Non-word repetition in Yélî Dnye

Alejandrina Cristia[1] & Marisa Casillas[2,3]

[1] Laboratoire de Sciences Cognitives et de Psycholinguistique, Département d'Etudes cognitives, ENS, EHESS, CNRS, PSL University

[2] Max Planck Institute for Psycholinguistics

[3] University of Chicago

Author Note

Correspondence concerning this article should be addressed to Alejandrina Cristia, 29, rue d'Ulm, 75005 Paris, France. E-mail: alecristia@gmail.com

Abstract

In nonword repetition (NWR) studies, participants are presented auditorily with an item that is phonologically legal but lexically meaningless in their language, and asked to repeat this item as closely as possible. NWR scores are thought to reflect some aspects of phonological development, saliently a perception-production loop supporting flexible production patterns. In this study, we report on NWR results among children learning Yélî Dnye, an isolate spoken on Rossel Island in Papua New Guinea. This study contributes to four lines of research. First, we document that non-word items containing typologically frequent sounds are repeated without changes more often that non-words containing typologically rare sounds, above and beyond any within-language frequency effects. Second, we document rather weak effects of item length, contributing to mounting research suggesting that length effects may be language- or population-specific. Third, we do not find strong individual variation effects in this population. Fourth, we discuss how NWR scores can be compared across studies, languages, and populations, in order to shed light on the factors universally structuring variation in language development. Together, these data provide a unique view of online phonological processing in an understudied language while making preliminary connections between language development and cross-linguistic features.

Keywords: phonology, non-word repetition, development

Word count: 9,000 words

Non-word repetition in Yélî Dnye

Introduction

Children's perception and production of phonetic and phonological units continues developing well beyond the first year of life, even extending into middle childhood (e.g., Hazan & Barrett, 2000). Much of the evidence for later phonological development comes from nonword repetition (NWR) tasks. In a NWR task, participants hear a short word-like form that is phonologically legal but lexically meaningless in the language(s) they are learning. After hearing this non-word, the participant's task is to try to immediately and precisely repeat it. NWR scores are thought to reflect long-term phonological knowledge (to perceive the item precisely despite not having heard it before) as well as online phonological working memory (to encode the item in the interval between hearing it and saying it back) and flexible production patterns (to produce the item precisely despite not having pronounced it before).

NWR has been used to seek answers to a variety of theoretical questions, including what the links are between phonology, working memory, and the lexicon (Bowey, 2001), and how extensively phonological constraints found in the lexicon affect online production (Gallagher, 2014). NWR is also frequently used in applied contexts, notably as a diagnostic tool for language delays and disorders (Estes, Evans, & Else-Quest, 2007). Since non-words can be generated in any language, it has attracted the attention of researchers working in multilingual and linguistically diverse environments, particularly in Europe (COST Action, 2009; Meir, Walters, & Armon-Lotem, 2016).

In the present study, we use NWR to investigate the phonological development of children learning Yélî Dnye, an isolate language spoken in Papua New Guinea (PNG) that has a large and unusually dense phonological inventory. The study was designed to contribute to four aspects of our understanding of phonological development.

First, we included a subset of non-word items with typologically rare and/or challenging sounds to ask whether these rare sounds are disadvantaged in the perception-production loop involved in NWR. Previous work using NWR has preferred relatively universal and early-acquired phonemes (with the possible exception of Gallagher, 2014), likely as a way to separate phoneme pronunciation from broader syllable structure and word-level prosodic effects (Gallon, Harris, & Van der Lely, 2007). Here, we investigate repetition of non-word items containing cross-linguistically common and cross-linguistically rare phonetic targets.

Second, we varied the length (in syllables) of non-words to contribute to growing research looking at the impact of word length on NWR repetition, and what this may reflect about phonological development. Our reading of previous NWR research is that there are variable effects of length between populations. For instance, Jaber-Awida (2018) reports an average of 95% correct repetition for items 1 or 2 syllables long among Arabic-speaking children at about 5 years of age, but 55% for items 3 or 4 syllables long. In contrast, Piazzalunga, Previtali, Pozzoli, Scarponi, and Schindler (2019) observe nearly no change in performance in similarly-aged Italian learners, with a score of 84% for 1-2 syllables versus 85% for 3-4 syllables. It is possible that differences are due to a host of variables, including the modal length of words in the language and/or in child-directed speech in that culture. That said, there could be other causal pathways: Research on adults suggests that illiterate Portuguese speakers repeat monosyllabic non-words just as accurately as literate adults, whereas scores are much lower among illiterate than literate speakers for items 3 or 4 syllables long (de Santos Loureiro et al., 2004). Given that both groups of adults speak the same language (Brazilian Portuguese), then perhaps differences in repetition accuracy reveal differences in how flexible the perception-production loop is. Given the paucity of evidence looking at this question, we do not approach this issue within a hypothesis-testing framework but sought instead to provide one more piece of data on the question, which may be re-used in future meta- or mega-analytic approaches.

Third, there are ongoing discussions as to what the key factors structuring individual variation are. Although the ideal systematic review is missing, a recent paper comes close with a rather extensive review on the literature looking at correlations between NWR scores and a variety of child-level variables (Farabolini, Rinaldi, Caselli, & Cristia, 2021). In a nutshell, most evidence is mixed, suggesting that consistent individual variation effects may be small, and more data is needed to estimate their true size. For this reason, we descriptively report association strength between NWR scores and child age, sex, birth order, and maternal education.

Fourth, these data contribute to the small literature using this task with non-Western, non-urban populations, speaking a language with a moderate to large phonological inventory (see Maddieson, 2005 for a broad classification of languages based on inventory size). Indeed, NWR has seldom been used outside of Europe and North America (with exceptions including Gallagher, 2014; Cristia, Farabolini, Scaff, Havron, & Stieglitz, 2020), outside urban settings (except for in Cristia et al., 2020), nor with languages having large phonological inventories. There are no theoretical reasons to presume that the technique will not generalize to these new conditions. That said, Cristia et al. (2020) recently reported relatively lower NWR scores among the Tsimane', a non-Western rural population, interpreting these findings as consistent with the hypothesis that lower levels of infant-directed speech and/or low prevalence of literacy in a population could lead to population-level differences in NWR scores. In view of these results, it is important to bear in mind that NWR is a task developed in countries where literacy is widespread, and it is considered an excellent predictor of reading, for instance better than rhyme awareness (e.g., Gathercole, Willis, & Baddeley, 1991). Therefore, it may not be a general index of phonological development, but more specifically reflect certain skills. Indeed, Cristia et al. (2020) present the task as being a good index of the development of "short-hand-like" representations specifically, which could thus miss, for example, more holistic representations.

Before going into the details of our study design we first give an overview of Yélî Dnye phonology as well as a brief ethnographic review of the developmental environment on Rossel Island. As discussed above, NWR has been almost exclusively used in urban, industrialized populations, so we provide this additional ethographic information to contextualize the adaptations we have made in running the task and in gathering age and other demographic information, compared to what is typical in urban, industrialized settings.

Yélî Dnye phonology. Yélî Dnye is an isolate language (presumed Papuan) spoken by approximately 7,000 people residing on Rossel Island, an island found at the far end of the Louisiade Archipelago in Milne Bay Province, Papua New Guinea. The Yélî sound system, much like its baroque grammatical system (Levinson, 2020), is unlike any other in the region.

With only four primary places of articulation (bilabial, alveolar, post-alveolar, and velar) and no voicing contrasts, the phonological inventory is remarkably packed with acoustically similar segments. The core oral stop set includes both singleton (/p/, /t/, /ṭ/, and /k/) and doubly-articulated (/tp/, /ṭp/, /kp/) segments, with full nasal equivalents (/m/, /n/, /ṇ/, /ŋ/, /nm/, /ṇm/, /ŋm/), and with a substantial portion of them contrastively pre-nasalized or nasally released (/mp/, /nt/, /ṇṭ/, /ŋk/, /nmtp/, /ṇmṭp/, /ŋmkp/, /ṭṇ/, /kŋ/, /ṭpṇm/, /kpŋm/). A large number of this combinatorial set can further be contrastively labialized, palatalized on release, or both (e.g., /pʲ/, /pʷ/, /pʲʷ/; /tpʲ/; /ṇmḍbʲ/; see Levinson (2020) for details).[1] The consonantal inventory also includes a number of non-nasal continuants (/w/, /j/, /ɣ/, /l/, /βʲ/, /lʲ/, /lβʲ/). Vowels in Yélî Dnye may be oral or nasal, short or long. The 10 oral vowel qualities, which span four levels of vowel height, (/i/, /ɯ/, /u/, /e/, /o/, /ə/, /ɛ/, /ɔ/, /æ/, /ɑ/) can be produced as short and long vowels, with 7 of these able to appear as short and long nasal vowels as well /ĩ/, /ũ/, /ə̃/, /ɛ̃/, /ɔ̃/, /æ̃/, /ɑ̃/).

In total, Yélî Dnye uses 90 distinctive segments (not including an additional 3 rarely used consonants), far outstripping the phonemic inventory size of other documented Papuan languages (Foley, 1986; Levinson, 2020) and includes at least two contrasts that had not previously been documented elsewhere (labial-coronal double-articulations with dental vs. post-alveolar coronal

[1] We use Levinson's (2020) under-dot notation (e.g., /ṭ/) to indicate the post-alveolar place of articulation; these stops are, articulatorily, somewhat variable in place, with at least some tokens produced fully sub-apically. In approximating cross-linguistic segment frequency below we use the corresponding retroflex for each stop segment (e.g., /ʈ/, /ʈp/, /ɳ/).

placement for both oral and nasal stops, Ladefoged & Maddieson, 1996; Maddieson & Levinson, n.d.).

Most words in Yélî Dnye are bisyllabic (~50%), with monosyllabic words (~40%) appearing most commonly after that, and with tri-and-above syllabic words appearing least frequently (~10%; based on >5800 lexemes in the most recent Yélî Dnye dictionary at the time of writing; Levinson, 2020). The vast majority of syllables use a CV format. A small portion of the lexicon features words with a final closed syllable, limited to codas of -/m/, -/p/, or -/j/ (e.g., “ndap” (Spondylus shell) /ṇṭæp/). However, in spontaneous speech an epenthetic /ɯ/ is often inserted after word-final coda, yielding a CV.CV structure instead (e.g., “ndap” pronounced “ndapî” /’ṇṭæ.pɯ/). This process is used frequently with English loan words that have a coda (e.g., “ponî” (phone) /’pɔ.nɯ/). An even smaller portion of the lexicon features words starting with a vowel (e.g., “ala” (here) /æ.’læ/), but these are limited to /æ/-. Finally, the lexicon features a handful of single-vowel grammatical morphemes (see Levinson (2020) for details).

The Yélî community. Most speakers of Yélî Dnye grow up speaking it monolingually until they begin attending school around the age of 7; school instruction is in English. While monolingual Yélî Dnye upbringing is common, multilingual families are not unusual, particularly in the region around the Catholic Mission—the same region in which the current data were collected—where there is a higher incidence of married-in mothers from other islands (Brown & Casillas, n.d.). Children in these multilingual families grow up speaking Yélî Dnye plus English, Tok Pisin, and/or another language(s) from the region.

Most Yélî people are swidden horticulturalists, raising a variety of tubers (e.g., yams, manioc, taro), coconut, banana, and more in their gardens. In addition to garden foods, Yélî children grow up eating fish and foraging for shellfish and nuts, which provide a regular source of extra protein. The typical household in our dataset includes seven individuals and is situated among a collection of 4 or more other households, with structures often arranged around an open grassy area. These household clusters are organized by patrilocal relation, such that they typically comprise a set of brothers, their wives and children, and their mother and father, with neighboring hamlets also typically related through the patriline.

Laying 250 nautical miles off the coast of mainland PNG and surrounded by a barrier reef, transport to and from the island is both infrequent and irregular. There is no electricity system (we use solar panels) or motor vehicles on the island, save dinghys with outboard motors that are primarily reserved for medical transport when there is diesel available. There is a medical radio connection at the health clinic, located at the Catholic Mission, via which news and messages are often passed to and from mainland PNG. There is also a cellular tower on the island, though it is often inoperative, usually for multiple months or years at a time. As such, access to news and outside connections, including international phone calls and digital exchanges that require significant data transfer, is typically unavailable. Our data collection is therefore limited to the duration of the researchers' on-island visits.

Despite restricted outside contact, formal education is a priority for many Yélî families. A recent study surveying more than 40 parents of young children in this region of the island found that nearly all parents completed 6 or more years of education, with 65% of mothers and 49% of fathers completing the full range of locally available education (~9 years), with around half of those parents having pursued an additional 2+ years of education on other islands in the region or on mainland PNG (Casillas, Brown, & Levinson, 2020). While there are a handful of local schools around the island, these may be well out of walking distance for many children (i.e., more than 1 hour on foot or by canoe each day). It is very common for households that are situated close to a school to billet their school-aged relatives during the weekdays (sometimes the weekends too, if the child's home is very far) for long segements of each school year. Combined with practices of collective child guardianship within each close-knit hamlet, the practical consequence of this billeting for researchers is that adult consent can often come from a combination of aunts, uncles, adult cousins, and grandparents standing in for the child's biological parents. On top of this, child assent is culturally pertinent, as child independence is encouraged and respected from toddlerhood onward (Brown & Casillas, n.d.).[2] Finally, as mentioned, the language of schooling in this region is English, and so the majority of teens and adults on the island speak some English—many speak it very fluently.

While work on Yélî language development is growing (e.g., Brown, 2011, 2014; Brown & Casillas, n.d.; Casillas et al., 2020; Liszkowski, Brown, Callaghan, Takada, & de Vos, 2012), our knowledge of children's linguistic development is quite limited, and research into their phonological development in particular has only just begun (e.g., Peute, Fikkert, & Casillas, n.d.). What we do know about the early language environment is that, while direct speech to children under 3;0 is relatively infrequent throughout the day (Casillas et al., 2020), shared caregiving practices and a near-universal fondness for social interaction with young children ensure that speech directed to children comes from all types of speakers: women, men, and other children (Bunce et al., n.d.; Casillas et al., 2020). While speech from women predominates in young children's speech environments on Rossel, as it does elsewhere (Bergelson et al., 2019; Bunce et al., n.d.), there is a significant and marked increase in child-directed speech from other children as infants get older (Bunce et al., n.d.; Casillas et al., 2020). This increase in child-directed speech from other children is attributed to the fact that, starting around age 2, children often spend large swaths of the day playing, swimming, and foraging for fruit, nuts, and shellfish in large (~10 members) independent and mixed-age child play groups (Brown & Casillas, n.d.; Casillas et al., 2020).

NWR design and analysis adaptations. In a basic NWR task, the participant listens to a production of a word-like form, such as /bilik/, and then repeats back what they heard without changing any phonological feature that is contrastive in the language. For instance, in English, a response of [bilig] or [pilik] would be scored as incorrect; a response [bi:lik], where the vowel is lengthened without change of quality would be scored as correct, because English does not have a general feature of contrastive vowel length. There is some variation in how past NWR studies have

[2]For these reasons, we were unable to record precise details for a few children, including child date of birth, parent age in years, and parent education.

designed the presentation procedure and structure of items. For example, while items are often presented orally by the experimenter (Torrington Eaton, Newman, Ratner, & Rowe, 2015), an increasing number of studies have turned instead to playing back pre-recorded stimuli in order to increase control in stimulus presentation (Brandeker & Thordardottir, 2015). Additionally, while some studies have used 10-15 non-words, others have employed up to 46 unique items (Piazzalunga et al., 2019). Authors also often modulate structural complexity, typically measured in terms of item length (measured in number of syllables) and/or syllable structure (open as opposed to closed syllables, Gallon et al., 2007).

Previous work typically steers clear of articulatorily and/or acoustically challenging sounds, but we included some in our experiment to more adequately represent Yélî Dnye's phonology. We ultimately used a relatively large number of items that would enable us to explore both variation in structural complexity and in more vs. less challenging sounds. However, aware that this large item inventory might render the task longer and more tiresome, we split items across children (see below). Naturally, designing the task in this way may make the study of individual variation within the population more difficult because different children are exposed to different items. However, a review of previous work on individual variation suggested to us that effects of individual differences in NWR are relatively small, and would not be detectable with the sample size that we could collect during our short visit. That said, we contribute to the literature by also reporting descriptive analyses of individual variation that could potentially be integrated in meta-analytic efforts.

Based on previous work, we looked at potential changes with age (Farmani et al., 2018; Kalnak, Peyrard-Janvid, Forssberg, & Sahlén, 2014; Vance, Stackhouse, & Wells, 2005). Previous work typically finds no significant differences as a function of maternal education (e.g., Farmani et al., 2018; Kalnak et al., 2014; Meir & Armon-Lotem, 2017) or child gender (Chiat & Roy, 2007). Although past research has not often investigated potential effects of birth order on NWR, there is a sizable literature on these effects in other language tasks (Havron et al., 2019), and therefore we report on these too.

Research questions. After some preliminary analyses to set the stage, we address the following questions:

- Does the cross-linguistic frequency of sounds in the stimuli predict NWR scores? Are rarer sounds more often substituted by commoner sounds?
- How does score change as a function of item length in number of syllables?
- Is individual variation in NWR scores attributable to child age, sex, birth order, and/or maternal education?

We had considered boosting the interpretational value of this evidence by announcing our analysis plans prior to conducting them. However, we realized that even pre-registering an analysis would be equivocal because we would not have enough power to look at all relationships of interest, in many cases possibly not enough to detect any of the known effects, given their variability across studies. To illustrate this issue, we portray in Figure 1 studies in which children's NWR scores were gathered between 4 and 12 years of age, and reported separately for items that are relatively short (1-2 syllables) versus longer items (3-4 syllables). Notice that the effect of stimulus length is minuscule among Italian children (Piazzalunga et al., 2019), but considerable among Tsimane' children (Cristia et al., 2020), where a drop of 40 percentage points is observed at all ages. A similar large difference in NWR scores for short versus long items was observed among Arabic children (Jaber-Awida, 2018). Even the effect of age is unstable in this sample of studies. Whereas it is quite clear that children's NWR scores increase in the Italian data, age effects are less stable among Tsimane' children. Therefore, all analyses in the present study are descriptive and should be considered exploratory.

Methods

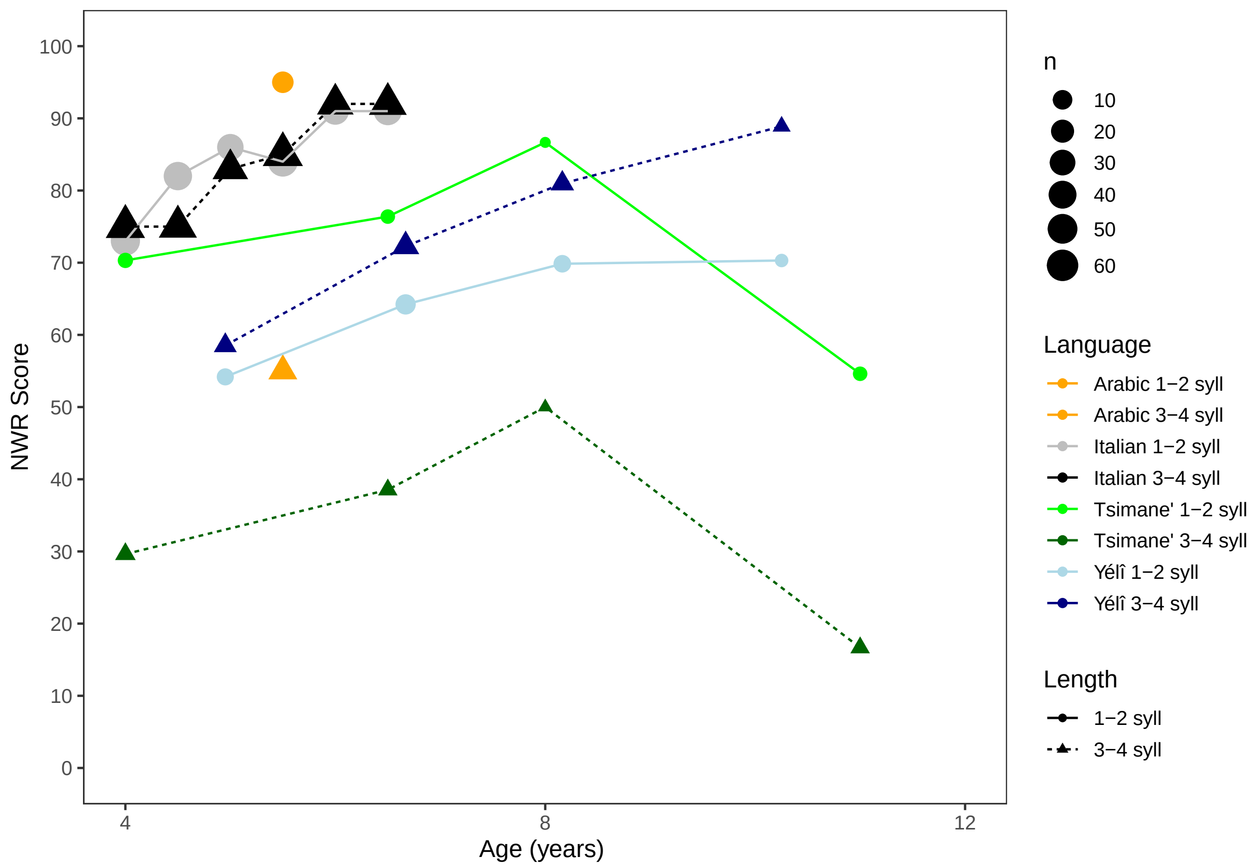

Figure 1. NWR scores as a function of age (in years) and item length for comparable studies. Arabic data from Jaber et al. (2018); Italian data from Piazzalunga et al. (2019); Tsimane' from Cristia et al. (2020); Yélî Dnye data from the present study.

Stimuli. Many NWR studies are based on a fixed list of 12-16 items that vary in length between 1 and 4 syllables, often additionally varying syllable complexity and/or cluster presence and complexity, and always meeting the condition that they do not mean anything in the target language (e.g., Balladares, Marshall, & Griffiths, 2016; Wilsenach, 2013). We kept the same variation in item length and requirement for not being meaningful in the language, but we did not vary syllable complexity or clusters because these are vanishingly rare in Yélî Dnye. We also increased the number of items an individual child would be tested on, such that a child would get up to 23 items to repeat (other work has also used up to 24-30 items: Jaber-Awida, 2018; Kalnak et al., 2014), with the entire test inventory of 44 items distributed across children.

A first list of candidate items was generated during a trip to the island in 2018 by selecting simple consonants (/p/, /t/, /ṭ/, /k/, /m/, /n/, /w/, /y/) and vowels (/i/, /o/, /u/, /a/, /e/) and combining them into consonant-vowel syllables, then sampling the space of 1- to 4-syllable sequences. These candidates were automatically removed from consideration if they appeared in Levinson's (2015) dictionary. The second author presented them orally to three local research assistants, all native speakers of Yélî Dnye, who repeated each form as they would in an NWR task and additinally let the experimenter know if the item was in fact a word or phrase in Yélî Dnye. Any item reported to have a meaning or a strong association with another word form or meaning was excluded.

A second list of candidate items was generated in 2019 by selecting complex consonants and systematically crossing them with all the vowels in the Yélî Dnye inventory to produce consonant-vowel monosyllabic forms. As before, items were automatically excluded if they appeared in the dictionary. Additionally, since perceiving vowel length in isolated monosyllables is challenging, any item that had a short/long lexical neighbor was excluded. Because there is still much to discover about the phonology and phonetics of Yélî Dnye (Levinson, 2020), it was also possible that we initially generated items with illegal, but currently undocumented constraints. Therefore, we made sure that the precise consonant-vowel sequence occurred in some real word in the dictionary (i.e., that there was a longer word included the monosyllable as a subsequence). These candidates were then presented to one informant, for a final check that they did not mean anything. Together with the 2018 selection, they were recorded, based on their orthographic forms, using a Shure SM10A XLR dynamic headband microphone and an Olympus WS-832 stereo audio recorder (using an XLR to mini-jack adapter) by the same informant, monitored by the second author for clear production of the phonological target. The complete recorded list was finally presented to two more informants, who were able to repeat all the items and who confirmed there were no real words present. Despite these checks, one monosyllable was ultimately frequently identified as a real word in the resulting data (intended "yî" /yɯ/; identified as "yi" /yi/, tree). This item is removed from the analyses below.

The final list includes: three practice items; 20 monoysllables containing sounds that are less frequent in the world's languages than singleton plosives; 8 bisyllables; 12 trisyllables; and 4 quadrisyllables (see Table 2).

Table 1

NWR stimuli in orthographic and phonological representations.

Practice		Monosyll		Bisyll		Trisyll		Tetrasyll	
Orth.	Phon.	Orth.	Phon.	Orth.	Phon.	Orth.	Phon.	Orth.	Phon.
nopimade	nɔpimæṭɛ	dp:a	ṭp̃æ	kamo	kæmɔ	dimope	ṭimɔpɛ	dipońate	ṭipɔnætɛ
poni	pɔni	dpa	ṭpæ	kańi	kæni	diyeto	ṭijɛtɔ	ńomiwake	nɔmiwækɛ
wî	wɯ	dpâ	ṭpɑ	kipo	kipɔ	meyadi	mɛjæṭi	todiwuma	tɔṭiwumæ
		dpê	ṭpə	ńoki	nɔki	mituye	mitujɛ	wadikeńo	wæṭikɛnɔ
		dpéé	ṭpeː	ńomi	nɔmi	ńademo	næṭɛmɔ		
		dpi	ṭpi	piwa	piwæ	ńayeki	næjɛki		
		dpu	ṭpu	towi	tɔwi	ńuyedi	nujɛṭi		
		gh:ââ	ɣ̃ɑː	tupa	tupæ	pedumi	pɛṭumi		
		ghuu	ɣuː			tiwuńe	tiwunɛ		
		kp:ââ	kp̃ɑː			tumowe	tumɔwɛ		
		kpu	kpu			widońe	wiṭɔnɛ		
		lv:ê	lβ̃ʲə			wumipo	wumipɔ		
		lva	lβʲæ						
		lvi	lβʲi						
		t:êê	t̃əː						
		tp:a	tp̃æ						
		tpâ	tpɑ						
		tpê	tpə						

A Praat script (Boersma & Weenink, 2020) was written to randomize this list 20 times, and split it into two sublists, to generate 40 different elicitation sets. The 40 elicitation sets are available online from osf.io/5qspb. The split had the following constraints:

- The same three items were selected as practice items and used in all 40 elicitation sets.
- Splits were done within each length group from the 2018 items (i.e., separately for 2-, 3-, and 4-syllable items); and among onset groups for the difficult monosyllables generated in 2019 (i.e., all the monosyllables starting with /tp/ were split into 2 sublists). Since some of these groups had an odd number of items, one of the sublists was slightly longer than the other (20 vs. 23).
- Once the sublist split had been done, items were randomized such that all children heard first the 3 practice items in a fixed order (1-, 2-, and 4-syllable), a randomized version of their sublist selection of difficult onset items, and randomized versions of their 2-syllable, then 3-syllable, and finally 4-syllable items.

Procedure. In adapting the typical NWR procedure for this context, we balanced three desiderata: That children would not be unduly exposed to the items before they themselves had to repeat them (i.e., from other children who had participated); that children would feel comfortable doing this task with us; and that community members would feel comfortable having their children do this task with us.

We ran children in batches, testing within a handful of hamlet clusters spread across the northeastern region of the island. Because space availability was limited in different ways from hamlet to hamlet, the places where elicitation happened varied across testing sites. We tested in four different sites, only making a single visit to each, conducting back-to-back testing of all eligible children present at the time of our visit in order to prevent the items from "spreading" between children through hearsay. In the first hamlet, we tested children in five different places, with some children being tested inside a house and others tested on the veranda. More information is available from the online supplementary materials. Whenever children living in the same household were tested, we tried to test children in age order, from oldest to youngest, to minimize intimidation for younger household members, and always using different elicitation sets.

We fitted the child with a headset microphone (Shure SM10A or WH20 XLR with a dynamic microphone on a headband, most children using the former) that fed into the left channel of a Tascam DR40x digital audio recorder. The headsets were designed for adult use and could not be comfortably seated on many children's heads without a more involved adjustment period. To minimize adjustment time, which was uncomfortable for some children given the proximity of the experimenter and equipment, we placed the headband on children's shoulders in these cases, carefully adjusting the microphone's placement so that it was still close to the child's mouth. A research assistant who spoke Yélî Dnye natively sat next to the child throughout the task to provide instructions and, if needed, encouragement. The research assistant coached the child throughout the task to make sure that they understood what they were expected to do. An experimenter (the first author) delivered the pre-recorded stimuli to the research assistant and the child over headphones.

The first phase of the experiment involved making sure the child understood the task. We explained the task and then orally presented the first practice item. At this point, many children did not say anything in response, which triggered the following procedure: First, the assistant insisted the child make a response. If the child still did not say anything, the assistant said a real word and then asked the child to repeat it, then another and another. If the child could repeat real words correctly, we provided the first training item over headphones again for children to repeat. Most children successfully started repeating the items at this point, but a few needed further help. In this case, the assistant modeled the behavior (i.e., the child and assistant would hear the item again, and the assistant would repeat it; then we would play the item again and ask the child to repeat it). A small minority of children still failed to repeat the item at this point. If so, we tried again with the second training item, at which point some children demonstrated task understanding and could continue. A fraction of the remaining children, however, failed to repeat this second training item, as well as the third one, in which case we stopped testing altogether (see Participants section for exclusions).

The second phase of the experiment involved running the child through the list of test items randomly assigned to them. This was done in the same manner as the practice items: the stimulus was played over the headphones, and then the child repeated it aloud. NWR studies vary in whether children are allowed to hear and/or repeat the item more than one time. We had a fixed procedure for the test items (i.e., the non-practice items) in which the child was alowed to make a further attempts if their first attempt was judged erroneous in some way by the assistant. The procedure worked as follows: When the child made an attempt, the assistant indicated to the experimenter whether the child's production was correct or not. If correct, the experimenter would whisper this note of correct repetition into a separate headset that fed into the right channel of the same Tascam recorder and the group moved on to the next item. If not, the child was allowed to try again, with up to five attempts allowed before moving on to the next item. Children were not asked to make repetitions if they did not produce a first attempt. In total, test sessions took approximately XX minutes on average, with XX of those minutes attributed to practice and XX to the actual test list.

Coding. The first author then annotated the onset and offset of all children's productions from the audio recording using Praat audio annotation software (Boersma & Weenink, 2020), then wrote and ran a script to extract these tokens, pairing them with their original auditory target stimulus, and writing these audio pairs out to short .wav files. The assistant then listened through all these productions, one target item at a time, with productions for each target item presented in a random order across children and repetitions. The assistant indicated in a notebook whether each production was a correct or incorrect repetition and orthographically transcribed the production, noting when the child uttered a recognizable word or phrase and adding the translation equivalent of that word/phrase into English. The assistant also provided some general examples of the types of errors children made without making specific reference to Yélî sounds or the items in the elicitation sets.

Analyses. In addition to a binary word-level exact repetition score we scored items in terms of the number of phonemes that could be aligned across the target and attempt, divided by the number of phonemes of whichever item was longer (the target or the attempt; see also Cristia et al., 2020). Although previous work does not use distance metrics, we additionally report these. For instance, recall our example of an English target being /bilik/ with an imagined response [bilig]. We would score this response as follows: at the whole item level this production would receive a score of zero (because the repetition is not exact); at the phoneme level this production would receive a score of 80% (4 out of 5 phonemes repeated exactly); and the phone-based Levenshtein distance for this production is 20% (because 20% of phonemes were substituted or deleted).

Additionally, we estimated the typological frequency of all phonological segments used in the target items using the PHOIBLE cross-linguistic phonological inventory database (Moran & McCloy, 2019). For each phone in our task, we extracted the number and percentage of languages noted to have that phone in its inventory. While PHOIBLE is an unprecedentedly comprehensive database, with phonological inventory data for over 2000 languages at time of writing, it is of course still far from complete, so estimated frequencies should be taken with a pinch of salt. In both PHOIBLE and UPSID—another large phonemic inventory database—note that nearly half of the segment types are only attested in one language (Steven Moran, personal communication; see also web.phonetik.uni-frankfurt.de/upsid_info.html). It is therefore unsurprising that we did not find three of our segments (/lβʲ/, /ṭp/, and /tp/) in the database. Following the general pattern that many phones only appear once cross-linguistically, even in large databases, we treat these unattested phones as having a frequency of 1, with a (rounded) percentage of 0%.

Finally, when describing children's patterns of errors, we take all repetitions of a given target into account. We describe the proportion of items where the change resulted in a real word (semantic errors); as well as deletions, additions, and substitutions.

Participants. This study was approved as part of a larger research effort by the second author. The line of research was evaluated by the Radboud University Faculty of Social Sciences

Ethics Committee (Ethiek Commissie van de faculteit der Sociale Wetenschappen; ECSW) in Nijmegen, The Netherlands (original request: ECSW2017-3001-474 Manko-Rowland; amendment: ECSW-2018-041). Participation was voluntary; children were invited to participate following indication of approval from an adult caregiver. Regardless of whether they completed the task, children were given a small snack as compensation. Children who showed initial interest but then decided not to participate were also given the snack.

We tested a total of 55 children from 38 families spread across four hamlets. We excluded test sessions from analysis for the following reasons: refused participation or failure to repeat items presented over headphones even after coaching (N = 8), spoke too softly to allow offline coding (N = 5), or were 13 years old or older (N = 2; we tested these teenagers to put younger children at ease). The remaining 40 children (14 girls) were aged 6.98 years (range 3.92-11.03 years). Among these children were 34 only exposed to Yélî Dnye at home and 6 children exposed to Yélî Dnye plus one or more other languages at home. Maternal years of education averaged 8.22 years (range 6-12 years).[3] In terms of birth order, 6 were first borns, 5 second, 2 third, 7 forth, 5 fifth, and 1 sixth.

Results

Preliminary analyses. We first checked whether whole-item NWR scores varied between first and subsequent presentations of an item by averaging word-level scores at the participant level separately for first attempts and subsequent repetitions. We excluded 1 child who did not have data for one of these two types. As shown in Figure 2, participants' mean word-level scores became more heterogeneous in subsequent repetitions. Surprisingly, whole-item NWR scores for

[3]Years of education is difficult to precisely pinpoint in this context. Schooling beyond grade 8 takes place off-island and enrollment is tied to an ability to pay that year's school fees, seasonal interruptions to school opening times, familial obligations, and more. We asked parents to report their highest completed level of education. We then record the number of years entailed by having completed that level under ideal conditions. This method results in many esimates around typical "graduation" times (e.g., graduating from primary education before starting secondary education).

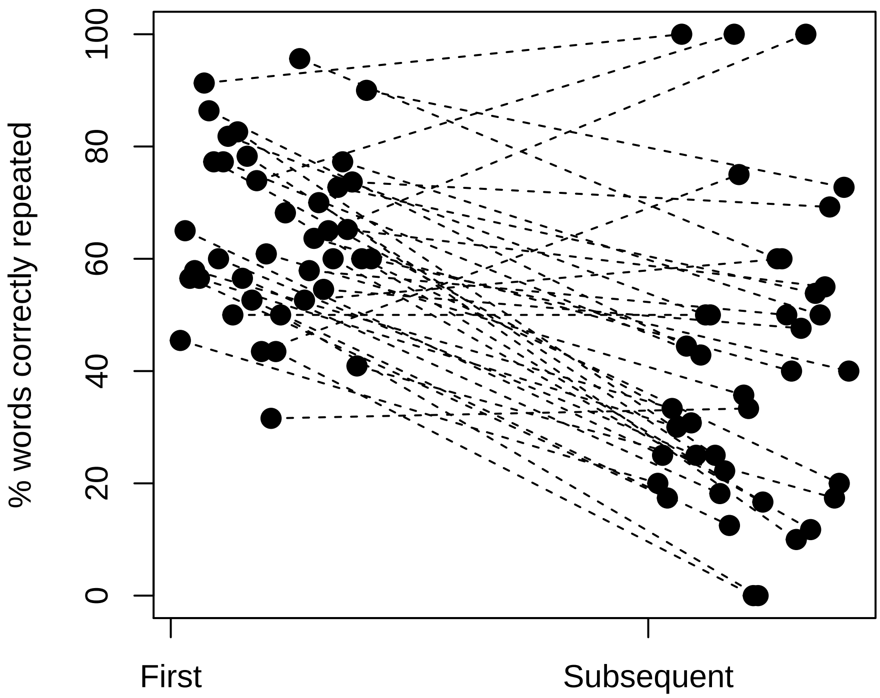

Figure 2. Whole-item NWR scores for individual participants averaging separately their first attempts and all other attempts.

subsequent repetitions (M = 40, SD = 26) were on average lower than first ones (M = 64, SD = 15), $t(38) = 6.28$, $p < 0.001$; Cohen's d = 1.14). Given uncertainty in whether previous work used first or all repetitions, and given that score here declined and became more heterogeneous in subsequent repetitions, we focus the remainder of our analyses only on first repetitions.

Taking into account only the first attempts, we averaged attempts by each of the 40 children who had data for first attempts; their ages ranged from 3 to 10 (M = 6.50, SD = 1.50). The overall NWR score was M = 64% (SD = 15%), Cohen's d = 4.36. Scores based on phonemes are even higher M = 78% (SD = 9%), Cohen's d = 8.55. The phoneme-based normalized Levenshtein distance was M = 22% (SD = 9%), meaning that about a fifth of phonemes were were substituted, inserted, or deleted (the normalized Levenshtein distance is the complement of phoneme-based scores).

NWR as a function of cross-linguistic phone frequency. We then analyzed variation in whole-item NWR scores as a function of the average frequency with which sounds composing individual target words are found in languages over the world. To look at this, we fit a mixed logistic regression in which the outcome variable was whether the non-word was correctly repeated or not. The fixed effect of interest was the average cross-linguistic phone frequency; we also included child age as a control fixed effect, and child ID & target ID as random variables.

We could include 850 observations, from 40 children producing in any given trial one of 41 potential target words. The analysis revealed a main effect of age ($ß = 0.33$, $SE\ ß = 0.13$, $p = 0.01$); and a significant estimate for the scaled average cross-linguistic frequency of phones in the target words ($ß = 0.82$, $SE\ ß = 0.18$, $p < 0.001$): Target words with phones found more frequently across languages had a higher proportion of words that were correctly repeated, as shown in Figure 3. Averaging across items and participants, the Pearson correlation between scaled average cross-linguistic phone frequency and whole-item NWR scores was $r(32) = 0.61$.

We next checked whether the association between whole-item NWR scores and cross-linguistic phone frequence could actually be due to frequency of the sounds within the language: One can suppose that sounds that occur more frequently across languages are also more frequent within a language, and therefore may be easier for children to represent and repeat. We estimated frequency of the phones present in the stimuli in a corpus of child-centered recordings (Casillas et al., 2020) by counting the number of word types in which they occurred, and applied the natural logarithm.[4] Here, unattested sounds were not considered (i.e., they were declared NA so that they do not count for analyses). Phone frequencies estimated from this corpus were correlated with cross-linguistic phone frequencies ($r(27) = 0.50$, $p < 0.01$). Moreover, averaging across items and participants, the Pearson correlation between scaled average corpus phone

[4]We also carried out this analysis using token phone frequency, but this measure was not correlated with whole-item NWR scores, and therefore the fact that it did not explain away the predictive value of cross-linguistic phone frequency was less informative than the relationship discussed in the main text, with type frequencies.

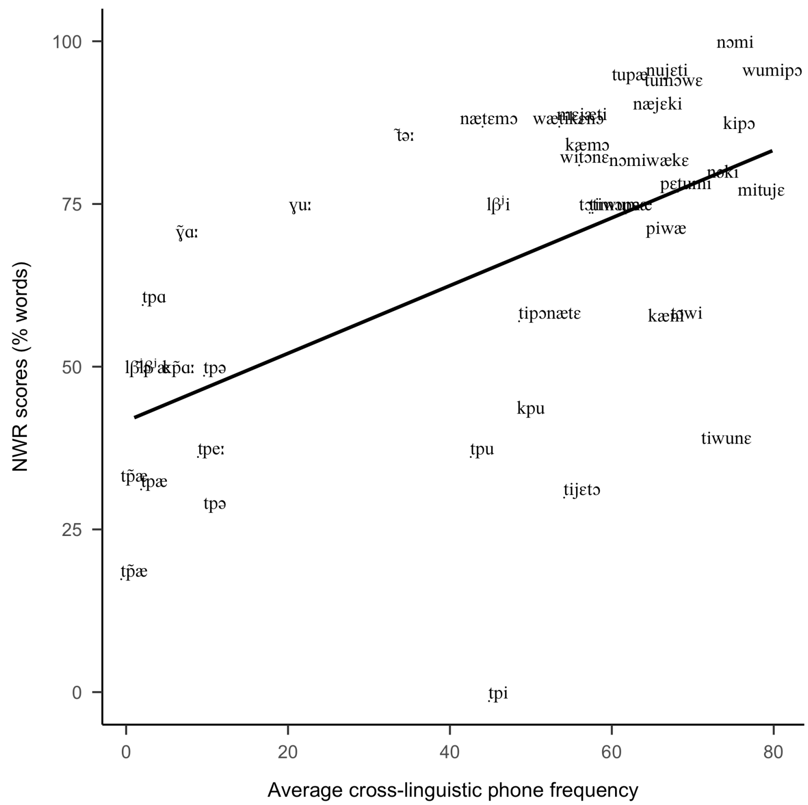

Figure 3. NWR scores for individual target words as a function of the average frequency with which each phone is found across languages.

frequency and whole-item NWR scores was $r = 0.46$. Therefore, we fit another mixed logistic regression, this time declaring as fixed effects both scaled cross-linguistic and corpus frequencies (averaged across all attested phones within each stimulus item), in addition to age. As before, the model contained random slopes for both child ID and target. In this model, both cross-linguistic frequency (ß $= 0.82$, SE ß $= 0.27$, $p < 0.01$) and age (ß $= 0.33$, SE ß $= 0.13$, $p = 0.01$) were significant predictors of whole-item NWR scores, but corpus frequency (ß $= 0.01$, SE ß $= 0.25$, $p = 0.98$) was not.

Patterns in NWR mispronunciations. Next, we investigated patterns of deletion and substitution. Deletions were relatively rare: there were only 7 instances of deleted vowels (~0.23% of all vowel targets), and 3 instances of deleted consonants (~0.12% of all consonant targets). Substitutions were somewhat more frequent, but still relatively rare. Nasal vowels were produced as oral vowels and vice versa (34 oral targets as nasal (~0.99% of targets) and 30 nasal targets produced as oral (~2.37% of targets)). Both oral and nasal vowels were also occasionally subsituted by vowels of a different quality that preserved the target oral/nature airflow (110 oral vowels (~1.86% of targets and 13 nasal vowels (~2.06% of targets). Among consonants, complex consonants were occasionally substituted by simpler ones (specifically [ṭp], [tp], [kp], [km], [kṇ], [mp], [ɣ], and [lβʲ] as [m], [n], [l], [w], [j], [w], [ṭ], [g], [p], [t], [k], [f], [h], and [tʃ]), 115 times, in approximately 1.96% of all complex consonant targets. Interestingly, simple consonants were sometimes also substituted by complex ones; this happened 2 times, in approximately 1.96% of all simple consonant targets. Finally, simple consonants were mispronounced as other simple consonants (68 instances, ~0.38% of all simple consonant targets) and complex consonants as other complex consonants (33 instances, ~1.02% of all complex consonant targets). Note that the substitutions included phones that are not native to Yélî Dnye but do occur in English (e.g., [tʃ]); as these data come from careful transcriptions by a native Yélî Dnye speaker who is very fluent in English, we take these segmental substitutions faithfully as an indication that several of our participants have mastered production of some English phones, possibly produced within whole English word forms (see below).

Finally, we looked into the frequency with which mispronunciations resulted in real words. In fact, two thirds of incorrect repetitions were recognizable as real words or phrases in Yélî Dnye or English: 63%. This type of analysis is seldom reported—we could only find one comparison point: Castro-Caldas, Petersson, Reis, Stone-Elander, and Ingvar (1998) found that illiterate European Portuguese adults' NWR mispronunciations resulted in real words in 11.16% of cases, whereas literate participants did so in only 1.71% of cases. The percentage we observe here is much higher than reported in Castro and colleagues' study, but we do not know whether age, language, or even test structure explains this difference.

NWR scores as a function of item length. Next, we inspected whether NWR scores varied as a function of word length (Table 2). Participants scored much lower on monosyllables than on

Table 2

NWR measured in whole-word scores, phoneme-based scores, and normalized Levenshtein Distance, separately for the four stimuli lengths.

Word	Phoneme	NLD
47 (22)	59 (17)	41 (17)
79 (22)	92 (9)	8 (9)
78 (19)	93 (7)	7 (7)
74 (32)	91 (12)	9 (12)

non-words of other lengths. This is likely due to the fact that the majority of monosyllables were designed to include sounds that are rare in the world's languages, which may indicate that they are hard to produce or perceive. Setting monosyllables aside, we observe the typical pattern of lower scores for longer items, although this is particularly salient for the whole-item scoring. While whole-item scoring is the most commonly reported NWR outcome, it is also the least forgiving. The pattern is less marked when other two scores are used, which are less sensitive to errors. Averaging across participants and items, the Pearson correlation between length (2-4 syllables) and whole-item NWR scores was $r(1) = -0.91$. In a generalized binomial mixed model, we included 479 observations, from 40 children producing, in any given trial, one of 24 potential target words. The analysis revealed a positive effect of age ($ß = 0.56$, SE $ß = 0.14$, $p < 0.001$) and a negative but non-significant estimate for target length in number of syllables ($ß = -0.15$, SE $ß = 0.33$, $p = 0.65$).

Factor structuring individual variation. Our final exploratory analysis assessed whether variation in scores was structured by factors that vary across individuals. As shown in Figure 4,

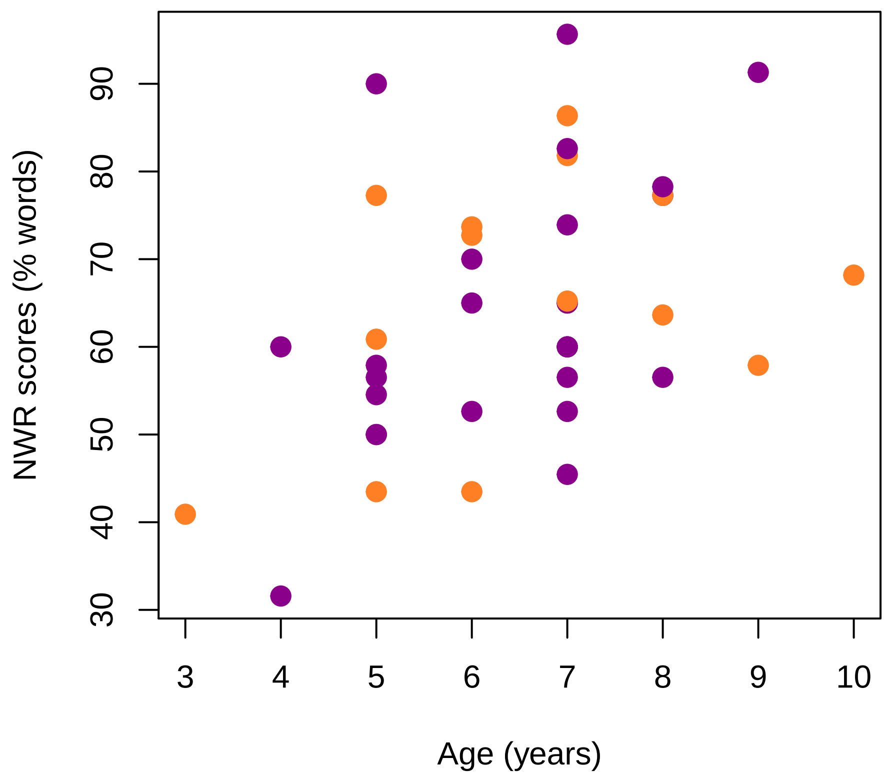

Figure 4. NWR whole-item scores for individual participants as a function of age and sex (purple = boys, orange = girls).

there was a greater deal of variance across the tested age range, with significantly higher NWR scores for older children (Spearman's rank correlation, given inequality of variance, rho (6,014.70) = 0.44, $p < 0.01$). In contrast, there was no clear association between NWR scores and sex (t (-0.29) = 27.56, p = 0.77), birth order (data missing for 15 children, rho = (3,441.90) = -0.18, p = 0.39), or maternal education (data missing for 0 children, rho (9,594.37) = 0.10, p = 0.54).

Discussion

We used non-word repetition to investigate phonological development in a language with a large phonological inventory (including some typologically rare segments). While the study, in itself, advances NWR research by demonstrating the successful application of this method in a rural, non-industrialized population (see also Cristia et al., 2020), the results also inform the larger body of NWR work regarding the influence of stimulus length, segmental frequency, age, and developmental context, where previous findings have often been mixed.

In our sample of 40 children between ages 3 and 10, the average item-level score was 64%, falling well within the expected range of NWR applications in previous studies on a variety of linguistic populations (see Figure 1). Segment-level analyses revealed that repetitions recovered an average of 78% of the segments in the original stimulus, indicating that inexact repetitions often constituted mostly accurate reroductions of the target. Complementary data from segment-based normalized Levenshtein distance reflect this same pattern, with an average of 22% of the stimulus segments getting substituted, inserted, or deleted upon repetition. Paired with our thorough training protocol, we take these NWR scores as indicating that (a) our adaptations to NWR for this context were successful, even given a number of non-standard changes to the training phase and to the design of the stimuli and (b) Yélî children show comparable performance to others tested on a similar task, despite the many linguistic, cultural, and socioeconomic differences between this and previously tested populations. Given successful (and comparable) execution of this task, we can dive deeper into effects of stimulus length, segmental frequency, age, and developmental context.

Item complexity. We investigated the effect of item complexity on NWR scores by varying both the number of syllables in the item and its average segmental frequency. Based on previous work, we had predicted that children would have higher NWR scores for shorter items. That said, previous work has shown both very small (Piazzalunga et al., 2019) and very large (Cristia et al., 2020; Jaber-Awida, 2018) effects of stimulus length and, further, the Yélî Dnye dictionary suggests that mono- and bi-syllabic words are nearly equally frequent in the current language, with trisyllabic and longer words making up a non-trivial 10% of the remaining words. Compare this to, for example, English, which is substantially more skewed toward monosyllabic word forms M2A: Alex I'm going off your note here ("Prediction for Yélî made before seeing the data: The length distribution in Yélî words is more balanced than that in English, and thus the score decline for poly- versus mono-syllables may be less pronounced than that for English.""). I don't have a reference for this, can you please finish the thought or nix this bit?. Setting aside our monosyllabic stimuli, which all contained typologically infrequent segments, we can examine effects of item length among the remaining stimuli, which range between 2 and 4 syllables long. While indeed NWR scores were overall lower for longer items (e.g., see Figure 1), the effect of item length was not significant in a statistical model that additionally accounted for age and random effects of item and participant. In light of mixed prior results of item length, we propose two possible (and non-mutually exclusive) explanations for this minimal impact of item length. First, further extensions of this type of analysis in more populations may reveal that, in general (and cross-linguistically), item length effects are variable between languages, potentially reflecting the distribution of word lengths in the ambient language and other (morpho-)phonological tendencies in the lexicon. Second, above and beyond these language-specific effects, the general impact of item length on NWR score may be relatively small, as shown in Piazzalunga et al.'s (2019) study on Italian and as borne out in the current dataset once controlling for other factors.

Our monosyllabic items included typologically rare segments so that we could test whether lower average segmental frequency is associated with lower NWR scores. Typologically common sounds are associated with higher performance on a handful of other tasks (REFS – M2A: Alex, I added this based on your note, where it sounded like you had some particular studies in mind?) though to our knowledge this has not yet been tested with non-word repetition. Regarding Yélî Dnye in particular, the phonemic inventory is both large and acoustically packed, in addition to containing several typologically infrequent (or unique) contrasts. We therefore expected to see that, while NWR scores would be lower for stimuli with lower average frequency, this effect would be relatively weak because the ambient language puts pressure on Yélî children to distinguish (perceptually and articulatorily) fine-grained phonetic differences in order to successfully communicate with others. Indeed we found a robust effect of average segmental frequency on NWR performance: Even accounting for age and random effects of item and participant, we see that target words with more frequent segments were repeated correctly more often. This effect is large, with a magnitude more than twice the size of the effect of participant age. This significant effect remains even once also accounting for the frequencies of these segments in Yélî Dnye child-directed speech, which are correlated with their typological frequencies. In sum, typological frequency effects, which have been found in other measurements of phonological processing. appear to strongly affect NWR performance, and do not appear mitigated by language-specific pressure to make finer-grained differences earlier in development.

With respect to the types of errors in repetition made, we did not see clear patterns to further guide our discussion: base rates of deletion and substitution were fairly low and the relative distribution of errors over, e.g., nasal vs. oral vowels and simple vs. complex consonants, revealed no remarkable bias in error types. That said, the lack of a difference could be due to relative imbalance across our stimuli in the use of these phonemic features (e.g., we included many more more oral than nasal targets) and future work should investigate such sources of error bias more systemtically. Some portion of the errors were introduced when the participant produced a real word (in Yélî Dnye or English) in response to the stimulus. Real-word repetitions here made up two thirds of errorful repetitions—this is quite high compared to past work (e.g., ???), but it is unclear what caused this pattern in the current study: Castro and colleagues' (???) study focused on adults rather than children, the task was administered by a team including a foreign, English-speaking researcher, and the particularities of the Yélî Dnye phonological inventory result in many true-word phonetic neighbors. Follow-up work exploring this type of error in children from other populations in addition to further work on Yélî children will clarify this effect.

Individual differences. A review of previous work (see Introduction) suggested that our anticipated sample size would not be sufficient to detect most individual differences using NWR. We give a brief overview of individual difference patterns of four types in the present data—age, sex, birth order, and maternal education—hoping that these findings can contribute to future meta-analytic efforts aggregating over smaller studies such as ours.

Following prior work, we expected that NWR scores would increase with participant age (Farmani et al., 2018; Kalnak et al., 2014; Vance et al., 2005). Indeed, age was significantly correlated with NWR score and also showed up as a significant predictor of NWR score when included as a control factor in the analyses of both item length and average segmental frequency. In brief, our results underscore the idea that phoonlogical development continues well past the first few years of life, extending into middle childhood and perhaps later (Hazan & Barrett, 2000).

In contrast, previous work shows little evidence for effects of maternal education (e.g., Farmani et al., 2018; Kalnak et al., 2014; Meir & Armon-Lotem, 2017) or participant gender (Chiat & Roy, 2007) on NWR scores. In addition to this prior work, education on Rossel Island, while generally highly valued, is not at all essential to ensuring one's success in society and may not be a reliable index of local socioeconomic variation. There is also limited variation in maternal education across the families in the region of the island where we sampled. We therefore expected little evidence for impact of either participant gender or maternal education in the present study. On the other hand, these predictors have established effects on other language development measures (REFS: M2A: Alex go ahead and pick your faves here). So to the extent that NWR scores share causal links to gender-based differences ein development and maternal linguistic input with these other language ouctome measures, we might then expect these factors to appear in NWR data. In fact, neither participant gender nor maternal education were correlated with NWR score in the current data.

Last but not least, we investigated whether birth order might affect NWR scores, as it does other language tasks, resulting in first-born children showing higher scores on standardized language tests than later-born children (Havron et al., 2019), presumably because later-born children receive a smaller share of maternal input than their older siblings. Given shared caregiving practices and the hamlet organization typical of Rossel communities, children have many sources of adult and older child input that they encounter on a daily basis and first born children quickly integrate with a much larger pool of both older and younger children with whom they partly share caregivers. Therefore we expected that any effects of birth order on NWR would be attenuated in this context. In line with this prediction, our descriptive analysis showed no correlation between birth order and NWR score.

Conclusions. While NWR can, in theory, be used to test a variety of questions about phonological development in any language, previous work has been primarily limited to a handful of related languages spoken in urban, industrialized contexts. The present study shows that, not only can NWR be adapted for very different populations than have previously been tested, but that effects of age and typological frequency may strongly influence phonological development across these diverse settings, while effects of item length, participant gender, maternal education, and birth order, may either have little impact on this facet of language development or have an impact that vaies depending on the linguistic, cultural, and sociodemographic properties of the population under study. Because these latter predictors strongly relate to other language outcomes, the present findings raise the issue of why NWR would pattern differently, what that could tell us about the relationship between lexical development, phonological development, and the input environment and, last but not least, what is implied about the joint applicability of these outcome measures as a diagnostic indicator for language delays and disorders. In the meanwhile, we take the present findings as robustly supporting the idea that phonological development continues well past early childhood and as yielding preliminary support for a connection between individual learners and global language patterns when it comes to acoustic and articulatory markedness.

Acknowledgments

We are grateful to the individuals who participated in the study. The collection and annotation of these recordings was made possible by Ndapw:éé Yidika, Taakêmê Ńamono, and Y:aaw:aa Pikuwa; with thanks also to the PNG National Research Institute, and the Administration of Milne Bay Province. AC acknowledges financial and institutional support from Agence Nationale de la Recherche (ANR-17-CE28-0007 LangAge, ANR-16-DATA-0004 ACLEW, ANR-14-CE30-0003 MechELex, ANR-17-EURE-0017) and the J. S. McDonnell Foundation Understanding Human Cognition Scholar Award. MC acknowledges financial support from an NWO Veni Innovational Scheme grant (275-89-033).

References

Balladares, J., Marshall, C., & Griffiths, Y. (2016). Socio-economic status affects sentence repetition, but not non-word repetition, in Chilean preschoolers. First Language, 36(3), 338–351. https://doi.org/10.1177/0142723715626067

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do North American babies hear? A large-scale cross-corpus analysis. Developmental Science, 22(1), e12724. https://doi.org/10.1111/desc.12724

Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer (Version 6.1.35). Retrieved from http://www.praat.org/

Bowey, J. A. (2001). Nonword repetition and young children's receptive vocabulary: A longitudinal study. Applied Psycholinguistics, 22(3), 441–469.

Brandeker, M., & Thordardottir, E. (2015). Language exposure in bilingual toddlers: Performance on nonword repetition and lexical tasks. American Journal of Speech-Language Pathology, 24(2), 126–138.

Brown, P. (2011). The cultural organization of attention. In A. Duranti, E. Ochs, & and Bambi B Schieffelin (Eds.), Handbook of Language Socialization (pp. 29–55). Malden, MA: Wiley-Blackwell.

Brown, P. (2014). The interactional context of language learning in Tzeltal. In I. Arnon, M. Casillas, C. Kurumada, & B. Estigarribia (Eds.), Language in interaction: Studies in honor of Eve V. Clark (pp. 51–82). Amsterdam, NL: John Benjamins.

Brown, P., & Casillas, M. (n.d.). Childrearing through social interaction on Rossel Island, PNG. In A. J. Fentiman & M. Goody (Eds.), Esther Goody revisited: Exploring the legacy of an original inter-disciplinarian (pp. XX–XX). New York, NY: Berghahn.

Bunce, J., Soderstrom, M., Bergelson, E., Rosemberg, C., Stein, A., Alam, F., … Casillas, M. (n.d.). A cross-cultural examination of young children's everyday language experiences.

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a papuan community. Journal of Child Language, XX, XX–XX.

Castro-Caldas, A., Petersson, K. M., Reis, A., Stone-Elander, S., & Ingvar, M. (1998). The illiterate brain. Learning to read and write during childhood influences the functional organization of the adult brain. Brain: A Journal of Neurology, 121(6), 1053–1063. https://doi.org/10.1093/brain/121.6.1053

Chiat, S., & Roy, P. (2007). The preschool repetition test: An evaluation of performance in typically developing and clinically referred children. Journal of Speech, Language, and Hearing Research, 50(2), 429–443.

COST Action. (2009). Language impairment in a multilingual society: Linguistic patterns and the road to assessment. Brussels: COST Office. Available Online at: Http://Www. Bi-Sli. Org.

Cristia, A., Farabolini, G., Scaff, C., Havron, N., & Stieglitz, J. (2020). Infant-directed input and literacy effects on phonological processing: Non-word repetition scores among the Tsimane'. PLoS ONE, 15(9), e0237702. https://doi.org/https://doi.org/10.1371/journal.pone.0237702

de Santos Loureiro, C., Braga, L. W., Nascimento Souza, L. do, Nunes Filho, G., Queiroz, E., & Dellatolas, G. (2004). Degree of illiteracy and phonological and metaphonological skills in unschooled adults. Brain and Language, 89(3), 499–502. https://doi.org/10.1016/j.bandl.2003.12.008

Estes, K. G., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. Journal of Speech, Language, and Hearing Research, 50(1), 177–195.

Farabolini, G., Rinaldi, P., Caselli, C., & Cristia, A. (2021). Non-word repetition in bilingual children: The role of language exposure, vocabulary scores and environmental factors. Speech Language and Hearing.

Farmani, H., Sayyahi, F., Soleymani, Z., Labbaf, F. Z., Talebi, E., & Shourvazi, Z. (2018). Normalization of the non-word repetition test in farsi-speaking children. Journal of Modern Rehabilitation, 12(4), 217–224.

Foley, W. A. (1986). The papuan languages of new guinea. Cambridge, UK: Cambridge University Press.

Gallagher, G. (2014). An identity bias in phonotactics: Evidence from Cochabamba Quechua. Laboratory Phonology, 5(3), 337–378. https://doi.org/10.1515/lp-2014-0012

Gallon, N., Harris, J., & Van der Lely, H. (2007). Non-word repetition: An investigation of phonological complexity in children with grammatical sli. Clinical Linguistics & Phonetics, 21(6), 435–455.

Gathercole, S. E., Willis, C., & Baddeley, A. D. (1991). Differentiating phonological memory and awareness of rhyme: Reading and vocabulary development in children. British Journal of Psychology, 82(3), 387–406.

Havron, N., Ramus, F., Heude, B., Forhan, A., Cristia, A., Peyre, H., & Group, E. M.-C. C. S. (2019). The effect of older siblings on language development as a function of age difference and sex. Psychological Science, 30(9), 1333–1343.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. Journal of Phonetics, 28(4), 377–396.

Jaber-Awida, A. (2018). Experiment in non word repetition by monolingual Arabic preschoolers. Athens Journal of Philology, 5(4), 317–334. https://doi.org/10.30958/ajp.5-4-4

Kalnak, N., Peyrard-Janvid, M., Forssberg, H., & Sahlén, B. (2014). Nonword repetition–a clinical marker for specific language impairment in swedish associated with parents' language-related problems. PloS One, 9(2), e89544.

Ladefoged, P., & Maddieson, I. (1996). The sounds of the world's languages (Vol. 1012). Blackwell Oxford.

Levinson, S. C. (2020). A grammar of yélî dnye, the papuan language of rossel island. Berlin, Boston: De Gruyter Mouton.

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & de Vos, C. (2012). A prelinguistic gestural universal of human communication. Cognitive Science, 36(4), 698–713. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Maddieson, I. (2005). Correlating phonological complexity: Data and validation. UC Berkeley PhonLab Annual Report, 1(1).

Maddieson, I., & Levinson, S. C. (n.d.). The phonetics of yélî dnye, the language of rossel island.

Meir, N., & Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic status (ses) and bilingualism on children’s vocabulary and verbal short-term memory. Frontiers in Psychology, 8, 1442.

Meir, N., Walters, J., & Armon-Lotem, S. (2016). Disentangling sli and bilingualism using sentence repetition tasks: The impact of l1 and l2 properties. International Journal of Bilingualism, 20(4), 421–452.

Moran, S., & McCloy, D. (Eds.). (2019). PHOIBLE 2.0. Jena: Max Planck Institute for the Science of Human History. Retrieved from https://phoible.org/

Peute, A. A. K., Fikkert, P., & Casillas, M. (n.d.). Early consonant production in Yélî Dnye and Tseltal.

Piazzalunga, S., Previtali, L., Pozzoli, R., Scarponi, L., & Schindler, A. (2019). An articulatory-based disyllabic and trisyllabic non-word repetition test: Reliability and validity in italian 3-to 7-year-old children. Clinical Linguistics & Phonetics, 33(5), 437–456.

Torrington Eaton, C., Newman, R. S., Ratner, N. B., & Rowe, M. L. (2015). Non-word repetition in 2-year-olds: Replication of an adapted paradigm and a useful methodological extension. Clinical Linguistics & Phonetics, 29(7), 523–535.

Vance, M., Stackhouse, J., & Wells, B. (2005). Speech-production skills in children aged 3–7 years. International Journal of Language & Communication Disorders, 40(1), 29–48.

Wilsenach, C. (2013). Phonological skills as predictor of reading success: An investigation of emergent bilingual Northern Sotho/English learners. Per Linguam: a Journal of Language Learning = Per Linguam: Tydskrif vir Taalaanleer, 29(2), 17–32. https://doi.org/10.5785/29-2-554