Supplementary materials to: Non-word repetition in children learning Yélî Dnye

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

[1] Laboratoire de Sciences Cognitives et de Psycholinguistique, Département d'Etudes Cognitives, ENS, EHESS, CNRS, PSL University
[2] Max Planck Institute for Psycholinguistics
[3] University of Chicago

Abstract

This is a supplementary material containing: 1. Explanation of a small-scale systematic review carried out to embed the main results in the NWR literature; 2. The use of this data to look at potential length effects on NWR as a function of language characteristics; 3. Length and age effects in this and previous work.

SM1: Small systematic review of previous NWR work

For many of our research questions, we wanted to know about previous work. It is common in psychology to do this via an unsystematic review – that is, by e.g. starting out with papers one knows and then reading other papers citing those, or by doing searches on google scholar or another search engine, or by reading someone else's review. It has been well-established that overviews based on such unsystematic searches do not generalize and have other flaws because they represent biased samples of the literature (Thomas-Odenthal, Molero, Does, & Molendijk, 2020) – for instance, famous studies are more likely to be found, even though there are no quality differences between more versus less famous studies.

So instead of relying on such an approach, we built on a systematic review that was carried out for a different paper: Cristia, Farabolini, Scaff, Havron, and Stieglitz (2020) "combined our previous knowledge of the literature and systematic searching to yield a sample of 17 studies that we could interrogate further." Specifically, the main goal of that systematic review was to check whether "effects of infant-directed input quantities may have been reported. Input variability has not been studied, to our knowledge, but two potential proxies of input have: Monolingual status, and socio-economic status. We set aside the literature comparing monolinguals against non-monolinguals because we reasoned that some readers may argue this contrast does not only show effects of input differences, but also interference across the languages being learned" (Cristia, 2020, personal records). Discovery combined an initial list produced by Gianmatteo Farabolini (based on readings of literature

on bilingual NWR mainly), with a scholar.google.com search carried out in incognito mode by Cristia, with keywords "non-word-repetition socio-economic-status." Studies finally included reported on children who were monolingual and between 3 and 7 years of age, and which used proportion of non-words repeated as the outcome metric.

We revisited this selection of articles in order to extract data on actual performance, taking into account also the length of the non-words used in the study. From that selection, only 8 articles reported on actual performance in the text, figures, or tables, fortunately representing a range of languages: Persian by Farmani et al. (2018); Israeli Arabic (Jaber-Awida, 2018); English (Vance, Stackhouse, & Wells, 2005); Slovak by Kapalková, Polišenská, and Vicenová (2013) and Polišenská and Kapalková (2014); Sotho by Wilsenach (2013); Swedish by Kalnak, Peyrard-Janvid, Forssberg, and Sahlén (2014) and Radeborg, Barthelom, SjöBerg, and Sahlén (2006). We extended that selection with 5 additional papers selected to increase the number of languages represented, with both languages sometimes described as having short words (Cantonese Stokes, Wong, Fletcher, & Leonard, 2006; Mandarin Lei et al., 2011); and others described as having long words (Italian Piazzalunga, Previtali, Pozzoli, Scarponi, & Schindler, 2019; Spanish Balladares, Marshall, & Griffiths, 2016). Finally, we also included Tsimane' data (Cristia, Farabolini, Scaff, Havron, & Stieglitz, 2020).

SM2: Assessing potential length effects

We looked at the subset of papers that reported NWR scores separately for different word lengths. These were: Israeli Arabic (Jaber-Awida, 2018); Cantonese (Stokes, Wong, Fletcher, & Leonard, 2006); English (Vance, Stackhouse, & Wells, 2005); Italian (Piazzalunga, Previtali, Pozzoli, Scarponi, & Schindler, 2019); and Tsimane' (Cristia, Farabolini, Scaff, Havron, & Stieglitz, 2020).

Our reading of that work is that, although there is cross-linguistic (or cross-sample) variation in length effects, these do not systematically line up with expected word length in different languages. For instance, the difference in NWR scores for 2- versus 3-syllable items (averaging across age groups) is largest in Tsimane' (~28%) and Arabic (~15%), which tend to have longer words, as does Italian, where the difference between 2- and 3-syllable items was only ~2%. Similarly, two languages that are often described as heavily biased towards monosyllables show diverse length effects (Cantonese ~8% versus English ~1%).

SM3: Integrating our data with that from other studies

For this analysis, we could include all studies that reported non-word repetition scores based on whole item scoring for at least some length, ideally separating children by age. Specifically, Arabic was represented by Jaber-Awida (2018); Cantonese by Stokes, Wong, Fletcher, and Leonard (2006); English by Vance, Stackhouse, and Wells (2005); Italian by Piazzalunga, Previtali, Pozzoli, Scarponi, and Schindler (2019); Mandarin by Lei et al. (2011); Spanish by Balladares, Marshall, and Griffiths (2016); Tsimane' by Cristia, Farabolini, Scaff, Havron, and Stieglitz (2020); and Yélî Dnye from the present study. Studies varied in the length of non-words that were considered; whenever results were reported separately for

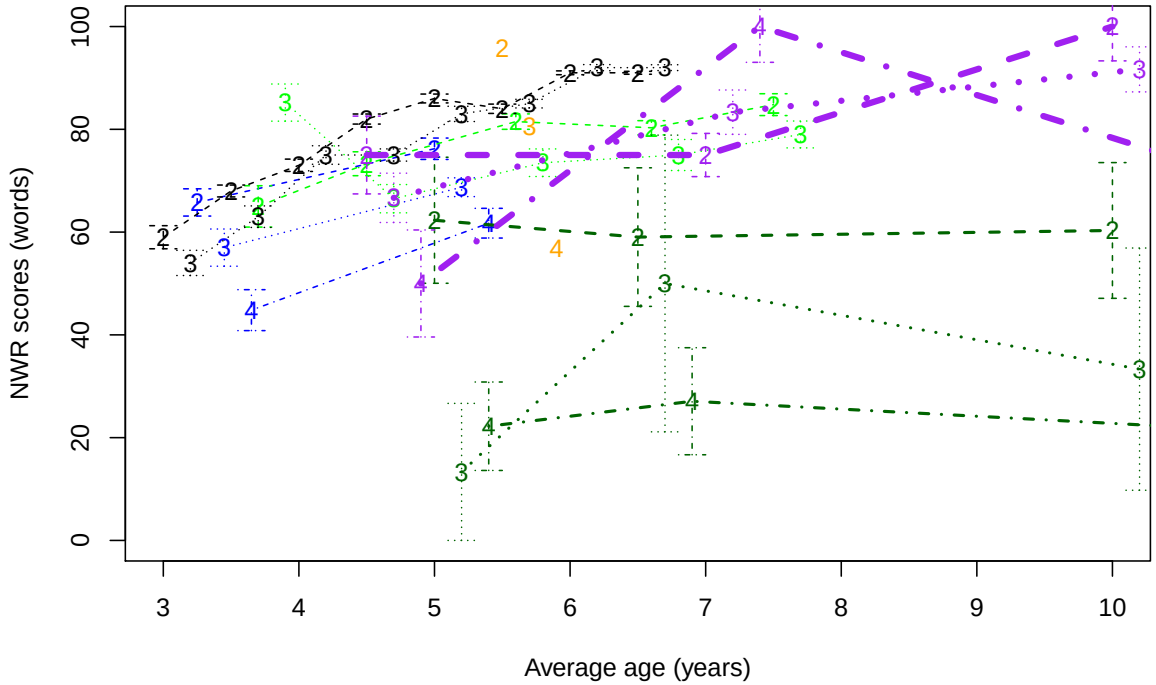

Figure 1. NWR scores as a function of age (in years) and item length for comparable studies (2-4 indicating number of syllables, 2=dashed, 3=dotted, 4=dotted and dashed). Jaber-Awida (2018) reported on 20 Israeli Arabic learners (orange); Piazzalunga et al. (2019) reported on groups of 24-60 Italian 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 shifted for ease of inspection of different lengths at a given age.

different lengths, we calculated overall averages based on lengths of 2 and 3 syllables, for increased comparability. Results separating different age groups are shown in Figure 2.

Several observations can be drawn from this figure. To begin with, we focus on the comparison between Yélî Dnye and Tsimane'. These two groups have been described as having roughly similar levels of child-directed speech, yet they exhibit very different results: Tsimane' shows lower overall NWR scores (and according to Figure 1, larger length effects). This suggests that the lower NWR scores found among the Tsimane' are due to long-term effects of lower levels of child-directed speech. Naturally, there is an alternative interpretation, namely that input estimation suggesting very slightly higher levels of child-directed speech among the Tsimane' than among Yélî Dnye learners is inaccurate. In fact, careful reading of previous reports highlight important methodological differences in how input quantity has been estimated across papers: Casillas, Brown, and Levinson (2020) hand-coded speech with the help of a native research assistant, and then summed all child-directed speech, which effectively establishes an upper boundary of the speech children could potentially process. Cristia, Dupoux, Gurven, and Stieglitz (2019) estimated quantities from behavioral

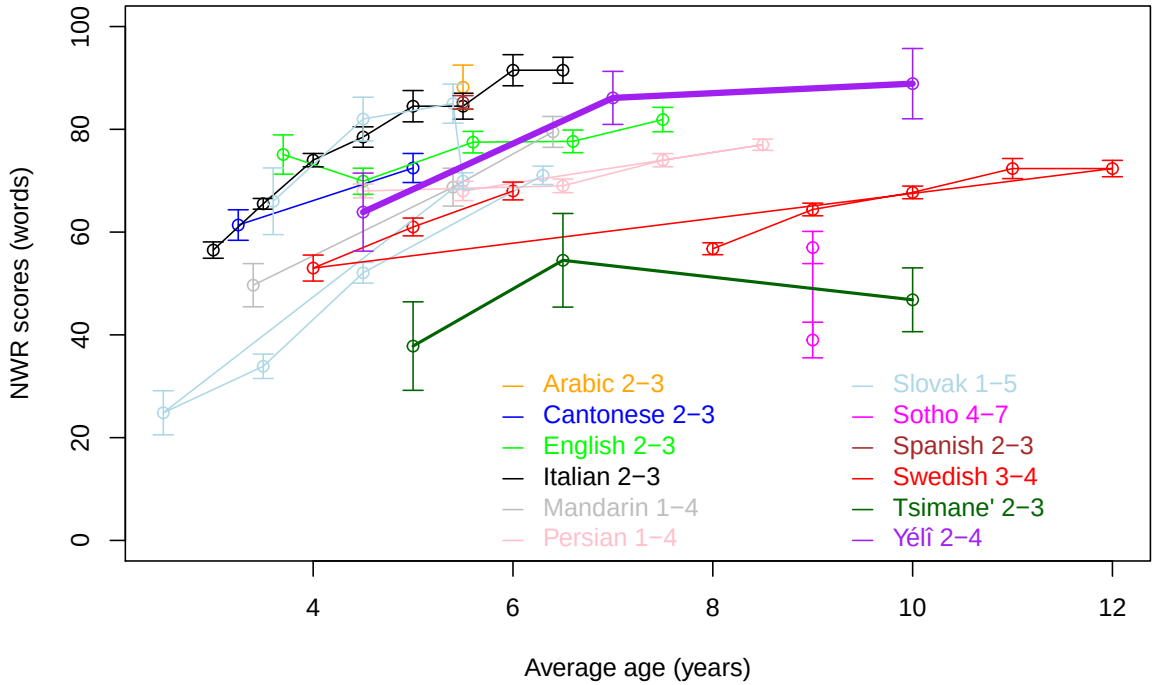

Figure 2. NWR scores as a function of age (in years), averaged across multiple non-word lengths, as a function of children's native languages. The legend indicates language and the length of non-words (in syllables). Central tendency is mean; error is one standard error.

observations on the frequency of child-directed one-on-one conversation, which is probably closer to a lower boundary. Finally, Scaff, Stieglitz, Casillas, and Cristia (2021) used human annotation for detecting speech but an automated temporal method for assigning speech as child-directed or not, in a way that could lead to over-estimation (because any speech by e.g. a female adult that was not temporally close to speech by others would count as child-directed). A final answer to the question of how much child-directed speech is afforded to Yélî and Tsimane' children must await fully comparable methods.

That said, Cristia, Farabolini, Scaff, Havron, and Stieglitz (2020) also pointed out another characteristic of the Tsimane' population, and this was the relatively low prevalence of literacy, and generally the variable access to formal education. This is a very different case from the Yélî population studied here, where nearly all adults have accumulated several years of schooling, and basic literacy in English (and sometimes Yélî Dnye) is widespread. If this second hypothesis holds, then this may mean that there are phonetic effects of learning to read in the input afforded to young children, and that this has consequences for young children's encoding and decoding of sounds in the context of NWR tasks. Notice that this is not the same as the oft-recorded effect of learning to read affecting NWR performance, illustrated for instance in the data for Sotho in Figure 2. These two data points have been gathered from two groups of children, all exposed mainly to Sotho, but children with higher NWR had been learning to read in Sotho, whereas those with lower scores were learning to read in English. What is at stake in our proposed alternative interpretation of the lower scores observed among the Tsimane' is related to literacy *in the broader population* (rather

than in the tested children themselves).

Although exciting, this hypothesis is only one of many. Another plausible explanation is that the Tsimane' results are not comparable to the previous body of literature, and specifically to our study. Cristia, Farabolini, Scaff, Havron, and Stieglitz (2020) administered the NWR in the form of a group game played outside, with a non-native experimenter providing the target, and each person of the group attempting it in their stead. This immediately means a number of important methodological differences with the standard implementation of NWR, where children are tested individually, they hear items spoken by a native speaker (often over headphones), the experimenter tends to belong to the same community as the children, and testing occurs in quiet conditions (with little background noise). Thus, a priority is for additional data gathered using this more novel testing paradigm in other populations, or from the Tsimane' using the more traditional paradigm.

Broadening our discussion to all of the studies in our literature review, we notice that there is rather wide variation of the range of NWR scores found across these samples, and that, in fact, the strength of age effects also varies. We performed some exploratory analyses to see whether features of the languages children were learning could be related to their overall NWR scores. We extracted the number of phonemes in the language from PHOIBLE and coded whether words in the language tended to be longer or shorter based on information in the papers or other sources. Neither of these two predictors explained variance in Figure 2. It is possible that average word length plays a role, but often researchers incorporate this into their design by including longer items when the native language allows this, with e.g. Sotho non-words having 4-7 syllables in length.

References

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

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

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-directed speech is infrequent in a forager-farmer population. *Child Development*, *90*, 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

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.

Jaber-Awida, A. (2018). Experiment in non word repetition by monolingual Arabic preschoolers. *Athens Journal of Philology*, *5*, 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.

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

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.

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*, 219–236.

Thomas-Odenthal, F., Molero, P., Does, W. van der, & Molendijk, M. (2020). Impact of review method on the conclusions of clinical reviews: A systematic review on dietary interventions in depression as a case in point. *PloS One*, *15*(9), e0238131.

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