

Non-word repetition in Yélî Dnye

Alejandrina Cristia¹ & Marisa Casillas²

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

² Max Planck Institute for Psycholinguistics

Author Note

All data are made available in a repository in the Open Science Framework. AC acknowledges the support of the Agence Nationale de la Recherche (ANR-17-CE28-0007 LangAge, ANR-16-DATA-0004, ANR-14-CE30-0003, ANR-17-EURE-0017); and the J. S. McDonnell Foundation Understanding Human Cognition Scholar Award.

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

Abstract

In nonword repetition (NWR) studies, participants are presented auditorily with an item that is phonologically legal but lexically meaningless in the local language. Accuracy is thought to reflect long-term phonological knowledge as well as online phonological working memory and flexible production patterns. 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. This study contributes to three lines of research. First, we document that non-word items containing typologically rare sounds are repeated accurately less often than non-words containing more common sounds. Second, we document rather weak effects of item length, contributing to other research suggesting that length effects may be language-specific. Third, we do not find strong individual variation effects in this population, contrary to previous results documenting strong age-related effects. Together, these data provide a unique view of online phonological processing in a seldom-studied language, and contribute to both typological and language acquisition research.

Keywords: phonology, non-word repetition, development

Word count: xxx words

28

Non-word repetition in Yélî Dnye

29 TODO Middy

- 30 • look through this manuscript for bits of text to be written by mc
- 31 • help with table 1 (formatting & making it appear, because it's broken by the phonological
- 32 representations)
- 33 • do you think Figure 2 should be done with phonetic symbols? If yes, could you try to
- 34 implement that?
- 35 • double check demo – probably start with the xlsx version that has the children's first names,
- 36 date of birth, and date of test. I looked again at my local files (still haven't checked the labs'
- 37 ones) and I'm surprised I have maternal education for more children than I have birth order
- 38 for... If you correct this, to generate the csv I also round the age, assign age.rounded to
- 39 reported age (in years) when DOB is missing, remove first names, DOB, DOTest and
- 40 age.precise before pushing to github
- 41 • double check my frequency entries in segments.xlsx using the search function in
- 42 <https://phoible.org/parameters>; take a look at <http://phoible.github.io/conventions/> in case
- 43 you see something about double articulation (I didn't find tp or lBj)
- 44 • double check phonological implementation of stimuli in stimuli.xlsx
- 45 • for both of the above, if you make any changes, they need to be re-saved as UTF16 txt file,
- 46 and then converted (see remember above)
- 47 • double check also stim_seg_freq – why are some items still with NA freq?
- 48 • consent, “correct”, “accuracy”, “performance”
- 49 • draft discussion

50 TODO later

- 51 • probably turn summary of types of error into a table; use proportion of correctly produced
- 52 rather than number; use first attempts rather than the whole data set DONE
- 53 • maybe add to the last ana the prevalence (ie out of all the gh which proportion get
- 54 transformed) DONE I think... (this item looks the same as one of the three above)
- 55 • maybe add analyses of score as a function of frequency (phones, diphones) in MC's corpus?
- 56 (mc suggested, but I don't think we need it...) - waiting until we talk about this

57 Introduction

58 Although infants begin to learn about their native language's phonology within the first year,
 59 many studies suggest that in perception and production, in phonetics and phonology, their
 60 knowledge continues to develop throughout childhood (e.g., Hazan & Barrett, 2000). One common
 61 task in this line of research is nonword repetition (NWR). In NWR studies, participants are
 62 presented auditorily with an item that is phonologically legal but lexically meaningless in the
 63 language children are learning. The child should immediately try to say it back without changing
 64 anything. Accuracy is thought to reflect long-term phonological knowledge (which allows the child
 65 to perceive the item accurately even though it is not a real word they have encountered before) as
 66 well as online phonological working memory (to encode the item in the interval between hearing it
 67 and saying it back) and flexible production patterns (to produce the item accurately even though it
 68 had never been pronounced before).

69 NWR has been used to seek answers to a variety of theoretical questions, including what the
 70 links between phonology, working memory, and the lexicon are (Bowey, 2001), and to assess the
 71 extent to which apparent phonological constraints found in the lexicon affect online production
 72 (Gallagher, 2014). It is also frequently used as for applications, notably as a diagnostic for
 73 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 (Action, 2009; Meir, Walters, & Armon-Lotem, 2016).

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. The present report uses these unique data to contribute to several lines of research. First, we made sure that some of the non-word items contained typologically rare and/or challenging sounds, so that we can contribute to the study of whether rare sounds are disadvantaged in perception and/or production, both in terms of overall accuracy of repetition and patterns of mispronunciation. Second, we look at the effects of item length, since previous NWR research has uncovered variable effects. Third, we contribute to the basic NWR literature by contributing an additional data point for children learning a rarely-studied language and culture, and further investigating whether there are structured sources that account for individual variation.

Yéli Dnye phonological overview. Yéli Dnye is an isolate language (presumed Papuan) spoken by approximately 7000 people residing on Rossel Island, which is found at the far end of the Louisiade Archipelago in Milne Bay Province, Papua New Guinea. The Yéli sound system, much like its baroque grammatical system (Levinson, n.d.), is unlike any other in the region.

With only four primary places of articulation (bilabial, alveolar, post-alveolar, and velar) and no voicing contrasts, the phonological inventory is remarkably packed with acoustically similar segments. The core oral stop set includes both singleton (/p/, /t/, /t̚/, and /k/) and doubly-articulated (/tp/, /t̚p/, /kp/) segments, with full nasal equivalents (/m/, /n/, /ɲ/, /ŋ/, /nm/, /ɲm/, /ŋm/), and with a substantial portion of these able to be contrastively pre-nasalized or nasally released (/mp/, /nt/, /ɲt̚/, /ŋk/, /nmtp̚/, /ɲmtp̚/, /ŋmkp̚/, /t̚ɲ/, /kɲ/, /tp̚ɲm̚/, /kp̚ɲm̚/). A large number of this combinatorial set can further be contrastively labialized or palatalized on release, or both (e.g., /pʲ/, /pʷ/, /pʲʷ/; /tp̚ʲ/; /ɲm̚ɸb̚ʲ/; see Levinson (n.d.) for details). The consonantal inventory also includes a number of non-nasal continuants (/w/, /j/, /ɣ/, /l/, /β̞ʲ/, /ʃ̞ʲ/, /lβ̞ʲ/). Vowels in Yéli 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 or long vowels, with 7 of these able to appear as short and long nasal vowels as well (/i/, /u/, /ə/, /ɛ/, /ɔ/, /æ/, /ɑ/).

In total, Yélî Dnye uses 90 distinctive segments (93 when including consonants that are extremely rarely used), far outstripping the phonemic inventory size of other documented Papuan languages (Levinson (n.d.); Foley 1986) and includes at least two contrasts not yet been documented elsewhere [labial-coronal double-articulations with dental vs. post-alveolar coronal placement in both oral and nasal stops, Maddieson & Levinson 2001; Ladefoged and Maddieson 1996].

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

While all consonants can appear in the first syllable of the word, in bisyllabic roots, the second syllable onset consonant is almost exclusively drawn from a subset of the simplex consonants (in order of descending frequency: /m/, /t/, /n/, /p/, /l/, /w/, /j/, /pʷ/, /ŋ/, /k/, /t/, /ʎ/, /mʷ/). In monosyllabic stems, oral vowels (short: 43% and long: 31%) are more common than nasal ones (short: 4% and long: 22%), with the first syllable of bisyllabic stems following a similar pattern. Most vowels in the second syllable of bisyllabic stems are also short and oral, with a

potential preference for rounding harmony between the first and second vowels. Stress typically falls on the first syllable of bisyllabic words, on the first and third syllables of tetrasyllabic words, and on the first or second syllable of trisyllabic words, depending on its vocalic features (see Levinson (n.d.) for details).

Intro to the people (mc) - please feel free to throw away anything that is not useful! Most speakers of the language grow up speaking it monolingually within their swidden horticulturalist communities until they begin attending school around the age of 7; school instruction is in English. While monolingual Yélî Dnye upbringing is common, multilingual families are not unusual, particularly in the region around the Catholic Mission (1953–present) where there appears to be a higher incidence of married-in mothers from other islands (Brown & Casillas, n.d.), the same region in which the current data were collected. Children in these multilingual families grow up speaking Yélî Dnye and English, Tok Pisin, and/or some other language(s) from the region.

Little is known about language development in children growing up in Rossel Island, a community of primarily subsistence farmers who tend to reside in close-knitted villages where child care is distributed across many individuals, and who typically speak Yélî Dnyé, a phonologically and lexically complex language.

- usually monolingual at home
- schooling in English but it starts at age XX, so not relevant here
- however, some use of English due to immigrants & children of immigrants
- children spend a lot of time with other children
- most parents are subsistence farmers
- parental education generally varies between XX and YY

Brief review of NWR for our purposes. There is some variation in the presentation procedure and structure of items found in previous NWR work. For example, items are often presented orally by the experimenter (Torrington Eaton, Newman, Ratner, & Rowe, 2015),

although an increasing number of studies have turned to playing back the stimuli in order to have greater control of the stability of the presentation (Brandeker & Thordardottir, 2015). Additionally, while some studies have used 10-15 non-words, others have employed up to 46 unique items (Piazzalunga, Previtali, Pozzoli, Scarponi, & Schindler, 2019). Often, authors 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, Harris, & Van der Lely, 2007).

Previous work seems to avoid difficult sounds, but we felt this was important to represent Yélî Dnye, so we also varied this factor. We designed a relatively large number of items but, aware that this may render the task longer and more tiresome, we split some of the items across children. This allowed us to get information about repetition accuracy of more items.

Naturally, designing the task in this way may render the study of individual variation within the population more difficult because different children are exposed to different items. However, a review of previous work on individual variation suggested to us that many individual differences effects are relatively small, and would not be detectable with the sample size that we could collect in a given visit.

That said, we contribute to the literature by also reporting descriptive analyses of individual variation that could potentially be integrated in meta-analytic efforts. Based on previous work, we looked at potential improvements with age (Farmani et al., 2018; Kalnak, Peyrard-Janvid, Forssberg, & Sahlén, 2014; Vance, Stackhouse, & Wells, 2005), and potential negative effects of bilingual exposure (Brandeker & Thordardottir, 2015; Meir & Armon-Lotem, 2017; Meir et al., 2016). Previous work typically finds no significant differences as a function of maternal education (e.g., Farmani et al., 2018; Kalnak et al., 2014; Meir & Armon-Lotem, 2017) or child gender (Chiat & Roy, 2007). Although previous research has not often investigated potential effects of birth order on NWR, there is a sizable literature on these effects in other language tasks (Havron et al., 2019), and therefore we report on those too.

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

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

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 given their variability across studies. To illustrate this, we portray studies in which children's NWR scores were gathered between 4 and 12 years of age, and reported separately for items that are relatively short (1-2 syllables) versus longer items (3-4 syllables) in Figure 1. Notice that the effect of stimulus length is minuscule among Italian children (Piazzalunga et al., 2019), but considerable among Tsimane' children (Cristia, Farabolini, Scaff, Havron, & Stieglitz, 2020), where a drop of 40 percentage points is observed at all ages. A similar difference in NWR scores for short versus long items was observed among Arabic children (Jaber-Awida, 2018). Even the effect of age is unstable in this sample. Whereas it is quite clear that children's NWR scores increase in the Italian data, age effects are less stable among Tsimane' children. Therefore, all analyses here 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

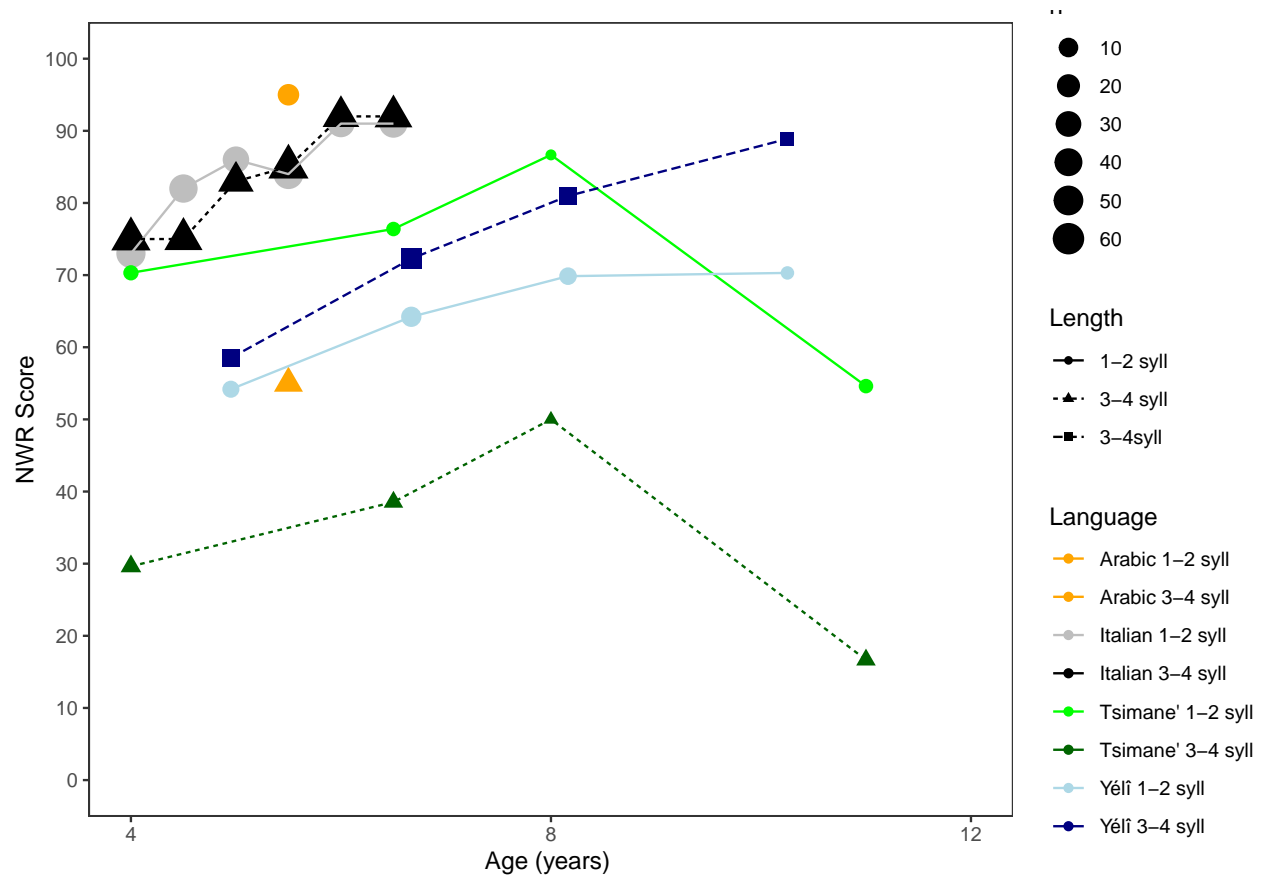


Figure 1. NWR scores as a function of age (in years) and item length for comparable studies. Arabic data from Jaber et al. (2018); Italian data from Piazzalunga et al. (2019); Tsimane' from Cristia et al. (2020); Yéli data from the present study.

and complexity, always meeting the condition that they do not mean anything in the target language (e.g., Balladares, Marshall, & Griffiths, 2016; Wilsenach, 2013). We kept the same variation in item length and the non-meaningfulness requirement, but we did not vary syllable complexity and clusters because these are vanishingly rare in Yéli Dnye. We also increased the number of items an individual child would be tested on, so that a child would get up to 23 items to repeat (note that up to other work has also used 24-30 items: Jaber-Awida, 2018; Kalnak et al., 2014), and we created more items and distributed them across children, so as to increase the coverage, and be able to study more items.

A first list of candidate items was generated in 2018 by selecting simple consonants (“p”,

“t”, “d”, “k”, “m”, “n”, “w”, “y”) and vowels (“i”, “o”, “u”, “a”, “e”) that were combined into consonant-vowel syllables, further sampling the space of 1- to 4-syllable sequences. These candidates were automatically checked against Levinson’s 2015 dictionary and removed from consideration if they appeared in the dictionary. The second author presented them orally to three local research assistants, who were asked to repeat them and further say whether they were real words. Any item for which two or more of the assistants reported them having a meaning or some form of association was excluded.

A second list of candidate items was generated in 2019 by selecting complex consonants and systematically crossing them with all the vowels in the Yélî inventory to produce consonant-vowel monosyllables. As before, items were automatically excluded if they appeared in the dictionary. Additionally, since hearing vowel length in monosyllables in 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 mc 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 mc** please fill in** and an Olympus XXX mc** 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.

The final list is composed of three practice items; 20 monosyllables containing sounds that are less frequent in the world’s languages than singleton plosives; 8 bisyllables; 12 trisyllables; and 4 quadrisyllables (see Table 2).

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 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 versus 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-syllables, then 3-syllables, and finally 4-syllable items.

Procedure. We tried to balance three desiderata: That children would not be unduly exposed to the items before they themselves had to repeat them; that children would feel comfortable doing this task with us; and that the community would feel safe with us doing this task with their children. Moreover, there were also some logistic constraints in terms of the space availability. As a result, the places where elicitation happened varied across the hamlets.

We visited four different hamlets once, and attempted to test all eligible children present at the time, to prevent the items “spreading” through hearsay. In the first hamlet, we tested children in five different places, with some children being tested inside a house and others tested on the veranda. The complete list of places and the ways in which they met the desiderata mentioned above can be found in the raw data, available from online supplementary materials.

The child was donned a headset (xx mc** please fill in** for most of the children, SHURE WH20 XLR headset with a dynamic microphone for the rest), recorded into the left channel into a

260 Tascam DR40x digital recorder. For most children, the headset could not stay comfortably on the
261 child's head, and thus it was placed on the child's shoulders, with the microphone carefully placed
262 close to the child's mouth. A local informant sat next to the child, to would provide the instructions
263 and, if needed, coach the child to make sure, using the three practice items as well as real words,
264 that they understood that the task was to repeat the items precisely without changing anything. An
265 experimenter (the first author) delivered the elicitation stimuli to the local informant and the child
266 over headphones.

267 The first phase was making sure the child understood the task. This was explained orally and
268 the first training item was presented. Often, children froze and did not say anything. If this
269 happened, then we followed this procedure. First, the informant insisted. If the child still did not
270 say anything, the informant asked the child to repeat a real word, and another, and another. If the
271 child could repeat these correctly, then we provided the recorded training item over headphones
272 again. Most children successfully started repeating the items presented over headphones at this
273 point; a few further needed the local informant to model the behavior (i.e., they would hear the
274 item again, and she would say it; then we would play it again, and ask the child to say it). A small
275 minority still failed to repeat the item after hearing it over headphones. If that occurred, we tried
276 with the second training item, at which point some children got it and could continue. A small
277 minority, however, failed to repeat this one, as well as the third training item, in which case we
278 stopped the test altogether.

279 NWR studies vary in whether children are provided with several opportunities to hear and
280 say the item. To have a fixed and clear procedure, we decided that items other than the initial three
281 training ones would not be repeated unless the child made an attempt to produce them. If this
282 attempt was judged correct by the local informant, then the experimenter would move on to the
283 next item (whispering this over a separate headset that was recording onto the right track of the
284 same Tascam). If the local informant heard a deviation, she indicated to the experimenter that the
285 item needed to be repeated, and up to 5 attempts were allowed.

Whenever siblings from the same family were tested, an attempt was made to test first the older and then the younger child, and always on different elicitation sets.

Coding. A script was written to randomize all tokens from all children, pairing each with the auditory target the child had been provided with. A native research assistant then listened through all productions of a given target (randomized across children and repetition order), and made a judgment of whether the item was correctly or incorrectly repeated. She additionally orthographically transcribed exactly what the child said, providing some examples of the types of errors children in general make (without making specific reference to Yélî sounds or the items in the elicitation sets).

Analyses. Some NWR studies employ phoneme-based scoring in addition to or instead of word-level accuracy (e.g., Cristia et al., 2020). We scored items in terms of the number of phonemes that could be aligned across the target and attempt, divided by the number of phonemes of whichever item was longer (the target or the attempt). Although previous work does not use distance metrics, we additionally report those.

Additionally, we looked up each phoneme in each target word in the Phoible database (Moran & McCloy, 2019). We entered the number of languages in which each phoneme was found, as well as what percentage of languages in the database that represented. We could not find a small number of phonemes in the database, and treat those as NA values that do not contribute to the target-level global average phone cross-linguistic frequency, although an alternative would be to attribute them a very small number.

Finally, for describing children's patterns of errors, all repetitions of a given target were taken into account. We describe the proportion of items where the change resulted in a real word (semantic errors); and classify the most common phonological errors.

Participants. This study was approved as part of a larger research effort by Marisa Casillas, with the research protocol having been approved by the IRB committee entitled The Ethics

Committee of the Faculty of Social Sciences (Ethiek Commissie van de faculteit der Sociale Wetenschappen; ECSW) of Radboud University, Nijmegen, The Netherlands. The approval for work outside of the Netherlands came as an amendment to the general research line on language development research (original request: ECSW2017-3001-474 Manko-Rowland; amendment: ECSW-2018-041). Participation was voluntary, with children being invited to come and participate. Regardless of how they performed, children were provided with a snack as compensation. Children who came up to participate but then refused were nonetheless provided with the snack.

A total 55 children were tested, from 34 of families, in five hamlets. Some children could not be included for the following reasons: refused participation or failed to repeat items presented over headphones even after coaching ($N = 0$), spoke too softly to allow offline coding ($N = 5$). In addition, 2 teenagers were tested to put younger children at ease; their data is not included in analyses below. The remaining 40 children (14 girls) were aged 6.96 years (range 3.90-11 years). There were 32 children exposed only to Yéli in the home, 6 children who were also exposed to another language in the home, and , 2 for whom this information was missing. Maternal years of education averaged 8.24 years (range 6-12 years; 2 children had this information missing).¹ In terms of birth order, 5 were first borns, 4 second, 2 third, 7 forth, 5 fifth, 1 sixth, and NA did not have this information.

Results

Preliminary analyses. We first checked whether accuracy varies between first and subsequent presentations of an item by averaging word-level accuracy at the participant level separately for first attempts and subsequent repetitions. We excluded 3 children who did not have data for one of these two types. As shown in Figure 1, participants' mean word-level accuracy became more heterogeneous in subsequent repetitions. Surprisingly, subsequent repetitions ($M = 39$, $SD = 26$) were on average less accurate than first ones ($M = 65$, $SD = 15$), $t(38) = 6.62$, p

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

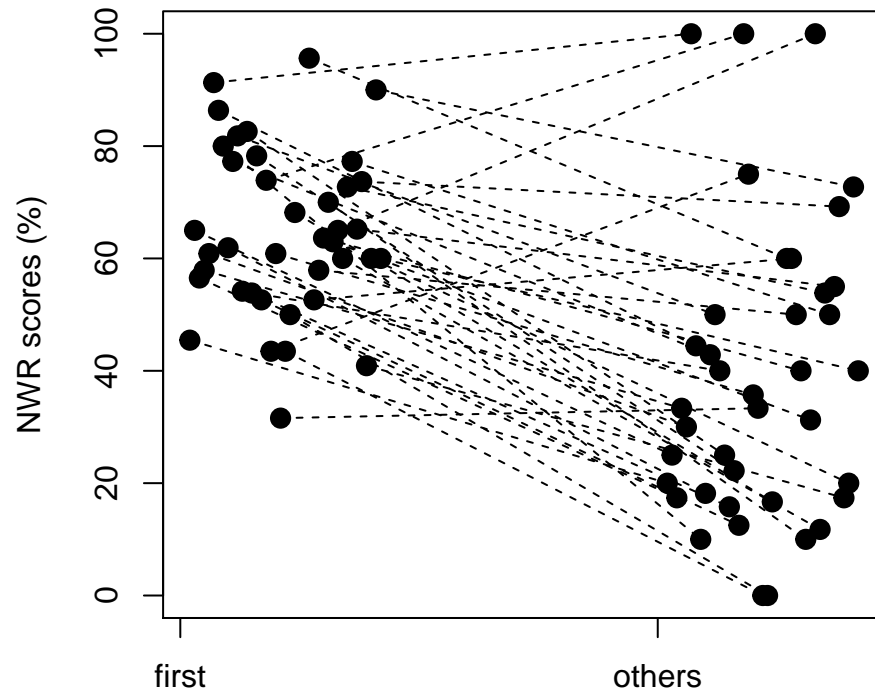


Figure 2. (#fig:Fig1-first_vs_others)NWR scores for individual participants averaging separately their first attempts and all other attempts.

= 0. Given the uncertainty in whether previous work used only the first or all repetitions, and since behavior degraded and became more heterogeneous in subsequent repetitions, the rest of the analyses focus on only the first repetitions.

Taking into account only the first attempts, we averaged attempts by each of the 24 children who had data for first attempts; their ages ranged from 3.90 to 10.20 ($M = 6.50$, $SD = 1.50$).

The overall NWR score was $M = 61\%$ ($SD = 15\%$). Scores based on phonemes are even higher $M = 78\%$ ($SD = 9\%$). The phoneme-based normalized Levenshtein distance was $M = 22\%$ ($SD = 9\%$), meaning that about a fifth of phonemes were substituted, inserted, or deleted. Notice that the normalized Levenshtein distance is the complement of phoneme-based scores.

NWR as a function of cross-linguistic phone frequency. In this analysis, we were interested in variation in performance 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 effects of interest was the average cross-linguistic phone frequency; we also included child age as a control fixed effect, and child ID & target ID as random variables.

We could include 885 observations (why – shouldn't it be 1k??), from 41 children producing in any given trial one of 41 potential target words. The analysis revealed a significant estimate for the average cross-linguistic frequency of phones in the target words ($\beta = 0.77$, $SE \beta = 0.19$, $p = 0$): Target words with phones found more frequently across languages had a higher proportion of words that were correctly repeated, as clear in Figure 2. Age was also a significant predictor ($\beta = 0.31$, $SE \beta = 0.13$, $p = 0.02$).

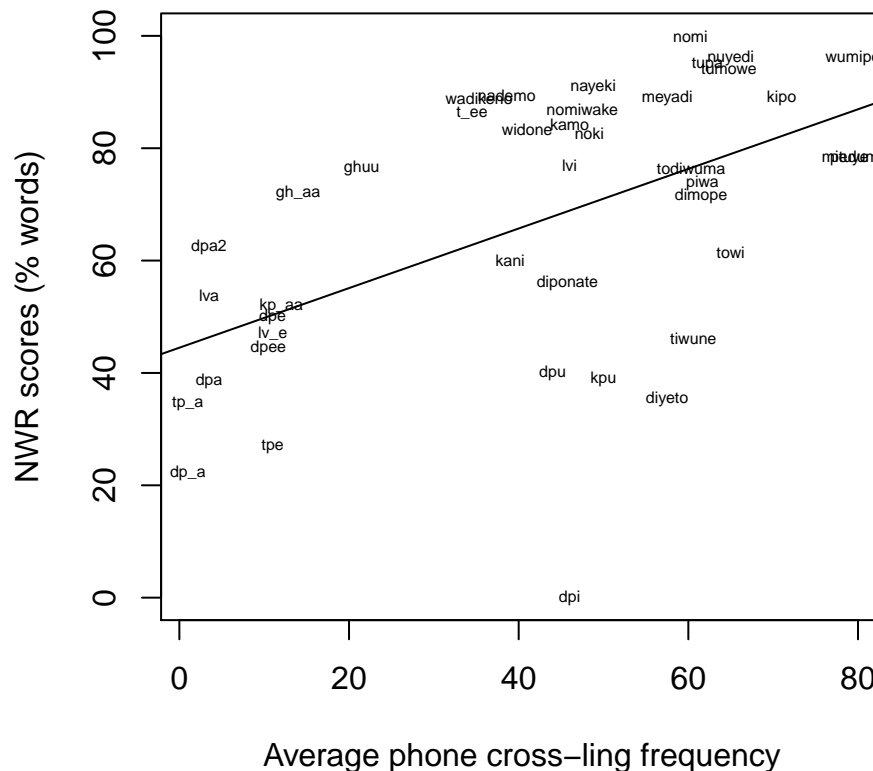


Figure 3. (#fig:Fig2-xling_freq)NWR scores for individual target words as a function of the average frequency with which each phone is found across languages.

We next checked whether this may actually be due to frequency of the sounds within the

language. We estimated frequency of the phones present in the stimuli in a corpus of child-centered recordings (???), and applied the natural logarithm. As before, unattested sounds were not considered (i.e., they were declared NA so that they do not count for analyses). Phone frequency estimated from this corpus were correlated with cross-linguistic phone frequencies ($r(23) = 0.65$, $p = 0$). Therefore, we fit another mixed logistic regression, this time declaring as fixed effects scaled cross-linguistic and corpus frequencies (averaged across all attested phones within each stimulus item), in addition to age. As before, the model contained random slopes for both child ID and target. In this model, both cross-linguistic frequency ($\beta = 0.50$, $SE \beta = 0.25$, $p = 0.04$) and age ($\beta = 0.31$, $SE \beta = 0.13$, $p = 0.02$) were significant predictors of accuracy, but corpus frequency ($\beta = 0.37$, $SE \beta = 0.24$, $p = 0.12$) was not.

use proportion of correctly produced rather than number; DONE

a2m use first attempts rather than the whole data set: having second thoughts. To be clear: the argument for using first attempts only is that all of the other analyses are on kids' first attempts, rather than all of their attempts, so we should use just first attempts here because it may confuse readers & it is more consistent. BUT there are fewer mistakes this way, which is Options 1 and 2 below has lots of zeros... CF Option 3 further down

a2m probably turn the paragraph on substitutions into a table: having second thoughts, the table would look just like a para, no?

Option 1: just numbers, first attempts only Next, we investigated patterns of deletion and substitution. Deletions were relatively rare, with only 7 vowels deleted, and 3 consonants.

As for substitutions, it was as common for a nasal vowel to be produced as an oral vowel as vice versa (0 oral target vowels produced as nasal vowels, 0 nasal target vowels produced as oral vowels). Substitutions in which the oral nature was preserved but the quality of the vowel was changed were a great deal more common than changes in quality among nasal vowels (43 oral vowels produced with a different quality; 0 nasal vowels produced with a different quality). As for

consonants, asymmetries were very marked with more complex consonants (specifically dptpkpkmknmbghlv) mispronounced as simple ones (specifically mnlwyvdgptkfhch, 0) than vice versa (0). Simple consonants were mispronounced as other simple consonants quite frequently (73 simple consonants mispronounced as other simple ones, compared to 0 complex ones).

Option 2: numbers and proportions, first attempts only Next, we investigated patterns of deletion and substitution. Deletions were relatively rare, with only 7 vowels deleted (about 0.54% of all vowel targets), and 3 consonants deleted (about 0.17% of all consonant targets).

As for substitutions, it was as common for a nasal vowel to be produced as an oral vowel as vice versa (0 oral target vowels produced as nasal vowels or about 0% out of all oral vowel targets, 0 nasal target vowels produced as oral vowels or about NA% out of all nasal vowel targets). Substitutions in which the oral nature was preserved but the quality of the vowel was changed were a great deal more common than changes in quality among nasal vowels (43 oral vowels produced with a different quality or about 0.29% out of all oral vowel targets; 0 nasal vowels produced with a different quality or about NA% out of all nasal vowel targets). As for consonants, asymmetries were very marked with more complex consonants (specifically dptpkpkmknmbghlv) mispronounced as simple ones (specifically mnlwyvdgptkfhch, 0 times or about NA% out of all complex consonant targets) than vice versa (0 times or about NA% out of all simple consonant targets). Simple consonants were mispronounced as other simple consonants quite frequently (73 simple consonants mispronounced as other simple ones or about 0.35% out of all simple consonant targets, compared to 0 complex ones or about NA% out of all complex consonant targets).

Option 3: numbers and proportions, ALL attempts Next, we investigated patterns of deletion and substitution. Deletions were relatively rare, with only 11 vowels deleted (about 0.25% of all vowel targets), and 8 consonants deleted (about 0.21% of all consonant targets).

As for substitutions, it was as common for a nasal vowel to be produced as an oral vowel as vice versa (52 oral target vowels produced as nasal vowels or about 2.01% out of all oral vowel

targets, 58 nasal target vowels produced as oral vowels or about 6.46% out of all nasal vowel targets). Substitutions in which the oral nature was preserved but the quality of the vowel was changed were a great deal more common than changes in quality among nasal vowels (205 oral vowels produced with a different quality or about 3.22% out of all oral vowel targets; 23 nasal vowels produced with a different quality or about 7.08% out of all nasal vowel targets). As for consonants, asymmetries were very marked with more complex consonants (specifically dptpkpkmknmbghlv) mispronounced as simple ones (specifically mnlwyvdgptkfhch, 250 times or about 5.57% out of all complex consonant targets) than vice versa (2 times about 5.57% out of all simple consonant targets). Simple consonants were mispronounced as other simple consonants quite frequently (129 simple consonants mispronounced as other simple ones or about 1.14% out of all simple consonant targets, compared to 55 complex ones or about 2.76% out of all complex consonant targets).

Finally, we looked into the frequency with which mispronunciations resulted in real words. Nearly all of them were: 97%.

TO DISCUSS TOGETHER this is higher than any previous report! can we be sure of it?

Accuracy as a function of item length. Next, we inspected whether accuracy varied as a function of word length. Results are shown on table XX. This table shows that monosyllables accuracy was much lower than other lengths. This is likely due to the fact that the majority of monosyllables included were chosen because they had sounds that are rare in the world's languages, which may indicate that they are hard to produce or to perceive.

Setting monosyllables aside, we observe the typical pattern of decreased accuracy for longer items, although this is particularly salient for the whole word scoring. This is the most commonly reported type of score, but it is also the least forgiving. The pattern is less marked when other two scores are used, which are less sensitive to errors.

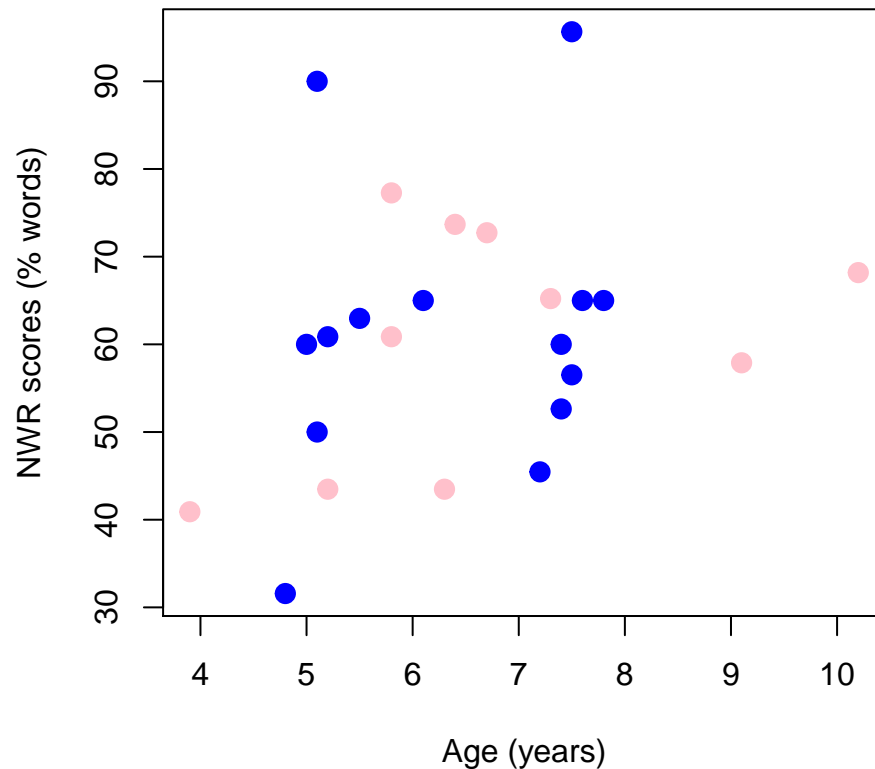


Figure 4. (#fig:Fig2-scores by age)NWR whole-word scores for individual participants as a function of age and sex (blue = boys, pink = girls).

Factor structuring individual variation. Our final exploratory analysis assessed whether variation in scores was structured by factors that vary across individuals. As shown in Figure 2, there was a greater deal of variance in earlier than later ages, with significantly higher NWR scores for older children (Spearman's rank correlation, given inequality of variance, $\rho(1,548.04) = 0.33$, $p = 0.12$). In contrast, there was no clear association between NWR scores, on the one hand, and sex ($t(0.18) = 21.34$, $p = 0.86$), birth order (data missing for 17 children, $2,678.47) = -0.16$, $p = 0.44$) or maternal education (data missing for 2 children, $2,091.92) = 0.09$, $p = 0.67$)

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 previous 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

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

References

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

- 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.
- Gallagher, G. (2014). An identity bias in phonotactics: Evidence from Cochabamba Quechua. *Laboratory Phonology*, 5(3), 337–378. <https://doi.org/10.1515/lp-2014-0012>
- Gallon, N., Harris, J., & Van der Lely, H. (2007). Non-word repetition: An investigation of phonological complexity in children with grammatical sli. *Clinical Linguistics & Phonetics*, 21(6), 435–455.
- Havron, N., Ramus, F., Heude, B., Forhan, A., Cristia, A., Peyre, H., & Group, E. M.-C. C. S. (2019). The effect of older siblings on language development as a function of age difference and sex. *Psychological Science*, 30(9), 1333–1343.
- Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28(4), 377–396.
- Jaber-Awida, A. (2018). Experiment in non word repetition by monolingual Arabic preschoolers.

Athens Journal of Philology, 5(4), 317–334. <https://doi.org/10.30958/ajp.5-4-4>

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

Levinson, S. C. (n.d.). *A grammar of yélî dnye, the papuan language of rossel island*. Cambridge, UK: Cambridge University Press.

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/>

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

- 559 emergent bilingual Northern Sotho/English learners. *Per Linguam: a Journal of Language*
560 *Learning = Per Linguam: Tydskrif vir Taalaanleer*, 29(2), 17–32.
561 <https://doi.org/10.5785/29-2-554>

Table 1

NWR stimuli in orthographic and phonological representations.

Practice		Monosyll		Bisyll		Trisyll		Tetrasyll	
Orth.	Phon.	Orth.	Phon.	Orth.	Phon.	Orth.	Phon.	Orth.	Phon.
nopimade	nɔpimæʔɛ	dp:a	t̪p̪æ	kamo	kæmɔ	dimope	ʔimɔpɛ	dipońate	ʔipɔnæʔɛ
poni	pɔni	dpa	t̪p̪æ	kańi	kæni	diyeto	ʔijetɔ	ńomiwake	nɔmiwæke
wî	wu	dpâ	t̪pa	kipo	kipɔ	meyadi	mɛjæʔi	todiwuma	tɔʔiwumæ
		dpê	t̪p̪ə	ńoki	nɔki	mituye	mitujɛ	wadikeńo	wæʔikɛnɔ
		dpée	t̪pɛ:	ńomi	nɔmi	ńademo	næʔemɔ		
		dpi	t̪pi	piwa	piwæ	ńayeki	næjekɪ		
		dpu	t̪pu	towi	tɔwi	ńuyedi	nujɛʔi		
		gh:ââ	ʔa:	tupa	tupæ	pedumi	pɛʔumi		
		ghuu	ʔu:			tiwuńe	tiwunɛ		
		kpu	kpu			tumowe	tumɔwɛ		
		lv:ê	lβ̥ʔ			widońe	wiʔɔnɛ		
		lva	lβ̥ʔæ			wumipo	wumipɔ		
		lvi	lβ̥ʔi						
		tp:a	t̪p̪æ						
		tpâ	t̪pa						
		tpê	t̪p̪ə						

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
47 (22)	59 (17)	41 (17)
79 (22)	94 (7)	6 (7)
78 (19)	94 (6)	6 (6)
75 (32)	93 (11)	7 (11)