Original Articles

Statistical learning and spelling: Evidence from Brazilian prephonological spellers☆

Rebecca Treiman[a,*], Cláudia Cardoso-Martins[b], Tatiana Cury Pollo[c], Brett Kessler[a]

[a] Department of Psychological and Brain Sciences, Washington University in St. Louis, United States
[b] Departamento de Psicologia, Universidade Federal de Minas Gerais, Brazil
[c] Departamento de Psicologia, Universidade Federal de São João Del–Rei, Brazil

ARTICLE INFO

ABSTRACT

We analyzed the spelling attempts of Brazilian children (age 3 years, 3 months to 6 years, 0 months) who were prephonological spellers, in that they wrote using letters that did not reflect the phonemes in the words they were asked to spell. We tested the hypothesis that children use their statistical-learning skills to learn about the appearance of writing and that older prephonological spellers, who have had on average more exposure to writing, produce more wordlike spellings than younger prephonological spellers. We found that older pre-phonological spellers produced longer spellings and were more likely to use letters and digrams in proportion to their frequency of occurrence in Portuguese. There were also some age-related differences in children's tendency to use letters from their own names when writing other words. The results extend previous findings with learners of English to children who are learning a more transparent orthography.

1. Introduction

The world is filled with patterns. People take advantage of these patterns by learning how often events occur and how often and under what circumstances they occur together. In this way, they develop a kind of mental statistics that helps them to deal with the environment. For example, someone who hears the word *abscond* for the first time may be able to determine its grammatical category based on the knowledge that two-syllable English words with stress on the second syllable are often verbs. Some patterns are explicitly taught, but others—such as this pattern about word stress—are not. In such cases, people pick up the patterns through implicit statistical learning.

Most studies of statistical learning have examined people's learning of artificial materials over short periods of time. For example, Chambers, Onishi, and Fisher (2003) constructed sets of spoken syllables in which some consonants occurred in the initial position but not the final position and other consonants showed the reverse pattern. Twenty-five such syllables were repeatedly played to infants over the course of 3–4 min. Infants' knowledge of the patterns was then tested by comparing the amount of time they spent listening to new syllables that conformed to the patterns and new syllables that did not conform. The results of this and other studies (e.g., Saffran, Aslin, & Newport, 1996) show that statistical learning is available even to infants and provide

some information about its properties. However, such studies may not fully capture statistical learning as it occurs in real life, where patterns may be probabilistic rather than all or none, exposure may be spaced rather than massed, the number of items to which learners are exposed may be large rather than small, and the delay between learning and testing may be long rather than short.

In the present study, we examined a real-life case of statistical learning: learning about the letter patterns in written words. The words of a language have characteristic lengths and letter patterns. In the written words of Portuguese, for example, the letters ‹a› and ‹o› are more common than the letters ‹e› and ‹i›. Among *digrams* (sequences of two adjacent letters), ‹ci› is more common than ‹cr›. Modern children see many examples of writing in their daily lives, giving them an opportunity to learn about such *graphotactic* patterns. In the present study, we assessed Brazilian preschoolers' knowledge about these patterns as reflected in their attempts to write words. Our main analyses involved *prephonological spellers*: children who use letters when asked to write, but not letters that make sense based on the sounds in the target items. Such a child might write ‹cicio› for *tartaruga* 'turtle', for example.

Despite their lack of knowledge of sound–letter relations, prephonological spellers appear to possess some knowledge about the letter patterns of their written language. For example, these children tend to use letters and digrams in proportion to the frequency with

which they occur in words of their language (Kessler, Pollo, Treiman, & Cardoso-Martins, 2013, for Brazilian prephonological spellers; Pollo, Kessler, & Treiman, 2009, for US and Brazilian prephonological spellers). The productions of prephonological spellers are also influenced by the characteristics of their given name, a word that is particularly important to them. These children use letters from their names rather often when asked to write other words—more often than expected based on the frequency of these letters in words in general (Treiman, Kessler, Boland, Clocksin, & Chen, 2018, for US prephonological spellers; see also Bloodgood, 1999, for US kindergartners and Both-de Vries & Bus, 2008, for Dutch children).

Older prephonological spellers have had, on average, more exposure to writing than younger prephonological spellers. Based on a statistical-learning view, we might expect to see age differences in the productions of prephonological spellers that reflect these differences in experience. Treiman et al. (2018) found some such differences in a cross-sectional study of US prephonological spellers who ranged in age from 3;2 (years; months) to 5;6. The spellings produced by older prephonological spellers were on average longer than those produced by younger ones, and older prephonological spellers were more likely than younger ones to use digrams in proportion to their frequency of occurrence in books. Repeated-letter digrams, such as ‹ee› in ‹coeec›, and alphabetic-order digrams, such as ‹ab› in ‹abeucy›, were less common among older prephonological spellers than among younger ones. Although older prephonological spellers were more likely than younger ones to use digrams in proportion to their frequency of occurrence in English, older and younger prephonological spellers were influenced to the same extent by the frequency of individual letters. The overall proportion of letters in children's spellings that came from their own names did not vary significantly as a function of age, but the position of the name from which these letters came did vary, such that younger prephonological spellers were especially likely to use the first letter of their name when writing other words. This difference may reflect children's increased knowledge of letters in the name beyond the first letters over the preschool years.

In the present study, we asked whether age differences in the characteristics of prephonological spelling are found in a different language, Portuguese. This is an important question because the orthography studied by Treiman et al. (2018), English, is often described as deep. That is, English has complex and sometimes inconsistent links between spellings and sounds. Portuguese orthography is more shallow. Many studies have shown that learning to spell and read is influenced

by the orthographic depth of a writing system (e.g., Serrano et al., 2011; Seymour, Aro, & Erskine, 2003). One might expect that learners of a deep orthography would rely heavily on visual patterns and that the learning of such patterns would be more important for a deep orthography such as English than for a shallower orthography such as Portuguese. On the other hand, orthographic depth might not influence the children of primary interest here: those who do not yet use letters for their sound values. In that case, we might see age-related effects in Brazilian prephonological spellers that are similar to those found among US prephonological spellers.

To address these issues, we combined data from previous studies conducted by our research group in Brazil. We conducted a preliminary analysis using the full set of preschool data to determine whether older preschoolers were less likely than younger ones to be prephonological spellers. In our main analyses, we examined just the prephonological spellers. We hypothesized that the productions of older prephonological spellers would differ in some ways from the productions of younger ones, reflecting the older children's greater experience with Portuguese writing in general and their own written names in particular.

2. Method

2.1. Participants

We used data from preschool children who had participated in studies that had been conducted by our research group in Belo Horizonte, Brazil. Table 1 provides information about the studies and the participants. Of the children whose data were included in the studies listed in the table, we used data from children who attended classes designed for 4- and 5-year-olds in private preschools and who produced at least six spellings. At the time the data were collected, children from middle- and upper-class families generally attended private preschools for several years before they reached the age of mandatory school attendance, 6 years. In the preschool classes that the children in our analyses attended, children are read to, learn to write their names, and learn about the alphabet, but they do not receive extensive literacy instruction. In the preliminary analysis that we report, we included all children who fit the criteria mentioned above, regardless of whether their spelling was prephonological. The first column of data in Table 1 shows, for each study and each time point, the number of such children. The second column of data shows the mean ages of these children. This full data set included data from 313 different children (154 girls). There

Table 1
Information about children contributing data to analyses.

Study	All spellers		Prephonological spellers	
	N	Mean age	N	Mean age
Cardoso-Martins,[a] Time 1	61	4;3	33	4;3
Cardoso-Martins,[a] Time 2	70	4;6	32	4;5
Cardoso-Martins,[a] Time 3	73	4;10	19	4;10
Cardoso-Martins,[a] Time 4	75	5;3	13	5;1
Cardoso-Martins,[a] Time 5	75	5;6	3	5;0
Cardoso-Martins,[a] Time 6	76	5;10	2	5;8
Cardoso-Martins et al. (2006), Exp. 2, Time 1	43	5;3	13	5;2
Cardoso-Martins et al. (2006), Exp. 2, Time 2	44	5;10	2	5;6
Pollo, Treiman, and Kessler (2008), Exp. 2[b]	46	5;5	7	5;4
Pollo, et al. (2009)	79	4;11	35	4;7
(Pollo, 2008, Exp. 2): subset of children from Pollo et al. (2009) at Time 2, 1 year later	33	5;5	2	5;6
Treiman et al. (2013), Exp. 2	68	4;11	17	4;5
Treiman et al. (2013) follow-up: subset of children from Exp. 2 of Treiman et al. at Time 2, 3 months later	23	4;10	9	4;6

[a] Cardoso-Martins (2013), Kessler et al. (2013), and Treiman et al. (2013) reported analyses of this longitudinal data set examining different questions than those addressed here.

[b] Includes 16 children whose data were excluded from published report because they did not write the letters of interest for those analyses in a dictation task.