The Independent Effect of Transition Probability on Speech Statistical Learning

Authors: Yu Wenbo, Wang Lu, Wang Tianlin, Liang Dandan, Liang Dandan

Date: 2025-04-25T00:00:00+00:00

Abstract

Experiment 1 employed a 2 (word length: two-syllable, three-syllable) × 3 (type of forced-choice pair: target word vs. cross-boundary word, target word vs. non-word, cross-boundary word vs. non-word) within-subjects design, and found that accuracy in three-syllable cross-boundary word vs. non-word forced-choice pairs was significantly higher than in the corresponding two-syllable pairs, while accuracy in three-syllable target word vs. cross-boundary word forced-choice pairs was marginally significantly lower than in the corresponding two-syllable pairs. In contrast to Experiment 1, which first combined syllables into words before synthesizing the artificial language, Experiment 2 directly concatenated the artificial language using syllables as units; the trials in the forced-choice task were completely identical to those in Experiment 1, and it was found that participants showed no learning effect. Since the contrasting conditions in Experiment 1 differed only in transitional probability, whereas Experiment 2 eliminated interference from syllable combination preferences, this study demonstrates the independent role of transitional probability in statistical learning.

Full Text

The Independent Effect of Transitional Probability on Verbal Statistical Learning

Wenbo Yu[1], Tianlin Wang[2], Hetong Qi[1], Dandan Liang[1],[3]

[1]School of Chinese Language and Culture, Nanjing Normal University
[2]School of Education, University at Albany, State University of New York
[3]Interdisciplinary Research Center for Linguistic Science, University of Science and Technology of China

Abstract

Experiment 1 employed a 2 (word length: disyllabic, trisyllabic) × 3 (forced-choice pair type: target word vs. partword, target word vs. nonword, partword vs. nonword) within-subjects design. The results revealed that accuracy in trisyllabic partword vs. nonword forced-choice pairs was significantly higher than in the corresponding disyllabic pairs, while accuracy in trisyllabic target word vs. partword pairs was marginally lower than in the corresponding disyllabic pairs. Unlike Experiment 1, which combined syllables into words before synthesizing the artificial language, Experiment 2 directly concatenated the artificial language at the syllable level while using identical test trials as Experiment 1, and found that participants showed no learning effect. Since the comparisons in Experiment 1 differed only in transitional probability, while Experiment 2 eliminated interference from syllable combination preferences, this study demonstrates the independent role of transitional probability in statistical learning.

Keywords: Transitional Probability, Word Frequency, Statistical Learning, Forced-Choice Task

Introduction

Statistical learning refers to the process by which individuals discover statistical information from temporal and spatial input and use this information to learn novel patterns (Saffran et al., 1996; Frost et al., 2020; Yu et al., 2021a, 2021b). Considered a crucial human cognitive ability, statistical learning has strong associations with spoken word segmentation and lexical-semantic acquisition (Estes et al., 2007, 2015; Newport, 2016; Raviv & Arnon, 2018; Saffran & Kirkham, 2018; Bogaerts et al., 2020; Siegelman, 2020). Although statistical learning mechanisms manifest across different modalities, research in the auditory modality has been most extensive (Mirman et al., 2008; Wang & Saffran, 2014; Gómez et al., 2017).

Forward transitional probability (TP) is one of the core probabilistic cues in statistical learning. Researchers hypothesize that the human brain can track transitional relationships between syllables, forming more robust memory representations for syllable combinations with higher transitional probabilities, thereby achieving learning. In constructing experimental materials, target words are pseudo-randomly concatenated according to specific rules and appear as cohesive units. The transitional probabilities between adjacent syllables within these words are high. As illustrated in Figure 1 [Figure 1: see original paper], where each letter represents a syllable, the transitional probabilities of AB and BC within the target word ABC are 1. In contrast, syllable combinations formed at the boundaries between two target words are called partwords. Since target word ABC can be followed randomly by other target words JKL, GHI, or DEF, the transitional probability of the first two syllables in the partword CJK is 1/3, making the overall transitional probability of the partword (0.67) lower than that of target words. After the learning phase, participants' learning outcomes are assessed. If participants can distinguish target words from partwords or nonwords (nonwords are constructed by combining syllables from different target words, with internal transitional probabilities and word frequencies of 0), a learning effect is considered to have occurred.

Figure 1 [Figure 1: see original paper] Schematic diagram of verbal statistical learning materials. Experiments with infants, adults, and second language learners have all validated the statistical learning mechanism hypothesis (Saffran et al., 1996; Erickson, Thiessen, & Estes, 2014; Estes et al., 2015; Estes & Lew-Williams, 2015; Potter, Wang, & Saffran, 2017). However, few studies have addressed whether individuals rely solely on transitional probability for segmentation and learning. As shown in the figure above, although partwords (e.g., CJK) have lower transitional probabilities than target words (e.g., ABC), they also appear more frequently in the artificial language. Assuming an artificial language synthesized from four meaningless trisyllabic target words, the transitional probability of partwords formed at word boundaries is 0.33. When each target word appears 100 times, partwords (CJK, HID) appear $100 \times 0.33 = 33$ times in the artificial language. Thus, target words and partwords differ on both transitional probability and word frequency, making it impossible to determine whether transitional probability alone can influence statistical learning effects.

Some studies have attempted to examine the effect of transitional probability on learning while controlling for word frequency (Aslin et al., 1998; Estes, 2012). In infant experiments, researchers typically halve the presentation frequency of half the target words during artificial language construction, then use only partwords formed from high-frequency target words and low-frequency target words as test materials in the head-turn preference procedure. Since these items have identical word frequencies in the artificial language, dishabituation effects would demonstrate that individuals can complete statistical learning tasks relying solely on transitional probability. Additionally, Endress and Langus (2016) found through comparative analyses that transitional probability likely exerts a stronger effect on statistical learning than frequency. However, these limited studies have several shortcomings. First, most research uses target words of equal length (disyllabic or trisyllabic) when constructing artificial languages, which does not reflect natural language where word lengths vary (Frost et al., 2020) and may cause participants to develop rhythmic expectations early in the learning phase, easily discovering that all words share the same length and thereby inflating learning effects (Hoch et al., 2013). Second, artificial languages composed of equal-length target words severely restrict the range of transitional probabilities available for testing. For example, with four trisyllabic target words, all partwords have a transitional probability of 0.67 ((1+0.33)/2), while target words have a transitional probability of 1, allowing examination of only this single contrast in forced-choice tasks. However, if transitional probability can indeed independently affect learning effects, participants' accuracy should differ across forced-choice pairs composed of target words, partwords, and nonwords, though no empirical research has yet investigated this. Finally, Yu et al. (2021b) examined the independent effect of transitional probability using a between-subjects design, reducing transitional probabilities by adding filler words to ensure options differed only in transitional probability, not word frequency. Although increased transitional probability facilitated statistical learning effects in adult experiments, the high and low transitional probability conditions were between-subjects variables. The high-transitional-probability condition had shorter learning durations, while the low-transitional-probability condition had longer learning times due to added filler words, exposing target words to greater interference and inhibition in memory. Consequently, the explanation for lower learning effects is not unique.

This study addresses these issues by designing target words of mixed lengths. First, we used disyllabic and trisyllabic target words to construct a mixed-length artificial language. In the forced-choice task, we created target words, partwords, and nonwords of both lengths, pairing them under the same length constraint to form three types of forced-choice pairs (target word vs. partword, target word vs. nonword, partword vs. nonword). Ensuring equal length within each forced-choice trial is necessary because Mandarin native speakers have rhythmic expectations and prefer disyllabic structures (Yu et al., 2021b; Pei, 2016). Second, we focused on comparing participants' accuracy across forced-choice pairs of different lengths because in disyllabic target word vs. partword pairs, the two options differ on both word frequency and transitional probability, but the word frequency difference persists in trisyllabic pairs. Therefore, by comparing accuracy differences across lengths for the same pair type, we can isolate the independent effect of transitional probability (see Table 1 for word frequencies and transitional probabilities of the six forced-choice options). Finally, both experiments in this study used within-subjects designs with identical learning durations, eliminating additional variables from differential learning times. In Experiment 1, we synthesized the artificial language using target words as units. Although the artificial language consisted of meaningless syllables, certain syllable combinations might be more acceptable or familiar to participants in the Mandarin context. This specific preference[1] could potentially confound experimental effects. Therefore, Experiment 2 directly synthesized the artificial language at the syllable level using the same rules (Toro et al., 2011), with forced-choice tasks using identical materials as Experiment 1. If participants in Experiment 2 still showed learning effects similar to Experiment 1, this would demonstrate that the effects stemmed from our designed target words and partwords being more compatible with Mandarin characteristics and more familiar to participants. Conversely, if no learning effects emerged, it would indicate that without word-level probabilistic relationships, participants could not distinguish between word types, further confirming that the experimental effects in Experiment 1 resulted from processing probabilistic relationships.

[1]In articles published in foreign journals, this phenomenon is referred to as "arbitrary listening preferences" or "general preferences for certain syllable strings," which relates to participants' language experience but encompasses a broader range.

Experiment 1

Participants

Forty participants took part in the experiment, including 15 males, with ages ranging from 19 to 25 years. All participants were native Mandarin speakers. The experiment required participants to learn an artificial language they had never encountered before. To exclude potential influences of second language experience and musical training, no participants majored in foreign languages or music. Finally, to prevent participants from guessing the experimental purpose, psychology majors were also excluded. Prior to the experiment, participants signed informed consent forms and received modest compensation afterward. This study was approved by the university ethics committee (××$×$2022060023).

Experimental Design

We employed a learning-test paradigm to examine participants' verbal statistical learning abilities. The experimental design was a 2 (word length: disyllabic, trisyllabic) × 3 (forced-choice pair type: target word vs. partword, target word vs. nonword, partword vs. nonword) within-subjects design. In the specific analyses, we first examined whether participants' accuracy on forced-choice pairs differed from chance level (0.5), then conducted ANOVA to test the independent effect of transitional probability. The experimental materials, data, and code have been uploaded to GitHub: https://github.com/wenboyu0803/independent-effect-of-TP.

Materials and Procedure

In the learning phase, we constructed target words from 10 syllables to synthesize the artificial language. All syllables were selected from Yu et al. (2021b) and recorded by a female native Mandarin speaker. After standardization, each syllable had a duration of 300 ms, intensity of 70 dB, and fundamental frequency of 266 Hz. Following previous research designs, these syllables were phonotactically legal in Mandarin and meaningless when pronounced with the first tone. We then randomly combined the 10 syllables into 4 target words (Gómez et al., 2017; Yu et al., 2021b) with no temporal intervals between syllables[2]. According to established statistical learning paradigms (Saffran et al., 1996), we concatenated the artificial language pseudo-randomly such that each target word could not immediately follow itself, and the probability of occurrence for other target words was equal (1/3 in this study). Each target word appeared 100 times in the artificial language, with uniform distribution across the first and second

[2]In Indo-European languages, statistical learning requires that syllable combinations forming target words be meaningless. However, in the context of Mandarin, syllables correspond to characters, so each syllable must be meaningless. Additionally, research indicates that participants track both segmental and suprasegmental structures, so Mandarin studies typically fix tone. Combined, these requirements limit the pool of suitable syllables to approximately 20.

halves. The artificial language lasted 5 minutes. We then created 4 partwords and 4 nonwords. Partwords contained syllables spanning the boundaries of two consecutive target words and thus had transitional probabilities greater than zero. Nonwords consisted of syllable combinations that never appeared adjacent in the artificial language, giving them transitional probabilities of zero. In constructing forced-choice pairs, to ensure that options within each trial differed only in transitional probability, all pairs consisted of options of equal length. Additionally, each pair was presented twice with the order of options reversed to control for order effects. The transitional probabilities and frequencies of the six word types in forced-choice pairs are shown in Table 1 .

Table 1 Transitional probabilities and frequencies of the six option types in the experiment
Disyllabic target words: remei, rouse

The experimental procedure was presented using E-Prime 3. Participants wore headphones and adjusted the computer volume to between 30% and 40% at their discretion. The procedure included practice and formal experimental phases. Before the practice phase, the experimenter explained the task requirements and instructions, played 5 seconds of the artificial language, and then presented 3 forced-choice trials to familiarize participants with the procedure. Materials from the practice phase did not appear in the formal experiment. In the formal experiment, participants listened to the artificial language for 5 minutes, then completed 24 forced-choice trials. In each trial, the two options were presented auditorily with a 500 ms interval between them. After both options played, a prompt appeared on screen asking participants to select the more familiar word by pressing "1" or "2". The entire experiment lasted approximately 10 minutes. A schematic of the procedure is shown in Figure 2 [Figure 2: see original paper].

Results

Statistical Learning Effects Under Mixed Word Length Conditions Figure 2 [Figure 2: see original paper] Schematic diagram of experimental procedure. In statistical learning research, learning is first assessed by comparing forced-choice accuracy to chance level. Partwords and nonwords are not the artificial words participants are meant to learn; in the first two forced-choice pair types of this study, they represent incorrect responses. However, in partword vs. nonword pairs, since partwords have higher transitional probabilities and appear in the artificial language, if participants select partwords more frequently than nonwords, this also indicates statistical learning.

Data analysis was conducted using R (version 4.3.1). One-sample t-tests revealed that accuracy in target word vs. partword pairs was marginally higher than chance, $t(39) = 1.97$, $p = 0.056$, $M = 0.57$, 95% CI for the difference: [-0.00, 0.14], $d = 0.31$. Accuracy in target word vs. nonword pairs was significantly higher than chance, $t(39) = 4.53$, $p < 0.001$, $M = 0.63$, 95% CI: [0.07, 0.19], $d = 0.72$. Accuracy in partword vs. nonword pairs was marginally higher than chance, t(39) = 1.86, p = 0.070, M = 0.55, 95% CI: [0.00, 0.10], d = 0.23. Overall accuracy was significantly higher than chance, t(39) = 3.71, $p < 0.001$, M = 0.55, 95% CI: [0.04, 0.13], d = 0.59. Based on effect sizes and confidence intervals, participants were able to form memory representations of structures with transitional probabilities (target words and partwords) during learning, demonstrating statistical learning effects, as shown in Figure 3 [Figure 3: see original paper].

The Impact of Transitional Probability on Learning Effects Figure 3 [Figure 3: see original paper] Statistical learning effects under mixed word length conditions. A two-way within-subjects ANOVA[3] revealed a marginally significant main effect of forced-choice pair type (F(1.75, 68.13) = 3.15, p = 0.056, ^{2}g = 0.02), a non-significant main effect of word length (F(1, 39) = 1.39, p = 0.245), and a significant interaction between forced-choice pair type and word length (F(1.83, 71.31) = 4.80, p = 0.013, ^{2}g = 0.03). Post-hoc tests on the main effect of pair type showed that accuracy in target word vs. nonword pairs was significantly higher than in partword vs. nonword pairs (t(39) = 2.68, p = 0.032). Simple effects analysis of the interaction revealed that in target word vs. partword pairs, accuracy in the disyllabic condition was marginally higher than in the trisyllabic condition (t(39) = 1.71, p = 0.095). However, in partword vs. nonword pairs, accuracy in the disyllabic condition was significantly lower than in the trisyllabic condition (t(39) = -2.60, p = 0.013). In target word vs. nonword pairs, the simple effect of word length was not significant (t(39) = -1.35, p = 0.185). Additionally, in the disyllabic condition, accuracy in both target word vs. partword pairs (t(39) = 2.34, p = 0.093) and target word vs. nonword pairs (t(39) = 2.52, p = 0.048) were (marginally) significantly higher than in partword vs. nonword pairs. In the trisyllabic condition, only target word vs. partword pairs showed significantly lower accuracy than target word vs. nonword pairs (t(39) = -2.72, p = 0.029); other differences were non-significant. Bonferroni correction was applied to all tests involving three or more levels. Participants' accuracy and standard deviations for the six forced-choice pair types are presented in Table 2 and Figure 4 [Figure 4: see original paper].

Table 2 Accuracy and standard deviations for the three forced-choice pair types

Target word vs. partword		Target word vs. nonword		Partword vs. nonword	
Disyllabic	Trisyllabic	Disyllabic	Trisyllabic	Disyllabic	Trisyllabic

Notably, participants' accuracy in trisyllabic target word vs. partword pairs and disyllabic partword vs. nonword pairs did not differ significantly from chance

[3]For Experiment 1, we also conducted linear mixed-effects modeling with logistic regression, including trial and participant as random intercepts. This analysis revealed a significant fixed effect: participants were 0.90 times as likely to respond correctly in trisyllabic partword vs. nonword pairs compared to disyllabic target word vs. partword pairs, consistent with the ANOVA interaction results. Since the statistical learning field prefers using group means to assess learning effects, this paper primarily reports t-test and ANOVA results. Relevant code is available on GitHub.

level, t(39) = 0.57, M = 0.53, p = 0.570; t(39) = -0.40, M = 0.49, p = 0.689, respectively, indicating no learning effects for these two pair types.

Discussion

In forced-choice pairs composed of target words vs. partwords and partwords vs. nonwords, participants' accuracy hovered around 0.55, marginally different from chance level. However, in target word vs. nonword pairs, accuracy reached 0.63, significantly above chance. Repeated measures ANOVA revealed marginally significant differences across the three forced-choice pair types, with accuracy in target word vs. nonword pairs significantly higher than in partword vs. nonword pairs. This pattern arises because target words, partwords, and nonwords show progressively decreasing frequencies and transitional probabilities in the learning materials. From the perspective of statistical learning memory models (Thiessen & Erik, 2017; Isbilen et al., 2020, 2022; Lukics & Lukács, 2021), high-transitional-probability structures—that is, target words—appear more frequently in the artificial language, providing more opportunities for representation. Partwords appear less frequently with relatively lower transitional probabilities, resulting in weaker memory representations. Nonwords never appear in the artificial language, making them least familiar and least recognizable. These factors collectively contribute to accuracy differences across forced-choice pair types.

Compared to previous research, Experiment 1 improved upon experimental materials by using different transitional probabilities for disyllabic (0.33) and trisyllabic (0.67) partwords while keeping their frequencies equal in the artificial language. When paired with same-length target words, both transitional probability and word frequency differed. However, when comparing accuracy across lengths for the same pair type, only transitional probability varied. Disyllabic partwords with lower transitional probabilities caused less interference to disyllabic target words, enabling participants to easily select target words. Consequently, accuracy in disyllabic target word vs. partword pairs was marginally higher than in trisyllabic pairs. Similarly, since disyllabic partwords had lower transitional probabilities and trisyllabic partwords had higher ones, participants more easily selected correct answers when the latter were paired with nonwords (which have transitional probability and frequency of 0), resulting in significantly higher accuracy for trisyllabic partword vs. nonword pairs than for disyllabic ones. Thus, the interaction effect in Experiment 1 further demonstrates that when word frequency differences are held constant, transitional probability can influence learning effects.

In Experiment 2, we synthesized the artificial language directly at the syllable level using identical forced-choice materials. If participants' performance in Experiment 1 resulted from preferences for our randomly combined target words and partwords, they would similarly show apparent "learning effects" in Experiment 2. Conversely, if no learning effects emerged, forced-choice accuracy would not differ significantly from chance level (0.5).

Experiment 2

Participants

The same recruitment criteria as Experiment 1 were applied. Thirty-eight new participants took part in Experiment 2, including 10 males, with ages ranging from 19 to 26 years. Due to a program error, one participant's data were overwritten, leaving behavioral data from 37 participants for analysis.

Experimental Design, Materials, and Procedure

We again employed a learning-test paradigm, keeping the independent and dependent variables identical to Experiment 1. The artificial language in the learning phase was synthesized using the 10 meaningless syllables from Experiment 1 as basic units. Each syllable could not be immediately repeated, but could be followed by any of the other nine syllables, resulting in an artificial language composed of 1,000 syllables. The forced-choice trials in Experiment 2 were identical to those in Experiment 1. Calculation revealed that all options appeared no more than 20 times in the artificial language, representing a very low proportion. The experimental procedure was also presented using E-Prime 3, including practice and formal phases, with details consistent with Experiment 1.

Results

One-sample t-test results showed that participants' overall forced-choice accuracy was not significantly higher than chance level, $t(36) = -1.59$, $M = 0.48$, $p = 0.121$. Additionally, accuracy in target word vs. partword pairs was not significantly above chance, $t(36) = 0.61$, $M = 0.53$, $p = 0.549$. Accuracy in target word vs. nonword pairs was not significantly above chance, $t(36) = -1.23$, $M = 0.46$, $p = 0.226$. Accuracy in partword vs. nonword pairs was also not significantly above chance, $t(36) = -1.54$, $M = 0.53$, $p = 0.133$. Subsequent analyses incorporating the word length variable revealed that participants' forced-choice accuracy did not exceed chance level in any of the six conditions (all $ps > 0.05$). Accuracy and standard deviations are presented in Table 2.

Experiment 2 used an artificial language synthesized at the syllable level to examine whether participants would still show greater familiarity for the "correct answers" from Experiment 1 when probabilistic information was absent. The results showed that forced-choice accuracy did not exceed chance level either overall or in any of the six conditions, indicating that the learning effects observed in Experiment 1 did not stem from specific preferences for target words or partwords, but rather from tracking and processing transitional probability information during the learning phase.

General Discussion

Previous statistical learning research has focused on examining whether individuals can achieve statistical learning or word segmentation using statistical information, with less attention paid to which specific statistical cues individuals utilize. The present study aimed to investigate whether variations in transitional probability affect learning effects, thereby verifying its independent contribution to statistical learning. In Experiment 1, we constructed target words, partwords, and nonwords of different lengths. Results showed that participants performed well on the forced-choice task, demonstrating statistical learning effects. Simultaneously, differences in transitional probability between options in different forced-choice pair types indeed influenced participants' accuracy. Experiment 2, using an artificial language constructed at the syllable level, revealed that the learning effects observed in Experiment 1 did not result from preferences for our randomly combined target words and partwords. Combined, these results demonstrate that transitional probability can influence verbal statistical learning outcomes independently of word frequency.

Statistical Learning Effects

Classic statistical learning tasks typically use only disyllabic words (Mirman et al., 2008; Estes et al., 2015; Gómez et al., 2017) or trisyllabic words (Saffran et al., 1996; Wang & Saffran, 2014) to synthesize artificial languages when testing whether individuals can segment speech streams using only syllable-level transitional probabilities. However, recent scholars have proposed that to better approximate natural language with variable word lengths (Saffran & Kirkham, 2018; Frost et al., 2020) and avoid rhythmic expectation effects (Hoch et al., 2013), experimental materials should use words of unequal lengths (Yu et al., 2021b). In Experiment 1, we found that participants' overall accuracy across all forced-choice trials was significantly higher than chance, reaching 0.58, thus demonstrating learning effects consistent with previous research (Estes et al., 2015; Palmer & Mattys, 2016; Gómez et al., 2017; Yu et al., 2021b). This indicates that even in more complex linguistic environments, individuals can still segment words by tracking probabilistic information within the language. The design of Experiment 1 further validates the explanatory adequacy of memory models for statistical learning effects. As shown in Table 1 and Figure 4, we used six word types to construct forced-choice pairs and found that accuracy was higher for pairs with larger differences in transitional probability and word frequency (e.g., trisyllabic target word vs. nonword pairs), but lower for pairs with smaller differences, even failing to differ significantly from chance (e.g., disyllabic partword vs. disyllabic nonword). Compared to previous research, this study provides more refined experimental evidence for memory model theories of statistical learning.

The Independent Effect of Transitional Probability

Statistical learning theory originates from Harris's (1954, 1955) induction of morpheme distribution patterns in natural spoken language. Harris proposed that individuals could acquire language or segment speech streams through discovery procedures, with the key to these procedures being the processing of statistical information in speech. Although both word frequency and transitional probability can be considered probabilistic information, they differ essentially. For high-transitional-probability syllable combinations, the preceding syllable can predict the following syllable's occurrence. However, high-frequency syllable combinations may not have high transitional probabilities, as the preceding syllable could combine with other syllables to form words that appear multiple times in the linguistic environment. Therefore, transitional probability serves as a better cue for identifying word boundaries in speech streams. In this study, although the disyllabic condition was only marginally significantly higher than the trisyllabic condition in target word vs. partword pairs, with effect sizes not reaching medium magnitude, one-sample t-test results showed that participants' accuracy in trisyllabic forced-choice pairs (0.53) did not differ significantly from chance. This indicates that under this condition, no learning effect remained—participants struggled to distinguish between trisyllabic target words and trisyllabic partwords. Correspondingly, the learning effect remained robust in the disyllabic forced-choice task (accuracy reaching 0.61), suggesting that participants could effectively differentiate target words from partwords in this pair type. Therefore, from the perspective of learning outcomes, we can conclude that after controlling for word frequency, syllable-level transitional probability can still influence verbal statistical learning effects.

Endress and Mehler (2009) created "phantom words" in an artificial language—words that never appeared as syllable combinations but had the same transitional probabilities as target words. Results showed that forced-choice accuracy for phantom words vs. target words did not differ from chance, yet participants showed greater preference for phantom words than for partwords. This suggests that participants first use transitional probability as the criterion for word segmentation and storage, with frequency information being secondary (Perruchet & Poulin-Charronnat, 2012). The independent role of transitional probability in statistical learning can also be evidenced in natural language. Spoken language lacks clear word boundaries, making syllable-level statistical information crucial for word segmentation. In an English spoken corpus[4], "there is" and "there are" appear with frequencies of 458.6 and 433.82 per million words, respectively, representing high-frequency structures. However, the probability of "is" following "there" is not high[5], indicating low syllable-level transitional probability. Although "there is" has high frequency, its low transitional probability makes it more likely to be segmented as two separate words rather than as two syllables within a single word.

[4]Data source: British National Corpus Online service, http://bncweb.lancs.ac.uk
[5]Data source: British National Corpus Online service, http://bncweb.lancs.ac.uk

Research Implications

Our findings contribute to understanding Chinese word boundary issues from a language usage perspective. From a linguistic standpoint, words are defined as the smallest sound-meaning units that can be used independently. They represent both the basic unit stored in individuals' minds and indivisible structures in sentence construction. In English, linguistically defined words correspond one-to-one with psychological word concepts; for example, "a," "apply," and "red" are both grammatically defined words and the most directly perceived word structures from a usage perspective. In Mandarin Chinese, structures like "dǎqiú" (play ball) and "xīnruǎn" (softhearted) are considered phrases or word groups grammatically, but are psychologically words for native Mandarin speakers[6]. For instance, Cai et al. (2010) included frequency data for such phrasal structures as "dǎqiú" and "chīfàn" (eat rice) in their corpus of Chinese spoken language frequencies compiled from television drama subtitles. The syllable "dǎ" can be followed by syllables like "jié" (rob), "tīng" (listen), "zhōng" (hit), "rù" (enter), "diànhuà" (telephone), forming serial verb, verb-complement, and verb-object structures. In fact, the transitional probability between the two syllables in "dǎqiú" is not high. Since our study demonstrates that predictive relationships between syllables (transitional probability) in spoken language processing can influence word boundary identification, individuals likely segment only "dǎ" as a structural unit for processing[7] rather than treating "dǎqiú" as a holistic unit. Our results can provide reference for the discrepancy between linguistic theory and language usage perspectives on Chinese word boundary demarcation.

References

庞加光，张韧.（2022）. 基线/阐释模型下的汉语离合词现象研究. 外国语（上海外国语大学学报)(02),72-83.

裴雨来.（2016）. 汉语的韵律词. 北京：北京语言大学出版社.

闫国利，张兰兰，孙莎莎，白学军.(2013). 汉语"主观词"的表征及其加工. 心理学报 (04),379-390.

于文勃，王璐，程幸悦，王天琳，张晶晶，梁丹丹.（2021a）. 语言经验对概率词切分的影响. 心理科学进展，29(05)：787-795.

于文勃，王璐，瞿邢芳，王天琳，张晶晶，梁丹丹.（2021b）. 转换概率和词长期待对语音统计学习的影响. 心理学报，53(06)：565-574.

朱佳蕾，刘凤樨.(2020). 离合词与词法句法的分工——再议"同源宾语说". 当代语言学(03),317-334.

[6] In related research, scholars refer to these as "subjective words" (Yan et al., 2013), "separable words" (Zhu & Liu, 2020; Pang & Zhang, 2022), or "prosodic words" (Pei, 2016).

[7] Cai et al. (2010) includes frequency data for "dǎ" as an individual word.

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

Bogaerts, L., Frost, R., & Christiansen, M. H. (2020). Integrating statistical learning into cognitive science. Journal of Memory and Language,115,104167.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. PLoS ONE, 5(6).

Endress, A. D., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. Journal of Memory and Language, 60(3), 351−367.

Endress, A. D., & Langus, A. (2016). Transitional probabilities count more than frequency, but might not be used for memorization. Cognitive Psychology, 92, 37.

Erickson, L. C., Thiessen, E. D., & Estes, K. G. (2014). Statistically coherent labels facilitate categorization in 8-month-olds. Journal of Memory and Language, 72, 49-58.

Estes, K. G. (2012). Infants generalize representations of statistically segmented words. Frontiers in Psychology, 3, 447.

Estes, K. G., & Lew-Williams, C. (2015). Listening through Voices: Infant Statistical Word Segmentation across Multiple Speakers. Developmental Psychology, 51(11), 1517–1528.

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.

Estes, K. G., Gluck, C. W., & Bastos, C. (2015). Flexibility in statistical word segmentation: finding words in foreign speech. Language Learning and Development, 11(3), 252-269.

Frost, R., Armstrong. B. C., & Christiansen. M. H. (2020). Statistical learning research: A critical review and possible new directions. Psychological Bulletin, 145(12), 1128-1153.

Gómez, D. M., Mok, P., Ordin, M., Mehler, J., & Nespor, M. (2017). Statistical speech segmentation in tone languages: The role of lexical tones. Language & Speech, 61(1), 84–96.

Harris, Z. S. (1954). Distributional structure. Word, 10(2-3), 146-162.

Harris, Z. S. (1955). From phoneme to morpheme. Language, 31(2), 190-222.

Hoch, L., Tyler, M. D., & Tillmann, B. (2013). Regularity of unit length boosts statistical learning in verbal and nonverbal artificial languages. Psychonomic Bulletin & Review, 20(1), Isbilen, E. S., Mccauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically induced chunking recall: a memory-based approach to statistical learning. Cognitive Science, 44(7).

Isbilen, E.S., Frost, R., Monaghan, P., & Christiansen, M. H. (2022). Statistically-based chunking of non-adjacent dependencies. PsyArXir.

Lukics, K. S., & Lukács, G. (2021) Tracking statistical learning online: Word segmentation in a target detection task. Acta Psychologica, 215(7):103271.

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. Cognition, 108(1), 271-280.

Newport, E. L. (2016). Statistical language learning: Computational, maturational, and linguistic constraints. Language and Cognition, 8(3), 447—461.

Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. The Quarterly Journal of Experimental Psychology, 69(12), 2390-2401.

Perruchet, P., & Poulin-Charronnat, B. (2012). Beyond transitional probability computations: Extracting word-like units when only statistical information is available. Journal of Memory and Language, 66(4), 807—818.

Potter, C. E., Wang, T., & Saffran, J. R. (2017). Second language experience facilitates statistical learning of novel linguistic materials. Cognitive science, 41(S4), 913-927.

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), e12593.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant Statistical Learning. Annual Review of Psychology, 69(1), 181–203.

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

Siegelman, N. (2020). Statistical learning abilities and their relation to language. Language and Linguistics Compass, (14), e12365.

Thiessen, E. D., & Erik, D. (2017). What's statistical about learning? insights from modelling statistical learning as a set of memory processes. Philosophical Transactions of the Royal Society B: Biological Sciences, 372.

Toro, J. M., Pons, F., Bion, R. A. H., & Sebastián-Gallés, N. (2011). The Contribution of Language-Specific Knowledge in the Selection of Statistically-Coherent Word Candidates. Journal of Memory and Language, 64(2), 171–180.

Wang, T. L., & Saffran, J. R. (2014). Statistical learning of a tonal language: the influence of bilingualism and previous linguistic experience. Frontiers in Psychology, 5.

The Independent Effect of Transitional Probability on Verbal Statistical Learning

Wenbo Yu[1], Tianlin Wang[2], Hetong Qi[1], Dandan Liang[1],[3]

[1]School of Chinese Language and Culture, Nanjing Normal University
[2]School of Education, University at Albany, State University of New York
[3]Interdisciplinary Research Center for Linguistic Science, University of Science and Technology of China

Abstract: In a typical SL task, participants are first exposed to a nonsensical artificial language for 5~10 mins and then asked to finish a 2 alternative forced choice task (2AFC). If the accuracy across participants is higher than chance level, it is assumed that learning has occurred. However, studies have also shown that factors other than TPs, such as word token frequency, also impact SL performance in such tasks. To date, these factors as well as their interactive effects remain under studied.

In this experiment we aimed to investigate whether TPs affect SL learning performance when controlling for target words' and partwords' token frequencies. In doing so, we created the artificial language by randomizing the order of two trisyllabic words and two disyllabic words.

During the 2AFC task, three types of items (target word, partword, and nonword) were paired together, with two items in equal length in each trial. There were 24 trials in the test. 40 native Mandarin monolinguals participated in the experiment; they first listened to the artificial language for 5mins and then finished the 2AFC task.

To investigate the independent effect of TP in SL, we subset the data by word length and found that participants' accuracy choosing trisyllabic target words over partwords was marginally lower than their choosing disyllabic target words over partword, which suggests that disyllabic words confer advantage in SL for this group of participants. In addition, participants' accuracy in choosing trisyllabic partwords over nonwords was significantly higher than that of disyllabic target partwords over nonwords.

A series of results across two behavior experiments highlight the unique contribution from TPs alone, since accuracy was assessed by controlling for word token frequency and word length.

Thus, the present study suggests that TP exerts effect on verbal SL performance independent of word token frequency.

Key words: transitional probability; tokens; statistical learning; 2-alternative forced choice task

Note: Figure translations are in progress. See original paper for figures.