# The ARC Toolbox: Artificial Languages with Rhythmicity Control

Lorenzo Titone[1], Nikola Milosevic[2], Lars Meyer[1,3]

1: Max Planck Research Group Language Cycles, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany.

2: Methods and Development Group Neural Data Science and Statistical Computing, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany.

3: Clinic for Phoniatrics and Pedaudiology, University Hospital Münster, 48149 Münster, Germany.

**Corresponding author:** Lorenzo Titone

**Address:** Stephanstrasse 1A, 04103 Leipzig, Germany.

**Phone number:** +49 341 9940-2586

**E-mail address:** titone@cbs.mpg.de

**Abstract**

Statistical learning is the ability to extract and retain statistical regularities from the environment. In language, extracting statistical regularities—so-called transitional probabilities, TPs—is crucial for segmenting speech and learning new words. To investigate whether neural activity synchronizes with these statistical patterns, so-called neural frequency-tagging paradigms have been employed. Here, brain activity is recorded while a stream of syllables is presented at one fixed rate (e.g., 3.3 Hz) and rhythmic patterns of low-TP events are present at a different rate (e.g., 1.1 Hz). Such studies report neural tracking of both syllables and TPs; that is, for both acoustic and abstract events, respectively. However, recent criticism pointed out that the possible confounding of acoustic and abstract rhythms limits the interpretability of this work. Here, we present a new Python toolbox that minimizes this confounding by creating *Artificial languages with Rhythmicity Controls* (ARC). ARC features a phonological rhythmicity index (PRI) to quantify the confounding periodicity of phonological features at the TP rate. It implements a pseudo-random-walk approach to ensure unconfounded TP stationarity throughout streams of arbitrary length. We provide a step-by-step guide to generate artificial pseudowords based on a pre-specified syllable inventory and create controlled lexicon streams. Simulation results show that ARC lexicons minimize PRI and TP variance, not only relative to sets of random syllable groups, but also to benchmarking sets of artificial languages used in the prior literature. ARC should be used by future work in order to suppress the confounding of acoustics and TPs in artificial language learning experiments.

## Introduction

Statistical learning is the ability to extract and retain regularities from the environment. Among its pivotal role for several cognitive functions (Bogaerts et al., 2020; Frost et al., 2019), statistical learning is known to be instrumental for speech segmentation and language acquisition. In fact, while segment boundaries are not acoustically marked in the speech envelope, they can be inferred from the transitional probabilities (TPs) between syllables. TPs are typically higher within a word than across words. Seminal work on language acquisition utilized artificial language learning paradigms (e.g., Saffran et al., 1996) to demonstrate that both infants and adults can acquire a new lexicon of artificial nonsense words solely by extracting syllable TPs embedded in a continuous stream. In this paradigm, TPs are manipulated such that trisyllabic chunks form artificial "words" (e.g., *golabu, padoti, bidaku, tupiro*) by virtue of having high TP within words and low TP across words. Explicit learning is behaviorally gauged by participants rating the artificial words as more familiar than trisyllabic chunks that violate the TPs.

More recently, electrophysiology studies employed so-called neural frequency-tagging (NFT) paradigms to assess whether the brain tracks regular patterns of TPs and if such tracking is associated with implicit or explicit learning of the artificial lexicon (Assaneo et al., 2019; Batterink & Paller, 2017; Buiatti et al., 2009; Cunillera et al., 2006, 2009; Getz et al., 2018; Henin et al., 2021; McNealy et al., 2006; Ordin et al., 2020). In these studies, brain activity is recorded while streams of syllables are presented at one fixed rate (e.g., 3.3 Hz). Crucially, the distribution of syllable TPs in the streams can be controlled experimentally such that one so-called *TP-random* stream presents a uniform TP distribution across syllable pairs, serving as a baseline for a so-called *TP-structured* stream that contains hidden rhythmic patterns of low-TP events marking boundaries of trisyllabic pseudowords at a different rate (e.g., 1.1 Hz). Strikingly, these studies revealed that the periodicity at the syllable rate (physically present in the stimulus) always associated with increased brain activity at this rate, but the periodicity at the pseudoword rate (putatively absent from the stimulus) was only mirrored in the brain activity for TP-structured streams, suggesting neural tracking of TPs, beyond acoustics.

Recent work raised concerns about the control of the stimuli used in these studies, specifically on the levels of phonotactics and acoustic regularities, which may confound the periodicity at the pseudoword rate (Pinto et al., 2022; van der Wulp, 2021). Indeed, NFT paradigms necessitate strict rhythmicity controls to ensure that only the variable of interest (e.g., TPs), and no other factors (e.g., phonological patterns), occur in the stimulus at a specific rate. In fact, when a stimulus contains multiple types of information at the same given rate, these become confounds that add up in the spectrum. Without disentangling statistical and acoustic regularities, it cannot be excluded that the observed neural responses at the pseudoword rate stem, at least partially, from tracking acoustic and phonological rhythms, rather than statistical ones. In addition, it should be considered that linguistic similarities between the artificial lexicon and the learners' native language could also influence their ability to learn the artificial lexicon (Schepens et al., 2020; Varela et al., 2024). Moreover, learners may take advantage of other sorts of regularities to infer lexical boundaries, including acoustic cues, such as stress (Hay & Saffran, 2012; Johnson & Jusczyk, 2001), phonological patterns (Onnis et al., 2005; Seidenberg et al., 2002), prosody (Kim et al., 2012), and phonotactics (van der Wulp, 2021).

To account for these confounding factors in artificial language learning studies, we present here an open-source Python toolbox to generate *Artificial languages with Rhythmicity Controls* (ARC). The ARC Toolbox offers the option to set multiple parameters to generate lexicons of variable sizes and streams of arbitrary length, while implementing phonological, statistical, and acoustic controls. In particular, ARC allows to tailor the generation of pseudowords to already existing languages, by taking into account their statistics (e.g., n-gram probabilities) and phonotactic rules. Decisively, we show that ARC suppresses acoustic and phonological rhythmicity at the rate of interest while providing full control over TPs. ARC is thus a useful tool to increase the systematicity in stimulus development and overcome current uncertainties in statistical learning studies.

## Methods

The ARC Toolbox implements a custom type system of linguistic elements. These custom types have been introduced for two main reasons: (i) to enable fast search over large combinatorial spaces of linguistic elements (mostly for controlled artificial lexicon generation) and (ii) to support the creation of TP– and rhythmicity-controlled streams with a large number of user-specified parameters (Figure 1). The Rhythmicity Control is one additional core feature of ARC. It is based on a universal matrix of binary phonological features for all phonemes. Hence, *Phonemes* are ARC's atomic linguistic elements, which other elements, like *Syllables*, *Pseudowords*, and *Lexicons* are based on. While ARC's type system was designed with the purpose of generating artificial lexicons embedded in TP- and rhythmicity-controlled streams, it is in principle independent of this use case.

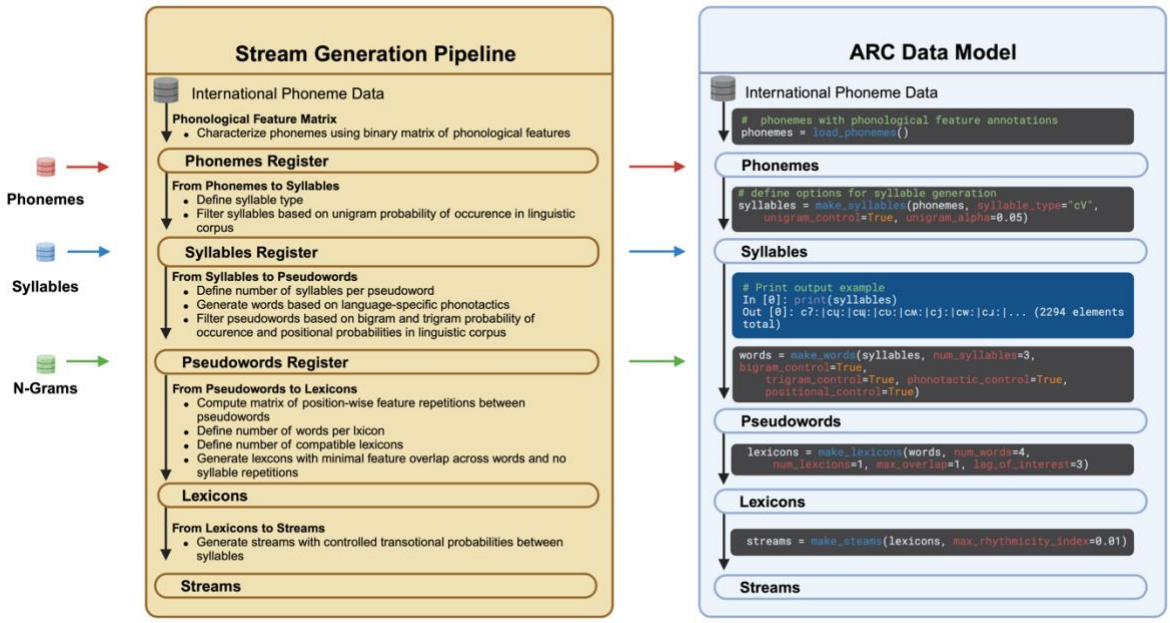


**Figure 1. Overview of the main modules in the ARC Toolbox.** The core stream generation pipeline (left box) concerns the generation of *Phonemes*, *Syllables*, *Words*, and *Lexicons Registers*. These elements are annotated with a set of binary phonological features for each phoneme. Optionally, ARC can use language-specific corpora (colored disks) to filter each register for frequency normalizations. Phonotactic constraints and rhythmicity controls can be imposed on syllable, word and lexicon generation, respectively. The ARC toolbox presents high-level functions (right box) to generate the linguistic inventories based on user-specified parameters that call low-level library functions. Created with BioRender.com.

*Step 1—from phonemes to syllables: frequency control*

Prior linguistic knowledge impacts statistical learning performance (Elazar et al., 2022; Siegelman et al., 2018; Varela et al., 2024). Thus, the first step to generate an unbiased artificial lexicon is to select an unbiased syllable inventory and ensure that phoneme combinations in the lexicon do not resemble real words in the learners' native language. While most prior studies relied on experts' knowledge to address this issue, others used linguistic corpora to exclude syllables associated with high frequency of use from the inventory (Kiai & Melloni, 2021), or to normalize syllable-level positional similarity relative to the known language (Fernandes et al., 2009; Varela et al., 2024). Expanding on this work, ARC offers a systematic method to control for n-gram and positional syllable and phoneme statistics.

First, ARC generates a syllable inventory by exploiting a binary matrix of phonological features for all phonemes, including class features (i.e., consonant, sonorant, syllable features), and phono-articulatory features (i.e., place, manner, and laryngeal features). Based on this matrix, ARC generates a syllable inventory with specific combinations of phonemes and/or phonological features. Hence, users may indicate a desired syllable structure (default = open CV, i.e., consonant-vowel syllables with long vowels) to construct the *Syllables Register*. Here, the default setting is recommended for experimental designs with isochronous input (e.g., NFT paradigms), since it minimizes the variability of phoneme duration across syllables. Additionally, users may provide a list of syllable tokens from a pre-acquired linguistic corpus in X-SAMPA or IPA format, enabling ARC to skim the pre-existing syllable inventory of a specific language based on the selected syllable type, rather than building the Syllables Register ex-novo, with no guarantee that the resulting syllables belong to any real language.

In addition to the choice of syllable structure, ARC may control for the unigram frequency of use of syllables in the inventory. If a corpus of syllable tokens associated with their frequency of use is provided, then token frequencies from the corpus will be log-transformed, normalized (z-scored), and tested against a uniform distribution. This test excludes tokens that significantly deviate from the uniform distribution (default $\alpha$ threshold = 0.05, two-tailed) and produces a frequency-matched Syllables Register that will serve as a basis to generate artificial pseudowords in the next step.

*Step 2—from syllables to pseudowords: phonotactic control*

Phonotactic rules in a given language can bias word segmentation. For instance, in languages such as Dutch, neighboring syllables with the same place of articulation more frequently occur across word boundaries than within words. The adherence to this phonotactic rule, also known as the Obligatory Contour Principle (OCP) was shown to influence word segmentation (Boll-Avetisyan & Kager, 2014) and to partially account for neural tracking of words in a recent reanalysis of a prior artificial language learning study (Batterink & Paller, 2017; van der Wulp, 2021). This result suggests that rhythmic variations in place of articulation may confound TP manipulations. Thus, researchers highlighted the need to better control phonotactics in future studies (Pinto et al., 2022; van der Wulp, 2021).

ARC offers a systematic method to generate controlled artificial lexicon streams with respect to phonotactics, by using the binary matrix of phonological features and controlling for the adherence to specific phonotactic rules, such as the OCP, during the process of pseudoword generation from the Syllables Register. To that end, ARC combines individual phonological features to extract phono-articulatory categories, such as the place and manner of articulation (Ramers, 1998). Based on this information, ARC restricts the possible syllable combinations that are allowed to form pseudowords, depending on their adherence to specific phonotactic rules. To accomplish that, ARC requires a specification of the desired pseudoword size (default = 3 syllables per pseudoword) and phonotactic control (default = OCP). With the default settings, ARC generates trisyllabic pseudowords according to the phonotactic rules of the OCP-manner, OCP-place, and OCP-vowel, meaning that the resulting pseudowords will be comprised of three syllables, all of which will differ in the manner and place of articulation of the consonants and in the identity of the vowels. Alternatively, the user may specify a different pseudoword size (e.g., 4 or 5 syllables per pseudoword), and adjust the level of flexibility in controlling for violation of the OCP rules (e.g., establishing that only consecutive syllables within a pseudoword must respect the rule). To date, we are not aware of evidence on the effects of other phonotactic violations on pseudoword tracking, beyond OCP. Therefore, we limited our phonotactic control to the OCP rule.

Since phonotactic constraints vary greatly across languages, providing a full set of phonotactic rules that could be applied to any language is beyond the scope of the present research. Nevertheless, we considered that some researchers may want to control for other phonotactic constraints, depending on the native language of their target population of learners. To that end, we offer the possibility to insert custom code blocks into ARC, and warmly invite fellow researchers to implement additional phonotactic rules within the ARC package, with the advantage of being able to enjoy its features and the aspiration of reaching the larger community.

After the generation of an inventory of phonotactically controlled artificial pseudowords with a fixed number of syllables, ARC may apply the same frequency control as in Step 1, to ensure that the pseudowords do not contain too frequent or rare phoneme combinations also across syllable boundaries. To that end, ARC requires the specification of bigram, trigram, and positional phoneme statistics from a speech corpus. Hence, ARC may filter the pseudoword inventory based on whether all phoneme bigrams and trigrams within each pseudoword are present in the corpus, or if any of those significantly deviates from the uniform distribution (bigram default α threshold = 0.05, two-tailed; trigram default = nonzero). Moreover, ARC may filter the pseudoword inventory based on positional statistics at the phoneme level. For instance, ARC can remove artificial pseudowords from the Pseudowords Register if they present a phoneme that rarely occupies a certain position in a pseudoword, based on the corpus (default α threshold = 0.05, one-tailed).

*Step 3—from pseudowords to lexicons: phonological control*

In generating lexicons of artificial pseudowords embedded in an isochronous stream of syllables, one needs to consider that the individual phonological features of the phonemes across pseudowords provide acoustic landmarks that the brain could track instead (or in addition to) statistical regularities (Di Liberto et al., 2015). While the periodic reoccurrence of phonological features is likely not an issue for studies that use natural language (due to its high variability), it is definitely important for artificial language learning studies that use NFT paradigms (due to the limited variability in the

linguistic items and the rigid temporal structure of the stimulus). Indeed, phonological information and sub-phonemic cues were previously argued to be essential for neural tracking of pseudowords (Chen et al., 2020).

In this context, we considered that if a phonological feature is present at the same rate of TP patterns, statistical and phonological cues cannot be disentangled. In other words, if neighboring pseudowords share a phonological feature at the same position, but not at all other positions, that feature will be *rhythmic*, confounding the TP rate. To suppress this effect, we reasoned that by limiting the opportunity for a feature to be present in the same position across pseudowords, its periodicity in the stream will be reduced. Hence, we exploited the binary matrix of phonological features to compute a matrix of position-specific feature overlap across pairs of pseudowords. In doing so, we considered different sets of phonological features for consonant and vowel phonemes, specifically: “sonorant”, “continuous”, “lateral”, “nasal”, “voiced”, “coronal”, “labial”, “high”, and “back” for the consonants; and “labial”, “high”, “low”, “back”, “tense”, and “long” for the vowels.

Since it is the syllable rate, not the phoneme rate, that is kept constant in the present design, the phonological rhythm must be controlled for at a fraction of that rate (e.g., 1/3 of the syllable rate for trisyllabic pseudowords). Hence, the binary vectors of phonological features for all phonemes in a syllable are lumped together to represent the set of relevant phonological features for an entire syllable. Then, ARC calculates the position-wise correlation matrix of phonological features across pairs of pseudowords. Based on this matrix, ARC employs a *lazy evaluation strategy* to build sets of artificial pseudowords while ensuring minimal repetitions of features across phonemes in the same position within a pseudoword (default = 0). In this context, lazy evaluation means that artificial lexicons of pseudowords of the requested size are only created when they are referenced in subsequent steps, such that lexicons with least feature repetitions are generated first. The feature overlap is the cumulative sum of repetitions of each feature at the rate of interest. If there is no pair of pseudowords with a minimum number of repetitions of phonological features, the upper bound of feature repetitions that two pseudowords are allowed to share increases by one unit.

This lazy evaluation strategy drastically reduces computation time relative to generating all possible pseudoword combinations and then sorting them. Users can specify the desired number of pseudowords per lexicon (default = 4) and indicate if repetitions of the same syllables are allowed (default = false). The outcome is an artificial lexicon in which the position-wise phonological feature overlap across minimal pairs of pseudowords is minimized (Figure 2).

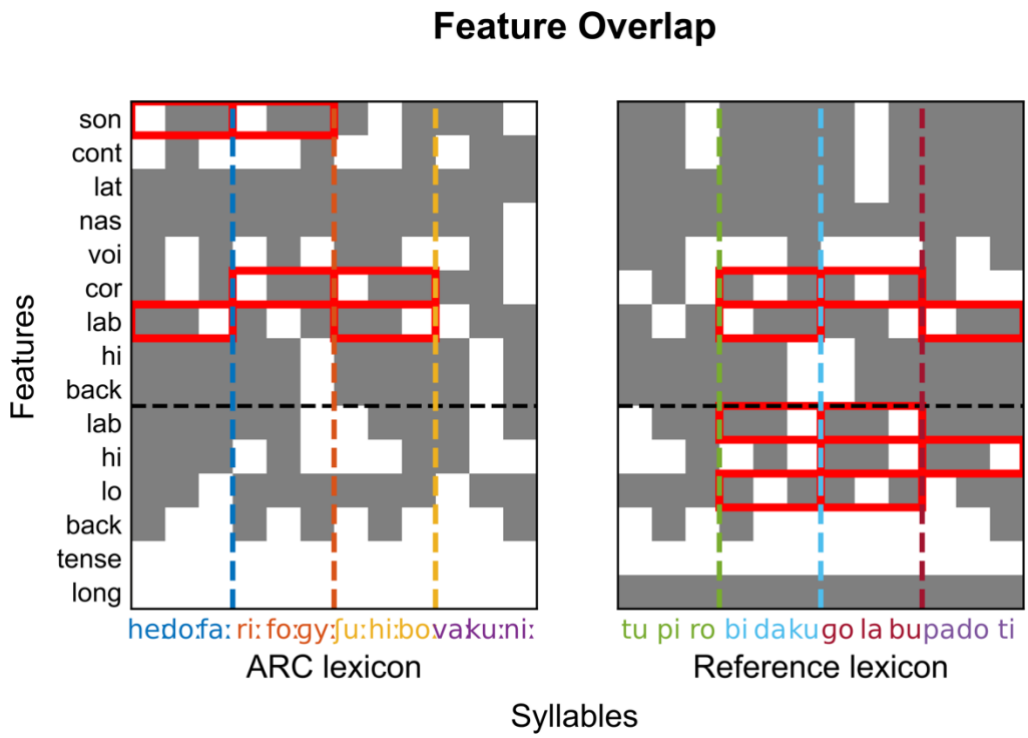


**Figure 2. Binary matrix of phonological feature overlap across minimal pairs of pseudowords.** Phonological features are plotted as a function of syllables and lexicons. The plots show the binary feature matrix for one example lexicon (left) produced with ARC (Appendix A) and one reference lexicon (right) taken from the prior literature (Saffran et al., 1996). Artificial pseudowords in the lexicons are color-coded on the abscissae, and pseudoword boundaries cut the matrices vertically (colored dotted lines). The horizontal line splits the feature labels associated to consonants (above) and vowels (below). Grey squares represent a minus (or a zero) in the binary feature vector of the syllable (i.e., the feature is absent); whereas white squares represent a plus (or a one) for that feature (i.e., the feature is present). The red rectangles highlight repetitive patterns of phonological features across minimal pairs of pseudowords. By assuming that all pseudowords will transition to one another in a stream, it is granted that any two pseudowords will be presented consecutively at times (even though they might appear distant from each other in the snippet of this figure). When this eventuality occurs, a feature is said to be rhythmic if it is present in two consecutive pseudowords at a fixed lag, resembling two cycles of an oscillation with period equal to the size of the pseudoword (e.g., [1, 0, 0, 1, 0, 0]; [0, 1, 0, 0, 1, 0]; or [0, 0, 1, 0, 0, 1]).

*Step 4—from lexicons to streams: TP control*

A common strategy to manipulate TPs in a syllable stream is to randomly shuffle a fixed number of syllables or pseudowords with the only constraint that they cannot be repeated consecutively. This process may lead to some transitions being more likely than others due to randomness. As a result, the TPs would be biased because tokens present different cue diagnosticity on the microscale of single items. Furthermore, under loose constraints, TPs may vary throughout the stream, possibly leading to spurious irregularities at the macroscale of the syllable stream. Importantly, statistical irregularities at micro– and macroscales are not exempt from having an impact on neural responses (Hasson, 2017). In fact, if TPs are variable and time-varying, some statistical boundaries might be perceived as more salient than others, with neural consequences. To suppress this bias, ARC operates strict statistical controls that could guarantee TP stationarity throughout the stream.

One additional caveat concerns the TP-random baseline condition. Recent studies pointed out that systematic acoustic differences at the pseudoword rate are actually visible in the acoustic spectra of TP-structured streams due to the fact that syllables reoccur at fixed positions in the sequence, but absent in TP-random streams in which syllables are randomly shuffled (Luo & Ding, 2020; Pinto et al., 2022). Pinto et al. (2022) addressed this issue by using a position-fixed baseline, where syllables in the TP-random condition are also constrained to occupy the same position within a pseudoword-size chunk. In doing so, the inevitable pseudoword-peak is introduced also in the baseline condition, thereby guaranteeing that both the TP-structured and the TP-random streams have the same acoustic profile. In that study, frequency-domain analyses of the brain responses revealed prominent peaks at the pseudoword rate (or its harmonics) for both streams, but higher peaks in the TP-structured relative to the TP-random streams (Pinto et al., 2022). These findings indicate that acoustic regularities do partially account for neural tracking at the pseudoword rate, but also that statistical learning effects are reliable if compared to an acoustically-matched baseline condition. Hence, this acoustic control is necessary for a safer interpretation of the neural responses at the pseudoword-rate in terms of statistical learning.

To precisely manipulate TPs between pseudowords in TP-structured streams and syllables in TP-random streams, ARC uses a *pseudo-random-walk* (PRW) approach. PRW guarantees that each token follows every other token equally often. In TP-structured streams, each of the $n$ pseudowords composing the lexicon is assigned to a token (from 0 to $n - 1$) and repeated a number of times (default = $n \times 15$). Then, to ensure that TPs remain constant throughout the stream, ARC iteratively samples new tokens based on a memory matrix of TPs. More specifically, the sampling of new tokens is restrained among those that have the lowest TP to the previous token based on the memory matrix. Moreover, every token must be used once before any other could be repeated one more time than all others, and consecutive repetitions of the same token are not allowed. In doing so, ARC produces streams in which the TPs at pseudoword boundaries always tend to exactly $1/(n - 1)$. Then, ARC decomposes each pseudoword-token into a fixed sequence of syllable-tokens, such that the syllable TPs within one pseudoword are always 1.

ARC adopts a similar approach for controlling TPs in the TP-random streams as well. However, in this condition each token maps directly to a syllable (from 0 to $n \times m - 1$, where $m$ is the number of syllables per pseudoword). To replicate recent evidence that repetitions of sounds at fixed positions induce acoustic regularities regardless the statistical structure, ARC defines two different functions to simulate both position-random and position-fixed TP-random baseline streams. In the TP-random position-random stream, consecutive repetitions of the same syllables are still not allowed; yet, all syllables must transition to each other homogenously and spread evenly across all possible positions within a pseudoword-size chunk. Under these constraints, TPs across all syllables always tend to exactly $1/(n \times m - 1)$ throughout the stream. On the other hand, syllable-tokens in the position-fixed baseline stream can occupy only one fixed position within a pseudoword-size token (e.g., [0, 1, 2], [3, 4, 5], [6, 7, 8], [9, 10, 11], for $n = 4$, and $m = 3$) and they must follow one another respecting these positions. Yet, syllable-tokens in the position-fixed streams are still constrained to always "walk" the viable path they have walked the least, based on the memory matrix. Thus, the syllable TPs always tend to exactly $1/n$ throughout the position-fixed stream.

Finally, to quantify rhythmicity of any given phonological feature in a stream, we developed a *phonological rhythmicity index* (PRI). Here, each token in TP-structured and TP-random streams is mapped to a syllable from the artificial lexicon and to its corresponding vector of binary phonological features. Then, the PRI is computed on the matrix of phonological features spanning the whole stream as the correlation between each phonological feature vector and binary convolutional kernels that resemble two cycles of an oscillation at the rate of interest (see also Figure 2). To obtain streams with minimal phonological rhythmicity, users may set the maximum PRI among all features to be lower than an arbitrary threshold (default = 0.1).

*Step 5—from streams to audios: acoustic control*

Pinto et al. (2022) adopted a position-fixed TP-random baseline to ensure that the TP-structured and TP-random streams present identical peaks at the pseudoword-rate in the envelope spectra. To show that position-random and position-fixed streams induce different acoustic peaks at pseudoword-rate, we conducted spectral analyses on all types of auditory streams based on one example ARC lexicon (Appendix A). To that end, we synthesized each individual syllable into separate audio files using Google-Text-to-Speech, and the Speech-Synthesis-Markup-Language (Taylor & Isard, 1997). Then, we edited each syllable using Praat (Boersma & Weenink, 2007) to equalize their duration to 300 ms and their amplitude to 65 dB, and concatenated them into continuous auditory streams of 216 seconds using the “wave” Python module. Finally, we computed the fast Fourier transform of the speech envelopes using MATLAB (The MathWorks, Inc.).

*Statistical analyses*

To assess the improvement of ARC in minimizing phonological rhythmicity, we simulated the generation of five TP-random position-random streams, five TP-random position-fixed streams, and five TP-structured streams for 21 controlled lexicons produced using ARC, 21 reference lexicons extracted from the prior literature (Assaneo et al., 2019; Batterink & Paller, 2017; Cunillera et al., 2006; Kiai & Melloni, 2021; McNealy et al., 2006), and 21 baseline lexicons produced as random

syllable combinations (Appendix B). Hence, we compared the distribution of PRI across features for each lexicon type and stream mode using one-sided Welch's independent sample t-tests.

To assess whether the PRW approach to stream generation produces a lower variance of syllable TPs relative to classical pseudo-randomization strategies used in the prior literature, we computed the transition matrix at each stream location to keep record of the evolution of TPs throughout the stream. The classical pseudo-randomization strategy entails shuffling of the tokens with the only constraint that no consecutive repetitions of the same token are allowed. Hence, for each stream mode, we conducted Levene's tests to compare the TP variance at each TP boundary (pseudoword boundaries in TP-structured streams and syllable boundaries in TP-random streams) among the two strategies.

## Results

Here, we present the output of ARC for the generation of artificial lexicons of four trisyllabic pseudowords based on a pre-specified corpus of German syllables. This output was obtained by using ARC with the default settings, thereby controlling for phoneme and syllable frequency statistics, phonotactics, phonological patterns, statistical, and acoustic regularities. Importantly, these settings implement the new PRI to minimize the presence of phonological regularities at the rate of interest, and the PRW to precisely manipulate syllable TPs within TP-structured and TP-random streams. Importantly, we show that ARC produces artificial lexicons with lower PRI and TP variance relative to benchmarking sets of stimuli from the previous literature.

### *Normalization of n-gram and positional phoneme statistics, and phonotactic controls*

A list of 6,397 syllable tokens paired to their frequency of use was acquired from a large dataset on conversational German (Schiel, 2010). To preserve phoneme timing constant across all syllables, we asked ARC to skim the inventory and retain only open CV syllables with long vowels, based on the binary feature matrix. This way, ARC retained 130 syllables matching the requirements. Then, we inspected the frequency distribution of the syllables in the corpus (Figure 3) and ran ARC unigram

frequency filter to obtain a more normal distribution of frequencies for the syllables in our inventory (see Step 1 in *Methods*). In doing so, frequencies were log-transformed, normalized, and tested for non-uniformity ($\alpha = 0.05$, two-tailed). Then, tokens that significantly deviated from uniformity were excluded, resulting in a Syllable Register of 76 open CV syllables with a more normal frequency distribution (Figure 3). This inventory was used as a basis to generate trisyllabic pseudowords that obeyed specific phonotactic constraints, based on the adherence to the OCP-place; OCP-manner; and OCP-vowel rules (see Step 2 in *Methods*). Next, we filtered the set of pseudowords using bigram, trigram, and positional phoneme statistics from a pre-acquired corpus (Arnold & Tomaschek, 2016) to ensure that no pseudoword contains too frequent or too rare phoneme combinations (Figure 3), resulting in a set of 1812 artificial pseudowords. Finally, we sampled a subset of pseudowords that obey these constraints (default Pseudoword Register sample size = 200).

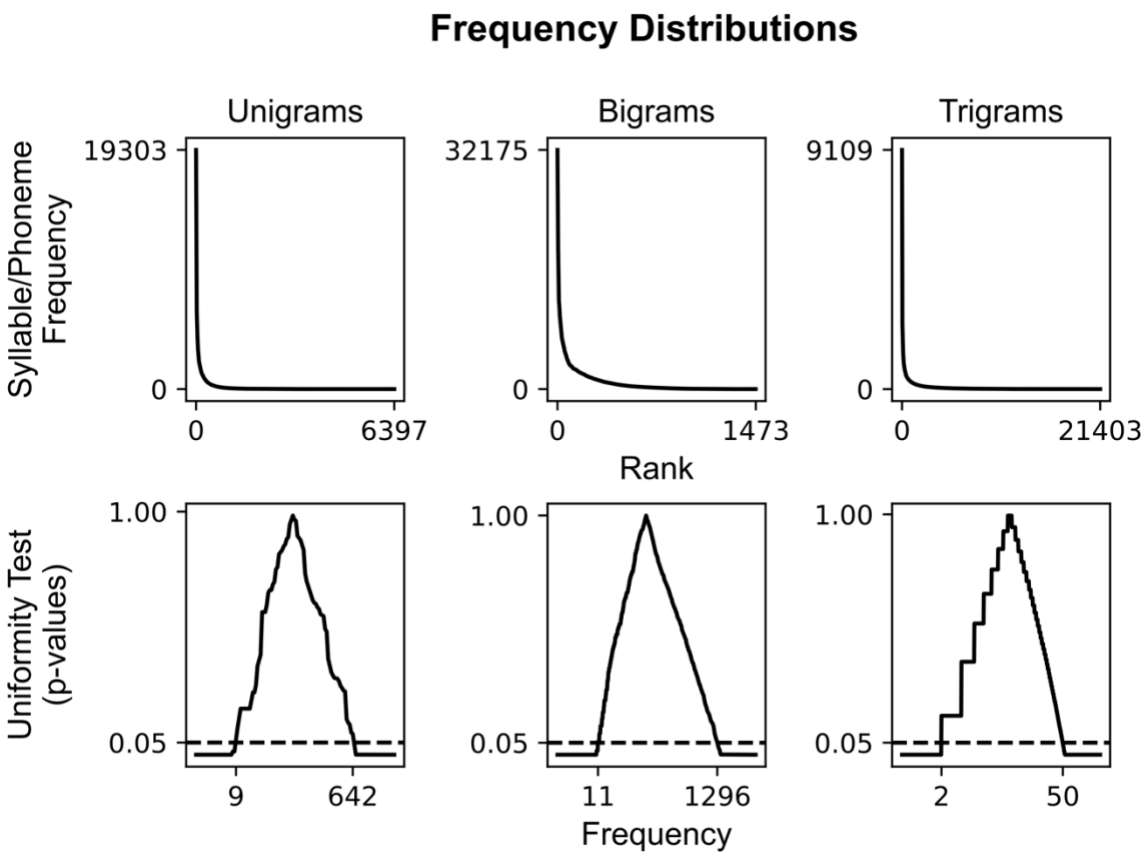


**Figure 3. Unigram, bigram and trigram probability distributions**. Top: Token frequency as a function of frequency-ranked tokens of syllables (left), phoneme bigrams (middle), and trigrams (right). Bottom: Distributions of p-values from the uniformity test as a function of token frequency. The horizontal line represents the alpha threshold (0.05).

*Lexicon generation with minimal phonological rhythmicity index*

To control for confounding phonological patterns at the rate of interest (i.e., the triplet rate where TP drops occur), we used the generator expression of ARC to bias the formation of lexicons towards having the smallest possible feature overlap at the TP rate (see Step 3 in *Methods*). This overlap is quantified at the stream level as the distribution of PRI across features (see Step 4 in *Methods*). To assess whether ARC's constraints on lexicon generation would reduce the PRI across features, we compared the distribution of PRI across phonological features for streams generated under ARC constraints, relative to one benchmark set of artificial lexicons used in previous studies, and to random lexicons with no phonological controls (Appendix B).

In particular, we considered the maximum PRI across all features as the most informative metric for assessing the potential confounding of phonological regularities. Consequently, for each of the 21 lexicons per category (ARC Controlled Lexicons; Reference Lexicons; and Random Lexicons), we generated five streams of syllables per stream mode (i.e., TP-random position-random; TP-random position-fixed; and TP-structured). Then, we calculated the maximum PRIs for each sample and compared the PRI distribution across the different sets of lexicon streams generated (Figure 4). One-sided independent sample Welch's t-tests revealed that ARC Controlled Lexicons produced a lower maximum PRI than both the Reference Lexicons and the Random Lexicons for TP-structured streams and TP-random position-fixed streams, but also that Reference Lexicons produced a lower maximum PRI compared to Random Lexicons for all stream modes (Table 1). These results indicate that while the streams originating from the Reference Lexicons reduced phonological rhythmicity compared to Random Lexicons, ARC Controlled Lexicons further reduce this confound systematically. Finally, to show that the reduction of PRI is consistent across features, we also depict the distributions of PRI for all individual phonological features (Figure S1). Here, Reference lexicons and Random lexicons present more outliers across features compared to ARC Controlled Lexicons. Hence, both Reference Lexicons and Random Lexicons present different features with high PRI values, but ARC Controlled Lexicons consistently limit the maximum PRI across all features.

**Table 1**

*Independent Pairwise Welch's T-Test Results of PRIs for all Stream Generation Modes.*

| Groups | TP Mode | T | dof | p-val | CI 95% | cohen-d | power |
|---|---|---|---|---|---|---|---|
| Controlled vs. Reference | TP-random position-random | 0.65 | 208 | 0.74 | [-inf 0.] | 0.09 | 0.01 |
| | TP-random position-fixed | -9.89 | 208 | 1.92E-19 | [-inf -0.08] | 1.36 | 1.00 |
| | TP-structured | -8.01 | 208 | 4.10E-14 | [-inf -0.08] | 1.11 | 1.00 |
| Controlled vs. Random | TP-random position-random | -0.48 | 208 | 0.31 | [-inf 0.] | 0.07 | 0.12 |
| | TP-random position-fixed | -10.26 | 208 | 1.53E-20 | [-inf -0.2] | 1.42 | 1.00 |
| | TP-structured | -15.39 | 208 | 1.82E-36 | [-inf -0.28] | 2.12 | 1.00 |
| Reference vs. Random | TP-random position-random | -1.06 | 208 | 0.14 | [-inf 0.] | 0.15 | 0.28 |
| | TP-random position-fixed | -5.47 | 208 | 6.52E-08 | [-inf -0.1] | 0.75 | 1.00 |
| | TP-structured | -8.63 | 208 | 7.87E-16 | [-inf -0.17] | 1.19 | 1.00 |

*Note.* We compare the distributions of the maximum feature PRIs of 105 streams for transition probability control conditions described in the main text. Numbers are rounded to two decimal places.

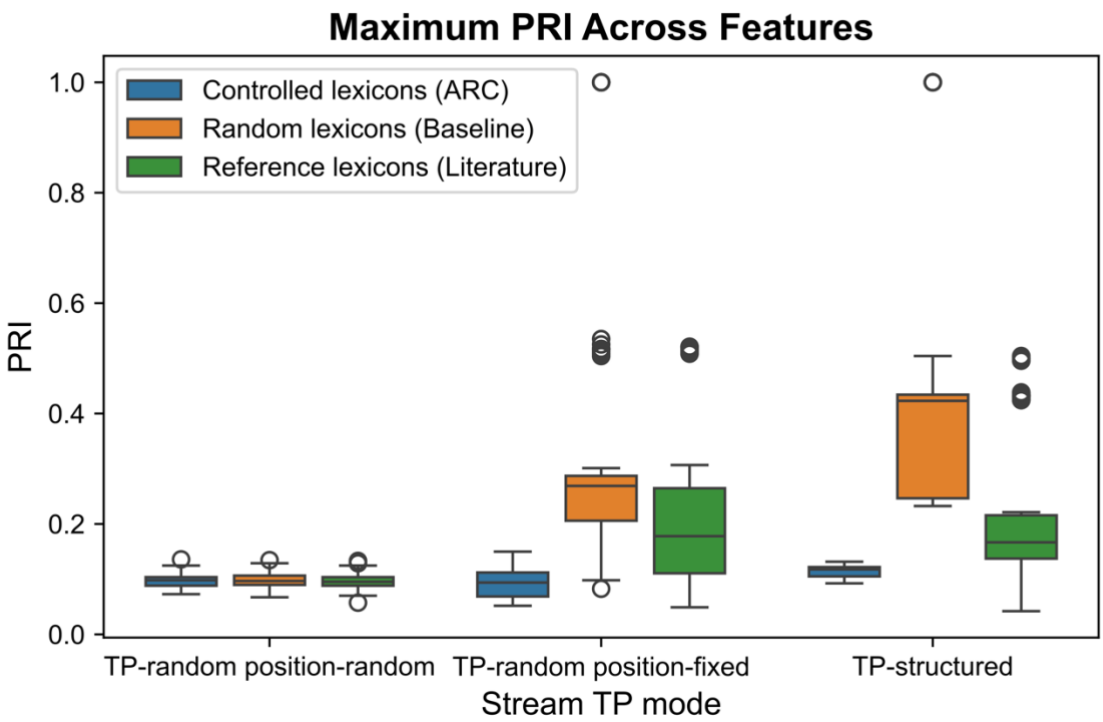


**Figure 4. Phonological Rhythmicity Index.** Boxplots of the distribution of maximum PRI across phonological features for the ARC streams and the benchmark streams (21 lexicons with 5 streams each, per category). Dots represent outliers.

*Stream generation with minimal TP variance*

ARC implements a PRW to stabilize TP variance (see Step 4 in *Methods*). Here, we aimed at showing that ARC's PRW stabilizes the TP values across the stream relative to a random shuffling method (Figure 5). Levene's tests between the distributions of TPs at all TP boundaries obtained from using ARC's PRW or the random shuffling method were significant for both the pseudoword-level in TP-structured streams ($F_{(1, 476)} = 35.99$, $p = 3.92 \times 10^{-9}$), and the syllable-level in TP-random streams ($F_{(1, 1436)} = 19.81$, $p = 9.21 \times 10^{-6}$). These results suggest that ARC's PRW reduces the TP variance in the streams compared to randomly shuffled streams, and indicate that ARC produces more stable TPs within each stream.

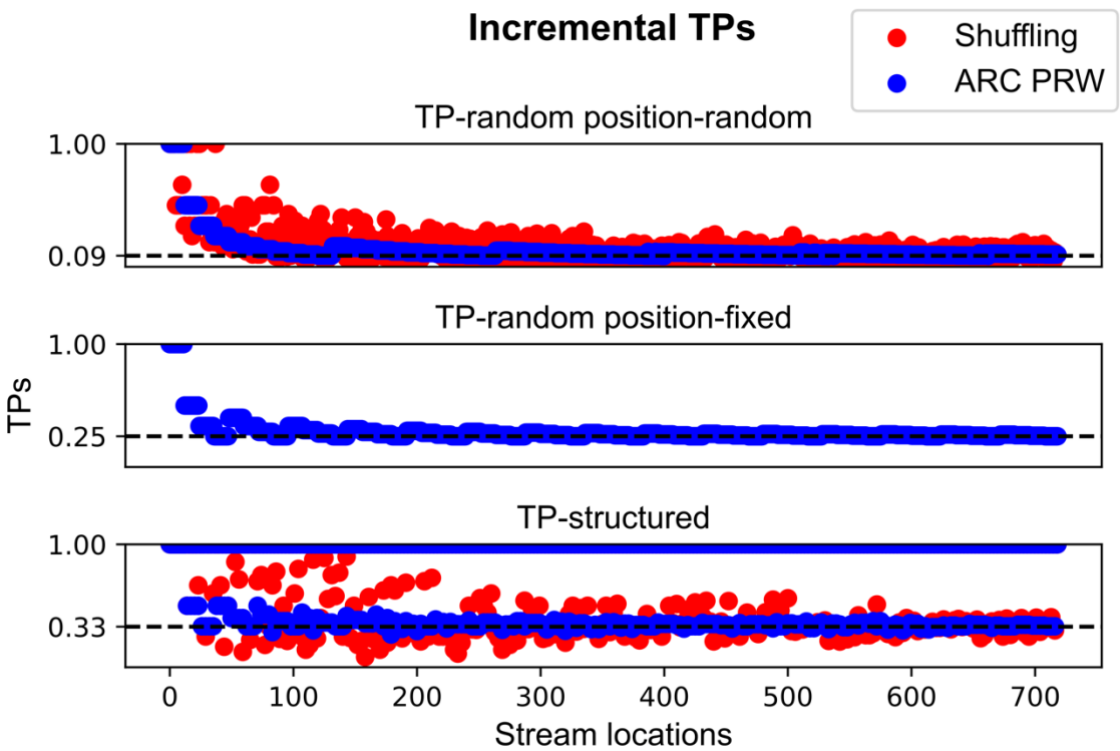


**Figure 5. TP variance.** Scatter plots of cumulative TPs at each transition boundary location in the streams (colored dots) for the position-random baseline (top); the position-fixed baseline (middle); and the TP-structured stream (bottom). The dotted line represents the expected TPs based on an ideal statistical manipulation.

*Acoustically-matched speech envelope spectra*

To investigate if the re-occurrence of syllables at the same position within a pseudoword constitutes a systematic acoustic difference across TP-stream modes that are visible in the frequency-domain, we

computed the envelope spectra for TP-random position-random, TP-random position-fixed, and TP-structured streams based on one example lexicon generated with ARC (Appendix A). The envelope spectra for the TP-random position-random stream shows no peak at the pseudoword rate (1.11 Hz); whereas the TP-random position-fixed stream presents a peak at that rate that is equivalent to the peak in the TP-structured stream (Figure 6). This finding indicates that systematic acoustic differences due to syllable positional constraints in artificial pseudowords are reflected in the envelope spectra.

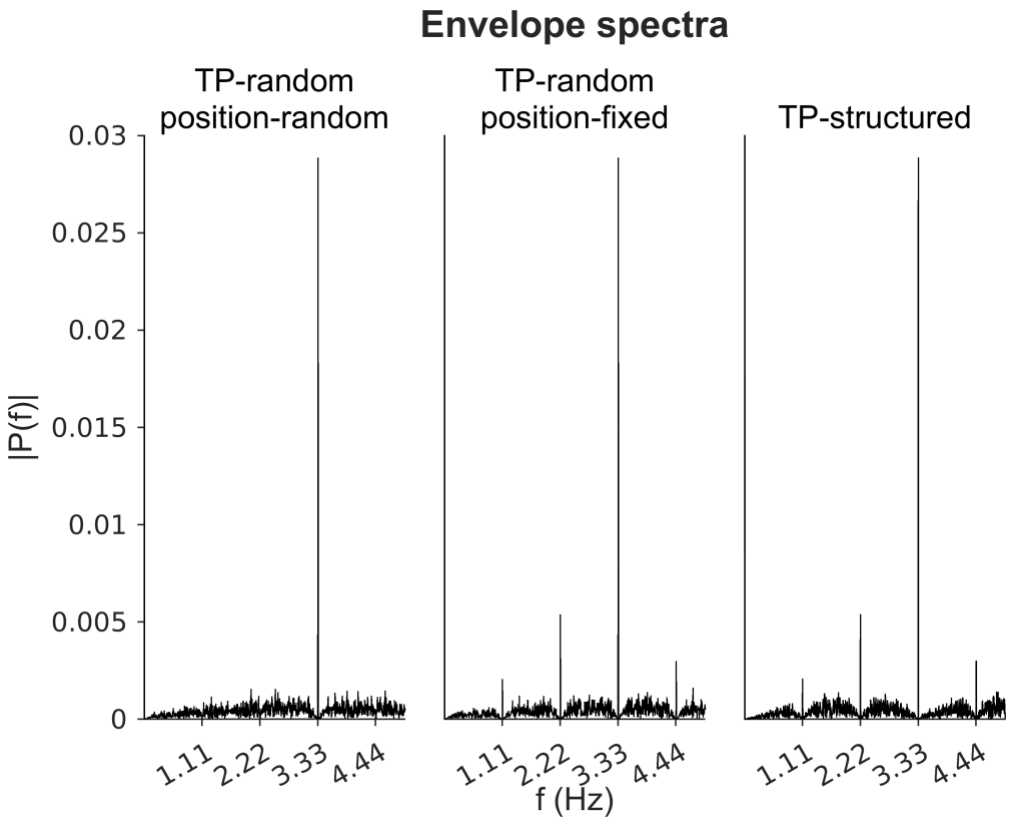


**Figure 6. Speech envelope spectra.** Power spectra from ARC's best lexicon stream envelopes. The TP-random position-random stream (left) has syllables that occupy all possible locations; its envelope spectrum does not display a peak at the pseudoword-rate (1.11 Hz). The TP-random position-fixed stream (middle) has syllables at fixed locations; its envelope spectrum displays power peaks at the pseudoword-rate (and its harmonics) equivalent to the TP-structured stream (right).

## Discussion

We presented a new toolbox (ARC) that controls for several confounding factors in artificial language learning studies, namely: prior knowledge of critical baseline frequency statistics in a given language; phonotactics; rhythmic phonological patterns; TP precision and stationarity; and acoustic regularities.

We provided a step-by-step guide to the use of ARC to generate lexicons (of arbitrary size) comprised of artificial pseudowords (of arbitrary syllable structure) embedded in TP-structured and TP-random streams (of arbitrary length). Finally, we showed that ARC controls for all these confounding factors relative to an unconstrained baseline lexicon generation pipeline and to a benchmark set of stimuli used in previous studies on linguistic statistical learning.

In sum, the ARC Toolbox constitutes a novel type system that provides a general framework, a data model, for linguistic objects. In this sense, ARC uses a *dual-type-system* to (a) represent elements (e.g., phonemes), and (b) contain multiple objects (e.g. registers to search into a dictionary). Hence, it can be flexibly used and adapted for multiple purposes that are optimized for, but not limited to, experimental design in linguistic statistical learning. In this particular domain, our results show that artificial languages generated with the ARC Toolbox reduce phonological and acoustic confounds. Below, ARC's core features are discussed highlighting advantages and limitations.

*ARC's inventories are tailored to the statistics of existing languages*

One core feature of the ARC Toolbox is the selection of various syllable types (e.g., open CV, CVC syllables, etc.) to create multi-syllabic pseudowords that form artificial lexicons of arbitrary size. The option to set the number and class of phonemes, syllables, and pseudowords per lexicon stream allows for flexible stimulus design. Importantly, ARC also offers the possibility to tailor the inventory of pseudowords to the statistics of existing languages based on corpus data. This entails filtering the inventory based on the frequency of use of syllables in a corpus, as well as the phoneme bigram, trigram, and positional probabilities. Thus, it moderates the interference effect that prior knowledge of native language phonemic representations may exert on artificial-language-learning performance (Elazar et al., 2022; Siegelman et al., 2018; Varela et al., 2024). While computational models held that frequency and transitional constraints are dissociable (Erickson & Thiessen, 2015), both of these sources of information are tracked by the brain (Tobia et al., 2012). Hence, studies that focus on neural tracking of TPs may also benefit from frequency controls that would mitigate this confound.

Furthermore, ARC implements phonotactic controls to ensure that artificial pseudowords obey the OCP, a phonotactic rule that was previously shown to partially account for neural tracking at the rate of statistical patterns (van der Wulp, 2021). While ARC's phonotactic controls are limited to OCP, we emphasize that researchers may add complementary code to control for phonotactic rules based on the native language of their target population.

*ARC suppresses acoustic and phonological confounds*

Acoustic and phonological confounds pose a challenge to the research in statistical learning (Luo & Ding, 2020; Pinto et al., 2022; van der Wulp, 2021; Varela et al., 2024). The ARC Toolbox addresses these confounds systematically by computing a dedicated PRI to quantify the periodicity of all phonological features at the rate of interest in the syllable streams; and by adopting a position-fixed baseline to eliminate acoustic differences between TP-structured and TP-random streams. Our results show that the lexicons generated with ARC reduce the presence of these confounding factors relative to a random selection of pseudowords, but also to a benchmarking set of stimuli that have been used in prior studies on auditory statistical learning. The lexicons employed in the prior literature certainly attained a better phonological rhythmicity control than our random baseline, which must be due to the expert conduct of the prior studies. Nevertheless, the lexicons generated with ARC produce a distribution of PRI across phonological features with an even lower mean and maximum relative to this benchmark. These results indicate that ARC drastically reduces the impact that the periodicity of phonological features may impose on the syllable streams. Indeed, previous studies suggest that low-frequency cortical tracking in speech may reflect phoneme-level processing (Di Liberto et al., 2015; but see also Daube et al., 2019). Furthermore, recent perspectives on language acquisition proposed that the infants' brain may track sustained phonological features in speech, as the permanency of such features through longer stretches of time matches the processing timescales of brain activity that they display at that age (Menn et al., 2023). Based on this work, it cannot be excluded that the sudden mapping of phonetic features to phonological categories can result in periodic activity at the rate of

repeating phonological categories that are regularly distributed in a sequence. In this context, we think that our PRI may be useful in future studies to test cortical tracking of specific phonological features. For instance, minimal edits to the function that generates artificial lexicons may lead to the creation of artificial speech streams that maximize the PRI of one, or more, phonological features, which could be decisive for the systematic assessment of cortical tracking of individual phonological features.

*ARC generates sequences with thorough statistical controls*

Beyond its utility for stimulus development in artificial language learning studies, ARC features a new pseudo-randomization method to precisely manipulate the TPs across abstract tokens in a stream. Contrary to classical random-walks, our PRW does not select new items fully at random based on a pre-specified TP distribution; rather, it limits next-token selection to always "explore" all possible paths evenly based on a memory matrix. This distinctiveness of our method is particularly important when there exist constraints on the desired length of a stream. Based on the Fundamental Theorem of Markov Chains, TPs derived from classical random walks will approach the stationary distribution as the number of elements in the sequence grows. Instead, our method stabilizes TPs already for short sequences, and it does so far more precisely than random shuffling approaches. Since TP stationarity has implications for the interpretability of statistical learning findings (Hasson, 2017), this method is recommended for any statistical learning design, even outside of the linguistic domain.

*Conclusions*

Prior linguistic knowledge, phonotactics, phonological regularities, and acoustic spectral differences are serious confounds that limit the interpretability of behavioral and neurobiological findings about linguistic statistical learning. To address these confounds, we presented a new Python toolbox that generates *Artificial languages with Rhythmicity Controls* (ARC). ARC can tailor the generation of artificial lexicons to the linguistic properties (e.g., phonotactics) of already existing languages, and minimize the confounds that are due to spurious rhythmicity of phonological features in acoustically matched TP-structured and TP-random streams with stationary TP distributions. These controls allow

for a safer assessment of the effects of statistical learning in the absence of acoustic, phonological and statistical confound; thus, making ARC particularly suitable for experimental designs that require thorough rhythmicity controls (e.g., NFT paradigms). Therefore, we recommend the use of ARC in future artificial language learning studies to rule out potential lurking variables from the implications of empirical findings on linguistic statistical learning.

**Acknowledgments**

The study was supported by the Max Planck Society through the award of the Independent Max Planck Research Group Language Cycles to LM. NM is supported by BMBF (Federal Ministry of Education and Research) through ACONITE (01IS22065), the Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI.) Leipzig, and the Max Planck IMPRS CoNI Doctoral Program. The authors declare no competing interests. LT and LM conceptualized research. LT and NM developed the software. LT and NM created visualizations. LM supervised research. LM acquired funding. LT wrote the manuscript. LT, NM, and LM edited the manuscript.

**Declarations**

**Funding:** The study was supported by the Max Planck Society through the award of the Independent Max Planck Research Group Language Cycles to LM. NM is supported by BMBF (Federal Ministry of Education and Research) through ACONITE (01IS22065), the Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI.) Leipzig, and the Max Planck IMPRS CoNI Doctoral Program.

**Conflict of interest statement:** The authors declare no competing interests.

**Ethics approval:** Not applicable

**Consent to participate:** Not applicable

**Consent for publication:** Not applicable

**Availability of data and materials:** Materials are available at: https://github.com/milosen/arc

**Code availability:** The code is available at the GitHub repository: https://github.com/milosen/arc

**Authors' contributions:** LT and LM conceptualized research. LT and NM developed the software. LT and NM created visualizations. LM supervised research. LM acquired funding. LT wrote the manuscript. LT, NM, and LM edited the manuscript.

## Open Practices Statement

Materials and code are available at https://github.com/milosen/arc.

## References


Arnold, D., & Tomaschek, F. (2016). *The Karl Eberhards Corpus of spontaneously spoken Southern German in dialogues-audio and articulatory recordings*.

Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature Neuroscience*, *22*, 627–632. https://doi.org/10.1038/s41593-019-0353-z

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004

Boersma, P., & Weenink, D. (2007). *PRAAT: Doing phonetics by computer (Version 5.3.51)*.

Bogaerts, L., Richter, C. G., Landau, A. N., & Frost, R. (2020). Beta-Band Activity Is a Signature of Statistical Learning. *The Journal of Neuroscience*, *40*(39), 7523–7530. https://doi.org/10.1523/JNEUROSCI.0771-20.2020

Boll-Avetisyan, N., & Kager, R. (2014). OCP-PLACE in Speech Segmentation. *Language and Speech*, *57*(3), 394–421. https://doi.org/10.1177/0023830913508074

Buiatti, M., Peña, M., & Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *NeuroImage*, *44*(2), 509–519. https://doi.org/10.1016/j.neuroimage.2008.09.015

Chen, Y., Jin, P., & Ding, N. (2020). The influence of linguistic information on cortical tracking of words. *Neuropsychologia*, *148*, 107640. https://doi.org/10.1016/j.neuropsychologia.2020.107640

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., Pujol, J., & Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *NeuroImage*, *48*(3), 541–553. https://doi.org/10.1016/j.neuroimage.2009.06.069

Cunillera, T., Toro, J. M., Sebastián-Galles, N., & Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues

on continuous speech segmentation: An event-related brain potential study. *Brain Research*, *1123*(1), 168–178. https://doi.org/10.1016/j.brainres.2006.09.046

Daube, C., Ince, R. A. A., & Gross, J. (2019). Simple Acoustic Features Can Explain Phoneme-Based Predictions of Cortical Responses to Speech. *Current Biology*, *29*(12), 1924-1937.e9. https://doi.org/10.1016/j.cub.2019.04.067

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. *Current Biology*, *25*(19), 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030

Elazar, A., Alhama, R. G., Bogaerts, L., Siegelman, N., Baus, C., & Frost, R. (2022). When the "Tabula" is Anything but "Rasa:" What Determines Performance in the Auditory Statistical Learning Task? *Cognitive Science*, *46*(2), e13102. https://doi.org/10.1111/cogs.13102

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. https://doi.org/10.1016/j.dr.2015.05.002

Fernandes, T., Kolinsky, R., & Ventura, P. (2009). The metamorphosis of the statistical segmentation output: Lexicalization during artificial language learning. *Cognition*, *112*(3), 349–366. https://doi.org/10.1016/j.cognition.2009.05.002

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128–1153. https://doi.org/10.1037/bul0000210

Getz, H., Ding, N., Newport, E. L., & Poeppel, D. (2018). Cortical tracking of constituent structure in language acquisition. *Cognition*, *181*, 135–140. https://doi.org/10.1016/j.cognition.2018.08.019

Hasson, U. (2017). The neurobiology of uncertainty: Implications for statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160048. https://doi.org/10.1098/rstb.2016.0048

Hay, J. F., & Saffran, J. R. (2012). Rhythmic Grouping Biases Constrain Infant Statistical Learning. *Infancy*, *17*(6), 610–641. https://doi.org/10.1111/j.1532-7078.2011.00110.x

Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., Doyle, W., Devinsky, O., & Melloni, L. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Science Advances*, *7*(8), eabc4530. https://doi.org/10.1126/sciadv.abc4530

Johnson, E. K., & Jusczyk, P. W. (2001). Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language*, *44*(4), 548–567. https://doi.org/10.1006/jmla.2000.2755

Kiai, A., & Melloni, L. (2021). What canonical online and offline measures of statistical learning can and cannot tell us. *bioRxiv*. https://doi.org/10.1101/2021.04.19.440449

Kim, S., Broersma, M., & Cho, T. (2012). The use of prosodic cues in learning new words in an unfamiliar language.

*Studies in Second Language Acquisition*, *34*(3), 415–444.

Luo, C., & Ding, N. (2020). Cortical encoding of acoustic and linguistic rhythms in spoken narratives. *eLife*, *9*. https://doi.org/10.7554/eLife.60433

MATLAB. (2024). *Version R20240a*. The MathWorks Inc.

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the Language Code: Neural Mechanisms Underlying Speech Parsing. *The Journal of Neuroscience*, *26*(29), 7629–7639. https://doi.org/10.1523/JNEUROSCI.5501-05.2006

Menn, K. H., Männel, C., & Meyer, L. (2023). Does Electrophysiological Maturation Shape Language Acquisition? *Perspectives on Psychological Science*, *18*(6), 1271–1281. https://doi.org/10.1177/17456916231151584

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, *53*(2), 225–237. https://doi.org/10.1016/j.jml.2005.02.011

Ordin, M., Polyanskaya, L., Soto, D., & Molinaro, N. (2020). Electrophysiology of statistical learning: Exploring the online learning process and offline learning product. *European Journal of Neuroscience*, *51*(9), 2008–2022. https://doi.org/10.1111/ejn.14657

Pinto, D., Prior, A., & Zion Golumbic, E. (2022). Assessing the Sensitivity of EEG-Based Frequency-Tagging as a Metric for Statistical Learning. *Neurobiology of Language*, *3*(2), 214–234. https://doi.org/10.1162/nol_a_00061

Ramers, K. H. (1998). *Einführung in die Phonologie*. Fink.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Schepens, J., Hout, R. van, & Jaeger, T. F. (2020). Big data suggest strong constraints of linguistic similarity on adult language learning. *Cognition*, *194*, 104056. https://doi.org/10.1016/j.cognition.2019.104056

Schiel, F. (2010, May). BAStat: New Statistical Resources at the Bavarian Archive for Speech Signals. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.

Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002). Does grammar start where statistics stop? *Science*, *298*(5593), 553–554.

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213. https://doi.org/10.1016/j.cognition.2018.04.011

Taylor, P., & Isard, A. (1997). SSML: A speech synthesis markup language. *Speech Communication*, *21*(1), 123–133. https://doi.org/10.1016/S0167-6393(96)00068-4

Tobia, M. J., Iacovella, V., Davis, B., & Hasson, U. (2012). Neural systems mediating recognition of changes in statistical regularities. *NeuroImage*, *63*(3), 1730–1742. https://doi.org/10.1016/j.neuroimage.2012.08.017

van der Wulp, I. (2021). *Word segmentation: TP or OCP? A re-analysis of Batterink & Paller (2017)* [Master's Thesis].

Varela, I. G., Orpella, J., Poeppel, D., Ripolles, P., & Assaneo, M. F. (2024). Syllabic rhythm and prior linguistic knowledge interact with individual differences to modulate phonological statistical learning. *Cognition*, *245*, 105737. https://doi.org/10.1016/j.cognition.2024.105737