1   Non-word repetition in Yélî Dnye

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

3   [1] Laboratoire de Sciences Cognitives et de Psycholinguistique, Département d'Etudes cognitives,

4   ENS, EHESS, CNRS, PSL University

5   [2] Max Planck Institute for Psycholinguistics

6   [3] University of Chicago

7   Author Note

8   Correspondence concerning this article should be addressed to Alejandrina Cristia, 29, rue

9   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. Our overarching goal is to reflect on how NWR scores can be compared across participants, studies, languages, and populations, in order to shed light on the factors universally structuring variation in language development. More specifically, this study contributes to three lines of research. First, we contribute to investigations on NWR across diverse languages, by documenting that, in Yélî Dnye, 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, contributing to mounting research suggesting that length effects may be language- or population-specific, we find rather weak effects of item length. Third, we add a datapoint on potential sources of individual variation effects, by establishing that in our sample age has a strong effect on NWR scores, whereas there are weak correlations with gender, maternal education, and birth order. 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

TODO Middy

- read over whole thing – does the logic sound ok?
- tell me if you think you CAN'T LIVE with my turning this in with figures 1 and 5 as they are (I promise I'll improve them in the revision stage)
- do you mind if I change all the figures to say NWR scores rather than NWR Scores? In the text, we use lower case.

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 between phonology, working memory, and the lexicon are (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

55 Action, 2009; Meir, Walters, & Armon-Lotem, 2016). In the present study, we use NWR to

56 investigate the phonological development of children learning Yélî Dnye, an isolate language

57 spoken in Papua New Guinea (PNG) that has a large and unusually dense phonological inventory.

58 The study was designed to contribute to four broad research areas.

59   The first research area is at the crossing of typology and phonological development. Previous

60 work using NWR has preferred relatively universal and early-acquired phonemes (with the possible

61 exception of Gallagher, 2014), in part as a way to separate phoneme pronunciation from broader

62 syllable structure and word-level prosodic effects (Gallon, Harris, & Van der Lely, 2007) and in

63 part because the test is sometimes used to measure working memory in the context of executive

64 functions (Mulder, Verhagen, Van der Ven, Slot, & Leseman, 2017) rather than purely language.

65 Here, we investigate repetition of non-word items containing cross-linguistically common and

66 cross-linguistically rare phonetic targets. Specifically, we included a subset of non-word items with

67 typologically rare sounds to ask whether these sounds are disadvantaged in the

68 perception-production loop involved in NWR.

69   Second, we varied the length (in syllables) of non-words to contribute to growing research

70 looking at the impact of word length on NWR repetition, and what this may reflect about

71 phonological development within specific languages. Some work documents much lower NWR

72 scores for longer, compared to shorter, items (e.g., among Cantonese-learning children; Stokes,

73 Wong, Fletcher, & Leonard, 2006), whereas differences are negligible in other studies (e.g., among

74 Italian learners; Piazzalunga, Previtali, Pozzoli, Scarponi, & Schindler, 2019). It is possible that

75 differences are due to language characteristics, including the modal length of words in the language

76 and/or in child-directed speech in that culture. In broad terms, one may expect languages with a

77 lexicon that is heavily biased towards monosyllables to show greater length effects than languages

78 where words are modally longer. To see whether there were broad generalizations that could be

79 drawn from previous literature fitting these predictions, we inspected NWR papers in a variety of

80 languages which reported NWR scores separately for different word lengths. We found data for

81 learners of Israeli Arabic Jaber-Awida (2018); Cantonese (Stokes et al., 2006); English (Vance,

82 Stackhouse, & Wells, 2005); Italian (Piazzalunga et al., 2019); and Tsimane' (Cristia, Farabolini,

83 Scaff, Havron, & Stieglitz, 2020); and integrated those data with Yélî Dnye results from the present
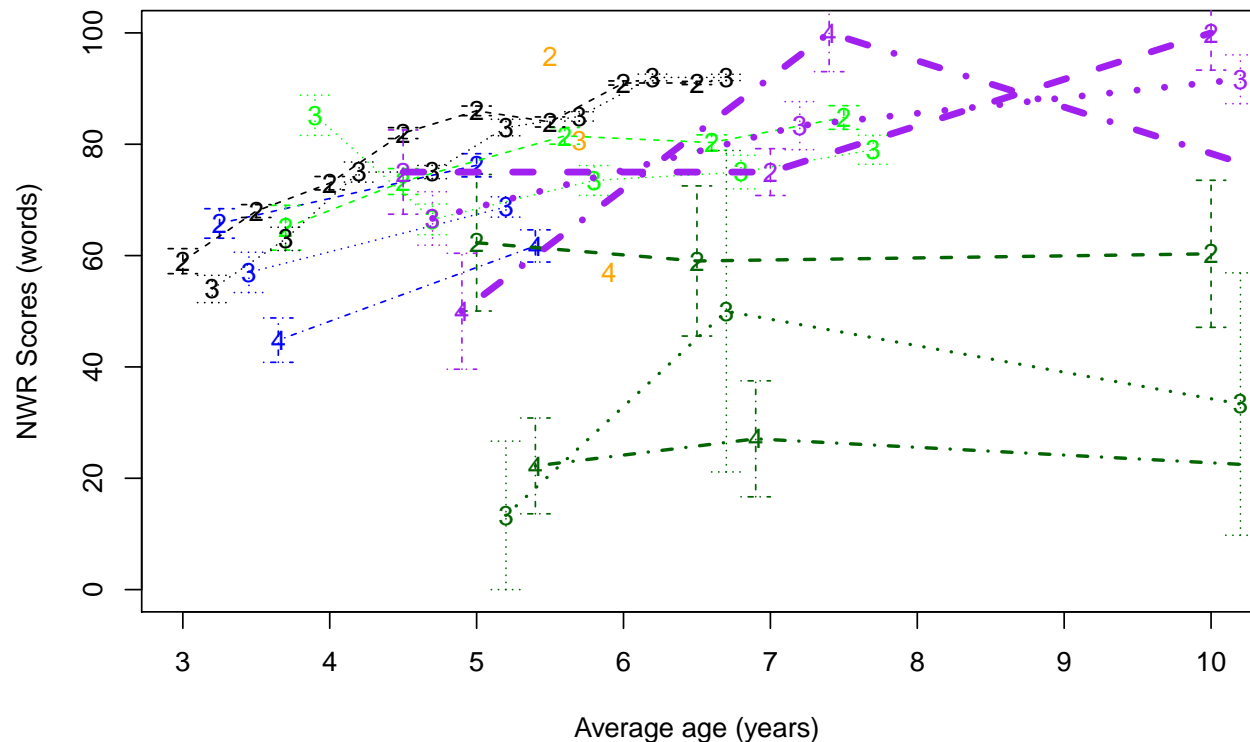
84 study in Figure 1.

Figure 1. NWR scores as a function of age (in years) and item length for comparable studies (2-4 indicating number of syllables, 2 = dashed, 3 = dotted, 4 = dotted and dashed). Jaber-Awida (2018) reported on 20 Israeli Arabic learners (orange); Piazzalunga et al. (2019) reported on groups of 24-60 Italian learners (black); Stokes et al. (2006) on 15 Cantonese learners (blue); Vance et al. (2005) on 17-20 English learners (light green); Cristia et al. (2020) reported on groups of 4-6 Tsimane' learners (dark green); the present study reports on groups of 8-19 Yélî Dnye learners (purple). Central tendency is the mean except for Italian and Yélî Dnye (median); error is one standard error. Age has been slightly jittered for ease of inspection of different lengths at a given age.

85   Our reading of this Figure is that, although there is cross-linguistic (or cross-sample)

86 variation in length effects, these do not systematically lign up with expected word length in

87 different languages. For instance, the difference in NWR scores for 2- versus 3-syllable items

88 (averaging across age groups) is largest in Tsimane' (~28%) and Arabic (~15%), which tend to

89 have longer words, as does Italian, where the difference between 2- and 3-syllable items was only

90 ~2%. Similarly, two languages that are often described as heavily biased towards monosyllables

91 show diverse length effects (Cantonese ~8% versus English ~1%). Given the paucity of research

92 looking at this question, and the diversity of current results, we do not approach this issue within a

93 hypothesis-testing framework but sought instead to provide one more piece of data on the question,

94 which may be re-used in future meta- or mega-analytic approaches.

95   The third research area we contribute data to relates to the possibility that individual

96 variation in NWR scores is structured. Although the ideal systematic review is missing, a recent

97 paper comes close with a rather extensive review of the literature looking at correlations between

98 NWR scores and a variety of child-level variables, including familial socio-economic status, child

99 vocabulary, and among multilingual children, levels of exposure to the language on which the

100 non-words are based (Farabolini, Rinaldi, Caselli, & Cristia, 2021). In a nutshell, most evidence is

101 mixed, suggesting that consistent individual variation effects may be small, and more data is

102 needed to estimate their true size. For this reason, we descriptively report association strength

103 between NWR scores and child age, sex, birth order, and maternal education. Based on previous

104 work, we looked at potential increases with age (Farmani et al., 2018; Kalnak, Peyrard-Janvid,

105 Forssberg, & Sahlén, 2014; Vance et al., 2005). Previous work typically finds no significant

106 differences as a function of maternal education (e.g., Farmani et al., 2018; Balladares, Marshall, &

107 Griffiths, 2016; Kalnak et al., 2014; Meir & Armon-Lotem, 2017) or child gender (Chiat & Roy,

108 2007). Although past research has not often investigated potential effects of birth order on NWR,

109 there is a sizable literature on these effects in other language tasks (Havron et al., 2019), and

110 therefore we report on these too.

111   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 urban settings in Europe and North America (with exceptions including Gallagher, 2014; Cristia et al., 2020), nor with languages having large phonological inventories [e.g., more than 34 consonants and 7 vowel qualities Maddieson (2013b);Maddieson (2013a); no exceptions to our knowledge]. 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 non-universal 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 phonological and phonetic representations. To our knowledge, there is little discussion of linguistic effects – i.e., of potential differences in NWR as a function of language typology – or cultural effects – i.e., of potential differences in NWR as a function of other differences across human populations, aside from Cristia et al. (2020)'s hypotheses just mentioned. Regarding potential language differences, we note that the very fact that studies compose items by varying syllable structure and word length, while prefering relatively simple and universal phones (notably relying on point vowels, simple plosives, and fricatives that are prevalent across languages, like /s/) may indicate a bias towards Indo-European languages, where syllable structure and word length are indeed important structural dimensions. This bias is, of course, implicit and unintentional, arising as researchers working in other languages attempt to build items that conform to the descriptions of the first people using the method, who tend to work on English.

139   Before going into the details of our study design we first give an overview of Yélî Dnye

140   phonology as well as a brief ethnographic review of the developmental environment on Rossel

141   Island. As discussed above, NWR has been almost exclusively used in urban, industrialized

142   populations, so we provide this additional ethographic information to contextualize the adaptations

143   we have made in running the task and collecting the data, compared to what is typical in commonly

144   studied sites, which are typically easily accessible. Laying 250 nautical miles off the coast of

145   mainland PNG and surrounded by a barrier reef, transport to and from Rossel Island is both

146   infrequent and irregular. International phone calls and digital exchanges that require significant

147   data transfer are typically not an option. Data collection is therefore typically limited to the

148   duration of the researchers' on-island visits.

149   Yélî Dnye phonology.   Yélî Dnye is an isolate language (presumed Papuan) spoken by

150   approximately 7,000 people residing on Rossel Island, an island found at the far end of the

151   Louisiade Archipelago in Milne Bay Province, Papua New Guinea. The Yélî sound system, much

152   like its baroque grammatical system (Levinson, 2020), is unlike any other in the region. In total,

153   Yélî Dnye uses 90 distinctive segments (not including an additional three rarely used consonants),

154   far outstripping the phonemic inventory size of other documented Papuan languages (Foley, 1986;

155   Levinson, 2020; Maddieson & Levinson, n.d.). Thus, with respect to our first research goal, Yélî

156   Dnye seemed is a good language to attempt an investigation on NWR with sounds varying in

157   cross-linguistic frequency because of its large inventory, which includes some rare sounds.

158   To provide some qualitative information on this inventory, we add the following

159   observations. With only four primary places of articulation (bilabial, alveolar, post-alveolar, and

160   velar) and no voicing contrasts, the phonological inventory is remarkably packed with acoustically

161   similar segments. The core oral stop system includes both singleton (/p/, /t/, /ṭ/, and /k/) and

162   doubly-articulated (/tp/, /ṭp/, /kp/) segments, with full nasal equivalents (/m/, /n/, /ṇ/, /ŋ/, /nm/, /ṇm/,

163   /ŋm/), and with a substantial portion of them contrastively pre-nasalized or nasally released (/mp/,

164   /nt/, /ṇṭ/, /ŋk/, /nmtp/, /ṇmṭp/, /ŋmkp/, /ṭn/, /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/; /ṭpʲ/; /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 seven of these able to appear as short and long nasal vowels as well /ĩ/, /ũ/, /ɔ̃/, /ɛ̃/, /ɔ̃/, /æ̃/, /ɑ̃/).

Regarding our second research goal, on the effect of non-word length on NWR, most Yélî Dnye words 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 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 CVC syllable, but these are limited to codas of -/m/, -/p/, or -/j/ (e.g., "ndap" /ṇtæp/ Spondylus shell) and are often resyllabified with an epenthetic /ɯ/ in spontaneous speech (e.g., "ndapî" /'ṇtæ.pɯ/). There are also a handful of words starting with /æ/ (e.g., "ala" /æ.'læ/ here) and a small collection of single-vowel grammatical morphemes (see Levinson (2020) for details).

Our knowledge of Yélî language development is growing (e.g., Brown, 2011, 2014; Brown & Casillas, n.d.; Casillas, Brown, & Levinson, 2020; Liszkowski, Brown, Callaghan, Takada, & de Vos, 2012), but research into Yélî phonological development has only just begun (e.g., Peute, Fikkert, & Casillas, n.d.). We hope the present study contributes to filling this gap. TODO incorporate brief summary of paper

The Yélî community.   Some aspects of the community are relevant for interpreting results found when addressing our thir research question, regarding sources of individual variation. Specifically, we investigated potential effects of age, gender, maternal education, and birth order.

---

[1]We use Levinson's (2020) under-dot notation (e.g., /ṭ/) to denote 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/).

188 There is nothing particular to note regarding age and gender, but we have some comments that

189 pertain to the other two factors.

190   The typical household in our dataset includes seven individuals (typically, a mixed sex

191 couple and children – their own and billeting others, as discussed in the next paragraph) and is

192 situated among a collection of four or more other households, with structures often arranged

193 around an open grassy area. These household clusters are organized by patrilocal relation, such that

194 they typically comprise a set of brothers, their wives and children, and their mother and father,

195 with neighboring hamlets also typically related through the patriline. Land attribution for building

196 one's home is decided collectively based on land availability, and typically does not take into

197 consideration an individual's desire to be close to a school.

198   Most Yélî parents are swidden horticulturalists, and those who are not may not reside in the

199 island. Within a group of households, it is often the case that most older adolescent and adults

200 spend their day tending to their gardens (which may not be nearby), bringing up water from the

201 river, washing clothes, preparing food, and engaging in other such activities, which leave them

202 little time to spend directly with the children in their household (other than infants). Starting

203 around age two years, children more often spend large swaths of their day playing, swimming, and

204 foraging for fruit, nuts, and shellfish in large (~10 members) independent and mixed-age child play

205 groups (Brown & Casillas, n.d.; Casillas et al., 2020). Formal education is a priority for Yélî

206 families, and many young parents have themselves pursued additional education beyond of what is

207 locally available (Casillas et al., 2020). Local schools are well out of walking distance for many

208 children (i.e., more than 1 hour on foot or by canoe each day), so it is very common for households

209 situated close to a school to billet their school-aged relatives during the weekdays for long

210 segments of the school year. Children start school often at around age six, although the precise age

211 depends on the child's apparent development.

212   Some general ideas regarding potential maternal education effects on our data may be drawn

213 from the observations above. To begin with, many of our participants above 6 years of age may not

be living with their birth mother but with other relatives, which may weaken maternal education

effects. Additionally, the importance given to formal education appears relatively stable over the

period that Rossel Island has been visited by language researchers (Steven Levinson and Penelope

Brown, about 20 years). Together with the fact that land attribution is essentially random with

respect to educational hopes, it seems to us that the length of formal education a given individual

may have is not necessarily a good index of their socio-economic status or other individual

properties, unlike what happens in industrialized sites, and variation may simply due to random

factors like living close to a school or having relatives there.

As for birth order, much of the work on birth order effects on cognitive development

(including language) has been carried out in the last 70 years and in agrarian or industrialized

settings (Barclay, 2015; Grätz, 2018), where nuclear families are more likely to be the prevalent

rearing environment (Lancy, 2015). It is possible that birth order effects are stronger in such a

setting, because much of the stimulation can only come from the parents, and when there are

multiple children, the inter-birth interval is small enough that older siblings may not be of an age

that allows them to contribute to their younger siblings' stimulation. This contrasts with this

picture just drawn in the Yélî community, where children regardless of their birth order in their

nuclear family will typically benefit from a rich and extensive socially stimulating setting,

surrounded by siblings, and cousins of several orders.

We add some observations that will help us integrate this study to the broader investigation

of NWR across cultures. As mentioned previously, there is one report of lower NWR scores

among the Tsimane', which the authors interpret as consistent with long-term effects of low levels

of infant-directed speech (Cristia et al., 2020). However, Cristia et al. (2020) also point out that

this is based on between-paper comparisons, and thus methods and a myriad other factors have not

been controlled for. The Yélî community can help us bring further light into this question because

direct speech to children under 3;0 is relatively infrequent in this community too (Casillas et al.,

2020). Although infant-directed speech has been measured in different ways among the Tsimane'

and the Yélî communities, our most comparable estimates at present suggests that Tsimane' young

children are spoken to about 4.2 minutes per hour (Scaff, Stieglitz, Casillas, & Cristia, 2021), and

Yélî children about 3.7 minutes per hour (Casillas et al., 2020). Thus, if input quantities in early

childhood are a major determinant of NWR scores, we should observe similarly low NWR scores

as in Cristia 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

contrastive vowel length. There is some variation in how past NWR studies have 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 (e.g., Cristia et al., 2020), 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 and

to contribute data on whether this affects repetition. 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, as discussed above, effects of individual

differences in NWR are probably relatively small, and thus we reasoned that they 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- or mega-analytic efforts.

Research questions.   After some preliminary analyses to set the stage, we perform statistical

analyses to inform answers to 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 do NWR scores 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?

Throughout these analyses and in the Discussion, we will also have in mind our fourth goal,

namely integrating NWR results across samples varying in language and culture.

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 the previously

discussed variability across studies. Therefore, all analyses in the present study are descriptive and

should be considered exploratory.

Methods

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 et al., 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 40 final 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 2- 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 a second trip to the island in 2019, when data were collected, 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" /yuɪ/; identified as "yi" /yi/, tree). Additionally, an error was made when

preparing files for annotation, resulting in two items being merged ("tpâ" /tpɑ/ and "tp:a" /tpæ̃/).

These three problematic items are not described here, and removed from the analyses below.

The final list includes three practice items and 40 test items (across infants): 16

monoysllables containing sounds that are less frequent in the world's languages than singleton

plosives; 8 bisyllables; 12 trisyllables; and 4 quadrisyllables (see Table 1).

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/dtxue/. 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 syllables), 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.

To inform our analyses, we estimated the typological frequency of all phonological segments present 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 the time of writing, it is of course still far from complete, which may mean that frequencies are estimates rather than precise descriptors). Note that nearly half of the segment types are only attested in one language (Steven Moran, personal communication). Extrapolating from this observation, we treat the three segments in our stimuli that were unattested in PHOIBLE (/lβʲ/, /ʈp/, and /tp/) as having a frequency of 1 (i.e., appearing in one language), with a (rounded) percentile of 0% (i.e., its cross-linguistic percentile is zero).

Additionally, we estimated frequency of the phones present in the target items 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.[2] Here, unattested sounds were not considered (i.e., they were declared NA so that they do not count for analyses).

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 tested in four different sites spread across the northeastern region of the island, 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.

---

[2]We also carried out analyses using token (rather than type) 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 Results section.

363 Whenever children living in the same household were tested, we tried to test children in age order,

364 from oldest to youngest, to minimize intimidation for younger household members, and always

365 using different elicitation sets. Because space availability was limited in different ways from

366 hamlet to hamlet, the places where elicitation happened varied across testing sites. More

367 information is available from the online supplementary materials.

368   We fitted the child with a headset microphone (Shure SM10A or WH20 XLR with a dynamic

369 microphone on a headband, most children using the former) that fed into the left channel of a

370 Tascam DR40x digital audio recorder. The headsets were designed for adult use and could not be

371 comfortably seated on many children's heads without a more involved adjustment period. To

372 minimize adjustment time, which was uncomfortable for some children given the proximity of the

373 experimenter and equipment, we placed the headband on children's shoulders in these cases,

374 carefully adjusting the microphone's placement so that it was still close to the child's mouth. A

375 research assistant who spoke Yélî Dnye natively sat next to the child throughout the task to provide

376 instructions and, if needed, encouragement. The research assistant coached the child throughout the

377 task to make sure that they understood what they were expected to do. An experimenter (the first

378 author) delivered the pre-recorded stimuli to the research assistant and the child over headphones.

379   The first phase of the experiment involved making sure the child understood the task. We

380 explained the task and then orally presented the first practice item. At this point, many children did

381 not say anything in response, which triggered the following procedure: First, the assistant insisted

382 the child make a response. If the child still did not say anything, the assistant said a real word and

383 then asked the child to repeat it, then another and another. If the child could repeat real words

384 correctly, we provided the first training item over headphones again for children to repeat. Most

385 children successfully started repeating the items at this point, but a few needed further help. In this

386 case, the assistant modeled the behavior (i.e., the child and assistant would hear the item again, and

387 the assistant would repeat it; then we would play the item again and ask the child to repeat it). A

388 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 going over the list of test items randomly

assigned to each child. 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 allowed to make 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 we 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 six

minutes, with the first minute attributed to practice and five minutes 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

ran a script to extract these tokens, pairing them with their original auditory target stimulus, and

writing these audio pairs out to .wav clips. The assistant then listened through all these paired

target-repetition clips randomized across children and repetitions, grouped such that all the clips of

the same target were listened to in succession. For each clip, the assistant indicated in a notebook

whether the child 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 was also provided with 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.   Previous work typically reports two scores: a binary word-level exact repetition score, and a phoneme-level score, defined as the number of phonemes that can be aligned across the target and attempt, divided by the number of phonemes of whichever item was longer (the target or the attempt; as in Cristia et al., 2020). Previous work does not use distance metrics, but we report these rather than the phoneme-level scores because they are more informative. To illustrate these scores, 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). Notice that the phone-based Levenshtein distance is the complement of the phoneme-level NWR score. An advantage of using phone-based Levenshtein distance is that it is scored automatically with a script, and it can then easily be split in terms of deletions and substitutions (insertions were not attested in this study).

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). As discussed in subsection "The Yélî community", the combination of collective child guardianship practices and common billeting of school-aged children for them to attend school is that adult consent often comes from a combination of aunts, uncles, adult cousins, and grandparents standing in for the child's biological parents. Child assent is also culturally pertinent, as independence is encouraged and respected from toddlerhood (Brown & Casillas, n.d.). Participation was voluntary; children were invited to participate following indication of approval

441 from an adult caregiver. Regardless of whether they completed the task, children were given a

442 small snack as compensation. Children who showed initial interest but then decided not to

443 participate were also given the snack.

444   We tested a total of 55 children from 38 families spread across four hamlets. We excluded

445 test sessions from analysis for the following reasons: refused participation or failure to repeat items

446 presented over headphones even after coaching (N = 8), spoke too softly to allow offline coding

447 (N = 5), or were 13 years old or older (N = 2; we tested these teenagers to put younger children at

448 ease). The remaining 40 children (14 girls) were aged from 3 to 10 years (M = 6.50 years, SD =

449 1.50 years). In terms of birth order, 6 were first borns, 5 second, 2 third, 7 forth, 5 fifth, and 1 sixth,

450 with birth order missing for 14 children. These children were tested in a remote hamlet, and we

451 unfortunately did not ask about birth order before leaving the site. Maternal years of education

452 averaged 8.22 years (range 6-12 years).[3] We also note that there were 34 only exposed to Yélî

453 Dnye at home, 6 children exposed to Yélî Dnye plus one or more other languages at home.[4]

454 Results

455   Preliminary analyses.   We first checked whether whole-item NWR scores varied between

456 first and subsequent presentations of an item by averaging word-level scores at the participant level

457 separately for first attempts and subsequent repetitions. We excluded 1 child who did not have data

458 for one of these two types. As shown in Figure 2, participants' mean word-level scores became

---

[3]We asked for mothers' highest completed level of education. We then record the number of years entailed by having completed that level under ideal conditions.

[4]Most speakers of Yélî Dnye grow up speaking it monolingually until they begin attending school around the age of 7 years; 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 other language(s) from the region.
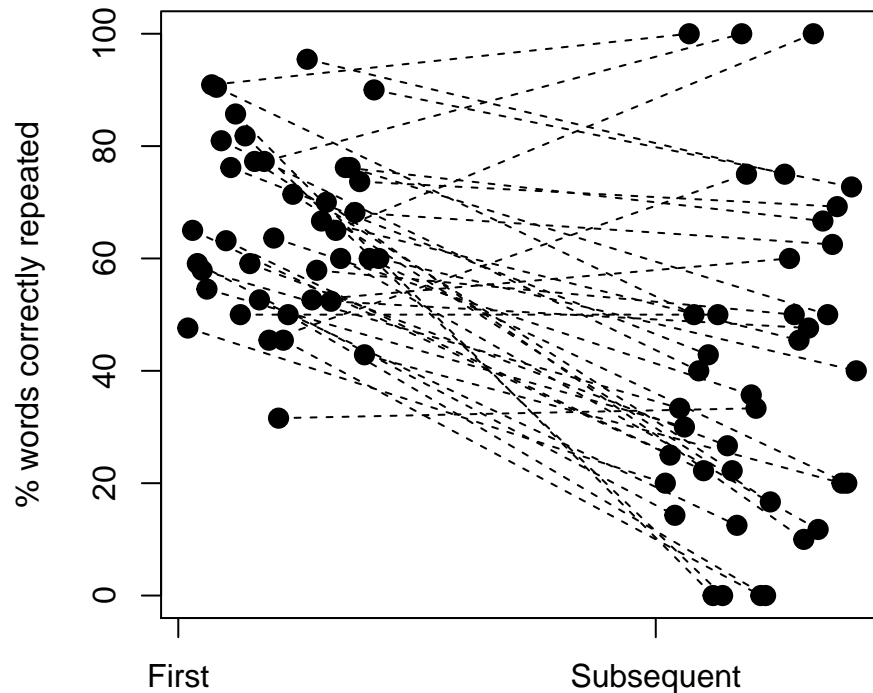
Figure 2. Whole-item NWR scores for individual participants averaging separately their first attempts and all other attempts.

more heterogeneous in subsequent repetitions. Surprisingly, whole-item NWR scores for subsequent repetitions (M = 40, SD = 28) were on average lower than first ones (M = 65, SD = 15), t(38) = 5.89, p < 0.001; Cohen's d = 1.13). 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, with the exception of qualitative analyses of substitutions.

Taking into account only the first attempts, we derived overall averages across all items. The overall NWR score was M = 65% (SD = 15%), Cohen's d = 4.39. The phoneme-based normalized Levenshtein distance was M = 21% (SD = 9%), meaning that about a fifth of phonemes were substituted, inserted, or deleted.

We also 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, test structure, or some other factor explains this difference, such as the particularities of the Yélî Dnye phonological inventory which lead any error to 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 may clarify this effect.

NWR as a function of cross-linguistic phone frequency.   Turning to our first research question, we 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 allowed slopes to vary over the random effects child ID and target ID.

We could include 826 observations, from 40 children producing in any given trial one of 40 potential target words. The analysis revealed a main effect of age (ß = 0.35, 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 = 0.78, SE ß = 0.19, p < 0.001): Target words with phones found more frequently across languages had higher correct repetition scores, as shown in Figure 3. Averaging across participants, the Pearson correlation between scaled average cross-linguistic phone frequency and whole-item NWR scores was r(38) = 0.54.

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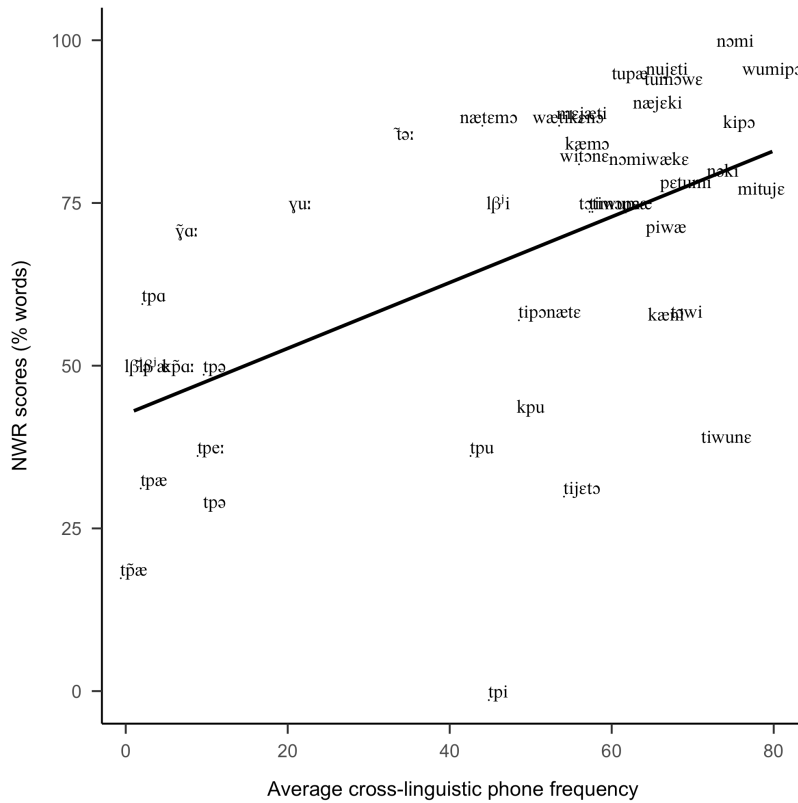
Figure 3. NWR scores for individual target words as a function of the average frequency with which each phone is found across languages.

frequent within a language, and therefore may be easier for children to represent and repeat because of the additional exposure. Phone corpus-based frequencies were correlated with phone cross-linguistic frequencies [$r(27) = 0.50$, $p < 0.01$]; and item-level average phone corpus-based frequencies were correlated with the corresponding cross-linguistic frequencies [$r(38) = 0.73$, $p < 0.001$]. Moreover, averaging across participants, the Pearson correlation between scaled average corpus phone frequency and whole-item NWR scores was $r(38) = 0.43$, $p < 0.01$. 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 phone frequency (ß = 0.78, SE ß = 0.27, $p < 0.01$) and age (ß = 0.35, SE ß = 0.13, $p < 0.01$) were significant predictors of whole-item NWR scores, but corpus phone

508 frequency (ß = 0.00, SE ß = 0.25, p = 0.99) was not.

509   Patterns in NWR mispronunciations.   We addressed our first research question in a second

510 way, by investigating patterns of error, looking at all attempts so as to base our generalizations on

511 more data. There were no cases of insertion, and deletions were very rare: there were only 12

512 instances of deleted vowels (~0.28% of all vowel targets), and 6 instances of deleted consonants

513 (~0.19% of all consonant targets). We therefore focus our qualitative description here on

514 substitutions: There were 820 cases of substitutions, ~16.95 of the 4839 phones found collapsing

515 across all children and target words, so that substitutions constituted the frank majority of incorrect

516 phones (~97.74 of unmatched phones). To inform our understanding of how cross-linguistic

517 patterns may be reflected in NWR scores, we asked: Is it the case that cross-linguistically less

518 common and/or more complex phones are more frequently mispronounced, and more frequently

519 substituted by more common ones than vice versa?

520   We looked for potential asymmetries in errors for different types of sounds in vowels by

521 looking at the proportion of vowel phones that were correctly repeated or not separately for nasal

522 and oral vowels. The nasal vowels in our stimuli occur in ~1.40% of languages' phonologies

523 (range 0% to 3%); whereas oral vowels in our stimuli occur in ~31.55% of languages' phonologies

524 (range 3% to 92%). As noted above, type frequency within the language is correlated with

525 cross-linguistic frequency, and thus these two types of sounds also differ in the former: Their type

526 frequencies in Yélî Dnye are: nasal vowels ~0.03‰ (range 0.00‰ to 0.05‰) versus oral ~0.23‰

527 (range 0.02‰ to 0.76‰).

528   We distinguished errors that included a change of nasality (and may or may not have

529 preserved quality), versus those that preserved nasality (and were therefore a quality error), shown

530 in Table 2. We found that errors involving nasal vowel targets were more common than those

531 involving oral vowels (35.90 versus 11.90). Additionally, errors in which a nasal vowel lost its

532 nasal character were 10 times more common than those in which an oral vowel was produced as a

533 nasal one. Note that this analysis does not tell us whether cross-linguistic or within-language

534 frequency is the best predictor, an issue to which we return below.

535   For consonants, we inspected complex ([t̪p], [tp], [kp], [km], [kn̪], [mp], [ɣ], and [lβʲ]) versus

536 simpler ones ([m], [n], [l], [w], [j], [w], [t̪], [g], [p], [t], [k], [f], [h], and [tʃ]), using the same logic:

537 We looked at correct phone repetition, substitution with a change in complexity category, or a

538 change within the same complexity category.[5] The complex consonants in our stimuli occur in

539 ~17.33% of languages' phonologies (range 0% to 78%); whereas simple consonants in our stimuli

540 occur in ~67.62% of languages' phonologies (range 13% to 96%). Again these groups of sounds

541 differ in their frequency within the language. Their type frequencies in Yélî Dnye are: complex

542 consonants ~0.04‰ (range 0.00‰ to 0.10‰) versus simple consonants ~0.32‰ (range 0.06‰ to

543 0.55‰).

544   Table 3 showed that errors involving complex consonants targets were more common than

545 those involving oral vowels (50.90 versus 8.20). Additionally, errors in which a complex

546 consonant was mispronounced as a simple consonant were quite common, whereas those in which

547 a simple consonant was produced as a complex one were vanishingly rare.

548   To address whether errors were better predicted by cross-linguistic or within-language

549 frequency, we calculated a proportion of productions that were correct for each phone (regardless

550 of the type of error or the substitution pattern). Graphical investigation suggested that in both cases

551 the relationship was monotonic and not linear, so we computed Spearman's rank correlations

552 between the correct repetition score, on the one hand, and the two possible predictors on the other.

553 Although we cannot directly test the interaction due to collinearity, the correlation with

554 cross-linguistic frequency [$r(319.72) = 0.76$, $p < 0.001$] was greater than that with within-language

555 frequency [$r(731.10) = 0.45$, $p = 0.05$].

─────────

[5]Note that the substitutions included phones that are not native to Yélî Dnye but do occur in English (e.g., [tʃ]). These data come from careful transcriptions by a native Yélî Dnye speaker who is very fluent in English. This result suggests that several of our participants have mastered production of some English phones, possibly produced within whole English word forms.

556   NWR scores as a function of item length.   We next turned to our second research question

557   by inspecting whether NWR scores varied as a function of word length (Table 4). In this section

558   and all subsequent ones, we only look at first attempts, for the reasons discussed previously.

559   Additionally, we noticed that participants scored much lower on monosyllables than on non-words

560   of other lengths. This is likely due to the fact that the majority of monosyllables were designed to

561   include sounds that are rare in the world's languages, which may be harder to produce or perceive,

562   as suggested by our previous analyses of NWR scores as a function of cross-linguistic phone

563   frequency and error patterns. Therefore, we set monosyllables aside for this analysis.

564   We observed the typical pattern of lower scores for longer items only for the whole-item

565   scoring, and even there differences were rather small. In a generalized binomial mixed model

566   excluding monosyllables, we included 479 observations, from 40 children producing, in any given

567   trial, one of 24 (non-monosyllabic) potential target words. The analysis revealed a positive effect

568   of age ($\beta = 0.56$, SE $\beta = 0.14$, $p < 0.001$) and a negative but non-significant estimate for target

569   length in number of syllables ($\beta = -0.15$, SE $\beta = 0.33$, $p = 0.65$).

570   Factor structuring individual variation.   Our final exploratory analysis assessed whether

571   variation in scores was structured by factors that vary across individuals, as per our third research

572   question. As shown in Figure 4, there was a greater deal of variance across the tested age range,

573   with significantly higher NWR scores for older children (Spearman's rank correlation, given

574   inequality of variance, rho (5,649.08) $= 0.47$, $p < 0.01$). In contrast, there was no clear

575   association between NWR scores and sex (Welch t (27.33) $= -0.60$, $p = 0.56$), birth order (data

576   missing for 15 children, rho $= (3,502.90) = -0.20$, $p = 0.33$), or maternal education (rho (
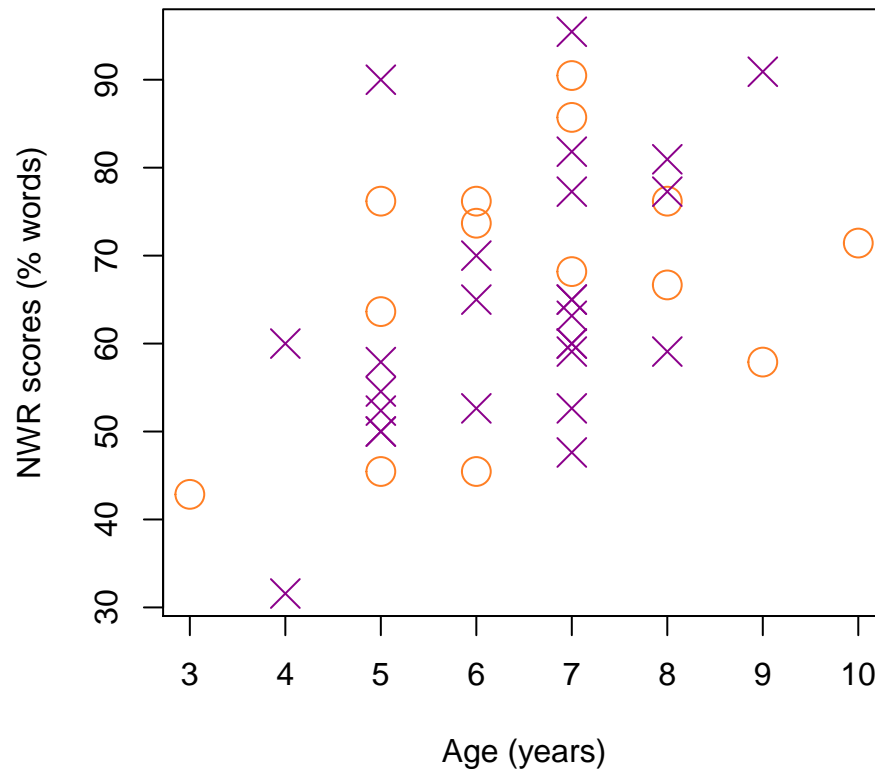
577   9,628.60) $= 0.10$, $p = 0.55$).

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

Discussion

We used non-word repetition to investigate phonological development in a language with a large phonological inventory (including some typologically rare segments). We aimed to provide additional data on two questions already visitied in NWR work, namely the influence of stimulus length and individual variation, and one research area that has received less attention, regarding the possible relationship between phone frequency and NWR scores. An additional overarching goal was to discuss NWR in the context of population and language diversity, since it is very commonly used to document phonological development in children raised in urban settings with wide-spread literacy, and has been less seldom used in non-European languages (but note there are exceptions, including work cited in the Introduction and other that will be discussed below). We consider implications of our results on each of these four research areas in turn.

589   Associations between NWR and cross-linguistic frequency.   Arguably the most innovative

590 aspect of our data relate to the inclusion of phones that are less commonly found across languages,

591 and rarely used in NWR tasks. Our monosyllabic items included typologically rare segments so

592 that we could test whether lower average segmental frequency is associated with lower NWR

593 scores. It would stand to reason that typologically common sounds are associated with higher

594 performance, but to our knowledge this has not yet been tested with NWR. There are some reasons

595 to believe that Yélî Dnye put that hypothesis to a critical test: The phonemic inventory is both

596 large and acoustically packed, in addition to containing several typologically infrequent (or unique)

597 contrasts. One could then predict that this effect should be relatively weak because the ambient

598 language puts pressure on Yélî children to distinguish (perceptually and articulatorily) fine-grained

599 phonetic differences in order to successfully communicate with others.

600   And yet, we found a robust effect of average segmental cross-linguistic frequency on NWR

601 performance: Even accounting for age and random effects of item and participant, we saw that

602 target words with typologically more common segments were repeated correctly more often. This

603 effect was large, with a magnitude more than twice the size of the effect of participant age.

604 Moreover, this significant effect remained even once also accounting for the frequencies of these

605 segments in Yélî Dnye children's input. An analysis of the substitutions made by children also

606 aligned with this interpretation, with more common sounds being substituted for less common ones.

607   We thus at present conclude that typological frequency of sounds is, to a certain extent,

608 mirrored in children's NWR, in ways that may not be due merely to how often those sounds are

609 used in the ambient language, and which are not erased by language-specific pressure to make

610 finer-grained differences early in development. We do not aim to reopen a debate on the extent to

611 which cross-linguistic frequency of occurrence can be viewed necessarily as reflecting ease of

612 perception or production (most often discussed in the case of phonotactic constraints on sequences,

613 e.g., Maddieson, 2009), but we do point out that this effect is interestingly different from effects

614 found in artificial language learning tasks (see Moreton & Pater, 2012 for a review) which are in

some ways quite similar to NWR. We believe that it may be insightful to extend the purview of

NWR from a narrow focus on working memory and structural factors to broader uses, including for

describing the fine-grained phonetic representations in the perception-production loop (as in e.g.,

Edwards, Beckman, & Munson, 2004).

Item Length.   We investigated the effect of item complexity on NWR scores by varying

both the number of syllables in the item. In broad terms, children should have higher NWR scores

for shorter items. That said, previous work summarized in the Introduction has shown both very

small (e.g., Piazzalunga et al., 2019) and very large (e.g., Cristia et al., 2020) effects of stimulus

length. Setting aside our monosyllabic stimuli (which contained typologically infrequent segments

with lower NWR scores, as just discussed), we examined effects of item length among the

remaining stimuli, which range between 2 and 4 syllables long. The effect of item length was not

significant in a statistical model that additionally accounted for age and random effects of item and

participant, and is small and inconsistent across ages (see Figure 1). We do not have a good

explanation for why samples in the literature vary so much in terms of the size of length effects,

but two possibilities are that this is not truly a length effect but a confound with some other aspect

of the stimuli, or that there is variation in phonological representations that is poorly understood.

We explain each idea in turn.

First, it remains possible that apparent length effects are actually due to uncontrolled aspects

of the stimuli. For instance, some NWR researchers model their non-words on existing words, by

changing some vowels and consonants – which could lead to fewer errors (since children have

produced similar words in the past); some researchers control tightly the diphone frequency of

subsequences in the non-words. Building on these two aspects that researchers often control, one

can imagine that longer items have fewer neighbors, and thus both the frequency with which

children have produced similar items and (relatedly) their n-phone frequency is overall lower. If

this idea is correct, a careful analysis of non-words used in previous work may reveal that studies

with larger length effects just happened to have longer non-words with lower n-phone frequencies.

641   Second, NWR is often described as a task that tests flexible perception-production, and as

642   such it is unclear why length effects should be observed at all. However, it is possible that NWR

643   relies on more specific aspects of perception-production, in ways that are dependent on stimulus

644   length. A hint in this direction comes from work on illiterate adults, who can be extremely accurate

645   when repeating short non-words, but whose NWR scores are markedly lower for longer items. In a

646   longitudinal study on Portuguese-speaking adults who were learning to read, Kolinsky, Leite,

647   Carvalho, Franco, and Morais (2018) found that, before reading training, the group scored 12.5%

648   on 5-syllable items, whereas after 3 months of training, they scored 62.5% on such long items,

649   whereas performance was at 100% for monosyllables throughout. Given that as adults they had

650   fully acquired their native language, and obviously they had flexible perception-production

651   schemes that allowed them to repeat new monosyllables perfectly, the change that occurred in

652   those three months must relate to something else in their phonological skills, something that is not

653   essential to speak a language natively. Thus, we hazard the hypothesis that sample differences in

654   length effects may relate to such non-essential skills. Since as stated this hypothesis is

655   under-specified, further conceptual and empirical work are both needed.

656   Individual differences.   Our review of previous work in the Introduction suggested that our

657   anticipated sample size would not be sufficient to detect most individual differences using NWR.

658   We give a brief overview of individual difference patterns of four types in the present data—age,

659   sex, birth order, and maternal education—hoping that these findings can contribute to future meta-

660   or mega-analytic efforts aggregating over studies.

661   In broad terms, we expected that NWR scores would increase with participant age, as this is

662   the pattern observed in several of the studies in Figure 1 (English Vance et al., 2005; Italian

663   Piazzalunga et al., 2019; Cantonese Stokes et al., 2006; but note Cristia et al., 2020 is an

664   exception). Indeed, age was significantly correlated with NWR score and also showed up as a

665   significant predictor of NWR score when included as a control factor in the analyses of both item

666   length and average segmental frequency. In brief, our results underscore the idea that phoonlogical

667 development continues well past the first few years of life, extending into middle childhood and

668 perhaps later (Hazan & Barrett, 2000).

669   In contrast, previous work shows little evidence for effects of maternal education (e.g.,

670 Farmani et al., 2018; Kalnak et al., 2014; Meir & Armon-Lotem, 2017) on NWR scores. We did

671 not expect large effects of maternal education in our sample because of two reasons: First,

672 education on Rossel Island is generally highly valued and so widespread that little variation is seen

673 there; second, formal education is not at all essential to ensuring one's success in society and may

674 not be a reliable index of local socioeconomic variation locally. In fact, maternal education

675 correlated with NWR score at about r~.1, which is small. Similarly, NWR scores may not vary

676 greatly with participant gender according to previous work (Chiat & Roy, 2007), and for that as

677 well we find effects of about that size.

678   Last but not least, we investigated whether birth order might affect NWR scores, as it does

679 other language tasks, resulting in first-born children showing higher scores on standardized

680 language tests than later-born children (Havron et al., 2019) and adults (in a battery including

681 verbal abilities, e.g., Barclay, 2015), presumably because later-born children receive a smaller

682 share of parental input and attention than their older siblings. Given shared caregiving practices

683 and the hamlet organization typical of Rossel communities, children have many sources of adult

684 and older child input that they encounter on a daily basis and first-born children quickly integrate

685 with a much larger pool of both older and younger children with whom they partly share caregivers.

686 Therefore we expected that any effects of birth order on NWR would be attenuated in this context.

687 In line with this prediction, our descriptive analysis showed a non-significant correlation between

688 birth order and NWR score. However, the effect size was larger than that found for the other two

689 factors and it is far from negligible, at r~.2 or Cohen's d~0.41. In fact, two large studies with

690 therefore precise estimates found effects of about d~.2 (Barclay, 2015; Havron et al., 2019), which

691 would suggest the effects we found are larger. We therefore believe it may be worth revisiting this

692 question with larger samples in similar child-rearing environments, to further establish whether

distributed child care indeed results in more even language outcomes for first- and later-born children.

NWR across languages and cultures.   The fourth research area to which we wanted to contribute pertained to the use of NWR across languages and populations, as when designing this study we wondered whether NWR was a fair test of phonological development. Although our data cannot answer this question because we have only sampled one language and population here, we would like to spend some time discussing the integration of these results to the wider NWR literature. It is important to note at the outset that we cannot obtain a final answer because integration across studies implies not only variation in languages and child-rearing settings, but also in methodological aspects including non-word length, non-word design (e.g., the syllable and phone complexity included in the items), and task administration, among others. Nonetheless, we feel the NWR task is prevalent enough to warrant discussion about this, as it is done for other tasks sometimes used to describe and compare children's language skills across populations, like the recent re-use of the MacArthur-Bates Communicative Development Inventory to look at vocabulary acquisition across multiple languages (Frank, Braginsky, Yurovsky, & Marchman, 2017).

At first sight, the range of performance we observed overlapped with previously observed levels of performance. Paired with our thorough training protocol, we had interpreted the NWR scores among Yélî Dnye learners as indicating that 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. Additionally, it seemed that Yélî children showed comparable performance to others tested on a similar task, despite the many linguistic, cultural, and socioeconomic differences between this and previously tested populations, unlike the case that had been reported for the Tsimane'.

To enrich this discussion, we looked for previous studies on monolingual children with normative development learning diverse languages, and entered them when they reported non-word repetition scores based on whole item scoring. We entered data from 14 studies (including ours),

719 presenting data from 12 languages. Specifically, Arabic was represented by Jaber-Awida (2018);

720 Cantonese by Stokes et al. (2006); English by Vance et al. (2005); Italian by Piazzalunga et al.

721 (2019); Mandarin by Lei et al. (2011); Persian by Farmani et al. (2018); Slovak by Kapalková,

722 Polišenská, and Vicenová (2013) and Polišenská and Kapalková (2014); Sotho by Wilsenach

723 (2013); Spanish by Balladares et al. (2016); Swedish by Kalnak et al. (2014) and Radeborg,

724 Barthelom, SjöBerg, and Sahlén (2006); Tsimane' by Cristia et al. (2020); and Yélî Dnye by the

725 present study. Studies varied in the length of non-words that were considered; whenever results

726 were reported separately for different lengths, we calculated overall averages based on lengths of 2

727 and 3 syllables, for increased comparability. Results separating different age groups are shown in 5.
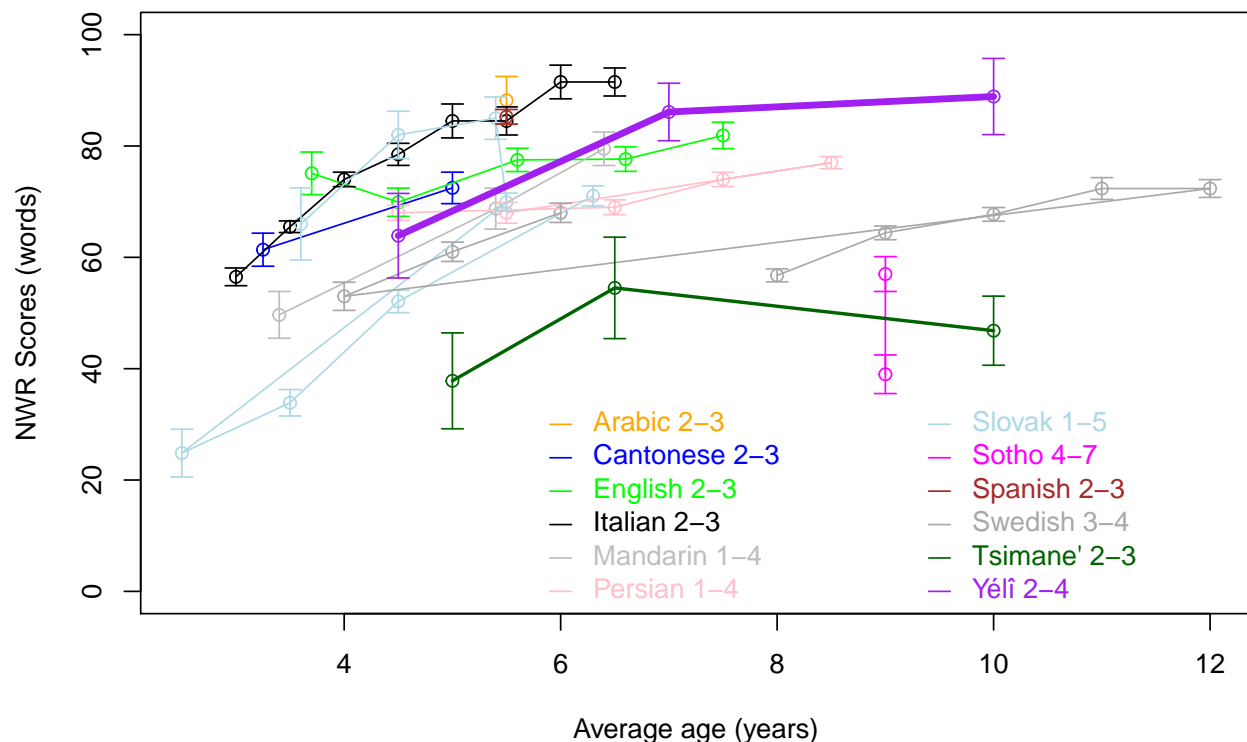

Figure 5. NWR scores as a function of age (in years), averaged across multiple non-word lengths, as a function of children's native languages. The legend indicates language and the length of non-words (in syllables). Central tendency is mean; error is one standard error.

728   Several observations can be drawn from this Figure. To begin with, we focus on the

729 comparison between Yélî Dnye and Tsimane'. These two groups have been described as having

730 roughly similar levels of child-directed speech, yet they exhibit very different results, with lower

731 overall NWR scores and (integrating with effect of length in 5) length effects. This may indicate

732 that the conclusion tentatively drawn in Cristia et al. (2020) about lower NWR scores consistent

733 with long-term effects of lower levels of child-directed speech was premature. Naturally, there is

734 an alternative interpretation, namely that input estimation suggesting very slightly higher levels of

735 child-directed speech among the Tsimane' than among Yélî Dnye learners is inaccurate. In fact,

736 careful reading of previous reports highlight important methodological differences in how input

737 quantity has been estimated across papers: Casillas et al. (2020) hand-coded speech with the help

738 of a native research assistant, and then summed all child-directed speech, which effectively

739 establishes an upper boundary of the speech children could potentially process. Cristia, Dupoux,

740 Gurven, and Stieglitz (2019) estimated quantities from behavioral observations on the frequency of

741 child-directed one-on-one conversation, which is probably closer to a lower boundary. Finally,

742 Scaff et al. (2021) used human annotation for detecting speech but an automated temporal method

743 for assigning speech as child-directed or not, in a way that could lead to over-estimation (because

744 any speech by e.g. a female adult that was not temporally close to speech by others would count as

745 child-directed). A final answer to the question of how much child-directed speech is afforded to

746 Yélî and Tsimane' children must await fully comparable methods.

747 That said, Cristia et al. (2020) also pointed out another characteristic of the Tsimane' culture,

748 and this was the relatively low prevalence of literacy, and generally the variable access to formal

749 education. This is a very different case from the Yélî population studied here, where all adults have

750 accumulated several years of schooling, and literacy is widespread. If this second hypothesis holds,

751 then this may mean that there are phonetic effects of learning to read in the input afforded to young

752 children, and that this has consequences for young children's encoding and decoding of sounds in

753 the context of NWR tasks. Notice that this is not the same as the oft-recorded effect of learning to

754 read affecting NWR performance, illustrated for instance in the data for Sotho in 5. These two data

755 points have been gathered from two groups of children, all exposed mainly to Sotho, but children

756 with higher NWR had been learning to read in Sotho, whereas those with lower scores were

757 learning to read in English. What is at stake in the second interpretation of the lower scores

758 observed among the Tsimane' is related to literacy in the broader population (rather than in the

759 tested children themselves).

760   Although exciting, this hypothesis is only one of many. Another plausible explanation is that

761 the Tsimane' results are not comparable to the previous body of literature, and specifically to our

762 study. Cristia et al. (2020) administered the NWR in the form of a group game played outside,

763 with a non-native experimenter providing the target, and each person of the group attempting it in

764 their stead. This immediately means a number of important methodological differences with the

765 standard implementation of NWR, where children are tested individually, they hear items spoken

766 by a native speaker (often over headphones), the experimenter tends to belong to the same

767 community as the children, and testing occurs in quiet conditions (with little background noise).

768 Thus, a priority is for additional data gathered using this more novel testing paradigm in other

769 populations, or from the Tsimane' using the more traditional paradigm.

770   Broadening our discussion to all of the studies in our literature review, we notice that there is

771 rather wide variation of the range of NWR scores found across these samples, and that, in fact, the

772 strength of age effects also varies. We performed some exploratory analyses to see whether

773 features of the languages children were learning could be related to their overall NWR scores. We

774 extracted the number of phonemes in the language from PHOIBLE and coded whether words in the

775 language tended to be longer or shorter based on information in the papers or other sources.

776 Neither of these two predictors explained variance in 5. It is possible that average word length

777 plays a role, but often researchers incorporate this into their design by including longer items when

778 the native language allows this, with e.g. Sotho non-words having 4-7 syllables in length. To be

779 more certain whether language characteristics do account for meaningful variation in NWR scores,

780 it will be necessary to design NWR tasks that are cross-linguistically valid. We believe this will be

781 excedingly difficult (or perhaps impossible), since it would entail defining a 10-20 set of items that

782 are meaningless in all of the languages as well as phonotactically legal. An alternative may be to

find ways to regress out some of these effects, and thus compare languages while controlling for choices of phonemes, syllable structure, and overall length of the NWR items. As for different strengths of age effects, here as well we are uncertain to what they may be due, but we do hope that these intriguing observations will lead others to collect and share NWR data.

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 potential association between individual learners' NWR and cross-linguistic phone frequency.

814   References

815   Balladares, J., Marshall, C., & Griffiths, Y. (2016). Socio-economic status affects sentence

816   repetition, but not non-word repetition, in Chilean preschoolers. First Language, 36(3),

817   338–351. https://doi.org/10.1177/0142723715626067

818   Barclay, K. J. (2015). A within-family analysis of birth order and intelligence using population

819   conscription data on swedish men. Intelligence, 49, 134–143.

820   Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer (Version 6.1.35).

821   Retrieved from http://www.praat.org/

822   Bowey, J. A. (2001). Nonword repetition and young children's receptive vocabulary: A

823   longitudinal study. Applied Psycholinguistics, 22(3), 441–469.

824   Brandeker, M., & Thordardottir, E. (2015). Language exposure in bilingual toddlers: Performance

825   on nonword repetition and lexical tasks. American Journal of Speech-Language Pathology,

826   24(2), 126–138.

827 Brown, P. (2011). The cultural organization of attention. In A. Duranti, E. Ochs, & and Bambi B

828   Schieffelin (Eds.), Handbook of Language Socialization (pp. 29–55). Malden, MA:

829   Wiley-Blackwell.

830 Brown, P. (2014). The interactional context of language learning in Tzeltal. In I. Arnon, M.

831   Casillas, C. Kurumada, & B. Estigarribia (Eds.), Language in interaction: Studies in honor

832   of Eve V. Clark (pp. 51–82). Amsterdam, NL: John Benjamins.

833 Brown, P., & Casillas, M. (n.d.). Childrearing through social interaction on Rossel Island, PNG. In

834   A. J. Fentiman & M. Goody (Eds.), Esther Goody revisited: Exploring the legacy of an

835   original inter-disciplinarian (pp. XX–XX). New York, NY: Berghahn.

836 Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Papuan

837   community. Journal of Child Language, XX, XX–XX.

838 Castro-Caldas, A., Petersson, K. M., Reis, A., Stone-Elander, S., & Ingvar, M. (1998). The

839   illiterate brain. Learning to read and write during childhood influences the functional

840   organization of the adult brain. Brain: A Journal of Neurology, 121(6), 1053–1063.

841   https://doi.org/10.1093/brain/121.6.1053

842 Chiat, S., & Roy, P. (2007). The preschool repetition test: An evaluation of performance in

843   typically developing and clinically referred children. Journal of Speech, Language, and

844   Hearing Research, 50(2), 429–443.

845 COST Action. (2009). Language impairment in a multilingual society: Linguistic patterns and the

846   road to assessment. Brussels: COST Office. Available Online at: Http://Www.bi-Sli.org.

847 Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-directed speech is infrequent in a

forager-farmer population. Child Development, 90(3), 759–773.

https://doi.org/10.1111/cdev.12974

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

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition, 47, 421–436.

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.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. Journal of Child Language, 44(3), 677–694.

Gallagher, G. (2014). An identity bias in phonotactics: Evidence from Cochabamba Quechua.

871   Laboratory Phonology, 5(3), 337–378. https://doi.org/10.1515/lp-2014-0012

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

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

878 Grätz, M. (2018). Competition in the family: Inequality between siblings and the intergenerational
879   transmission of educational advantage. Sociological Science, 5, 246–269.

880 Havron, N., Ramus, F., Heude, B., Forhan, A., Cristia, A., Peyre, H., & Group, E. M.-C. C. S.
881   (2019). The effect of older siblings on language development as a function of age
882   difference and sex. Psychological Science, 30(9), 1333–1343.

883 Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged
884   6–12. Journal of Phonetics, 28(4), 377–396.

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

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

890 Kapalková, S., Polišenská, K., & Vicenová, Z. (2013). Non-word repetition performance in
891   Slovak-speaking children with and without SLI: novel scoring methods. International
892   Journal of Language and Communication Disorders, 48(1), 78–89.
893   https://doi.org/10.1111/j.1460-6984.2012.00189.x

Kolinsky, R., Leite, I., Carvalho, C., Franco, A., & Morais, J. (2018). Completely illiterate adults can learn to decode in 3 months. Reading and Writing, 31(3), 649–677. https://doi.org/10.1007/s11145-017-9804-7

Lancy, D. F. (2015). The anthropology of childhood. Cambridge, UK: Cambridge University Press.

Lei, L., Pan, J., Liu, H., McBride-Chang, C., Li, H., Zhang, Y., … others. (2011). Developmental trajectories of reading development and impairment from ages 3 to 8 years in chinese children. Journal of Child Psychology and Psychiatry, 52(2), 212–220.

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. (2009). Phonology, naturalness and universals. Poznán Studies in Contemporary Linguistics, 45(1), 131–140.

Maddieson, I. (2013a). Consonant inventories. The World Atlas of Language Structures Online. Retrieved from https://wals.info/chapter/1

Maddieson, I. (2013b). Vowel quality inventories. The World Atlas of Language Structures Online. Retrieved from https://wals.info/chapter/2

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/

Moreton, E., & Pater, J. (2012). Structure and substance in artificial-phonology learning, part ii: Substance. Language and Linguistics Compass, 6(11), 702–718.

Mulder, H., Verhagen, J., Van der Ven, S. H., Slot, P. L., & Leseman, P. P. (2017). Early executive function at age two predicts emergent mathematics and literacy at age five. Frontiers in Psychology, 8, 1706.

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.

Polišenská, K., & Kapalková, S. (2014). Improving child compliance on a computer-administered nonword repetition task. Journal of Speech, Language and Hearing Research, 57(3).

Radeborg, K., Barthelom, E., SjöBerg, M., & Sahlén, B. (2006). A Swedish non-word repetition test for preschool children. Scandinavian Journal of Psychology, 47(3), 187–192. https://doi.org/10.1111/j.1467-9450.2006.00506.x

Scaff, C., Stieglitz, J., Casillas, M., & Cristia, A. (2021). Daylong audio recordings of young children in a forager-farmer society show low levels of verbal input with minimal age-related changes. Draft.

Stokes, S. F., Wong, A. M., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: The case of cantonese. Journal of Speech, Language, and Hearing Research, 49(2), 219–236.

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

Table 1

NWR stimuli in orthographic (Orth.) and phonological (Phon.) representations.

Practice		Monosyll		Bisyll		Trisyll		Tetrasyll	
Orth.	Phon.	Orth.	Phon.	Orth.	Phon.	Orth.	Phon.	Orth.	Phon.
nopimade	nɔpimæțɛ	dp:a	țp̃æ	kamo	kæmɔ	dimope	țimɔpɛ	dipońate	țipɔnætɛ
poni	pɔni	dpa	țpæ	kańi	kæni	diyeto	țijɛtɔ	ńomiwake	nɔmiwækɛ
wî	wɯɯ	dpâ	țpɑ	kipo	kipɔ	meyadi	mɛjæți	todiwuma	tɔțiwumæ
		dpê	țpə	ńoki	nɔki	mituye	mitujɛ	wadikeńo	wæțikenɔ
		dpéé	țpeː	ńomi	nɔmi	ńademo	nætɛmɔ		
		dpi	țpi	piwa	piwæ	ńayeki	næjɛki		
		dpu	țpu	towi	tɔwi	ńuyedi	nujɛți		
		gh:ââ	ɣ̃ɑː	tupa	tupæ	pedumi	pɛțumi		
		ghuu	ɣuː			tiwuńe	tiwunɛ		
		kp:ââ	kp̃ɑː			tumowe	tumɔwɛ		
		kpu	kpu			widońe	wițɔnɛ		
		lv:ê	lβ̃ə			wumipo	wumipɔ		
		lva	lβʲæ						
		lvi	lβʲi						
		t:êê	t̃əː						
		tpê	tpə						

Table 2

Number (and percent) of vowel targets that were correctly repeated (Corr.), deleted (Del.), or substituted, as a function of vowel type, and whether the error resulted in a nasality change (Nasal Err.) or only a quality change (Qual. Err.)

	Corr.	Del.	Nasal Err.	Qual. Err	% Corr.	% Del.	% Nasal Err.	% Qual Err.
Nasal Target	100	0	39	17	64.1	0.0	25.0	10.9
Oral Target	1992	12	52	205	88.1	0.5	2.3	9.1

Table 3

Number (and percent) of consonant targets that were correctly repeated (Corr.), deleted (Del.), or substituted, as a function of the complexity of the consonant, and whether the error resulted in a change of complexity (Cmpl Err.) or not (Othr Err.)

	Corr.	Del.	Cmpl Err.	Othr Err.	% Corr.	% Del	% Cmpl Err.	% Othr Err.
Complex Target	257	0	218	48	49.1	0.0	41.7	9.2
Simple Target	1425	6	2	120	91.8	0.4	0.1	7.7

Table 4

NWR means (and standard deviations) measured in whole-word scores and normalized Levenshtein Distance (NLD), separately for the four stimuli lengths.

	Word	NLD
1 syll	48 (22)	40 (18)
2 syll	79 (22)	8 (9)
3 syll	78 (19)	7 (7)
4 syll	74 (32)	9 (12)