

RESEARCH

Open Access



Combining statistics: the role of phonotactics on cross-situational word learning

Rodrigo Dal Ben^{1*} , Débora de Hollanda Souza¹ and Jessica F. Hay²

Abstract

Language learners can rely on phonological and semantic information to learn novel words. Using a cross-situational word learning paradigm, we explored the role of phonotactic probabilities on word learning in ambiguous contexts. Brazilian-Portuguese speaking adults ($N = 30$) were exposed to two sets of word-object pairs. Words from one set of labels had slightly higher phonotactic probabilities than words from the other set. By tracking co-occurrences of words and objects, participants were able to learn word-object mappings similarly across both sets. Our findings contrast with studies showing a facilitative effect of phonotactic probability on word learning in non-ambiguous contexts.

Keywords: Statistical learning, Phonotactic probability, Cross-situational word learning, Language learning, Word learning

Introduction

Most everyday word learning unfolds in phonologically rich and referentially ambiguous contexts (Quine, 1960). One phonological regularity that has been shown to influence word learning is phonotactic probability, which can be defined as positional statistics that represent how frequently phonological segments happen together in a given language (Vitevitch & Luce, 2004). Words with higher phonotactic probabilities contain common segments from a language, whereas words with lower phonotactic probabilities contain less frequent segments. For instance, due to variations in the phonotactic probabilities of their initial biphones, the word *fall* has an overall higher phonotactic probability than the word *tall* but has a lower

phonotactic probability than the word *call*.¹ One way to investigate the role phonotactic probabilities play in word learning is to create novel words with different degrees of phonotactic probabilities and then pair them with novel referents. By doing so, research has shown that, across the lifespan, words with higher phonotactic probabilities are learned faster and with greater accuracy than words with lower phonotactic probabilities (e.g., Benitez & Saffran, 2021; Estes & Bowen, 2013; Estes et al., 2011; Steber & Rossi, 2020; Storkel et al., 2013; Sundara et al., 2022; but see Cristia, 2018). Most of these studies use unambiguous word learning paradigms, in which one novel word is paired with one novel referent in a given trial. In natural contexts, however, word learning usually unfolds across ambiguous contexts (e.g., Clerkin et al., 2017). Thus, it is important to understand how phonotactic

*Correspondence: dalbenwork2@gmail.com

¹ Department of Psychology, Universidade Federal de São Carlos, São Carlos, SP, Brazil

Full list of author information is available at the end of the article

¹ Biphone log-based phonotactic probabilities calculated using Vitevitch and Luce (2004) online calculator: fall (/fsl/), 0.0050; tall (/tsl/), 0.0044; call (/kəl/), 0.0060. For details on how these values are calculated, see the Stimuli section.

probabilities impact word learning in these more ecologically relevant (i.e., ambiguous) contexts.

Work by Fitneva et al. (2009) has begun to shed some light on this issue by investigating how a related phonological cue, namely phonological typicality, impacts word learning in ambiguous contexts. Phonological typicality indexes how typical the phonology of a word is in relation to its lexical category (e.g., verbs, nouns, adjectives). It is calculated by computing the distance of each phonemes' features, in each word position, in relation to words from the same lexical category. Children (7-years-old) were presented with ambiguous trials with a novel word and two pictures depicting actions or objects. Novel words had different degrees of phonological typicality as verbs or nouns. Results showed that participants relied on typicality to make initial associations, choosing actions or objects as a function of the labels' typicality as verbs or nouns. Following their initial associations, they received feedback on their choices. Once feedback started, participants ignored typicality and relied on mutual exclusivity to map novel words. The authors argue that by relying on phonological information to make initial guesses, learners would be better situated to learn word-referent relations in complex ambiguous contexts, and this initial bias could have a cascading effect on word learning.

One semantic regularity that does not rely on contingent feedback and that can help solve referential ambiguity is the co-occurrence of words and referents. For instance, in a seminal study, Yu and Smith (2007) showed that by comparing word-referent co-occurrences across ambiguous trials, adults were able to solve the ambiguity of individual trials and learn novel words. Their paradigm is known as cross-situational word learning. Although the exact cognitive mechanism involved in this paradigm is still a matter of debate, with some defending a gradual aggregation of information and others a sequential hypothesis testing (for an overview, see Yurovsky & Frank, 2015), there is now evidence suggesting that infants (Smith & Yu, 2008), children (Vlach & DeBrock, 2017), and older adults (Peñaloza et al., 2017) can track word-referent co-occurrences to learn novel words (nouns and verbs; e.g., Fitneva & Christiansen, 2011) in ambiguous contexts (for a recent meta-analysis, (Dal Ben R, Souza DH, Hay JF: Cross-situational word learning: Systematic review and meta-analysis, unpublished)).

Most cross-situational word learning studies use novel words with legal phonotactics (e.g., McGregor et al., 2013; Smith & Yu, 2013), making them plausible labels. However, we do not know whether varying degrees of phonotactic probability impact cross-situational word learning (e.g., Alt et al., 2014). Uncovering any links between phonotactics and cross-situational word learning provides a more comprehensive understanding of

how different sources of statistical information may interact to promote or impair word learning (Bohn et al., 2021). This is especially important considering the balance between variability and consistency in word learning in natural environments (Braginsky et al., 2019) and the multimodal nature of statistical language learning (Saffran, 2020; Smith et al., 2018).

For instance, the phonotactic probability of a word might continue to impact its learnability under conditions of ambiguity. Alternatively, the increased task complexity resulting from the inherent ambiguity present in a cross-situational word learning may diminish any potential impact of phonotactics, driving learners to focus on co-occurrences rather than phonological information. To that end, we designed an exploratory study to investigate whether phonotactic knowledge, gathered in pre-experimental experience with natural language, would guide word learning in ambiguous contexts. Critically, all our stimuli had high phonotactic probabilities in Brazilian-Portuguese, the language of our participants. However, some words were slightly more probable than others. We decided to use such subtle differences in phonotactic probabilities for two reasons. First, by having high phonotactic probabilities, all stimuli were good label candidates, but some were better than others. Second, we found that these subtle differences were enough to boost or impair speech segmentation with the same population (Dal Ben et al., 2021) indicating that participants were able to perceive the phonotactic differences between words. Using the same stimuli allowed us to investigate whether or how phonotactic probabilities with known effects in an auditory statistical learning task would be integrated with semantic information to impact cross-situational word learning (e.g., Fitneva et al., 2009; Räsänen & Rasilo, 2015). Our integrative effort is in line with recent discussions on the scope and multimodality of statistical language learning (e.g., Saffran, 2014, 2020; Smith et al., 2014, 2018). Finally, the current study will increase the generalizability of previous findings (e.g., Yu & Smith, 2007; Chen & Yu, 2017) by replicating cross-situational word learning with Brazilian-Portuguese speaking adults.

Method

Participants

Thirty adults ($M_{age} = 22.23$ years ± 4.5 *SD*, 25 females), all Brazilian-Portuguese native speakers with no reported visual or auditory impairment, participated. They were recruited at the Universidade Federal de São Carlos and received no compensation for their participation (Ethics Committee approval #1.484.847, #3.085.914).

Given the absence of prior research that could inform a power analysis of the impact of phonotactic probabilities on cross-situational word learning, we opted for a sample

Table 1 Phonetic transcription (IPA), Phonotactic Probabilities (PP) of the set with pseudowords with highest possible phonotactic probabilities (PP+) and the set with the lower phonotactic probabilities (PP−)

PP+ ^a			PP− ^b		
	IPA	PP		IPA	PP
dini	[dʒini]	0.0090	nipe	[nipe]	0.0066
deta	[deta]	0.0085	tadi	[taʒi]	0.0074
pemi	[pemi]	0.0082	mide	[mide]	0.0075
sute	[sute]	0.0084	teba	[teba]	0.0074
viko	[viko]	0.0080	kosu	[kosu]	0.0078
bara	[bara]	0.0090	ravi	[ravi]	0.0073
Mean		0.0085	Mean		0.0073

^a Items with the highest possible phonotactic probabilities (before becoming words) in Brazilian-Portuguese

^b Items with slightly lower phonotactic probabilities, but that still had relatively high phonotactic probability

size that would allow us to capture cross-situational word learning at an above chance level, regardless of the potential effects of phonotactic probabilities. Our sample size was based on Yu and Smith (2007, Experiment 1, 2 × 2 condition), who reported a large effect size of $d = 4.37$ with a sample of 38 adults. A post hoc power analysis (one-sample t -test against chance, 0.25, with alpha at 0.05; Faul et al., 2007) estimated that our sample size provided more than 80% of power to detect cross-situational word learning at an above chance level.

Stimuli

Twelve novel words and twelve novel objects were used. Words came from Dal Ben et al. (2021). To ensure the tight control of phonotactic probabilities, words were created in three steps. First, the algorithm proposed by Vitevitch and Luce's (2004) was applied to a database of Brazilian-Portuguese biphones (Estivalet & Meunier, 2015) in the following way: Biphones' log (base 10) phonotactic probabilities were calculated by dividing the sum of the log frequency of each biphone (token) on each word position by the total log frequency of words (token) with biphones in that given position (e.g., log frequency of /mæ/ as the first biphone divided by the total log frequency of all words with at least one biphone). Log transformations were used because they were reported to better correlate with performance in linguistic tasks compared to raw frequency (Vitevitch & Luce, 2004). Second, a search engine was created to find and concatenate biphones. Using this engine, six novel disyllabic words with consonant–vowel structure (CVCV) and with the highest possible phonotactic probability in Brazilian-Portuguese were created (labeled PP+; Table 1). Finally, their biphones were recombined

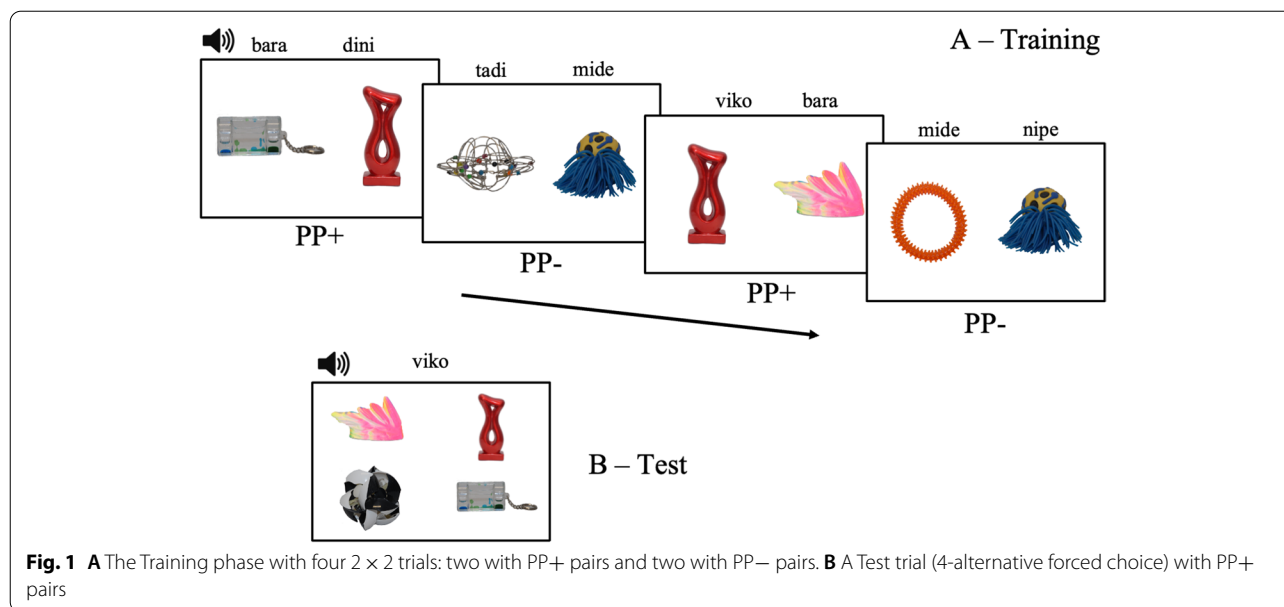
to create other six novel words that had slightly less probable, but still high, phonotactic probabilities (labeled PP−; Table 1). Third, both PP+ and PP− sets were recorded using MBROLA speech synthesizer with the female Brazilian-Portuguese database br4 (Dutoit et al., 1996). Each word lasted for 696 ms, had a mean F0 of 220 Hz, and a mean intensity of 77 dB.

Objects were 3D pictures from the NOUN database (Horst & Hout, 2016) with high levels of discriminability ($M = 90%$) and novelty ($M = 77%$). Words and objects were randomly paired. Also, to avoid spurious relations, a counterbalanced version of the pairs was created by switching objects across sets. Participants were randomly assigned to one of these versions. All stimuli are openly available at OSF (<https://osf.io/6fqzg/>).

Design

We used a cross-situational word learning design (Yu & Smith, 2007). The experimental task had two phases: Training and Test. During Training, participants were passively exposed to the 12 word–objects pairs across a series of ambiguous trials (2 × 2; Yu & Smith; Fig. 1). Training trials began with two objects displayed on a screen, side by side. After 950 ms of silence, a word corresponding to one of the objects was played (≈ 696 ms), followed by a silent pause of 700 ms; then another word was played (≈ 696 ms), followed by another silent pause of 950 ms (total duration ≈ 4 s; cf., Yu & Smith). Across trials, there was no reliable correspondence between the position of the objects (left or right) and the order that words were played (first or second). Each of the 12 word–object pairs was presented six times, for a total of 36 trials. Importantly, we used an interleaved presentation of PP+ and PP− word–object pairs. Trials contained words with either higher (PP+) or lower (PP−) phonotactics. Words from different sets never appeared together in the same trial. Thus, for each set, each label was contrasted with the other 5 labels (6 PP+ pairs and 6 PP− pairs).

During the Test phase, each trial began with four objects from the same stimuli set (either PP+ or PP−; never mixed) displayed on each corner of the screen. After 1 s of silence, a word was played (≈ 696 ms), and participants chose the matching object, with no time limit. Each of the 12 word–object pairs were tested twice, for a total of 24 trials (note that we conducted two sets of analyses, one with both trials and another with only the first test trial of each pair; see the Data analysis section). Between each Training and Test trials, a blank screen with a central cross was displayed for 1 s to reorient participants' gaze to the middle of the screen. Also, between Training and Test, two warm-up trials were conducted to familiarize participants with the structure of the upcoming Test trials.



Each warm-up trial displayed four known objects in each corner of the screen (i.e., a house, a duck, a ball, a cat), followed by an audio label of one of the objects (e.g., “House,” in Brazilian-Portuguese). The task was programmed and computer-administered using Psychopy (Peirce et al., 2019).

Procedure

Participants were seated in a sound-attenuated room in front of a 17" computer monitor and were fitted with headphones (AKG 240 powered by a Fiio E10K dac/amp). The experimental task began with instrumental music playing at the same volume as the subsequent experimental stimuli (77 dB). Participants were instructed to adjust the volume to a comfortable level. Participants were then instructed that they would: “hear novel words and see novel objects and that their task was to discover word-object relations” (all instructions are available at OSF: <https://osf.io/6fqzg/>). They were not told that each word corresponded to only one object. Next, the Training phase (as described in the Design section) began and lasted for approximately 3 min. In the subsequent warm-up trials (as described in the Design section), participants were instructed to select the matching object by pressing keys 1, 2, 3, or 4 on a custom keyboard, corresponding to objects in the upper left corner, upper right corner, lower left corner, or lower right corner, respectively. Finally, the Test phase (as described in the Design section) began and lasted for an average of 2 min. To ensure compliance to the instructions,

during the entire experiment, participants were monitored by a close-circuit television.

Data analysis

Our main dependent measure was accuracy (either correct or incorrect selections) during Test trials. However, testing each pair of stimuli twice could have provided participants with additional learning opportunities during Test phase. To account for that, and following literature (e.g., Yu & Smith, 2007), we conducted two sets of analyses: one with both test trials for each pair (24 trials, full dataset) and another with only the first test trial for each pair (12 trials, halved dataset).

For both set of analyses, trials in which reaction times were greater than 3 SDs away from the mean were excluded, as they were most likely the result of participant distraction. Across all participants, a total of 16 trials were excluded from the full dataset and 9 from the halved dataset (either way, 2% of the data). Next, we modeled our binomial (correct or incorrect) and repeated measures (either 24 or 12 trials per participant) using mixed logistic regressions. We used Frequentist and Bayesian approaches (lme4 and brms packages for R; Bates et al., 2015; Bürkner, 2018; R Core Team, 2017). The dependent variable was the selection of the target object during Test (either correct or incorrect). To assess the relationship between phonotactics and above chance word learning, our fixed effects were the chance level (logit of 0.25) and stimuli phonotactics (PP- or PP+). We started with the maximal random structure with pairs (stimuli) as random slopes and participants as

random intercepts (Barr et al., 2013). This model failed to converge for the Frequentist approach, but converged for the Bayesian approach, which we report. We then pruned the Frequentist model to include random intercepts for stimuli and participants, this model converged.

It is worth noting that the PP- set was the reference level in the models. Thus, the intercept measures the chances of selecting PP- pairs above chance level (0.25). The odds ratio for selecting PP+ pairs reflect a change in odds from this reference (i.e., PP- above chance level). To arrive at the odds of choosing PP+ pairs, we multiplied the intercept odds by the PP+ odds (Sommet & Morselli, 2017). Finally, given the exploratory nature of our investigation, we do not report *p*-values for our Frequentist analyses (Scheel et al., 2020). Scripts and data are openly available at OSF (<https://osf.io/6fqzg/>).

Results and discussion

Results from the full dataset (24 trials) and the halved dataset (12 trials) modeled by Frequentist or Bayesian models were comparable. Participants selected the correct objects above chance level for both PP- and PP+ pairs (Fig. 2). Furthermore, participants were much more likely to choose the correct objects rather than the incorrect ones for both PP- and PP+ pairs (Tables 2 and 3). The odds ratio for choosing PP+ pairs was just slightly higher than the odds ratio for choosing PP-. The complete models' outputs are available at OSF (<https://osf.io/6fqzg/>).

In contrast to evidence suggesting that words with higher phonotactic probabilities are learned faster and more accurately than words with lower phonotactic probabilities (e.g., Gonzalez-Gomez et al., 2013; Estes et al., 2011; Storkel, 2004; Storkel et al., 2013), we found only small differences between PP+ and PP- on our cross-situational word learning task. Indeed, our task was considerably more complex than word learning in unambiguous tasks. On top of having to track word-object co-occurrences across trials to solve referential ambiguity, participants also had to track two independent sets of pairs. Thus, it is unsurprising that participants focused on tracking word-object co-occurrences to solve ambiguities rather than phonological information. Furthermore, the overall accuracy in our task (~ 66%) is comparable to more complex cross-situational word learning studies (e.g., Chen & Yu, 2017; for a recent meta-analysis, (Dal Ben R, Souza DH, Hay JF: Cross-situational word learning: Systematic review and meta-analysis, unpublished)).

These preliminary results prompt a careful examination of the role of phonotactic probabilities in more complex learning environments, with multiple semantic and phonological regularities (Lany & Saffran, 2013; Saffran, 2014; Smith et al., 2014, 2018). Phonotactics might assume different roles depending on environmental complexity. If confirmed by future studies, our exploratory findings might add to the literature pointing to a hierarchical organization of statistical cues as

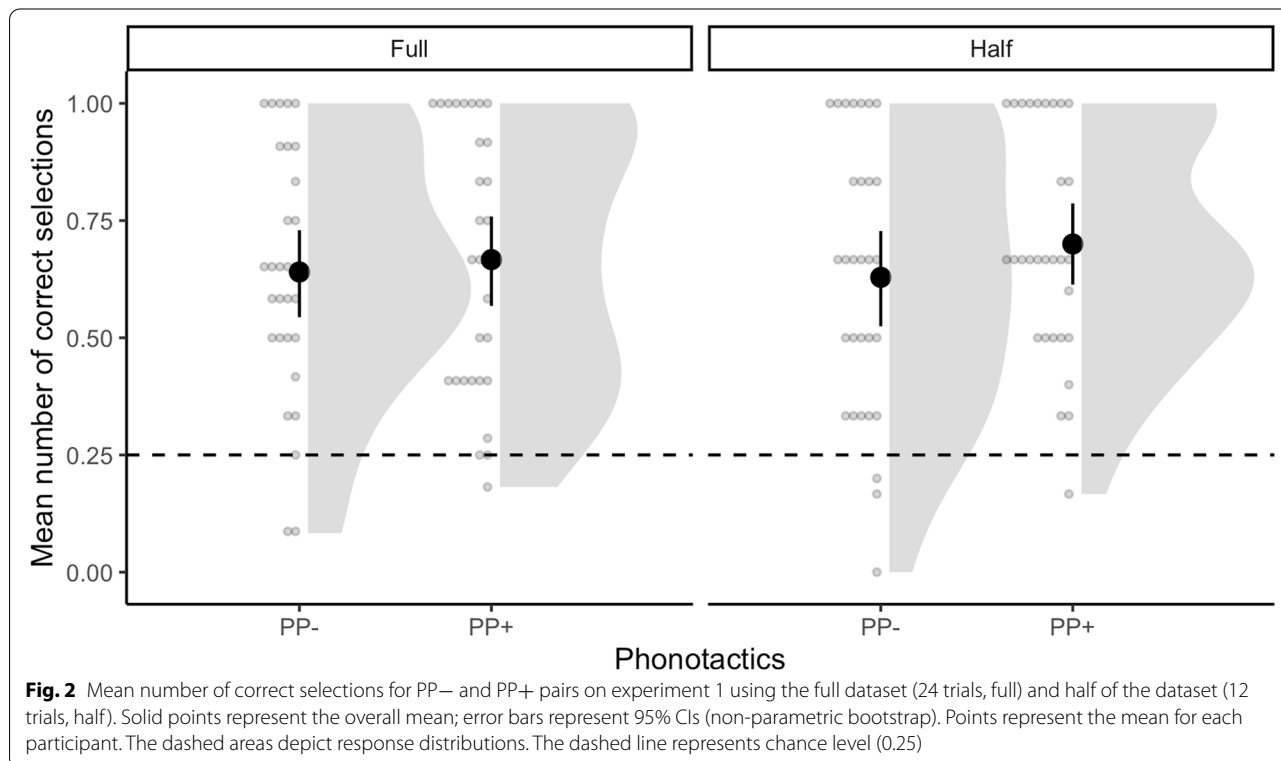


Fig. 2 Mean number of correct selections for PP- and PP+ pairs on experiment 1 using the full dataset (24 trials, full) and half of the dataset (12 trials, half). Solid points represent the overall mean; error bars represent 95% CIs (non-parametric bootstrap). Points represent the mean for each participant. The dashed areas depict response distributions. The dashed line represents chance level (0.25)

Table 2 Fixed and random effects and specifications for Frequentist models

Model specification (R)	Selection ~ chance level + phonotactics + (1 stimuli) + (1 participant)			
	Full dataset		Half dataset	
Fixed effects				
Predictors	Odds ratio	Confidence interval	Odds ratio	Confidence interval
PP– (intercept)	7.61	3.89–14.90	6.25	3.46–11.29
PP+	1.19	0.83–1.71	1.56	0.95–2.56
Random effects				
σ^2	3.29		3.29	
τ_{00} participants	2.35		1.45	
τ_{00} stimuli	0.21		0.09	
ICC	0.44		0.32	
N_{stimuli}	12		12	
$N_{\text{participants}}$	30		30	
Observations	704		351	
Marginal R^2 /conditional R^2	0.001/0.438		0.010/0.327	

Table 3 Fixed and random effects and specifications for Bayesian models

Model specification (R)	Selection ~ chance level + phonotactics + (stimuli participant)			
	Full dataset		Half dataset	
Fixed effects				
Predictors	Odds ratio	Credible interval	Odds ratio	Credible interval
PP– (intercept)	11.93	4.54–35.96	9.42	3.97–28.61
PP+	1.03	0.54–2.01	1.71	0.78–3.87
Random effects				
σ^2	3.29		3.29	
τ_{00} participants	5.43		4.10	
ICC	0.62		0.55	
$N_{\text{participants}}$	30		30	
Observations	704		351	
Marginal R^2 /conditional R^2	0.001/0.470		0.007/0.468	

a function of environmental complexity. For instance, in previous research using the same set of stimuli used here, we found that these small differences in phonotactic probabilities could boost or impair speech segmentation based on transitional probabilities (Dal Ben et al., 2021; see also Finn & Hudson Kam, 2008; Mersad & Nazzi, 2011). In the present study, however, phonotactic probabilities might have assumed a secondary role in contrast to word-object co-occurrences. This is in line with previous research suggesting that phonological cues might have limited impact on word learning in ambiguous contexts in comparison to mutual exclusivity based on feedback (Fitneva et al., 2009). On the co-occurrence side, complexity has also been shown to dynamically modulate cross-situational word learning, with learners aggregating information across less complex learning

environments and changing to tracking and testing a few label candidates when complexity increases (Yurovsky & Frank, 2015).

As for potential limitations of the current study, both of our stimuli sets (PP+, PP–) had high phonotactic probabilities; thus, words from both sets could be perceived as good label candidates. It is possible that we would have observed a greater influence of phonotactic probability on cross-situational word-learning if their differences had been more salient (e.g., Storkel et al., 2013) or if we had contrasted legal vs. illegal phonotactics (e.g., Estes & Bowen, 2013) or even if PP– and PP+ had been mixed in the same trials (rather than interleaved across trials). Another concern could be that participants may not have perceived the subtle differences in phonotactics between our stimuli. Although previous research

investigating speech segmentation with the same set of stimuli suggests participants are indeed sensitive to these subtle differences in phonotactics (Dal Ben et al., 2021), future studies should verify participants' sensitivity to these phonotactic differences using neurophysiological measures (e.g., EEG), which can reveal implicit perception. Moreover, our sample size was set so it provided enough power to detect an above chance word learning (based on effect sizes reported by Yu & Smith, 2007)—which happened regardless of phonotactic differences between stimuli. When testing the effects of different semantic conditions within a cross-situational word learning task, Chen and Yu (2017) reported a lower effect size than the one we used to inform our power analysis. Mindful of differences in task complexity, future studies could base their power analyses on effect sizes such as those reported by Chen and Yu to ensure enough statistical power to detect small differences between phonotactic probabilities. Finally, as with many studies in the adult psycholinguistic literature, there are at least two potential constraints to our study's generality (Simons et al., 2017). First, our participants were young college students from a single language background. It is possible that our findings may not generalize to different populations. That said, as the first study to investigate Brazilian-Portuguese speakers in a cross-situational word learning task, our findings do provide evidence for the generality of cross-situational word learning. Second, our study used simplified stimuli (isolated words and highly discriminable objects) in an experimental design. Word learning in everyday life is much more complex, building on several statistical and social cues. Future studies should extend the current research by employing more ecological designs for assessing word learning (e.g., Bergelson et al., 2019; Clerkin et al., 2017).

Mastering a language is a daunting task. Luckily, we can take advantage of environmental regularities to learn a great deal about language. The initial findings we present here move us closer to understanding how regularities in the linguistic input can interact to shape word learning in dynamic environments marked by the constant balance of variance and consistency (Bohn et al., 2021; Braginsky et al., 2019), and ultimately leading to our complex and fascinating linguistic repertoires.

Abbreviations

PP+: Made-up words with highest possible phonotactic probability in Brazilian-Portuguese; PP−: Made-up word with slightly less probable phonotactic probability in Brazilian-Portuguese when compared to PP+ words.

Acknowledgements

We are grateful to all participants. This work was supported by grants from FAPESP (#2015/26389-7, #2018/04226-7) and CAPES (#001) to RDB; from INCT-ECCCE (National Institute on Cognition, Behavior and Teaching; CNPq #573972/2008-7, #465686/2014-1, FAPESP #2008/57705-8, #2014/50909-8) to

DHS; and from the NICHD (#R01HD083312) to JFH. The funders had no role in study design, data collection, analysis and interpretation of the data, decision to publish, or preparation of the manuscript.

Authors' contributions

CRedit statement: RDB: conceptualization, methodology, software, investigation, visualization, data formal analysis, writing—original draft. DHS and JFH: conceptualization, writing—review and editing, supervision, project administration, funding acquisition. The author(s) read and approved the final manuscript.

Funding

This work was supported by grants from FAPESP (#2015/26389-7, #2018/04226-7) and CAPES (#001) to RDB; from INCT-ECCCE (National Institute on Cognition, Behavior and Teaching; CNPq #/2008-7, #465686/2014-1, FAPESP #2008/57705-8, #2014/50909-8) to DHS; and from the NICHD (#R01HD083312) to JFH. RDB is now at Concordia University. The funders had no role in study design, data collection, analysis and interpretation of the data, decision to publish, or preparation of the manuscript.

Availability of data and materials

The datasets analyses scripts, and materials can be found in OSF: <https://osf.io/6fqzg/>.

Declarations

Ethics approval and consent to participate

We declare that our study was approved by the Ethics Committee of the Universidade Federal de São Carlos (#1.484.847, #3.085.914) and that participants consented to participate.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Psychology, Universidade Federal de São Carlos, São Carlos, SP, Brazil. ²Department of Psychology, University of Tennessee, Knoxville, TN, USA.

Received: 19 May 2022 Accepted: 18 September 2022

Published online: 28 September 2022

References

- Alt, M., Meyers, C., Oglivie, T., Nicholas, K., & Arizmendi, G. (2014). Cross-situational statistically based word learning intervention for late-talking toddlers. *Journal of Communication Disorders*, 52, 207–220. <https://doi.org/10.1016/j.jcomdis.2014.07.002>.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>.
- Benitez, V. L., & Saffran, J. R. (2021). Two for the price of one: Concurrent learning of words and phonotactic regularities from continuous speech. *PLoS One*, 16(6), e0253039. <https://doi.org/10.1371/journal.pone.0253039>.
- Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, 22(1), e12715. <https://doi.org/10.1111/desc.12715>.
- Bohn, M., Tessler, M. H., Merrick, M., & Frank, M. C. (2021). How young children integrate information sources to infer the meaning of words. *Nature Human Behaviour*, 5(8), 1046–1054. <https://doi.org/10.1038/s41562-021-01145-1>.
- Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, 3, 52–67. https://doi.org/10.1162/opmi_a_00026.

- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1), 395. <https://doi.org/10.32614/RJ-2018-017>.
- Chen, C., & Yu, C. (2017). Grounding statistical learning in context: The effects of learning and retrieval contexts on cross-situational word learning. *Psychonomic Bulletin & Review*, 24(3), 920–926. <https://doi.org/10.3758/s13423-016-1163-x>.
- Clerkin, E. M., Hart, E., Rehag, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), 20160055. <https://doi.org/10.1098/rstb.2016.0055>.
- Cristia, A. (2018). Can infants learn phonology in the lab? A meta-analytic answer. *Cognition*, 170, 312–327. <https://doi.org/10.1016/j.cognition.2017.09.016>.
- Dal Ben, R., Souza, D. H., & Hay, J. F. (2021). When statistics collide: The use of transitional and phonotactic probability cues to word boundaries. *Memory & Cognition*. <https://doi.org/10.3758/s13421-021-01163-4>.
- Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vrecken, O. (1996). The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In *Proceedings of Fourth International Conference on Spoken Language Processing, ICSLP*, (pp. 3, 1393–3, 1396). <https://doi.org/10.1109/ICSLP.1996.607874>.
- Estes, K., & Bowen, S. (2013). Learning about sounds contributes to learning about words: Effects of prosody and phonotactics on infant word learning. *Journal of Experimental Child Psychology*, 114(3), 405–417. <https://doi.org/10.1016/j.jecp.2012.10.002>.
- Estes, K., Edwards, J., & Saffran, J. R. (2011). Phonotactic constraints on infant word learning. *Infancy*, 16(2), 180–197. <https://doi.org/10.1111/j.1532-7078.2010.00046.x>.
- Estivalet, G. L., & Meunier, F. (2015). The Brazilian Portuguese Lexicon: An instrument for psycholinguistic research. *PLoS One*, 10(12), 1–24. <https://doi.org/10.1371/journal.pone.0144016>.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>.
- Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, 108(2), 477–499. <https://doi.org/10.1016/j.cognition.2008.04.002>.
- Fitneva, S. A., & Christiansen, M. H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cognitive Science*, 35(2), 367–380. <https://doi.org/10.1111/j.1551-6709.2010.01156.x>.
- Fitneva, S. A., Christiansen, M. H., & Monaghan, P. (2009). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language*, 36(5), 967–997. <https://doi.org/10.1017/S0305000908009252>.
- Gonzalez-Gomez, N., Poltrock, S., & Nazzi, T. (2013). A "bat" is easier to learn than a "tab": Effects of relative phonotactic frequency on infant word learning. *PLoS One*, 8(3). <https://doi.org/10.1371/journal.pone.0059601>.
- Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, 48(4), 1393–1409. <https://doi.org/10.3758/s13428-015-0647-3>.
- Lany, J., & Saffran, J. R. (2013). Statistical learning mechanisms in infancy. In J. Rubenstein, & P. Rakic (Eds.), *Neural circuit development and function in the brain*, (vol. #volume#, pp. 231–248). Elsevier. <https://doi.org/10.1016/B978-0-12-397267-5.00034-0>.
- McGregor, K. K., Rost, G., Arenas, R., Farris-Trimble, A., & Stiles, D. (2013). Children with ASD can use gaze in support of word recognition and learning. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 54(7), 745–753. <https://doi.org/10.1111/jcpp.12073>.
- Mersad, K., & Nazzi, T. (2011). Transitional probabilities and positional frequency phonotactics in a hierarchical model of speech segmentation. *Memory and Cognition*, 39(6), 1085–1093. <https://doi.org/10.3758/s13421-011-0074-3>.
- Pearce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>.
- Peñaloza, C., Mirman, D., Cardona, P., Juncadella, M., Martin, N., Laine, M., & Rodríguez-Fornells, A. (2017). Cross-situational word learning in aphasia. *Cortex*, 93, 12–27. <https://doi.org/10.1016/j.cortex.2017.04.020>.
- Quine, W. A. O. (1960). *Word and object*. MIT press.
- R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Räsänen, O., & Rasilo, H. (2015). A joint model of word segmentation and meaning acquisition through cross-situational learning. *Psychological Review*, 122(4), 792–829. <https://doi.org/10.1037/a0039702>.
- Saffran, J. (2014). Sounds and meanings working together: Word learning as a collaborative effort. *Language Learning*, 64(s2), 106–120. <https://doi.org/10.1111/lang.12057>.
- Saffran, J. R. (2020). Statistical language learning in infancy. *Child Development Perspectives*, 14(1), 49–54. <https://doi.org/10.1111/cdep.12355>.
- Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2020). Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science*, 174569162096679. <https://doi.org/10.1177/1745691620966795>.
- Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, 12(6), 1123–1128. <https://doi.org/10.1177/1745691617708630>.
- Smith, L., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, 22(4), 325–336. <https://doi.org/10.1016/j.tics.2018.02.004>.
- Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106(3), 1558–1568. <https://doi.org/10.1016/j.cognition.2007.06.010>.
- Smith, L., & Yu, C. (2013). Visual attention is not enough: Individual differences in statistical word-referent learning in infants. *Language Learning and Development*, 9(1), 25–49. <https://doi.org/10.1080/15475441.2012.707104>.
- Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word-referent learning. *Trends in Cognitive Sciences*, 18(5), 251–258. <https://doi.org/10.1016/j.tics.2014.02.007>.
- Sommet, N., & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using Stata, R, Mplus, and SPSS. *International Review of Social Psychology*, 30(1), 203–218. <https://doi.org/10.5334/irsp.90>.
- Steber, S., & Rossi, S. (2020). So young, yet so mature? Electrophysiological and vascular correlates of phonotactic processing in 18-month-olds. *Developmental Cognitive Neuroscience*, 43, 100784. <https://doi.org/10.1016/j.dcn.2020.100784>.
- Storkel, H. L. (2004). Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighborhood density. *Journal of Speech, Language, and Hearing Research*. <https://doi.org/10.1092-4388/04/4706-1454>.
- Storkel, H. L., Bontempo, D. E., Aschenbrenner, A. J., Maekawa, J., & Lee, S.-Y. (2013). The effect of incremental changes in phonotactic probability and neighborhood density on word learning by preschool children. *Journal of Speech, Language, and Hearing Research*, 56(5), 1689–1700. [https://doi.org/10.1044/1092-4388\(2013\)12-0245](https://doi.org/10.1044/1092-4388(2013)12-0245).
- Sundara, M., Zhou, Z. L., Breiss, C., Katsuda, H., & Steffman, J. (2022). Infants' developing sensitivity to native language phonotactics: A meta-analysis. *Cognition*, 221, 104993. <https://doi.org/10.1016/j.cognition.2021.104993>.
- Vitevitch, M. S., & Luce, P. A. (2004). A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, 36(3), 481–487. <https://doi.org/10.3758/BF03195594>.
- Vlach, H. A., & DeBrock, C. A. (2017). Remember dax? Relations between children's cross-situational word learning, memory, and language abilities. *Journal of Memory and Language*, 93, 217–230. <https://doi.org/10.1016/j.jml.2016.10.001>.
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18(5), 414–420. <https://doi.org/10.1111/j.1467-9280.2007.01915.x>.
- Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition*, 145, 53–62. <https://doi.org/10.1016/j.cognition.2015.07.013>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.