge about all the potential mappings across multiple learning trials, even under the situation where they had to assign a tentative referent to a word under uncertainty. The only difference was in the performance on the R-switch probes. The participants were disrupted to a greater extent in recovering the referent from their previous incorrect mapping in verb learning than in noun learning. In addition, they failed to perform above chance on the very first R-switch probe in verb learning with two intervening trials. Theoretically speaking, tracking verb referents under uncertainty across multiple scenes is more demanding than tracking noun referents. A novel verb can denote a cluster of elements in an event, such as the result, the tool,

the manner, or the direction of an action (Childers, 2011). In principle, there are more than two potential construals derivable from two dynamic scenes upon hearing a verb. Suppose one event involves a rabbit stamping on its shadow and making its shadow larger and rounder in size and the other involves a fish blowing out yellow light rays. The meaning of the verb can be the stamping action, the resultative state of the shadow or the two elements in combination in the first event. It can be the blowing action, the emission of the rays, or the two elements in combination in the second event. What is more, in the present study, the performers of the actions changed across observations, making it difficult to abstract the consistency of the event elements (Earles & Kersten, 2017; Maguire et al., 2008). When the exposure and the probe trials were not in an immediate succession, there was a high likelihood that the participants lost track of the event referent not initially chosen, as evidenced by the at-chance performance on the first R-switch probe for a verb. Losing track of a verb's referent on the first R-switch probe might prompt the participants to encode the alternative referent more elaborately for the remaining task. This may explain why the performance on the first R-switch probe of verb learning with intervening intervals did not exceed chance while the overall mean performance did.

One may argue that the noticeable poor performance in certain conditions of CSWL is not a matter of losing track. Rather, it can be that the participants were flexible in recruiting different mechanisms in response to different task demands (Yurovsky & Frank, 2015). When the word-referent co-occurrences are not complex, learners attend to the statistics structure of the stimuli and track multiple referents. When the co-occurrences are complex, learners track one referent at a time. However, this account cannot easily explain why only verb learning on the first R-switch probe with intervening intervals was at chance. First, if the participants perceived the difficulty representing word-referent co-occurrences in verb learning at the very beginning of the task and started out the verb learning task by tracking only one event from the outset, they should have also performed at chance on the first R-switch probe when there were no intervening intervals. However, the consecutive interval group performed above chance on the first R-switch probe, indicating they did not start out the verb learning task by tracking one event referent only. Second, if the participants did not perceive the difficulty at the beginning of the task but sensed the difficulty when they were given an R-switch probe following two intervening trials, they should have shown the evidence of tracking only one referent in the later parts of the verb learning task. However, the overall mean performance for verb learning was above chance either with or without intervening intervals. These results indicated that the participants tracked all possible mappings in verb learning at the outset. They only lost track on the very first R-switch probe.

Taken together, the associative-learning model is more parsimonious and coherent than the hypothesis-testing model in accounting for the performance pattern in noun and verb learning by the FL learners in the present study. The at-chance performance was evident only on the first R-switch probe in verb learning following intervening intervals. All the other performances in verb learning and all the performances in noun learning were above what was expected by chance. These results indicate that FL learners tracked all the word-referent co-occurrences not only in

noun learning but also in verb learning, though they sometimes lost track when the situations were intervened by other trials in verb learning.

### Learner Variables

The three learner variables examined in the present study have limited predictive value of CSWL performance in adult FL learners. English vocabulary and statistical learning were not associated with either noun learning or verb learning. Phonological STM was moderately correlated with verb learning, but not with noun learning. The null correlation between phonological STM and CSWL of nouns is similar to the results on child FL learners (Hu, 2017). The new finding in the present study is the correlation between phonological STM and verb learning performance. In particular, backward digit span predicted verb learning after partialling out differences due to forward digit span, suggesting that learning verbs via CSWL by adult FL learners may require the more active processing component of phonological STM over and above passive storage. This is not to say that noun learning is not resources-dependent. Participants' performance on both nouns and verbs was significantly lower when memory load increased from null to two intervening trials, indicating that both noun and verb learning drew on memory resources. It is that individual learner differences in phonological STM came to play a role only when the demands to package, encode, and track word-referent pairings were at the edge of overwhelming FL learners' memory capacities. As noted earlier, CSWL of verbs is more complicated than CSWL of nouns. The complexity in verb learning is multiplied beyond simply linking one word to two possible referents. Verb learning via CSWL requires packaging multiple perceptual elements from a dynamic scene and aligning the yet-uncertain elements with a novel word from one observation to another. The mapping could be one-to-many perceptual elements packaged from each of the two dynamic scenes. The complexity in CSWL of verbs can require active efforts to encode and track word-scene mappings and may explain its unique association with the processing component of phonological STM.

The lack of correlation between statistical learning and CSWL performance might be in part due to the aspect of statistical structure captured by the statistical learning task in the present study. In the present study, the statistical learning task required extracting a phonological pattern (AAB or ABB) from a string of sound sequence. In contrast, the CSWL task required integrating co-occurrence information from two separate observations. The extraction and integration of statistical information are considered to be two independent processes (Erickson & Thiessen, 2015; Thiessen et al., 2013), which may account for the lack of the correlations between statistical learning and CSWL performance in the present study.

The lack of correlations between English vocabulary and CSWL performance in adult FL learners is interesting. In the present study, participants differed considerably in their English vocabulary scores (ranging from 8 to 168). Yet such a wide range of variability in English vocabulary was not associated with differences in CSWL. These results suggest that CSWL may be a fundamental learning ability,

which does not rely on complex linguistic knowledge (Poepsel & Weiss, 2016; Yu & Smith, 2012).

## Conclusion

The results of the current study are more consistent with the associative-learning view than with the simple hypothesis-testing view of CSWL by adult FL learners. Adult FL learners encoded not only the information about the selected mapping but also "an approximation to the co-occurrence statistics" (Yurovsky & Frank, 2015), even when they were instructed to make an overt decision about the word's referent at each observation. They made the strongest association with the proposed referent and weaker association with the alternative referent. Encoding all the potential co-occurrences facilitated recovering from incorrect mappings (as shown by the R-switch probes) and disambiguating another novel word in subsequent learning (as shown by the F-switch probes). They sometimes failed to track all the potential co-occurrences when the complexity of the co-occurrences was high, for example, when learning verbs with intervening trials. Verb learning (but not noun learning) was specifically correlated with the processing component of phonological STM, probably due to the intricacies of verb semantics in CSWL. To adult FL learners, CSWL via tracking co-occurrences appears to be a fundamental learning mechanism, unaffected by statistical learning and English vocabulary levels.

The results of the present study cannot be generalized to CSWL by child FL learners. Children and adults learn languages differently (e.g., Fitneva & Christiansen, 2017; Hudson Kam & Newport, 2005; Ramscar et la., 2013; Roembke et al., 2018). CSWL may operate with a different mechanism in young children, more in an item-based manner than in a system-based manner (Aravind et al., 2018; Stevens et al., 2017; Trueswell et al., 2013; Woodard et al., 2016). Indeed, the study was originally taken to investigate CSWL by child FL learners. However, we terminated data collection with children due to COVID-19 lockdown of elementary schools in Taiwan. Further work needs to examine whether child FL learners employ a different mechanism in CSWL. In addition, CSWL in the present study involves mapping between one word and two referents. It is necessary to examine whether a hypothesis-testing model or a hybrid of two models comes to operate when the mapping complexity increases, for example, when two novel words co-occur with two novel referents.

Finally, even though the results of the current study demonstrate that FL learners capitalize on the co-occurrence pattern in resolving referential ambiguity, opportunities for true cross-situational learning are rare in FL classrooms, where vocabulary instruction typically involves one-to-one mapping. We may question whether we have long underestimated FL learners' ability to tolerate and solve referential ambiguity, thus biasing FL vocabulary instruction almost exclusively toward memorization of one-to-one mapping between word and referent in the classroom. In light of the evidence shown in the present study and considering the growing popularity of immersion or bilingual type of education in many non-English speaking countries, it should be promising for future studies to further delineate the mechanism

FL learners use to monitor information across moments of exposure and to resolve referential ambiguity in and across individually ambiguous situations.

## Appendix A

Nonce words for noun learning.

Target nouns: borale, pintry, hebat, trafer, chesus, zonet, unber, griced, haser, dattle, chantef, doron, crinklo, ferpent, songa, royat, namper, motate.

F-switch probes: pango, benor, liden, fruin, carep, jatal.

## Appendix B

Nonce words for verb learning.

Target verbs: tasher, cament, debine, hortal, larned, gelon, narker, burver, selped, fespit, meaped, claced, plimat, mennel, abron, davest, garol, pravest.

F-switch probes: ragon, pisgress, erpand, hitfer, kanced, inchem.

**Author Contribution** Not applicable.

**Funding** The study was supported by a grant from the Ministry of Science and Technology, Taiwan, R. O. C. (MOST 108–2410-H-845–013-MY2).

**Data Availability** The datasets generated during and/or analyzed during the current study are available from the author on reasonable request.

### Declarations

**Ethics Approval** Ethical approval for this study was obtained from the institutional review board at University of Taipei (IRB-2019–033).

**Competing Interests** The author declares no competing interests.

## References

Aravind, A., de Villiers, J., Pace, A., Valentine, H., Golinkoff, R., Hirsch-Pasek, K., Iglesias, A., & Wilson, M. S. (2018). Fast mapping word meanings across trials: Young children forget all but their first guess. *Cognition, 177*, 177–188.

Aslin, R. N. (2017). Statistical learning: A powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science, 8*(1–2), e1373.

Brooks, P. J., Kwoka, N., & Kempe, V. (2017). Distributional effects and individual differences in L2 morphology learning. *Language Learning, 67*, 171–207.

Chen, C. H., Gershkoff-Stowe, L., Wu, C. Y., Cheung, H., & Yu, C. (2017). Tracking multiple statistics: Simultaneous learning of object names and categories in English and Mandarin speakers. *Cognitive Science, 41*, 1485–1509.

Childers, J. B. (2011). Attention to multiple events helps two-and-a-half-year-olds extend new verbs. *First Language, 31*, 3–22.

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology, 38*, 967–978.

Childers, J. B., Paik, J. H., Flores, M., Lai, G., & Dolan, M. (2017). Does variability across events affect verb learning in English, Mandarin, and Korean? *Cognitive Science, 41*, 808–830.
Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge University Press.
Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test-Third edition: Manual*. Circle Pines, MN: American Guidance Services.
Earles, J. L., & Kersten, A. W. (2017). Why are verbs so hard to remember? Effects of semantic context on memory for verbs and nouns. *Cognitive Science, 41*, 780–807.
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.
Fitneva, S. A., & Christiansen, M. H. (2017). Developmental changes in cross-situational word learning: The inverse effect of initial accuracy. *Cognitive Science, 41*, 141–161.
Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23*, 83–94.
Gleitman, L., & Gleitman, H. (1992). A picture is worth a thousand words, but that's the problem: The role of syntax in vocabulary acquisition. *Current Directions in Psychological Science, 1*, 31–35.
Hamada, M. (2014). The role of morphological and contextual information in L2 lexical inference. *The Modern Language Journal, 98*, 992–1005.
Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*, 303–317.
Hu, C. F. (2012). Fast mapping and deliberate word-learning by EFL children. *The Modern Language Journal, 96*, 439–453.
Hu, C. F. (2017). Resolving referential ambiguity across ambiguous situations in young foreign language learners. *Applied Psycholinguistics, 38*, 633–656.
Hu, C. F., & Maechtle, C. (2021). Construction learning by child learners of foreign language: Input distribution and learner factors. *The Modern Language Journal, 105*, 335–354.
Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*, 151–195.
Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R. M., & Shigematsu, J. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development, 79*, 979–1000.
Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology, 48*, 171–184.
Kidd, E., & Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development, 87*, 184–193.
Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences, 22*, 154–169.
Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science, 11*, 628–634.
Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*(5398), 77–80.
Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science, 18*, 387–391.
Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20*, 121–157.
Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences, 108*, 9014–9019.
Miller, K. M., Price, C. C., Okun, M. S., Montijo, H., & Bowers, D. (2009). Is the N-back task a valid neuropsychological measure for assessing working memory? *Archives of Clinical Neuropsychology, 24*, 711–717.
Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning, 62*, 302–331.
Monaghan, P., Mattock, K., Davies, R. A., & Smith, A. C. (2015). Gavagai is as gavagai does: Learning nouns and verbs from cross-situational statistics. *Cognitive Science, 39*, 1099–1112.
Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *Tesol Quarterly, 37*, 645–670.

Poepsel, T. J., & Weiss, D. J. (2016). The influence of bilingualism on statistical word learning. *Cognition, 152*, 9–19.
Ramscar, M., Dye, M., & Klein, J. (2013). Children value informativity over logic in word learning. *Psychological Science, 24*, 1017–1023.
Roembke, T. C., & McMurray, B. (2016). Observational word learning: Beyond propose-but-verify and associative bean counting. *Journal of Memory and Language, 87*, 105–127.
Roembke, T. C., & McMurray, B. (2021). Multiple components of statistical word learning are resource dependent: Evidence from a dual-task learning paradigm. *Memory & Cognition, 49*, 984–997.
Roembke, T. C., Wiggs, K. K., & McMurray, B. (2018). Symbolic flexibility during unsupervised word learning in children and adults. *Journal of Experimental Child Psychology, 175*, 17–36.
Schwanenflugel, P. J., Stahl, S. A., & Mcfalls, E. L. (1997). Partial word knowledge and vocabulary growth during reading comprehension. *Journal of Literacy Research, 29*, 531–553.
Scott, R. M., & Fisher, C. (2012). 2.5-Year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition, 122*, 163–180.
Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods, 49*, 418–432.
Stevens, J. S., Gleitman, L. R., Trueswell, J. C., & Yang, C. (2017). The pursuit of word meanings. *Cognitive Science, 41*, 638–676.
Sunderman, G., & Kroll, J. F. (2009). When study-abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics, 30*, 79–99.
Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645–659.
Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin, 139*, 792–814.
Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology, 66*, 126–156.
Van Zeeland, H. (2014). Lexical inferencing in first and second language listening. *The Modern Language Journal, 98*, 1006–1021.
Vlach, H. A., & DeBrock, C. A. (2017). Remember dax? Relations between children's cross-situational word learning, memory, and language abilities. *Journal of Memory and Language, 93*, 217–230.
Vlach, H. A., & Sandhofer, C. M. (2014). Retrieval dynamics and retention in cross-situational statistical word learning. *Cognitive Science, 38*, 757–774.
Woodard, K., Gleitman, L., & Trueswell, J. (2016). Two-and three-year-olds track a single meaning during word learning: Evidence for propose-but-verify. *Language Learning and Development, 12*, 252–261.
Yu, C., & Smith, L. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*, 414–420.
Yu, C., & Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review, 119*, 21–39.
Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition, 145*, 53–62.
Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review, 21*, 1–22.