¹ Non-word repetition in Yélî Dnye

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

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

⁴ cognitives, ENS, EHESS, CNRS, PSL University

⁵ [2] Max Planck Institute for Psycholinguistics

⁶ Author Note

¹¹ Correspondence concerning this article should be addressed to Alejandrina Cristia, 29,

¹² rue d'Ulm, 75005 Paris, France. E-mail: alecristia@gmail.com

13 Abstract

14 ADD

15 *Keywords:* phonology, non-word repetition, development

16 Word count: xxx words

TODO

- add table r/d by stim length & age

- double check demo: exact age missing: 78, 82, 83, 115, 119; rough age missing 78, there's a mixture of BGMF in sex, first born is missing, round age

- discussion

TODO later

- ask Middy about consent, correct, accuracy

- ask Middy to look through

- probably turn summary of types of error into a table; use proportion of correctly produced rather than number; use first attempts rather than the whole data set

- maybe add to the last ana the prevalence (ie out of all the gh which proportion get transformed)

Introduction

Although infants begin to learn about their native language's phonology within the first year, many studies suggest that in perception and production, in phonetics and phonology, their knowledge continues to develop throughout childhood (e.g., Hazan & Barrett, 2000). One common task in this line of research is nonword repetition (NWR). In NWR studies, participants are presented auditorily with an item that is phonologically legal but lexically meaningless in the language children are learning. The child should immediately try to say it back without changing anything. Accuracy is thought to reflect long-term phonological knowledge (which allows the child to perceive the item accurately

even though it is not a real word they have encountered before) as well as online phonological working memory (to encode the item in the interval between hearing it and saying it back) and flexible productiong patterns (to produce the item accurately even though it had never been pronounced before).

NWR has been used to seek answers to a variety of theoretical questions, including what are the links between phonology, working memory, and the lexicon (Bowey, 2001), as well as for applications, notably using nwr as a diagnostic for language delays and disorders (Estes, Evans, & Else-Quest, 2007). Since non-words can be generated in many languages, it has been used acros a wide array of languages, particularly in Europe (Meir, Walters, & Armon-Lotem, 2016,@is08042009language).

In this study, we report on NWR results among children learning Yêly Dnyé, an isolate spoken in Rossel Island, PNG, with an unusually dense phonological inventory. In our implementation of NWR, we made sure that some of the items contained typologically rare and/or challenging sounds. MISSING: WHY DO WE DO THIS? WHAT IS THE POINT?

Intro to the language ((???)) - please feel free to throw away anything that is not useful!

- complexity in the vowel system
- complexity in the consonant system
- word shapes
- typical word length
- although not the focus of this paper, high use of suppletion in verbal paradigms, other features of language, see Levinson XXX for details

Intro to the people ((???)) - please feel free to throw away anything that is not useful! Little is known about language development in children growing up in Rossel Island, a community of primarily subsistence farmers who tend to reside in

⁶⁴ close-knitted villages where child care is distributed across many individuals, and who

⁶⁵ typically speak Yélî Dnyé, a phonologically and lexically complex language.

⁶⁶ • usually monolingual at home

⁶⁷ • schooling in English but it starts at age XX, so not relevant here

⁶⁸ • however, some use of English due to immigrants & children of immigrants

⁶⁹ • children spend a lot of time with other children

⁷⁰ • most parents are subsistence farmers

⁷¹ • parental education generally varies between XX and YY

⁷² **Brief review of NWR for our purposes.**   There is some variation in the

⁷³ presentation procedure and structure of items found in previous NWR work. For example,

⁷⁴ items are often presented orally by the experimenter (Torrington Eaton, Newman, Ratner, &

⁷⁵ Rowe, 2015), although an increasing number of studies have turned to playing back the

⁷⁶ stimuli in order to have greater control of the stability of the presentation (Brandeker &

⁷⁷ Thordardottir, 2015). Additionally, while some studies have used 10-15 non-words, others

⁷⁸ have employed up to 46 unique items (Piazzalunga, Previtali, Pozzoli, Scarponi, & Schindler,

⁷⁹ 2019). Often, authors modulate structural complexity, typically measured in terms of item

⁸⁰ length (measured in number of syllables) and/or syllable structure (open as opposed to

⁸¹ closed syllables, Gallon, Harris, & Van der Lely, 2007).

⁸² Previous work seems to avoid difficult sounds, but we felt this was important to

⁸³ represent Yélî Dnye, so we also varied this factor. We designed a relatively large number of

⁸⁴ items but, aware that this may render the task longer and more tiresome, we split some of

⁸⁵ the items across children. This allowed us to get information about repetition accuracy of

⁸⁶ more items.

⁸⁷ Naturally, designing the task in this way may render the study of individual variaiton

⁸⁸ within the population more difficult because different children are exposed to different items.

89 However, a review of previous work on individual variation suggested to us that many

90 individual differences effects are relatively small, and would not be detectable with the

91 sample size that we could collect in a given visit.

92   That said, we contribute to the literature by also reporting descriptive analyses of

93 individual variation that could potentially be integrated in meta-analytic efforts. Based on

94 previous work, we looked at potential improvements with age (Farmani et al., 2018; Kalnak,

95 Peyrard-Janvid, Forssberg, & Sahlén, 2014; Vance, Stackhouse, & Wells, 2005), and potential

96 negative effects of bilingual exposure (Brandeker & Thordardottir, 2015; Meir &

97 Armon-Lotem, 2017; Meir et al., 2016). Previous work typically finds no significant

98 differences as a function of maternal education (e.g., Farmani et al., 2018; Kalnak et al.,

99 2014; Meir & Armon-Lotem, 2017) or child gender (Chiat & Roy, 2007). Although previous

100 research has not often investigated potential effects of birth order on NWR, there is a sizable

101 literature on these effects in other language tasks (Havron et al., 2019), and therefore we

102 report on those too.

103   **Research questions.**   We sought to address the following questions:

104   • What is the overall repetition score, according to a variety of relevant metrics (whole

105   word, phoneme based, distance)?

106   • How does score change as a function of item complexity (number of syllablex, sound

107   complexity)?

108   • How frequent are errors that result in real words? Is that a function of item

109   complexity?

110   • What are some common patterns of phonological errors?

111   • Is variation attributable to child age, sex, birth order, monolingual status, and/or

112   parental education?

113   In view of the hypothesis-driven nature of this work, 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 do not have enough power to look at all relationships of interest, and often to detect any of the known effects. For instance, an effect of stimulus length and age has been reported as significant sometimes, but Tables XX and XX show that even these effects can be quite small (data from Cristia, Farabolini, Scaff, Havron, & Stieglitz, 2020). **add table with r/ds from previous work**) Moreover, by virtue of this being a language-specific study, it is unclear that our method (including saliently the items we used) is comparable in precision to previous NWR studies. Therefore, all analyses here are descriptive and should be considered exploratory.
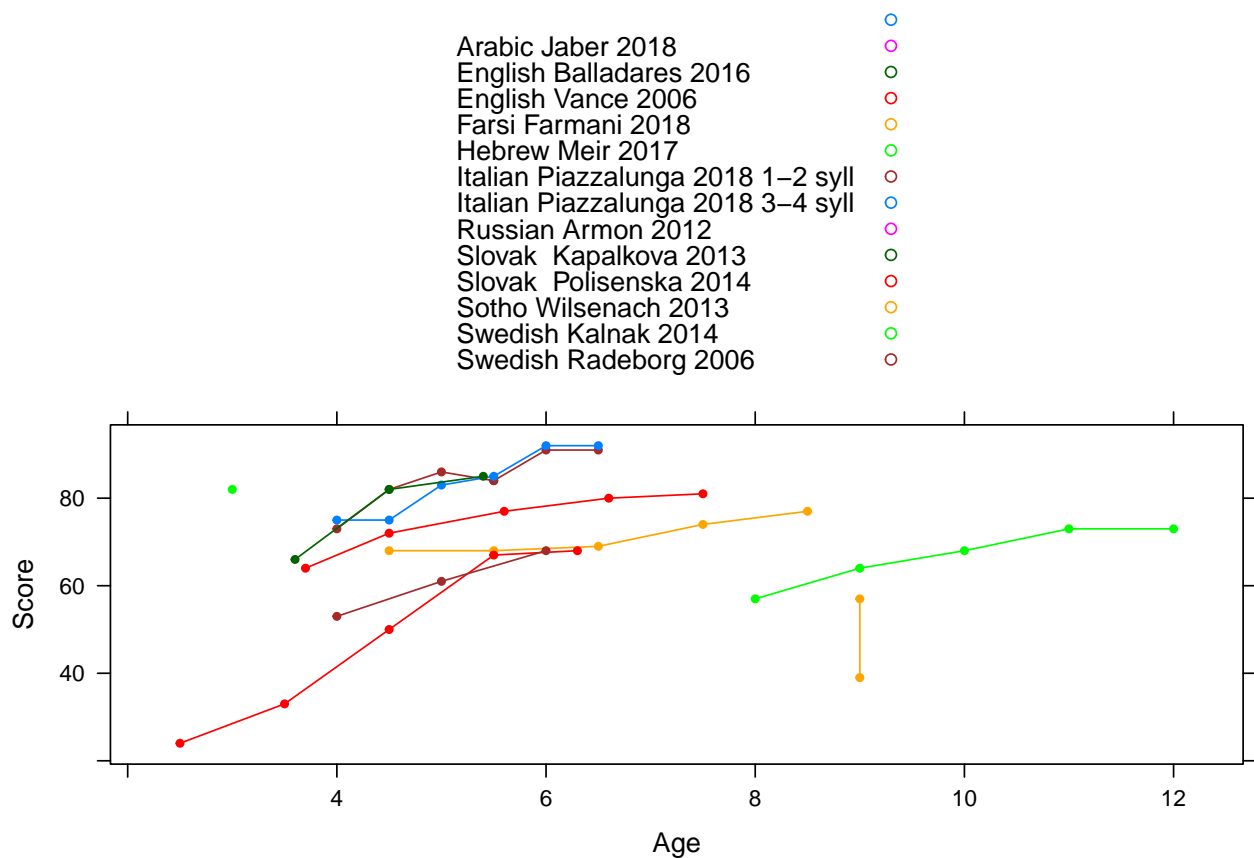
Figure 1. Proportion of non-words correctly repeated as a function of age (in years), study (first author and year regardless of the number of authors), and, when available, shorter versus longer non-words.

Methods

¹²⁵   **Stimuli.**   Many NWR studies are based on a fixed list of 12-16 items that vary in

¹²⁶ length between 1 and 4 syllables, often additionally varying syllable complexity and/or

¹²⁷ cluster presence and complexity, always meeting the condition that they do not mean

¹²⁸ anything in the target language (e.g., Balladares, Marshall, & Griffiths, 2016; Wilsenach,

¹²⁹ 2013). We kept the same variation in item length and the non-meaningfulness requirement,

¹³⁰ but we did not vary syllable complexity and clusters because these are vanishingly rare in

¹³¹ Yély Dnye. We also increased the number of items an individual child would be tested on, so

¹³² that a child would get up to 23 items to repeat (note that up to other work has also used

¹³³ 24-30 items: Jaber-Awida, 2018; Kalnak et al., 2014), and we created more items and

¹³⁴ distributed them across children, so as to increase the coverage, and be able to study more

¹³⁵ items.

¹³⁶   A first list of candidate items was generated in 2018 by selecting simple consonants

¹³⁷ ("p", "t", "d", "k", "m", "n", "w", "y") and vowels ("i", "o","u", "a","e") that were

¹³⁸ combined into consonant-vowel syllables, further sampling the space of 1- to 4-syllable

¹³⁹ sequences. These candidates were automatically checked against Levinson's 2015 dictionary

¹⁴⁰ and removed from consideration if they appeared in the dictionary. The second author

¹⁴¹ presented them orally to three local research assistants, who were asked to repeat them and

¹⁴² further say whether they were real words. Any item for which two or more of the assistants

¹⁴³ reported them having a meaning or some form of association was excluded.

¹⁴⁴   A second list of candidate items was generated in 2019 by selecting complex

¹⁴⁵ consonants and systematically crossing them with all the vowels in the Yélî inventory to

¹⁴⁶ produce consonant-vowel monosyllables. As before, items were automatically excluded if

¹⁴⁷ they appeared in the dictionary. Additionally, since hearing vowel length in monosyllables in

¹⁴⁸ isolation is challenging, any item that had a short/long real word neighbor was filtered out.

Since the phonology and phonetics of Yélî is still in the process of being described (CITE **???** please fill in), there could have been undocumented constraints that rendered items illegal. 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 presented to one informant, for a final check that they did not mean anything. Together with the 2018 selection, they were recorded using a headset **XXX (???) please fill in** and an Olympus **XXX (???) please fill in** from the written form presented together with the same item orally (by the second author). The complete recorded list was finally presented to two more informants, who could repeat all the items and who confirmed there were no real words. Even so, there was one monosyllable that was often identified as a real word (intended "yî" /yXX/; identified as "yi" /yi/, *tree*). This item is removed from analyses below.

(???) can you please add the IPA in // for all of the items in the excel then save as tsv? The final list is composed of 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).

##		Type	Orthographic	Phonological	Type	Orthographic	Phonological	Type
##	28	practice	nopimade	NA	1	dp:a	NA	2
##	32	practice	poni	NA	1	dpa	NA	2
##	43	practice	wî	NA	1	dpâ	NA	2
##	1	<NA>	<NA>	NA	1	dpê	NA	2
##	2	<NA>	<NA>	NA	1	dpéé	NA	2
##	3	<NA>	<NA>	NA	1	dpi	NA	2
##	4	<NA>	<NA>	NA	1	dpu	NA	2
##	5	<NA>	<NA>	NA	1	gh:ââ	NA	2
##	6	<NA>	<NA>	NA	1	ghuu	NA	<NA>

```
## 7     <NA>        <NA>           NA    1      kp:ââ         NA <NA>

## 8     <NA>        <NA>           NA    1       kpu          NA <NA>

## 9     <NA>        <NA>           NA    1      lv:ê          NA <NA>

## 10    <NA>        <NA>           NA    1       lva          NA <NA>

## 11    <NA>        <NA>           NA    1       lvi          NA <NA>

## 12    <NA>        <NA>           NA    1      t:êê          NA <NA>

## 13    <NA>        <NA>           NA    1      tp:a          NA <NA>

## 14    <NA>        <NA>           NA    1       tpâ          NA <NA>

## 15    <NA>        <NA>           NA    1       tpê          NA <NA>

##     Orthographic Phonological Type Orthographic Phonological Type Orthographic

## 28      kamo          NA    3      dimope        NA    4      diponate

## 32      kani          NA    3      diyeto        NA    4      nomiwake

## 43      kipo          NA    3      meyadi        NA    4      todiwuma

## 1       noki          NA    3      mituye        NA    4      wadikeno

## 2       nomi          NA    3      nademo        NA <NA>        <NA>

## 3       piwa          NA    3      nayeki        NA <NA>        <NA>

## 4       towi          NA    3      nuyedi        NA <NA>        <NA>

## 5       tupa          NA    3      pedumi        NA <NA>        <NA>

## 6       <NA>          NA    3      tiwune        NA <NA>        <NA>

## 7       <NA>          NA    3      tumowe        NA <NA>        <NA>

## 8       <NA>          NA    3      widone        NA <NA>        <NA>

## 9       <NA>          NA    3      wumipo        NA <NA>        <NA>

## 10      <NA>          NA <NA>        <NA>        NA <NA>        <NA>

## 11      <NA>          NA <NA>        <NA>        NA <NA>        <NA>

## 12      <NA>          NA <NA>        <NA>        NA <NA>        <NA>

## 13      <NA>          NA <NA>        <NA>        NA <NA>        <NA>

## 14      <NA>          NA <NA>        <NA>        NA <NA>        <NA>
```

```
## 15          <NA>        NA <NA>        <NA>        NA <NA>        <NA>
##    Phonological
## 28          NA
## 32          NA
## 43          NA
## 1           NA
## 2           NA
## 3           NA
## 4           NA
## 5           NA
## 6           NA
## 7           NA
## 8           NA
## 9           NA
## 10          NA
## 11          NA
## 12          NA
## 13          NA
## 14          NA
## 15          NA
```

A Praat script 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 https://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

Table 1

NWR stimuli in orthographic

and phonological

representations.

Orthographic	Phonological
dp:a	NA
dpa	NA
dpâ	NA
dpê	NA
dpéé	NA
dpi	NA
dpu	NA
gh:ââ	NA
ghuu	NA
kp:ââ	NA
kpu	NA
lv:ê	NA
lva	NA
lvi	NA
t:êê	NA
tp:a	NA
tpâ	NA
tpê	NA
kamo	NA
kani	NA
kipo	NA
noki	NA
nomi	NA
piwa	NA

228   in 2019 (i.e., all the monosyllables starting with tp were split into 2 sublists). Since

229   some of these groups had an odd number of items, one of the sublists was slightly

230   longer than the other (20 versus 23).

231 • once the sublist split had been done, items were randomized such that all children

232   heard first the 3 practice items in a fixed order (1, 2, and 4 syllables), a randomized

233   version of their sublist selection of difficult onset items, and randomized versions of

234   their 2-syllables, then 3-syllables, and finally 4-syllable items.

235   **Procedure.**   We tried to balance three desiderata: That children would not be

236 unduly exposed to the items before they themselves had to repeat them; that children would

237 feel comfortable doing this task with us; and that the community would feel safe with us

238 doing this task with their children. Moreover, there were also some logistic constraints in

239 terms of the space availability. As a result, the places where elicitation took place varied

240 across the hamlets.

241   We visited four different hamlets once, and attempted to test all eligible children

242 present at the time, to prevent the items "spreading" through hearsay. In the first hamlet,

243 we tested children in five different places, with some children being tested inside a house and

244 others tested on the veranda. The complete list of places and the ways in which they met

245 the desiderata mentioned above can be found in the raw data, available from online

246 supplementary materials.

247   The child was donned a headset (**xx (???) please fill in** for most of the children,

248 SHURE WH20 XLR headset with a dynamic microphone for the rest), recorded into the left

249 channel into a Tascam DR40x digital recorder. For most children, the headset could not stay

250 comfortably on the child's head, and thus it was placed on the child's shoulders, with the

251 microphone carefully placed close to the child's mouth. A local informant sat next to the

252 child, to would provide the instructions and, if needed, coach the child to make sure, using

253 the three practice items as well as real words, that they understood that the task was to

²⁵⁴ repeat the items precisely without changing anything. An experimenter (the first author)

²⁵⁵ delivered the elicitation stimuli to the local informant and the child over headphones.

²⁵⁶   The first phase was making sure the child understood the task. This was explained

²⁵⁷ orally and the first training item was presented. Often, children froze and did not say

²⁵⁸ anything. If this happened, then we followed this procedure. First, the informant insisted. If

²⁵⁹ the child still did not say anything, the informant asked the child to repeat a real word, and

²⁶⁰ another, and another. If the child could repeat these correctly, then we provided the recorded

²⁶¹ training item over headphones again. Most children successfully started repeating the items

²⁶² presented over headphones at this point; a few further needed the local informant to model

²⁶³ the behavior (i.e., they would hear the item again, and she would say it; then we would play

²⁶⁴ it again, and ask the child to say it). A small minority still failed to repeat the item after

²⁶⁵ hearing it over headphones. If that occurred, we tried with the second training item, at which

²⁶⁶ point some children got it and could continue. A small minority, however, failed to repeat

²⁶⁷ this one, as well as the third training item, in which case we stopped the test altogether.

²⁶⁸   NWR studies vary in whether children are provided with several opportunities to hear

²⁶⁹ and say the item. To have a fixed and clear procedure, we decided that items other than the

²⁷⁰ inital three training ones would not be repeated unless the child made an attempt to

²⁷¹ produce them. If this attempt was judged correct by the local informant, then the

²⁷² experimenter would move on to the next item (whispering this over a separate headset that

²⁷³ was recording onto the right track of the same Tascam). If the local informant heard a

²⁷⁴ deviation, she indicated to the experimenter that the item needed to be repeated, and up to

²⁷⁵ 5 attempts were allowed.

²⁷⁶   Whenever siblings from the same family were tested, an attempt was made to test first

²⁷⁷ the older and then the younger child, and always on different elicitaiton sets.

278   **Coding.**   A script was written to randomize all tokens from all children, pairing each

279   with the auditory target the child had been provided with. A native research assistant then

280   listened through all productions of a given target (randomized across children and repetition

281   order), and made a judgment of whether the item was correctly or incorrectly repeated. She

282   additionally transcribed exactly what the child said, providing some examples of the types of

283   errors children in general make (without making specific reference to Yélî sounds or the

284   items in the elicitation sets).

285   **Analyses.**   Some NWR studies employ phoneme-based scoring in addition to or

286   instead of word-level accuracy (e.g., Cristia et al., 2020). We scored items in terms of the

287   number of phonemes that could be aligned across the target and attempt, divided by the

288   number of phonemes of whichever item was longer (the target or the attempt). Although

289   previous work does not use distance metrics, we additionally report those.

290   Finally, for describing children's patterns of errors, all repetitions of a given target were

291   taken into account. We describe the proportion of items where the change resulted in a real

292   word (semantic errors); and classify the most common phonological errors.

293   **Participants.**   This study was approved as part of a larger research effort by **(???)**

294   **add info from Middy's FAIR reply**. Participation was voluntary, with children being

295   invited to come and participate. Regardless of how they performed, children were provided

296   with a snack as compensation. Children who came up to participate but then refused were

297   nonetheless provided with the snack.

298   A total 55 children were tested, from 34 of families, in five hamlets. Some children

299   could not be included for the following reasons: refused participation or failed to repeat

300   items presented over headphones even after coaching (N=0), spoke too softly to allow offline

301   coding (N=5). In addition, 2 teenagers were tested to put younger children at ease; their

302   data is not included in analyses below. The remaining 40 children (14 girls) were aged 6.96

303 years (range 3.92-11.03 years). There were 32 children exposed only to Yélî in the home, 6

304 children who were also exposed to another language in the home, and , 2 for whom this

305 information was missing. Maternal years of education averaged 8.24 years (range 6-12 years;

306 2 children had this information missing).[1] In terms of birth order, 0 were first borns, 5

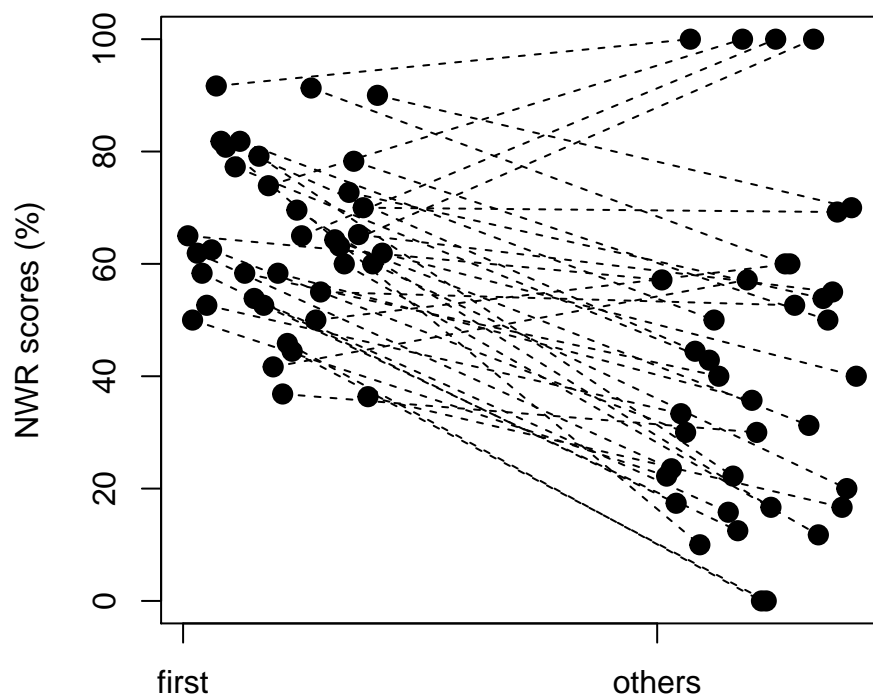307 second, 4 third, 2 forth, 6 fifth, 5 sixth, and 1 did not have this information.

Results

Figure 2. (#fig:Fig1-first_vs_others)NWR scores for individual participants averaging separately their first attempts and all other attempts.

309 **Preliminary analyses.** We first checked whether accuracy varies between first and

310 subsequent presentations of an item by averaging word-level accuracy at the participant level

311 separately for first attempts and subsequent repetitions. We excluded 5 children who did not

312 have data for one of these two types. As shown in Figure 1, participants' mean word-level

313 accuracy became more heterogeneous in subsequent repetitions. Surprisingly, subsequent

---

[1]Education is often reported in even years because people typically complete two-year cycles.

314   repetitions (M = 42, SD = 28) were on average less accurate than first ones (M = 64, SD =

315   15), t(36) = 5.32, p = 0. Given the uncertainty in whether previous work used only the first

316   or all repetitions, and since behavior degraded and became more heterogeneous in

317   subsequent repetitions, the rest of the analyses focus on only the first repetitions.

318   **Overall repetition accuracy.**   Taking into account only the first attempts, we

319   averaged attempts by each of the 24 children who had data for first attempts; their ages

320   ranged from 3.92 to 10.20 (M = 6.50, SD = 1.50).

321   The overall NWR score was M = 60% (SD = 14%). Scores based on phonemes are

322   even higher M = 76% (SD = 9%). The phoneme-based normalized Levenshtein distance was

323   M = 24% (SD = 9%), meaning that about a fifth of phonemes were were substituted,

324   inserted, or deleted. Notice that the normalized Levenshtein distance is the complement of

325   phoneme-based scores.

326   **Accuracy a function of item length.**   We then inspected whether accuracy varied

327   as a function of word length. Results are shown on table XX. This table shows that

328   monosyllables accuracy was much lower than other lengths. This is likely due to the fact that

329   the majority of monosyllables included were chosen because they had sounds that are rare in

330   the world's languages, which may indicate that they are hard to produce or to perceive.

331   Setting monosyllables aside, we observe the typical pattern of decreased accuracy for

332   longer items, although this is particularly salient for the whole word scoring. This is the

333   most commonly reported type of score, but it is also the least forgiving. The pattern is less

334   marked when other two scores are used, which are less sensitive to errors.

335   **Patterns of errors.**

336   • How frequent are errors that result in real words? Is that a function of item

337   complexity? What are some common patterns of phonological errors?

Table 2

NWR scores measured in

whole word accuracy,

phoneme-based accuracy, and

normalized Levenshtein

Distance, separately for the

four stimuli lengths.

Word	Phoneme	NLD
48 (21)	60 (16)	40 (16)
80 (24)	93 (10)	7 (10)
78 (21)	94 (7)	6 (7)
66 (31)	83 (19)	17 (19)

338   We investigated patterns of deletion and substitution. Deletions were relatively rare,

339   with only 28 vowels deleted, and 8 consonants.

340   As for substitutions, it was as common for a nasal vowel to be produced as an oral

341   vowel as vice versa (54 oral target vowels produced as nasal vowels, 58 nasal target vowels

342   produced as oral vowels). Substitutions in which the oral nature was preserved but the

343   quality of the vowel was changed were a great deal more common than changes in quality

344   among nasal vowels (202 oral vowels produced with a different quality; 23 nasal vowels

345   produced with a different quality). As for consonants, asymmetries were very marked with

346   more complex consonants (specifically dptpkpkmknmbgh) mispronounced as simple ones

347   (specifically mnlwyvdgptkfhch, 251) than vice versa (2). Simple consonants were

348   mispronounced as other simple consonants quite frequently (135 simple consonants

349   mispronounced as other simple ones, compared to 58 complex ones).

350   **probably turn this into a table; use proportion of correctly produced**

351 **rather than number; use first attempts rather than the whole data set**

352   Finally, we looked into the frequency with which mispronunciations resulted in real

353 words. Nearly all of them were: 96%.

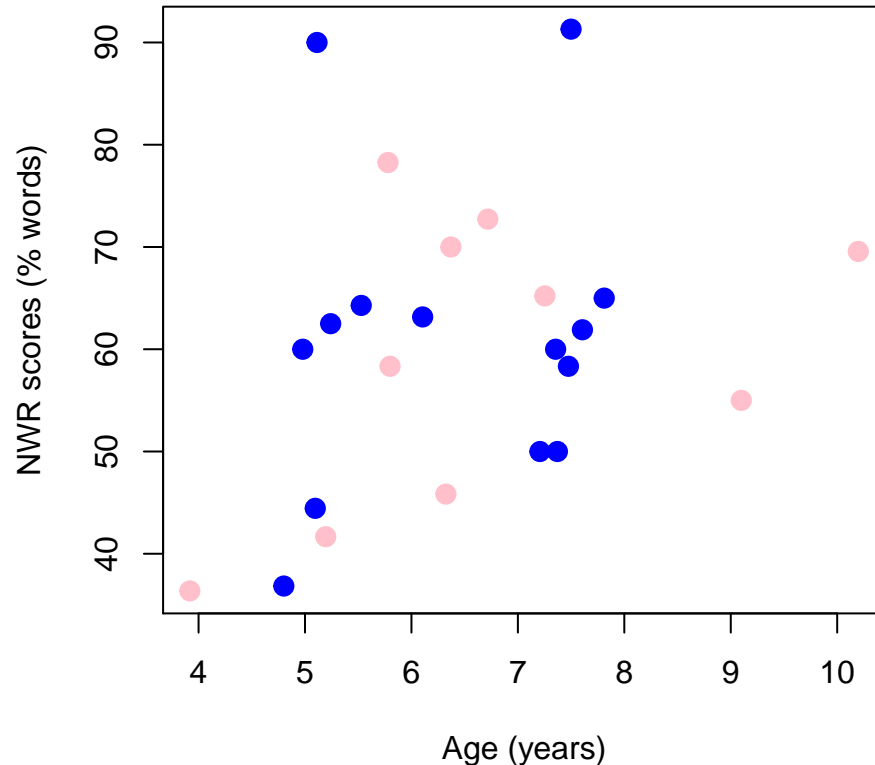354   ** this is higher than any previous report! can we be sure of it?**

Figure 3. (#fig:Fig2-scores by age)NWR whole-word scores for individual participants as a
function of age and sex (blue = boys, pink = girls).

355   **Factor structuring individual variation.**   Our final exploratory analysis assessed

356 whether variation in scores was structured by factors that vary across individuals. As shown

357 in Figure 2, there was a greater deal of variance in earlier than later ages, with significantly

358 higher NWR scores for older children (Spearman's rank correlation, given inequality of

359 variance, rho (1,535.50) = 0.33, p = 0.11). In contrast, there was no clear association

360 between NWR scores, on the one hand, and sex (t (0.33) = 20.09, p = 0.75), birth order

361 (2,595.77) = -0.13, p = 0.55) or maternal education (2,038.91) = 0.11, p = 0.60)

Discussion

- What is the overall repetition accuracy (whole word, phoneme based, distance)?

- How does this change as a function of item complexity (number of syllables, sound complexity)?

1. Children are more accurate for mono-syllables than longer items

2. The length distribution in Yélî words is more balanced than that in English, and thus the performance decline for poly- versus mono-syllables may be less pronounced than that for English. **Check for work on European languages that may have looked into this**

3. Similarly, we do not know of NWR research that manipulates the difficulty of the sounds that are included in the items, but word naming and other research suggests that children are more accurate when producing easy and/or typologically common sounds than difficult and/or typologically rare sounds [CITE]. Therefore, we expect higher accuracy for items with common sounds than in those with rare sounds.

4. The Yélî sound inventory is very large and compressed, with many similar sounds that are acoustic and articulatory neighbors. Therefore, this may constitute a pressure for children to have finer auditory skills (and perhaps more precise articulations) than children speaking languages with a simpler inventory. As a result, differences between easier and harder items may be smaller in this work than in other research. **no work looking at consonants & vowels? no work looking at nasal vowels in particular?**

   (MC: but we can try and do a cursory analysis based on the corpora we have from Steve and my transcription of naturalistic interactions!)

- How frequent are errors that result in real words? Is that a function of item complexity?

- Is individual variation explainable by child age, sex, birth order, monolingual status, and/or parental education?

3. Children's accuracy increases with child age.

4. Non-monolingual Yélî children are less accurate than monolingual ones when tested on the society-dominant language (we did not test any non-dominant language)

5. As revious NWR evidence on this is mixed, but general findings on language development suggest that children whose mothers are more educated are more accurate than children whose mothers are less educated.

6. To our knowledge, there is no previous NWR work on this, but other research suggests that first-born children should outperform later-born children

7. Anecdotally Yélî children grow up in close-knitted communities and thus may receive significant portions of their language input from people not in their nuclear family (or at least from people other than their mothers, who tend to be the non-native speakers). If so, the difference between monolinguals and not monolinguals may be smaller than that found in other work . That said, one recent study on the same population shows that most child-directed input in the first 2 years does come from the mother , so in so far as this input has a crucial formational role, then there may still be a performance gap between these two groups.

8. In the Rossel community, formal education plays an extremely minor role in ensuring individual's success, is not a good index of relative socio-economic status, and furthermore there is only a narrow range of variation in maternal educational attainment. This may lead to no or only very small advantages for children whose

mothers are more educated, provided that the causal chain between maternal

education and child language is via SES more broadly. However, if education directly

boosts maternal verbal skills and the incidence of verbal behavior (as suggested by

CITE), then we should still see a difference along this factor.

9. One main causal path between birth order and language development is via parental

input (CITE). Given our arguments above for how mothers may not be as important

among Rossel people than in other places, then the performance gap between first

borns and later borns may be smaller.

Acknowledgments

<div align="center">

References

</div>

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

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.

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.

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'. *Preprint*.

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.

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.

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.

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.

IS0804, C. A. (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.*

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.

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.

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