The Development of Color Terms in Shipibo-Konibo Children

Danielle Kellier[1], *Martin Fortier*[2], Maria Fernández Flecha[3], & Michael C. Frank[4]

[1] University of Pennsylvania

[2] PSL Research University

[3] Pontificia Universidad Católica del Perú

[4] Stanford University

Author Note

* these authors contributed equally.

MCF was supported by a Jacobs Foundation Fellowship. We are especially grateful to Paul Kay for assistance throughout the study.

Correspondence concerning this article should be addressed to Martin Fortier*, Postal address. E-mail: martin.fortier@ens.fr

Abstract

Enter abstract here. Each new line herein must be indented, like this line.

Keywords: keywords

Word count: X

The Development of Color Terms in Shipibo-Konibo Children

TO BE PASTED FROM GOOGLE DOC

Experiment 1

In our first experiment, our goal was to replicate and update the characterization of the adult SK color system given by the World Color Survey. We were further interested in the use of Spanish terms as language contact and multilingualism have increased in the years since the original World Color Survey work.

Methods

Participants. We recruited 39 adult participants (7 men). Most of participants were from SK communities of the Middle Ucayali region (from Yarinacocha, San Francisco, and Nueva Betania), but some of them were from communities of the Lower (Paoyhan) and Upper (Puerto Belen) Ucayali. In Yarinacocha (a small town located in the vicinity of Pucallpa), participants were recruited in Bena Jema, a neighborhood where most of the inhabitants are SK. All the other places where participants were recruited were native community villages exclusively inhabited by SK people. Overall, the sample included both somewhat urbanized SK (Yarinacocha and San Francisco) as well as SK individuals who were still used to more traditional activities and regular contact with the surrounding rainforest (Nueva Betania, Paoyhan, and Puerto Belen).

The median self-reported age for participants was 38 years with a range between 20 to 64 years of age (SD = 13.60yo). Regarding occupations, 41% of the women were homemakers (33% overall) and another 41% were artisans (33%). Three of the 7 men were horticulturalists (43%, 8% overall). Four women (12%, 10% overall) and three men (43%, 8% overall) identified as students.

Although all adult participants spoke Shipibo-Konibo as a first language, all were bilingual to a substantial degree. All reported an introduction to the Spanish language before early adolescence (M = 7.80yo, SD = 2.90y). Participant age and reported age of introduction to Spanish were positively correlated; younger participants reported learning Spanish at an early age although all participants reported introductions before early adolescence ($r = 0.43$, $p = 0.01$).

Materials

We used the 330 Munsell color chips as stimuli for the study. However, only 165 chips were used for each single participant (see below). These chips were exactly those used to collect data for the WCS. Individual color chips were 2 cm x 2.5 cm.

Procedure

In order to make sure that the natural light intensity would not vary much between participants, the experiment took place indoors, near a window or a door. The study was conducted entirely in the SK language.

Our procedure was similar to that used in the WCS (see Kay, Berlin, Maffin, Merrifield, & Cook, 2009, pp. 585–591). Participants were seated in front of the experimenter and introduced to the whole procedure and the general goal of the study. Then the primary procedure involved presenting participants with a color chip and asking them: “What is the color of this chip?”[1] and recording their response or responses.

One major difference between the WCS procedure and ours is that, in the WCS, the experimenter was expected to brief participants so that they would only provide basic color terms during the task (e.g., “blue” as opposed to “navy blue” or “sky-like”). However, we

[1] The SK word for color that we used was the Spanish word *color*. In general, the SK language includes some castillanisms that are well-known by all speakers; color is one of them.

found it rather difficult to help participants understand in a few sentences what a basic color term was.[2] Thus, we opted to let participants provide any term they wished. If they did not provide a basic color term, we would ask further questions to elicit a basic color term. For example, if the participant provided the term “blood-like” (a non-basic color term) when presented with a red color chip, the experimenter would ask: “Do you know of any other word to refer to the color of this chip?” If the participant subsequently responded “dark red” (another non-basic color term), the experimenter would further ask: “How would you refer to this color with only one word?” Eventually, the participant would say “red” (a basic color term).

For some chips, participants would provide a basic color term at once; but for others, they would first provide one or two non-basic terms before actually providing a basic term. When participants did not provide a basic color term after three trials (i.e., two follow-up questions), no further questions was asked, and the experimenter proceeded to the next chip. This method was more effortful and time-consuming than the WCS procedure, but it improved the fluency and the intuitiveness of the task for participants.

A second difference between our procedure and that of the WCS concerned the number of chips each participant was presented with. In the WCS, every participant was expected to provide color terms for each of the 330 chips of the set. As we were afraid that doing so would take too long and that participants would find the task tedious, we decided that the set of chips would be split in two (even and uneven numbers) and that every participant would be randomly ascribed to one of the two subsets. As a result, each participant was presented with only 165 chips.

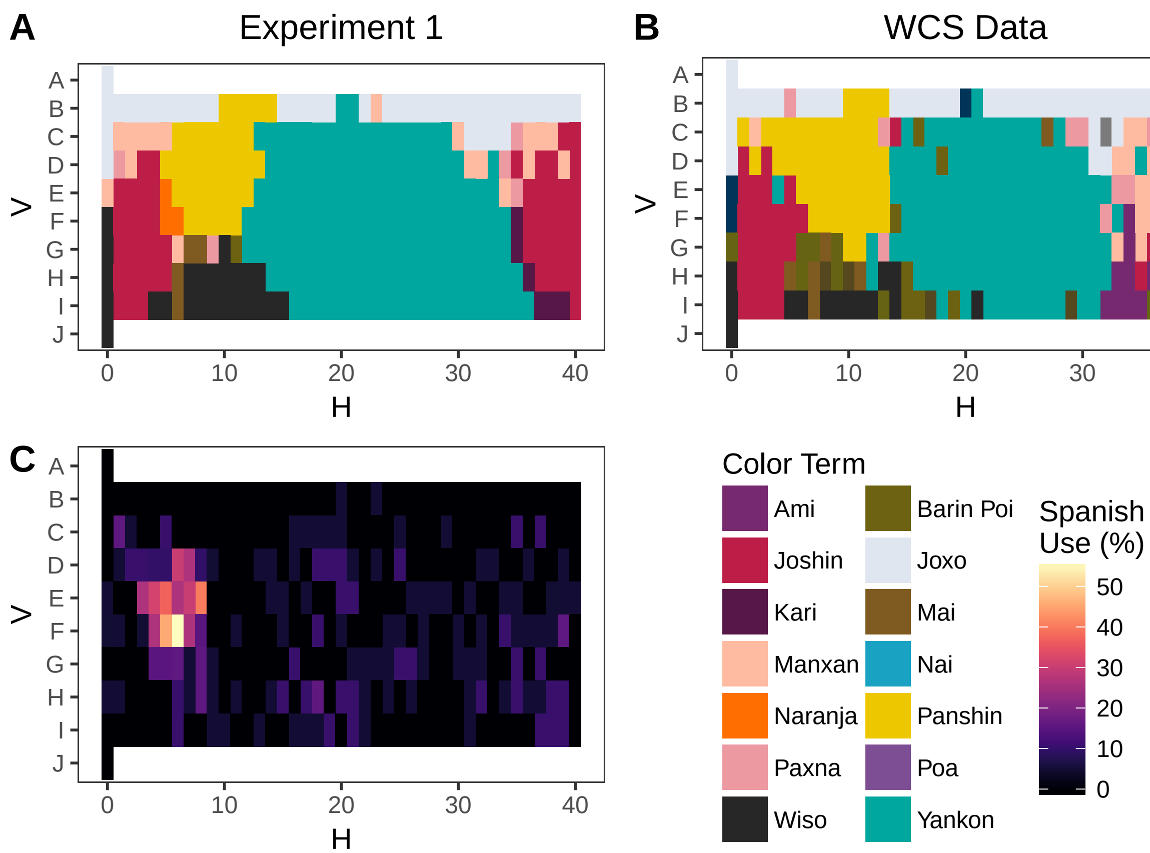

Figure 1. (A and B) Plots of the modal term given for a particular chip. Color coordinates were represented in 2-D Munsell space. Modal responses were given by SK adults during (A) the original World Color Survey and during (B) our Experiment 1. (C) Heat map of prevalence of Spanish-language responses during Experiment 1. Legends for all three subplots located in the bottom-right quadrant.

Results and Discussion

Broadly speaking, our results were quite similar to the WCS findings. Figure 1 shows a comparison between our data (Panel A) and the WCS (panel B). The basic level colors in our data were quite similar, as well. All participants described at least 1 chip with the following set of color terms: light/white ("joxo"), dark/black ("wiso"), yellow ("panshin"),

[2] Indeed, as Berlin & Kay (2009: 587-589) acknowledge, there is no straightforward necessary and sufficient criteria for the "basicness" of a color term.

red ("joshin"), and green/blue ("yankon"). Most (79%) participants also used described at least 1 chip as faded or "manxan", referring to a chip's saturation. In terms of overall popularity, participants on average described 32% of chips as "yankon" (SD = 10%) followed by "joshin" (M = 12%, SD = 6%), "joxo" (10%, 5%), "panshin" (9%, 4%), "manxan" (7%, 7%), and "wiso" (6%, 4%).

One departure from the Berlin-Kay data was that 59% of adults described at least 1 chip using a Spanish-language color term, accounting for 4% of all responses (Figure 1, Panel C). In particular, Spanish use reached as high as 55% when participants were asked to label chips that English speakers would consider to be orange, or "naranja" in Spanish. However, there was a high amount of variability in Spanish use between subjects (M = 4%, SD = 12%) which neither participant age ($p = 0.87$) nor reported age of Spanish introduction ($p = 0.56$) failed to predict. Some subjects never responded in Spanish whereas one participant used Spanish labels for 71% of all trials despite all sessions being conducted entirely in the Shipibo-Konibo language. While we can only speculate as to this participant's motivations, it seems likely that they were more familiar with Spanish color vocabulary or viewed Spanish color terms as more precise descriptors.

Participants on average described 69% of chips using a SK basic color term like "yankon" (SD = 22%). Some participants described chips using SK ad-hoc color terms, such as "nai" or *sky* for blue chips (M = 11%, SD = 12%), or ad hoc terms referring to saturation or luminosity of a chip, such as "manxan" (M = 7%, SD = 7%). Virtually all instances where a participant responded in Spanish involved a Spanish basic color term such as "rojo" (M = 4%, SD = 10%). In other words, participants typically only responded in Spanish to label chips into basic categories; they relied on Shipibo-Konibo for other descriptors.

Given these data, we moved on to exploring the development of SK color vocabulary in childhood. Experiment 2 tests production and comprehension of SK color terms using

Table 1

Demographics of participants in Experiment 2.

Age Group	N	Male
5	3 (5%)	1 (33%)
6	8 (14%)	3 (38%)
7	12 (21%)	4 (33%)
8	15 (26%)	5 (33%)
9	10 (18%)	5 (50%)
10	4 (7%)	2 (50%)
11	5 (9%)	3 (60%)

SK-prototypical color chips; Experiment 3 tests children in Spanish using Spanish-prototypical chips.

Experiment 2

In Experiment 2, we tested children on their production and comprehension skills with a set of chips representing the prototypical colors for common SK color terms.

Methods

Participants. The Pontificia Universidad Católica del Perú's Institutional Review Board approved our study protocol. We recruited 57 5- to 11-year-old children (23 boys). Table 1 shows the distribution of ages and genders. Fifteen children were recruited from neighborhoods in Yarinacocha, in the Pucallpa region of Peru, as well as in 42 children from Bawanisho, a native community settled along the Ucayali River, south of Pucallpa. Children were recruited either through their parents or through local schools. When recruited at school, consent for participation was collected from both the teachers and the parents; otherwise, only consent from the parents was collected.

Materials. Based on findings of Experiment 1, we selected out 8 color chips that were prototypical instances of prominent SK color terms. These color chips were blue (WCS n°1), green (WCS n°234), red (WCS n°245), white (WCS n°274), yellow (WCS n°297), black (WCS n°312), greeny-yellow (WCS n°320), and purple (WCS n°325). These color chips were exactly the same as those used in Experiment 1; the only difference was that adult participants in Experiment 1 were presented with these chips along the rest of their assigned 165 chip set. Child participants only had these 8 chips.

Procedure. The production and comprehension tasks were both conducted in SK. In both tasks, children were seated in front of the experimenter. A table on which the color chips were display stood between them. The production task was always performed before the comprehension task.

Production task. The procedure was very similar to that of Experiment 1. Children were first introduced to the whole procedure and the general goal of the study. It was specified that they would be expected to provide color terms in SK (and not in Spanish). Children were then asked: "What is the color of this chip?". As with adults, we used follow-up questions to elicit basic color terms when the terms children initially provided were not basic. When children provided Spanish color terms, the experimenter would write down their response but further ask: "What is the name of this color in SK?" When children replied "I don't know" to this prompt, the experimenter would not ask further questions and would move forward to the next color chip. As a result, responses of some children include only non-basic SK color terms or Spanish color terms. In total, we collected production data for 8 color chips. For each chip, the data include either one response (when children provided a SK basic color term in the first trial) or two or three responses (when children's initial responses were either non-basic and/or in Spanish).

Further, Experiment 1 showed that for some of these color terms, only one response was accurate, while for others, several responses were equally correct. For example, responses during Experiment 1 to a particular purple chip ranged from red to blue with some using the terms ami ("flower") or pua ("yam") as common descriptors. Accuracy was coded based on the results derived from Experiment 1: if at least 15% of participants in Experiment 1 labeled a chip with a particular term, we considered a trial to be correct if the child made the same pairing, regardless of whether the term as a basic or ad-hoc color term.

Comprehension task. The 8 color chips of the production task were simultaneously displayed in front of the children. The experimenter would then ask: "Can you give me the [color] chip?" In total, the comprehension of 9 SK color terms was tested. The choice of these terms was based on the findings of Experiment 1. Not all of them were basic, but all of them stood out as being prominent in the SK color system. The 9 terms used as prompts included: yankon ("green/blue"), joshin ("red"), panshin ("yellow"), joxo ("white), wiso ("black"), nai ("blue"), and barin poi ("greeny-yellow"). In addition, as Experiment 1 revealed that two non-basic terms are widely used to refer to green and purple, two words were used to test comprehension of each of these two colors: pei/xo ("green") and ami/pua ("purple").

When the experimenter asked children to pick up a color that was instantiated by several chips, we followed the following procedure. The experimenter would ask: "Can you give me the [color] chip?" Children would then pick up a chip. The response would be registered and the chip be taken out of the table. As a result, only 7 chips would be remaining on the table. The experimenter would subsequently ask: "Can you give me another [color] chip?". Children would then pick up a new chip. The response would be registered and the chip be taken out of the table. The experimenter would then ask the same question again until a total of as many times as there were correct instances. Thus, for "yankon", four chips would be elicited, whereas for "joshin" or "pei/xo", two chips would be elicited. Like the preceding production task, accuracy was scored based on responses given in Experiment 1. Similar to the production task, a child's choice for a particular chip was deemed accurate if at least 15% of Experiment 1 participants made the same chip-label association. Unlike the production task, as some trials could have multiple pairings, accuracy was scored as an average, rather than dichotomous. For instance, if a child correctly chose 3 out of 4 chips for the "yankon" trial, they received a score of 0.75.

Results and Discussion

Production. Children's production accuracy increased substantially across nearly all color chips in the age range that we tested. Figure 2, top panel shows the accuracy of children's first production, both in SK (solid line) and in either language (dashed line). To quantify these developmental trends, we fit two generalized linear mixed effects models, one for the accuracy of SK production and one for the accuracy of production in either language. Both of these predicted accuracy as a function of the child's age, and included random intercepts for color chip and for participant, as well as a random slope of age by color chip. Age was a significant predictor in both models: $\beta = 1.05$, SE = 0.28, $p = 0$ and $\beta = 1.11$, SE = 0.23, $p < .0001$.

Over a quarter (28%) of all responses were given in Spanish, and the distribution of Spanish responses was non-random. Children tended to respond in Spanish when presented with a chip with low naming consensus among adult participants in Experiment. As an exploratory analysis, we attempted to quantify low naming consensus using naming entropy (following E. Gibson et al., 2017). We computed the naming entropy for each chip by computing the probabilities for each chip c to be named with a particular label l ($p(l \mid c)$) and then taking $H(c) = -\sum p(l \mid c) \log[p(l \mid c)]$ (see inset entropy values by chip in Figure 2).

To assess the hypothesis that naming entropy in adults was related to Spanish use in children, we fit a mixed effects model predicting Spanish responses as a function of age, entropy of the chip's naming distribution for adults, and their interaction. We included random intercepts for color chip and for participant, but our model did not converge with a random slope term and so we pruned this term following our lab's standard operating procedure. We found a reliable effect of entropy ($\beta = -6.09$, SE = 2.38, $p = 0.01$) and an interaction between age and entropy ($\beta = -3.97$, SE = 1.49, $p = 0.01$), suggesting that adults' uncertainty regarding naming was related to children's likelihood of producing Spanish labels.

One reason to use Spanish would be if children fail to recall the proper SK color term but do know the proper mapping in the Spanish. But another possibilitiy is that children may have more imprecise representations and choose to respond with a same-language but adjacent color term (such as "joshin" for a *panshin*-colored chip). In our next analysis, following Wagner, Dobkins, and Barner (2013), we aggregate across color chips and examine the pattern of children's first responses, categorizing them as same-language, adjacent, and different-language. This analysis is shown in Figure 3, left panel.

We fit a mixed-effects model predicting correct performance with predictors specified as above, but including only random intercepts for participants due to convergence issues). We found a significant improvement in accuracy scores when we allowed different-language but corresponding responses ($p < 0.001$) but no significant change when allowing for same-language but adjacent responses ($p = 0.409$). This result suggests that children's incorrect responding was not due to imprecise knowledge of SK terms.

Comprehension. Children's accuracy in the comprehension task increased substantially across nearly all color chips in the age range that we tested. Figure 3, top panel shows the accuracy of children's first production, both in for strict accuracy (solid line) and including chips for adjacent colors (dashed line). To quantify these developmental trends, we fit two generalized linear mixed effects models, one for the accuracy of SK production and one for choosing the accurate or adjacent chips. Both of these predicted

Table 2

Demographics of participants in Experiment 3.

Age Group	N	Male
5-years-old	2 (4%)	1 (50%)
6-years-old	2 (4%)	0 (0%)
7-years-old	11 (24%)	4 (36%)
8-years-old	9 (20%)	1 (11%)
9-years-old	11 (24%)	4 (36%)
10-years-old	8 (17%)	3 (38%)
11-years-old	3 (7%)	3 (100%)

accuracy as a function of the child's age, and included random intercepts for color chip and for participant, as well as a random slope of age by color chip. Age was a significant predictor in both models: $\beta = 0.60$, SE = 0.18, $p = 0.00$ and $\beta = 0.67$, SE = 0.19, $p < .0001$.

Experiment 3

Noting the level of bilingualism in the SK population, we designed Experient 3 as its complement. Due to the length of these experiments, however, as well as the task demands involved in testing the same children sequentially in both languages, we chose to perform this next experiment with a separate group of children. In Experiment 3, we tested children entirely in Spanish with a set of chips representing prototypical colors for the Spanish color system.

Participants. As with Experiment 2, our protocol received ethical approval from Pontificia Universidad Católica del Perú's Institutional Review Board. Children were recruited in a SK neighborhood of Yarinacocha (Bena Jema) as well as in Bawanisho. As before, children were recruited either through their parents or through the local school. When recruited at school, consent for participation was collected from both the teachers and the parents; otherwise, only consent from the parents was collected. Data were collected from a total of 46 children (16 boys) between the ages of 5 and 11 years old.

Materials. Even though participants in Experiment 3 were instructed to give color terms in SK, some Spanish color terms were provided (this was especially true of younger adult participants, who were more proficient in Spanish). Based on these data and on previous studies of Spanish color systems, we singled out 11 color chips that were prototypical instances of prominent Peruvian Spanish color terms. These color chips were grey (WCS n°46), pink (WCS n°65), orange (WCS n°121), green (WCS n°234), red (WCS n°245), brown (WCS n°266), white (WCS n°274), blue (WCS n°291), yellow (WCS n°297), black (WCS n°312) and purple (WCS n°325). These color chips were exactly the same as those used in Experiment 1; the only difference was that while 330 chips were used in Experiment 1, only 11 of them were used in Experiment 3. Six chips were shared between Experiment 2 and Experiment 3.

Procedure. Since SK children are not very fluent in Spanish, the production and comprehension tasks were both conducted in SK, and Spanish was only used for color terms (i.e., Spanish color terms were embedded in SK sentences). As in Experiment 2, the production task was always performed before the comprehension task.

Production task. The procedure was the same as that of Experiment 2. Children were first introduced to the whole procedure and the general goal of the study. It was specified that they would be expected to provide color terms in Spanish (and not in SK). Children were then asked: “what is the color of this chip?” When children provided SK color terms, the experimenter would write down their response but further ask: “what is the name of this color in Spanish?” When children replied “I don’t know” to this prompt, the experimenter would not ask further questions and would move forward to the next color chip. As a result, responses by some children include only non-basic Spanish color terms or SK color terms. For each chip, the data include either one response (when children provided a Spanish basic color term in the first trial) or two or three responses (when children's initial responses were either non-basic and/or in SK).

Comprehension task. The procedure was identical to that of the comprehension task of Experiment 2, with the exception of the set of chips and labels. In total, the comprehension of 11 Spanish color terms was tested. The choice of these terms was based on previous studies examining Spanish color terms as well as on Experiment 1. The 11 terms used as prompts included: blanco ("white"), verde ("green"), rojo ("red"), amarillo ("yellow"), azul ("blue"), negro ("black"), naranja ("orange"), gris ("grey"), morado ("purple"), marrón ("brown"), and rosa ("pink"). Since each color term was instantiated by only one color chip, no term required the special procedure that was followed in Experiment 2 for the ambiguous terms.

Results and Discussion

Production. The results of the production task are shown in Figure 2, bottom panel. Qualitatively, we saw smaller developmental effets. As in Experiment 2, we fit two generalized linear mixed effects models, one for the accuracy of SK production and one for the accuracy of production in either language. Both of these predicted accuracy as a function of the child's age, and included random intercepts for color chip and for participant, as well as a random slope of age by color chip. Age was not a significant predictor in either model: $\beta = 0.32$, SE = 0.20, $p = 0.11$ and $\beta = 0.43$, SE = 0.16, $p < .0001$.

Similar to Experiment 2, over a quarter of all responses (M = 28%, SD = 18%) were given in another language (Shipibo in this case). There was significant variation in language-switching with some children completing the entire task in Spanish while others responded to upwards of 59% of trials in Shipibo.In addition, similar to Experiment 2, we found that participants tended to respond in Shipibo when presented with items that had low entropy among SK adults during Experiment 1 ($p = 0.006$). This suggests that participants across Studies 2 and 3 preferred to respond in Shipibo when presented with a high-consensus chip and in Spanish when shown a low-consensus chip.

Also following our analysis in Experiment 2, we adopted alternative scoring to accommodate language-switching from Spanish to Shipibo-Konibo and same-language adjacent responses. Results are shown in Figure 3, right panel. Using a mixed-effects model, we did not find that age explained a significant amount of the variation seen in accuracy ($p = 0.124$), in concordance with earlier analyses. However, we did find that participants made use of *both* alternative strategies, either providing SK responses ($p < 0.001$) or same-language, adjacent responses ($p = 0.002$). In other words, in both Experiment 2 and 3, we find frequent use of language switching but only Experiment 3 shows significant use of djacent terms as well.

We speculate that the findings of Experiment 3 – the lack of developmental increases and the increasing use of adjacent Spanish terms – are a function of the nature of second-language exposure in Spanish. SK children are often exposed to Spanish at a young age, but they do not receive any formal Spanish education until later in adolescence. With a limited knowledge of Spanish color terms, children may spontaneously provide Spanish color terms during the SK-language Experiment 2 for those mappings they know but may still struggle to succeed during Spanish-language Experiment 3. More generally, we see children relying on a mixture of strategies to communicate colors even in the absence of complete knowledge in either language.

Comprehension

Unlike the production task for Experiment 3, children's accuracy in the comprehension task increased substantially across nearly all color chips in the age range that we tested. Figure 3, top panel shows the accuracy of children's first production, both in for strict accuracy (solid line) and including chips for adjacent colors (dashed line). To quantify these developmental trends, we fit two generalized linear mixed effects models, one for the accuracy of SK production and one for choosing the accurate or adjacent chips. Both of these predicted accuracy as a function of the child's age, and included random intercepts for color chip and for participant, as well as a random slope of age by color chip. Age was a significant predictor in both models: $\beta = 0.64$, SE $= 0.22$, $p = 0.00$ and $\beta = 0.49$, SE $= 0.17$, $p < .0001$.

General Discussion

TO BE PASTED FROM GOOGLE DOC

References

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., … Conway, B. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, *114*(40), 10785–10790.

Kay, P., Berlin, B., Maffin, L., Merrifield, W. R., & Cook, R. (2009). *The world color survey*. Stanford, CA: Center for the Study of Language; Information.

Wagner, K., Dobkins, K., & Barner, D. (2013). Slow mapping: Color word learning as a gradual inductive process. *Cognition*, *127*, 307–317.

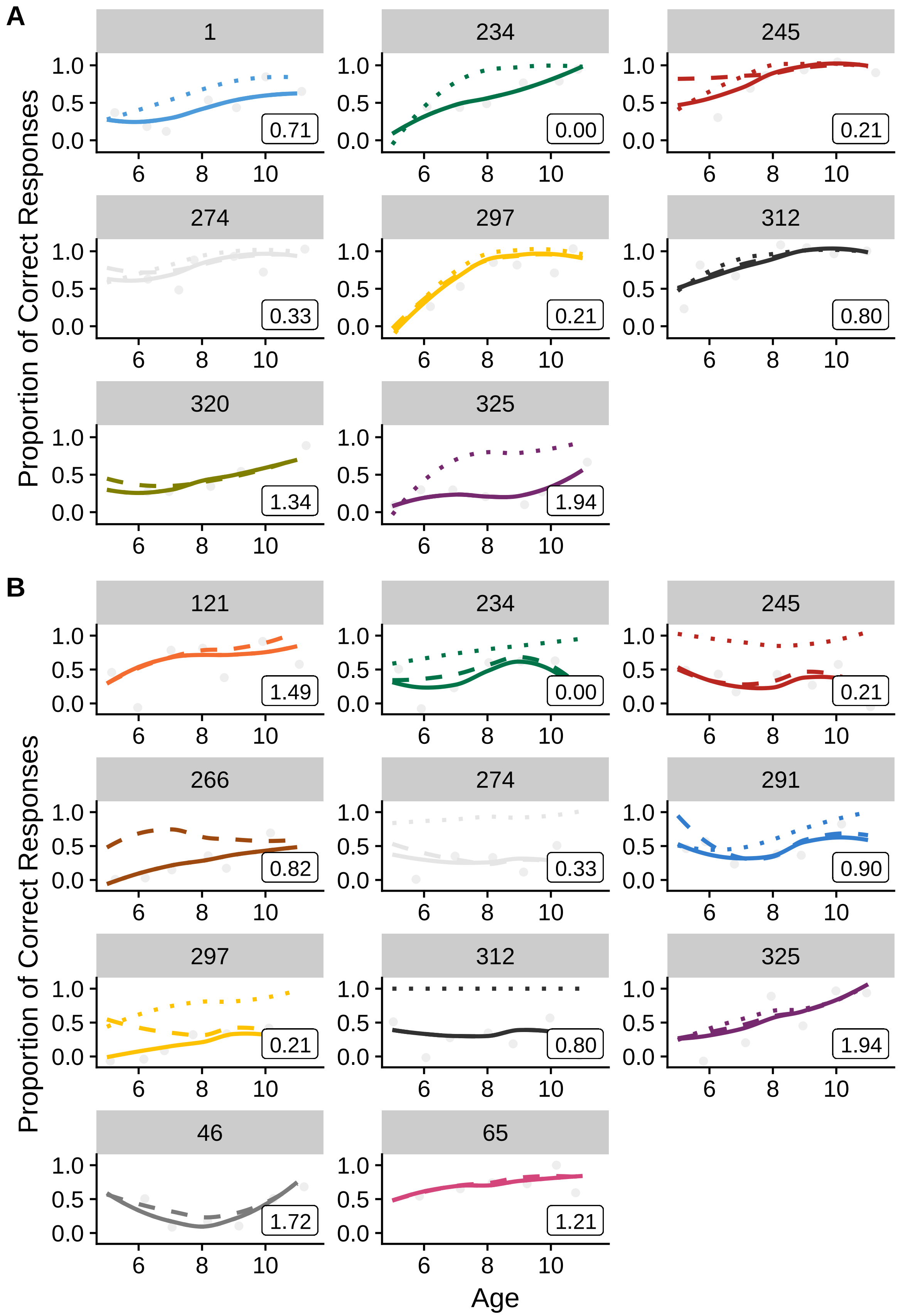

Figure 2. (#fig:prod_childfigure)(A and B) A comparison of children's performance during the production task in Studies 2 (top) and 3 (bottom). Solid or dotted lines represent

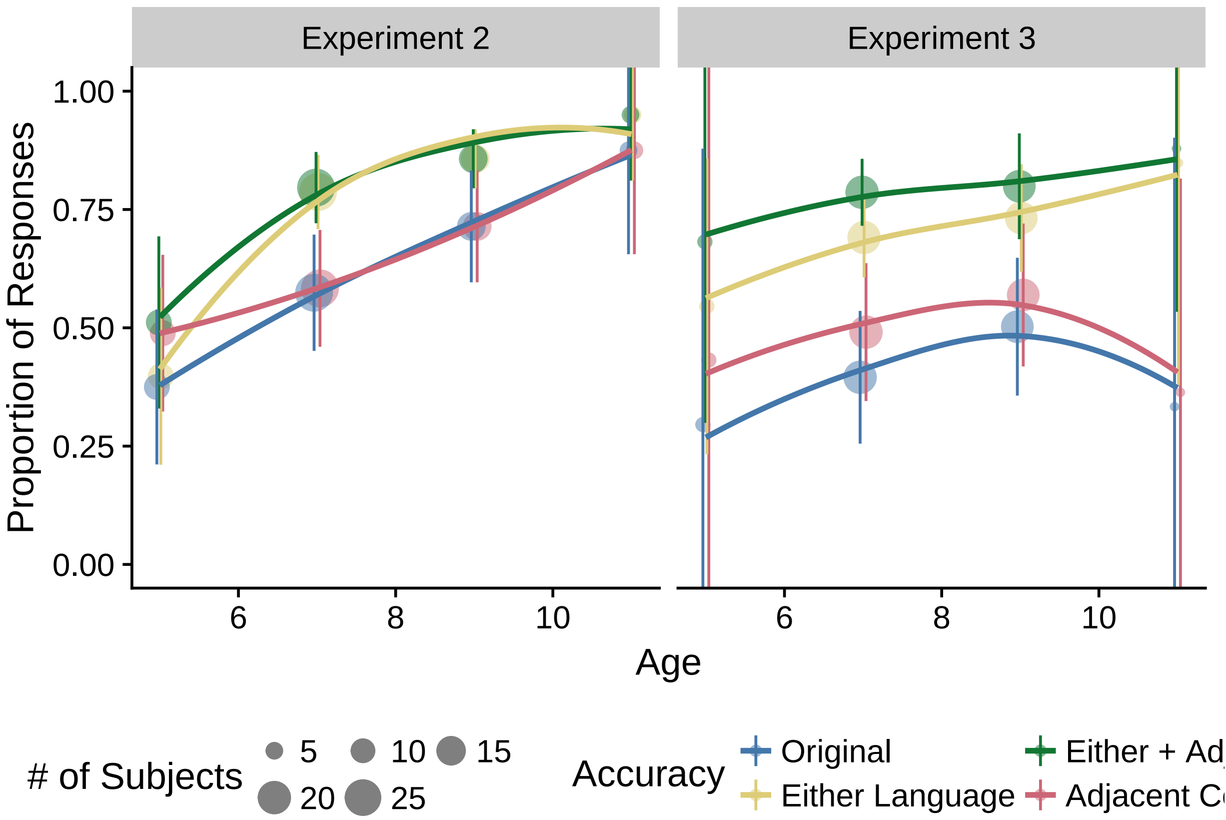

Figure 3. (#fig:study23accuracyplots_prod)Proportion of accurate responses when applying different accuracy criteria, by age and experiment. Points show the mean for a 2-year age group (chosen arbitrarily for visualization) with 95% confidence intervals. Lines show a loess smoothing function.

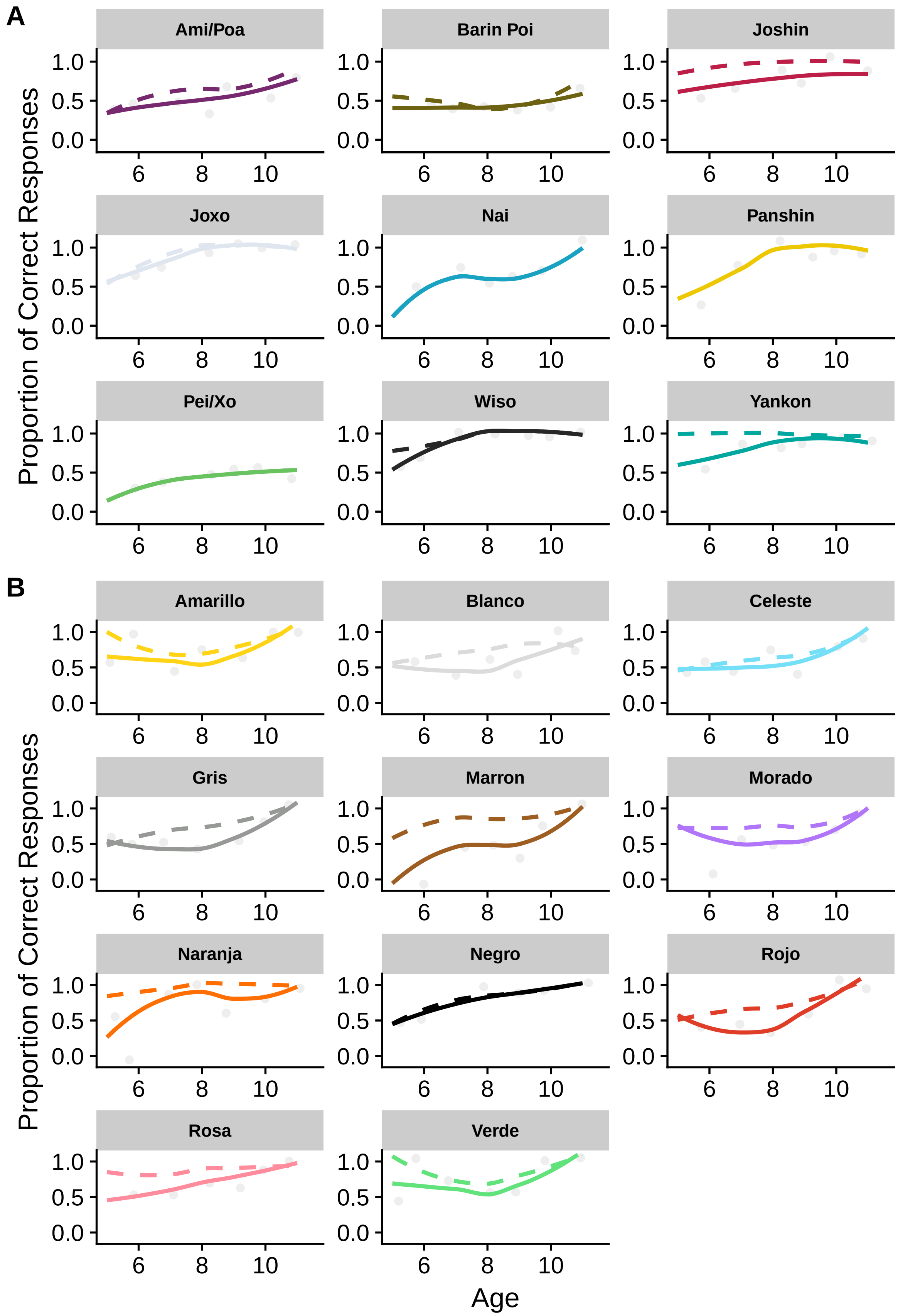

Figure 4. (#fig:comp_childfigure)(A and B) A comparison of children's performance during the comprehension tasks in Studies 2 (top) and 3 (bottom). Solid or dotted lines represent

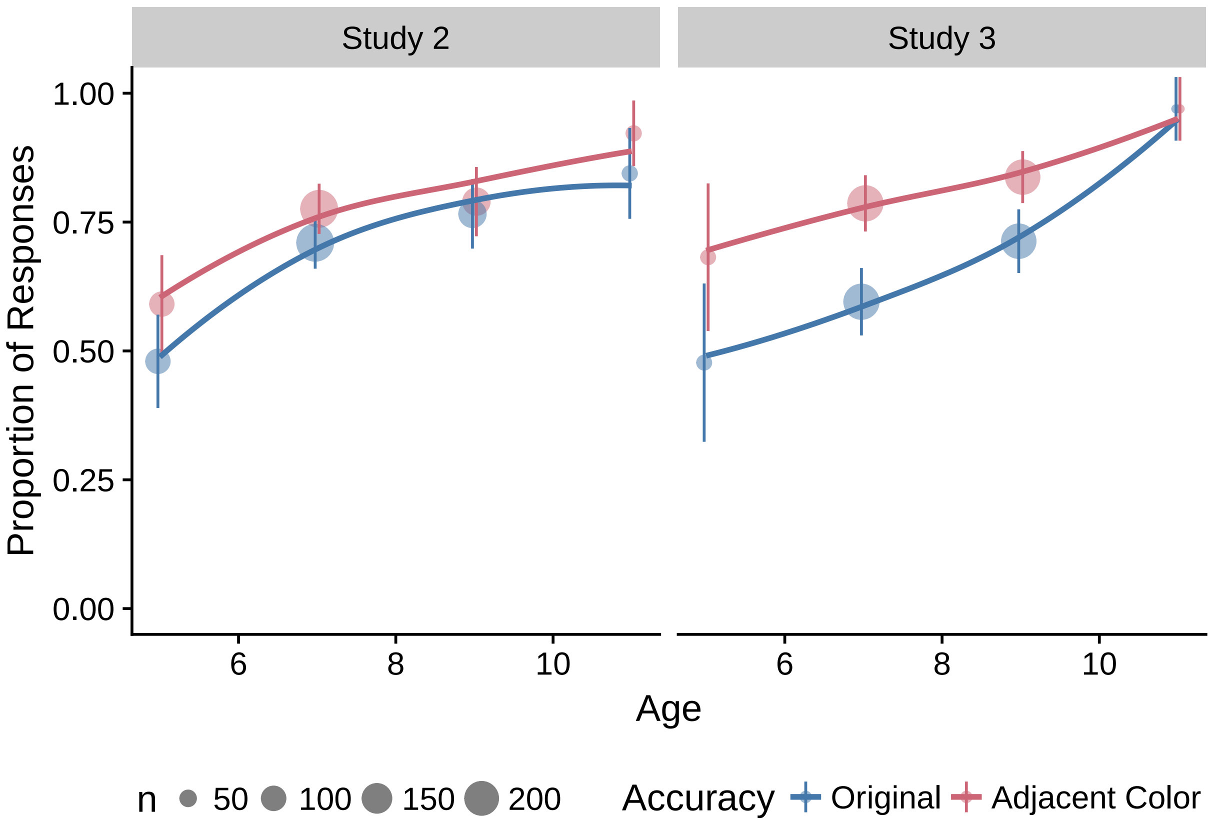

Figure 5. (#fig:study23accuracyplots_comp)Proportion of accurate responses when applying different accuracy criteria, by age and experiment. Points show the mean for a 2-year age group (chosen arbitrarily for visualization) with 95% confidence intervals. Lines show a loess smoothing function.