Iconic Artificial Language Learning in the Field: An Experiment with San Martín Peras Mixtec Speakers

Naomi Tachikawa Shapiro (naomi.shapiro@ru.nl)
Centre for Language Studies, Radboud University, 6525 HT Nijmegen, The Netherlands

Andrew A. Hedding (ahedding@uw.edu)
Department of Linguistics, University of Washington, Seattle, WA 98195 USA

Shane Steinert-Threlkeld (shanest@uw.edu)
Department of Linguistics, University of Washington, Seattle, WA 98195 USA

Abstract

The present study examines the feasibility of conducting *iconic* artificial language learning (ALL) experiments in a fieldwork setting. We taught the pictographic language from Shapiro and Steinert-Threlkeld (2023) to speakers of San Martín Peras Mixtec in Oaxaca, Mexico. In a qualitative analysis, we explore whether these speakers display similar word-ordering behaviors to those observed among other populations, while developing insights for future ALL field experiments. We show that iconic ALL offers a promising path forward for including understudied communities in the cognitive sciences.

Keywords: artificial language learning; field experiments; Mixtec; iconicity; word order; modifiers; scope-isomorphism; typology; language universals

Introduction

Artificial language learning (ALL) has emerged as a prominent method for addressing one of the central problems in the cognitive sciences: To what extent do crosslinguistic commonalities reflect universal features of human cognition? Studying how humans learn and use artificial languages has the potential to reveal learning and representational biases that may have shaped how languages evolved (Culbertson, 2023; Fedzechkina, Newport, & Florian Jaeger, 2016; Folia, Uddén, De Vries, Forkstam, & Petersson, 2010).

Much of this work, however, has focused on speakers of English and other Indo-European languages, often from so-called WEIRD[1] societies (Henrich, 2020; Henrich, Heine, & Norenzayan, 2010a, 2010b), as part of an Anglocentrism that runs deep in the cognitive sciences (Blasi, Henrich, Adamou, Kemmerer, & Majid, 2022). In general, language communities with few speakers and little socio-political influence—also known as "small language" communities (Wagers & Chung, 2023)—are often left out of psycholinguistics research, despite their languages making up the majority of the world's typological diversity. For ALL, this may have a particularly pernicious effect: Findings meant to attest universality may in fact reflect "transfer" artifacts from sampling populations whose languages possess the crosslinguistic features in question (cf. Majid, 2023).

To expand the reach of the ALL paradigm, Shapiro and Steinert-Threlkeld (2023) introduced *iconic* ALL, replacing traditional nonce inventories with pictographic lexicons. Such an approach has several benefits. Firstly, it enables using the same linguistic stimuli with diverse language populations, facilitating crosslinguistic investigations that both measure and control for transfer effects. This can help bolster claims of universality from the ALL paradigm. Secondly, iconic stimuli render ALL experiments more accessible by making the hurdle of learning new lexical elements simpler. This accessibility, in turn, can enable experiments with shorter and less arduous training blocks, which may prove especially valuable when working with small language communities in the field.

To date, the iconic ALL approach has successfully replicated well-known modifier-ordering findings with English speakers (Shapiro & Steinert-Threlkeld, 2023). Motivated by that proof-of-concept, we report on the feasibility of using iconic ALL in a fieldwork setting. We conducted the modifier-ordering experiment from Shapiro and Steinert-Threlkeld (2023) with speakers of San Martín Peras Mixtec, an Otomanguean language spoken in Oaxaca, Mexico. To our knowledge, this is the first psycholinguistic experiment conducted with speakers of any Mixtec language and the first to use iconic ALL in the field.

In the next section, we present ALL research on modifier orders in the Noun Phrase (NP) and the hypothesized "scope-isomorphism" universal. Thereafter, we introduce Mixtec languages and speakers before describing the participants in our study. We then turn to our iPad-based experiment, detailing the design decisions we made to make the study more field-friendly. The paper concludes with a qualitative analysis of our results with $n = 7$ speakers, where we highlight the promise of iconic ALL in the field while sharing methodological lessons for future work.

ALL Work on Scope-Isomorphic NPs

Recent ALL work has sought to uncover a universal bias for *scope-isomorphism*, wherein morphosyntactic structures mirror their semantic scope relations (e.g., Culbertson & Adger, 2014; Martin, Abels, Adger, Kanampiu, & Culbertson, 2024; Martin, Holtz, Abels, Adger, & Culbertson, 2020; Martin, Ratitamkul, Abels, Adger, & Culbertson, 2019). In the case of complex NPs, the fact that demonstratives (Dem) scope over numerals (Num) and numerals over adjectives (Adj) might help explain typological facts: While the literature has attested a wide variety of NP-internal word orders, the scope-isomorphic orders N-Adj-Num-Dem and Dem-Num-Adj-N

[1] "Western, Educated, Industrialized, Rich, and Democratic"

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

account for nearly half of the world's languages, with the other scope-conforming orders accounting for much of the remaining diversity (Cysouw, 2010; Dryer, 2018).

Inspired by Poverty of the Stimulus arguments, several studies have taught participants single-modifier NPs in an artificial language (e.g., just N-Adj or N-Num), then observed how the participants ordered *multiple* modifiers (e.g., N-Adj-Num or N-Num-Adj). Crucially, the linguistic stimuli taught to the participants did not specify whether the artificial language adhered to scope-isomorphic word orders, as the participants only ever saw one modifier within any given NP. Since the participants had to generalize what they learned during training to more complex cases in the critical trials, this style of ALL has been called the Extrapolation Paradigm.

When the participants had to produce the novel multi-modifier NPs, Culbertson and Adger (2014), Martin et al. (2024, 2020, 2019), and Shapiro and Steinert-Threlkeld (2023) all found that they tended to order the words in ways that reflected semantic scope. While the studies were mostly conducted with English speakers, Martin et al. (2024, 2019) additionally worked with Thai and Kîîtharaka speakers.

Notably, Kîîtharaka is a Bantu language of Kenya that has a non-isomorphic word order (N-Dem-Num-Adj). That Martin et al. (2024) still found a preference for scope-isomorphism suggests that the bias exists even among speakers of non-isomorphic languages and may therefore be universal. The present study shares Martin et al.'s motivation to conduct ALL studies in the field, but differs by using iconic rather than auditory stimuli. We return to their work in the Discussion.

Mixtec Languages & Speakers

Mixtec, also known as Tu'un Nda'vi or Tu'un Savi, is a group of languages within the Otomanguean family spoken by approximately half a million people in the Mexican states of Oaxaca, Puebla, and Guerrero, and by more in diaspora communities across Mexico and the United States. Though language contact has led to many Spanish loanwords, Mixtec languages are substantially different from Spanish in virtually all aspects of their syntax, morphology, and phonology. Mixtec speakers are organized along a dialect continuum and there is no clear consensus on the number of distinct Mixtec varieties that exist (80 are recognized by the Mexican government; Instituto Nacional de Lenguas Indígenas, 2008). Exhibiting considerable variation, many varieties are not mutually intelligible. For these reasons, the name of a town or municipality is typically used as a modifier to localize the particular variety (e.g., "San Martín Peras Mixtec").

The present study was conducted with speakers living in the town of Ahuejutla (∼2,000 residents), located in the municipality of San Martín Peras (∼ 11,500 residents), shown in Figure 1. Within that municipality, 97% of the population speaks an indigenous language and 60% speak Spanish, meaning that roughly 37% of the population is monolingual (Instituto Nacional de Estadística y Geografía, 2020). Currently, San Martín Peras Mixtec (henceforth, SMPM) is used in nearly all aspects of community life with the exception of school, which is taught exclusively in Spanish. Most residents have a very limited understanding of English, perhaps knowing only a few words.

Figure 1: San Martín Peras (red) within Oaxaca, Mexico.

Fifty-eight percent of the population above the age of 15 is literate, and literacy is correlated with age (Instituto Nacional de Estadística y Geografía, 2020). Though a standard orthography has been developed to write Mixtec languages (Ve'e Tu'un Savi [Academia de la Lengua Mixteca], 2022), it is not taught in school and it is not yet in wide use within Ahuejutla. Consequently, most literate residents tend to read and write exclusively in Spanish.

As in Spanish, adjectives follow nouns in SMPM (1), while numerals precede the noun (2). SMPM lacks most nominal morphology (e.g., number is not overtly marked on nouns).

(1) *tása kuá'à*
mug red
'red mug'

(2) *ivi tomi tujŭ*
two feather black
'two black feathers'

Unlike in Spanish, SMPM demonstratives double as locatives (e.g., *yó'o* meaning *this/here*) and surface postnominally, yielding the scope-isomorphic order Num-N-Adj-Dem. (In Spanish, the order is Dem-Num-N-Adj.)

Participants & Community Engagement

The present study was conducted in August 2023 during a field trip that included another iconic ALL experiment as well as separate projects to document and describe SMPM. Participants were recruited via word-of-mouth and by a loudspeaker announcement from the town council. Those recruited participated in a wide range of elicitation and documentation tasks, at which point they were invited to participate in the ALL studies. Participants were paid 50 pesos (roughly 3 USD) to participate in each experiment.

Out of eight speakers who participated in the present study, three were female and five were male. All of the participants were bilingual in SMPM and Spanish, though they were monolingual in SMPM at least until age 5. They each reported using SMPM on a regular basis in the community and with family, and most mentioned using Spanish with outsiders.

Table 1: Iconic lexicon with English / SMPM translations.

Noun	Adj	Num
ball / pelóta	*red / kuá'à*	*two / ivi*
feather / tomi	*black / tujǔ*	*three / oni*
mug / tása		

They all could read and write in Spanish. No participants reported familiarity with English.

The participants ranged between 18 and 71 years old. With the exception of one older participant, they were typically in their 20s or 30s. This self-selection bias was largely due to various community obligations of older community members (e.g., fixing roads, gathering food, cooking) that kept them from participating. Moreover, the nature of the task may have made it more appealing to younger participants, as older community members tend to have less experience with digital technology. (All of the participants had used a smartphone before and four had previously used an iPad.)

In general, community obligations made large-scale recruitment within a short window relatively challenging. However, once the participants agreed to participate, they uniformly completed the task and expressed willingness to complete similar tasks in the future.

Methods

We adapted the experiment from Shapiro and Steinert-Threlkeld (2023). The participants learned bare and single-modifier NPs and were then prompted to produce two-modifier NPs. Following Martin et al. (2020), each trial depicted a girl, whom we named *Yolanda*, standing behind a table (cf. Figure 2). The participants were tasked with describing objects on the table (e.g., a red mug or two black feathers) using Yolanda's pictographic language.

Artificial Language

Shown in Table 1, we adopted a subset of the iconic lexicon from Shapiro and Steinert-Threlkeld (2023), only teaching the participants nouns ($\times$3), adjectives ($\times$2), and numerals ($\times$2). The participants were randomly assigned to one of two conditions, learning either noun-initial or noun-final NPs. This differentiated the language from SMPM, where numerals and adjectives straddle the noun. As we unpack in the Discussion, including both conditions further controls for ordering strategies where participants preserve the relative linearization of modifiers found in SMPM, ordering them Num-Adj, regardless of the noun's placement.

Procedure

We adapted the jsPsych (de Leeuw, 2015) interface from Shapiro and Steinert-Threlkeld (2023). Our study consisted of three training blocks (15 trials) and one testing block (12 trials), totaling four blocks (27 trials). CSS animations guided the participants through the study.

We deployed the experiment on GitHub Pages and made the website available for offline usage through simple JavaScript.[2] Conducting the experiment from an iPad made the experiment portable and easy to adapt on the fly, provided intermittent WiFi. Results were downloaded as JSON files and stored in the Jayson app until they could be transferred to a computer.

Verbal Instructions At the onset of the experiment, the participants listened to instructions pre-recorded in SMPM by a native speaker, as scripted in (3). While the instructions played, the participants were shown a picture of Yolanda.

(3) SMPM instructions: *Ká'àn Yolanda tù'un xí'in ñà tsiáà nǔnàna ñà'a. Tsiáa nà míí tù'un nà táko'va ntsíkùn tǎ'án míí á. Ntù sákua'ǎ ntù xà kóto'và inka yivǐ tù'un nà. Kônì ntù tyintsie ún ntù. Koñǔn kuà'a nùnána Yolanda xíto'ni ñá ñà'a íntosǒ nùjǔ mésa. Tá iin iin mía nàna, kônì ntù kuntàà ini ntù: ¿xá ká'ún kà'àn míí Yolanda nàní míí ñà'a tù'un míí ñá?*

"Yolanda speaks a language that is written with pictures of things. Her language is written in a consistent way. We are studying how other people learn her language. We want you to help us. You will see various pictures of Yolanda with things lying on a table. For each picture, we want to find out: How do you think Yolanda would call each thing in her language?"

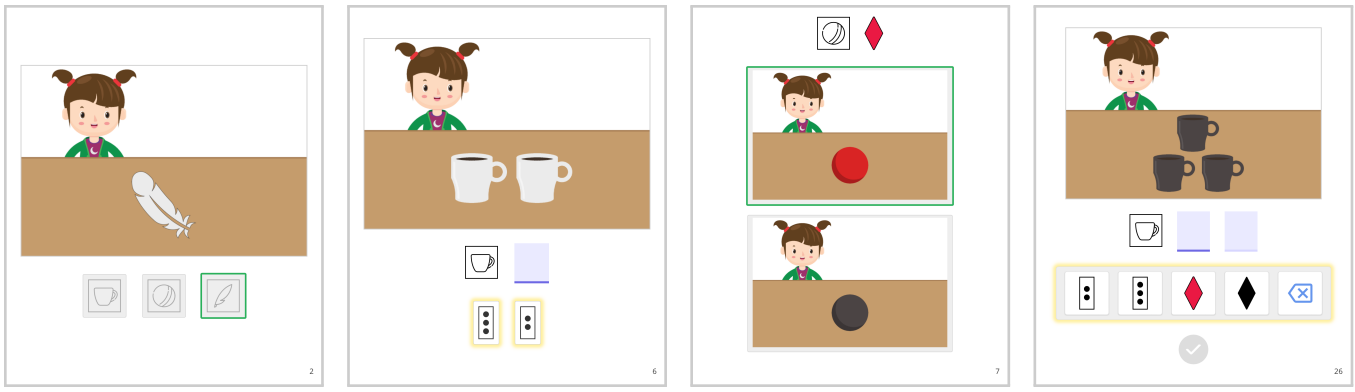
Demo Block We introduced a demo block (8 trials) at the start of the experiment to familiarize the participants with the experiment's forced-choice and production-style tasks. First, the experimenter walked the participant through one trial per each of the four trial types (described in the next section). The four trials then repeated, giving the participant an opportunity to step through the trials with the experimenter's guidance. The demo trials only questioned the noun glyphs.

We purposefully designed the interface with easy-to-reference bright colors. This facilitated the experimenter guiding the participants through the demo trials (e.g., 'tap a yellow one' to make a selection or 'tap the blue one' [a backspace key] to remove an item).

Training Blocks The training blocks drew on the task types from Shapiro and Steinert-Threlkeld (2023) and used an "active learning" design where the participants had to infer the meanings of the glyphs. Animations informed the participants whether they had answered a trial correctly and further prompted them to fix any mistakes.

Block 1 focused on *noun learning* via three "glyph-selection" trials (one per noun; Figure 2a). Block 2 then turned to *modifier learning*, introducing the four modifiers—as well as the condition's word order—through eight forced-choice trials. For each modifier, the participants completed a "cloze"

[2]GitHub repo: https://github.com/tsnaomi/iconic-all-cogsci-2024.



(a) Block 1 glyph-selection task (b) Block 2 cloze task (c) Block 2 picture-selection task (d) Block 4 production task

Figure 2: Experiment trials with visual stimuli from Martin et al. (2020). The trials depicted in (b)–(d) come from the noun-initial condition. In trials (a) and (c), the correct answers are marked in green.

task (Figure 2b) and a "picture-selection" task (Figure 2c).

Finally, Block 3 introduced the "production" task, where the participants practiced producing a *one-modifier NP* for each of the four modifiers. This block was meant to reinforce the condition's word order and familiarize the participants with the production task format. The trials presented the participants with a picture, for which they had to produce a caption using a pressable keyboard on the screen. The keyboard included the relevant noun, the target modifier and the other modifier of the same type, and a "backspace" key on the far right. The lexical keys matched the condition's word order on alternating trials. Participants could only submit a response once they entered the correct number of glyphs.

Testing Block Block 4 tested the participants on 12 critical *two-modifier NPs* (3 nouns × 4 modifier pairs) using the production task format (Figure 2d). In contrast to Block 3, the caption supplied the participants with the noun. The keyboard keys were presented in scope-isomorphic order on alternating trials. Once again, the participants could only submit after entering the correct number of glyphs. To minimize exposition during the experiment and to simplify the overall experience for the participants, the testing block flowed seamlessly from the training block. To keep the experience consistent, all of the critical trials were marked "correct" from the participants' perspective (i.e., the responses turned green), regardless of whether they had entered the relevant glyphs. We revisit this "positive feedback" decision in the Discussion.

Preliminary Results

For inclusion in our initial analysis, we required participants to produce at least nine (out of 12; 83%) analyzable two-modifier NPs, such that their responses included the correct modifiers (in either order). Seven out of eight participants met this threshold, each producing 12 (out of 12; 100%) analyzable critical NPs and taking on average 7.2 minutes to complete the experiment (range: 4.1–11.7). The one remaining participant produced zero analyzable critical NPs and took 16.7 minutes to complete the experiment (noun-final condition).

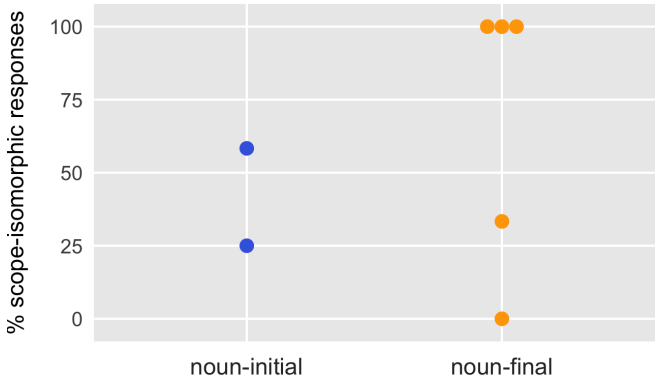


Figure 3: Per-participant percentage of responses that were scope-isomorphic. Each dot represents a participant.

Out of the 84 analyzable two-modifier NPs, 59.6% were scope-isomorphic (noun-initial: 41.7%; noun-final: 66.7%). Figure 3 visualizes the per-participant percentage of responses that were scope-isomorphic. Three of the participants uniformly produced scope-isomorphic NPs, while one participant consistently preferred non-isomorphic NPs. Given our small sample size, we do not perform a regression analysis to see whether the participants significantly favored scope-isomorphic NPs.

Discussion

Shapiro and Steinert-Threlkeld (2023) proposed that iconic ALL can help broaden language representation in psycholinguistics, leading to a clearer understanding of cognitive universals (cf. Majid, 2023). Pictographic lexicons may prove more accessible to diverse groups, since their meanings are intuitive, they need not conform to the phonological systems of participating speakers, nor do they require carefully composed audio recordings. They further open up the possibility of using the same artificial language with different language

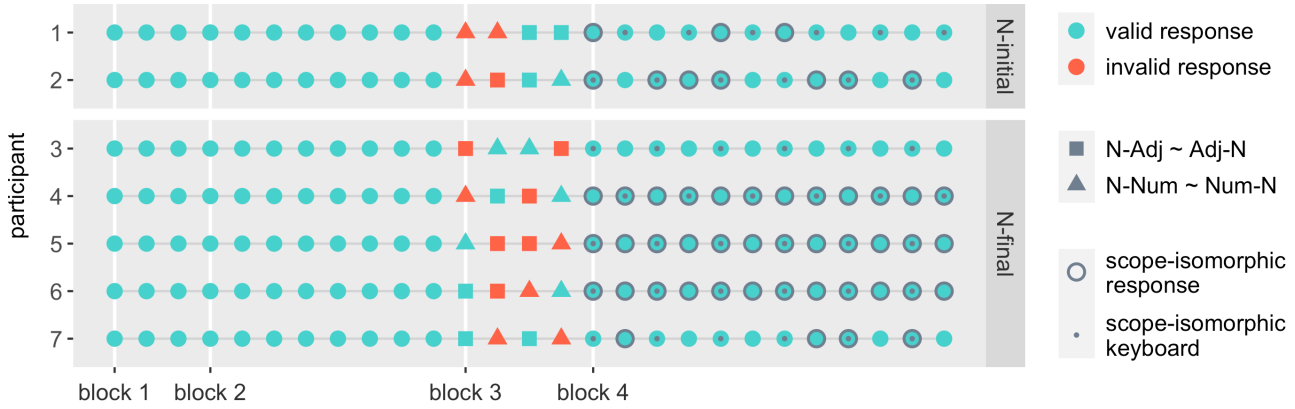


Figure 4: Trial-by-trial responses for the $n = 7$ participants included in the analysis. "Invalid responses" included incorrect glyphs and word orders (where applicable). In the error-prone Block 3, the participants practiced producing one-modifier NPs.

populations, enhancing experimental control in crosslinguistic comparisons. Shapiro and Steinert-Threlkeld thus called for future work to vet iconic ALL among typologically, culturally, and sociologically diverse communities.

We therefore explored the feasibility of conducting iPad-based iconic ALL experiments in the field, examining the modifier-ordering behaviors of SMPM speakers in Oaxaca, Mexico. Though we refrain from drawing conclusions about whether this community prefers scope-isomorphic NPs, the present small-scale study demonstrates the accessibility and scalability of iconic ALL for addressing these types of questions with indigenous communities often excluded from consideration in psycholinguistics research. In this Discussion, we focus on methodological insights from our study.

Iconic ALL & the Extrapolation Paradigm

Seven out of eight participants met the inclusion criteria for our analysis, producing analyzable responses for *all* 12 critical trials. For comparison, in their modifier-ordering study with Kîîtharaka speakers, Martin et al. (2024) excluded the majority of their participants for not producing enough on-task responses or for struggling to remember lexical items during the critical trials.[3] Our high inclusion rate suggests than iconic ALL studies may scale well in the field compared to traditional ALL formats. This is likely due to pictographic systems being easier and faster to learn than nonces. Indeed, upon examining the participants' trial-by-trial responses (Figure 4), we found that they got *zero* items incorrect during the first two training blocks, which were dedicated to noun and modifier learning.

However, it seems our participants struggled to learn the noun-initial and noun-final word orders. The qualifying participants only made mistakes during Block 3, where they had to produce one-modifier NPs. These were predominantly ordering errors (13 out of 15), where the participants reversed the condition's word order.[4] Since they were not required to produce the nouns in the critical trials, this raises the question whether the participants fully learned the artificial language. Future work might extend Block 3 to reinforce the word order, while requiring participants to provide the full two-modifier NPs—including the noun—in Block 4.

Our decision to supply the nouns in Block 4 was made to simplify the experiment: Unaware of prior psycholinguistic work with Mixtec communities and how members might respond to the study, we tried to make it as accessible as possible. For this reason also, we provided "positive reinforcement" during the critical trials, which allowed the testing block to flow seamlessly from the training blocks. A downside of this decision, however, is that some participants may have committed to the Adj∼Num order they entered on the first critical trial once they saw it marked "correct". Such responses would not bear on a scope-isomorphism bias. Studies typically get around this issue by telling participants that they will be tested on new items (e.g., longer phrases).

This highlights a key tension of doing these types of studies in the field. Central to the Extrapolation Paradigm is the need to *test* participants on items they hadn't seen during training. Crucially, that this testing phase is often made explicit can be at odds with conducting culturally-sensitive experiments in a fieldwork setting. Nonchalance with test-like activities may be specific to university students and participants crowdsourced online, whereas such experiences can be fraught for members of small language communities (Wagers & Chung, 2023). Experiments often presuppose that participants are familiar or comfortable with being tested. In a community where this presupposition is not met, test-like tasks can be uncomfortable

[3]While the vocabulary size in Martin et al. (2024) was the same as ours, they taught demonstratives instead of numerals. Another difference is that their critical trials further challenged participants to produce the nouns in addition to the two modifiers (all orally). Lastly, our participants took on average 7.2 minutes to complete the entire experiment, while their training phase was spread over two days.

[4]The "word order" errors were balanced across the modifier types: seven with numerals (2 noun-initial, 5 noun-final) and six with adjectives (1 noun-initial, 5 noun-final). Upon re-analyzing the data from Shapiro and Steinert-Threlkeld (2023), we found that their Block 3 was likewise the most error-ridden, suggesting that this block is primarily responsible for instilling the artificial word order.

for participants and have the potential to harm relationships between the research team and the community.

A possible solution is to conduct extrapolation experiments in two distinct stages, as is canon, but to reframe the testing instructions in terms of asking for guidance: For example, "We're unsure how Yolanda would describe the following pictures. What do you think?" This premise would enable withdrawing feedback during the critical trials without it being jarring. Furthermore, it implies *truthfully* that there are no right answers and that the participants are in a position to help us. This setup might pair especially well with iconic lexicons, because the participants are unlikely to fret about glyph meanings, compared to recalling traditional nonces.

Transfer Effects & Scope-Isomorphism

An ever-present concern is that participants will develop strategies based on subverting the word order of their primary language, a form of transfer effect. For instance, some of our participants may have preserved the *relative* order of modifiers found in SMPM (Num-N-Adj), placing the numeral before the adjective, regardless of where the noun appeared in the artificial language. Conveniently, this would predict a non-isomorphic preference in the noun-initial condition (N-Num-Adj) and a scope-isomorphic preference in the noun-final condition—where the latter responses are scope-isomorphic (Num-Adj-N) for reasons *unrelated* to semantic scope. Our preliminary results are not incompatible with this story: In our oral debriefing questionnaire, three noun-final participants did note placing the 'numeral first', with one going as far to say 'like the order in Spanish'. It could be that this prediction would be borne out with more data.

Noun-final scope-isomorphic NPs also align with the linear order of nominal modifiers in Kîîtharaka (N-Dem-Num-Adj). Since prior work (Martin et al., 2020) has shown English speakers to suppress the linear order of English modifiers during ALL, Martin et al. (2024) argued that their Kîîtharaka speakers likewise ignored the linear order of Kîîtharaka modifiers in their study. Yet, behaviors during these studies could be influenced by task familiarity, differing notions and subsequent accommodation of experimenter expectations, as well as culturally-mediated factors (Wagers & Chung, 2023). In short, it may be that language users vary cross-culturally in the strategies they recruit during ALL, posing problems for claims of universality. This ambiguity underscores the importance of measuring transfer effects and asking participants explicitly about their thoughts and ordering strategies, especially in the field where participants may behave less like we expect (if we're accustomed to university students and online participants).

Although no clear scope-isomorphism preference surfaced in our small sample, extant research suggests that such a bias may be universal (Culbertson & Adger, 2014; Martin et al., 2024, 2019; Saldana, Oseki, & Culbertson, 2021), predicting that a refined experiment and more data from SMPM speakers will support this generalization. We will revisit this question in the future. If SMPM speakers genuinely prefer scope-isomorphic NPs (both noun-initially *and* noun-finally), we will have one more piece of evidence of this bias being universal (though NPs in SMPM are already scope-isomorphic, inviting *structural* transfer). Either outcome would reaffirm the importance of studying typologically and culturally diverse communities in crosslinguistic investigations before positing cognitive universals (Majid, 2023). Moreover, it would be interesting to explore whether—and why—populations might vary in their ALL strategies and the degree to which transfer effects emerge in these types of investigations.

Digital Literacy

We conducted the present study on an iPad, using an interface that relied on CSS animations. Pew Research Center has found that, in many emerging and developing countries, more than half of the population uses smartphones, with computer access being much rarer (Silver et al., 2019). This is consistent with what we observed in Ahuejutla, where smartphones were quite common and, anecdotally, few had personal computers. Together, these observations suggest it will be easier for speakers in similar communities to engage with activities on a touchscreen, compared to computer-based experiments.

At the same time, device familiarity is not guaranteed and is further correlated with younger populations (Silver et al., 2019). Digital interfaces may thus be more foreign for older members in these communities, making it harder for them to follow along in digitized experiments and disinclining them from participating. Notably, we struggled with recruiting older participants in the present study and the one participant whose data we excluded from the analysis was our one older participant (71 years old). While iPad experiments are more portable and more easily revised in the field, this sociological dimension presents a trade-off. For example, older members of San Martín Peras tend to be more monolingual in SMPM, which may be essential for some research questions.

Note that the iconic methodology does not require an iPad nor a digital interface. While digital literacy may have limited the present data collection, future studies might experiment with iconic *analog* materials for older participants.

Conclusion

We conducted an iconic ALL experiment with speakers of San Martín Peras Mixtec, illustrating the potential of this methodology in a fieldwork setting. Running a simplified version of the modifier-ordering study from Shapiro and Steinert-Threlkeld (2023), we showed that SMPM speakers engaged with the pictographic stimuli in meaningful ways. We laid the groundwork for future experiments with SMPM speakers while sharing the lessons we've learned along the way. Importantly, the use of iconic ALL in the field can help pave the way for (i) expanding language representation in the cognitive sciences, (ii) comparing transfer effects across diverse communities, and (iii) more thoroughly testing claims of universality. We will pursue these avenues in future work.

Acknowledgements

We gratefully acknowledge the residents of Ahuejutla for welcoming us and allowing us to conduct this study in their community. In particular, we thank Juan Carlos Espinosa, Juan Gracida Ortiz, Margarita Cruz Salazar, and Nacho Martínez for support while in Ahuejutla. In addition, Natalia Gracida Cruz provided invaluable advice and expertise during the preparation of this experiment. We also thank Ben Eischens for logistical and moral support during our time in Oaxaca. Lastly, we are grateful to the UW Meeting on Syntax and Semantics as well as the *CogSci* reviewers for their insightful feedback. Financial support for this project came in part from a Jacobs Research Fund grant awarded to the second author.

References

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, *26*(12), 1153–1170. doi: 10.1016/j.tics.2022.09.015

Culbertson, J. (2023). Artificial language learning. In J. Sprouse (Ed.), *Oxford Handbook of Experimental Syntax* (pp. 271–300). Oxford University Press. doi: 10.1093/oxfordhb/9780198797722.013.9

Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, *111*(16), 5842–5847. doi: 10.1073/pnas.1320525111

Cysouw, M. (2010). Dealing with diversity: Towards an explanation of NP-internal word order frequencies. *Linguistic Typolgy*, *14*, 253–286. doi: 10.1515/lity.2010.010

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavioral Research Methods*, *47*(1), 1–12. doi: 10.3758/s13428-014-0458-y

Dryer, M. S. (2018). On the order of demonstrative, numeral, adjective, and noun. *Language*, *94*(4), 798–833. doi: 10.1353/lan.2018.0054

Fedzechkina, M., Newport, E. L., & Florian Jaeger, T. (2016). Miniature artificial language learning as a complement to typological data. In L. Ortega, A. E. Tyler, H. I. Park, & M. Uno (Eds.), *The Usage-Based Study of Language Learning and Multilingualism* (pp. 211–232). Georgetown University Press. doi: 10.1353/book45841

Folia, V., Uddén, J., De Vries, M., Forkstam, C., & Petersson, K. M. (2010). Artificial language learning in adults and children. *Language Learning*, *60*(s2), 188–220. doi: 10.1111/j.1467-9922.2010.00606.x

Henrich, J. (2020). *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous*. Farrar, Straus and Giroux.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). Most people are not WEIRD. *Nature*, *466*, 29. doi: 10.1038/466029a

Henrich, J., Heine, S. J., & Norenzayan, A. (2010b). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. doi: 10.1017/S0140525X0999152X

Instituto Nacional de Estadística y Geografía. (2020). *Censo de Población y Vivienda.*

Instituto Nacional de Lenguas Indígenas. (2008). *Catalogo de las Lenguas Indígenas Nacionales: Variantes Lingüísticas de México con sus autodenominaciones y referencias geoestadísticas.*

Majid, A. (2023). Establishing psychological universals. *Nature Reviews Psychology*, *2*(4), 199–200. doi: 10.1038/s44159-023-00169-w

Martin, A., Abels, K., Adger, D., Kanampiu, P. N., & Culbertson, J. (2024). A universal cognitive bias in word order: Evidence from speakers whose language goes against it. *Psychological Science*, *35*. doi: 10.1177/09567976231222836

Martin, A., Holtz, A., Abels, K., Adger, D., & Culbertson, J. (2020). Experimental evidence for the influence of structure and meaning on linear order in the noun phrase. *Glossa*, *5*(1), 1–21. doi: 10.5334/gjgl.1085

Martin, A., Ratitamkul, T., Abels, K., Adger, D., & Culbertson, J. (2019). Cross-linguistic evidence for cognitive universals in the noun phrase. *Linguistics Vanguard*, *5*(1), 20180072. doi: 10.1515/lingvan-2018-0072

Saldana, C., Oseki, Y., & Culbertson, J. (2021). Cross-linguistic patterns of morpheme order reflect cognitive biases: An experimental study of case and number morphology. *Journal of Memory and Language*, *118*, 104204. doi: 10.1016/j.jml.2020.104204

Shapiro, N. T., & Steinert-Threlkeld, S. (2023). Iconic artificial language learning: A conceptual replication with English speakers. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45).

Silver, L., Smith, A., Johnson, C., Taylor, K., Jiang, J., Anderson, M., & Rainie, L. (2019). Mobile connectivity in emerging economies. *Pew Research Center*.

Ve'e Tu'un Savi [Academia de la Lengua Mixteca]. (2022). *Norma de Escritura del Tu'un Savi (Idioma Mixteco)*. Instituto Nacional de Lenguas Indígenas.

Wagers, M., & Chung, S. (2023). Language-processing experiments in the field. In J. Sprouse (Ed.), *Oxford Handbook of Experimental Syntax* (pp. 491–512). Oxford University Press. doi: 10.1093/oxfordhb/9780198797722.013.15