UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Numbers and counting: Silent gesture and artificial language learning do not always reflect typological patterns

Permalink

https://escholarship.org/uc/item/1nr0d12r

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

Authors

Antono, Gregory
Yim, Ashley
Chambers, Craig
et al.

Publication Date

2025

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Numbers and counting: Silent gesture and artificial language learning do not always reflect typological patterns

Gregory Antono
Dept. of Linguistics
University of Toronto
gregory.antono@mail.utoronto.ca

Ashley Yim
Ontario Institute for Studies in Education
University of Toronto

Craig Chambers
Dept. of Psychology
University of Toronto

Daphna Heller
Dept. of Linguistics
University of Toronto

Abstract

In classifier languages, a sequence consisting of a noun (N), a numeral (Num), and a numeral classifier (CL) could in principle occur in one of six possible word orders. However, the cross-linguistic distribution of these word orders is highly uneven. Specifically, classifier languages tend to use Num-CL orders and, furthermore, N-medial orders are completely unattested in the world's languages. We use an artificial language learning paradigm (Experiment 1) and a silent gesture paradigm (Experiment 2) to test the hypothesis that typological patterns arise from cognitive biases at the level of individual speakers. In contrast to studies that examined coarser grained ordering effects, our results do not align with typological preferences. We consider the possibility that cognitive biases might not play a role in "finer grained" ordering phenomena involving units such as classifiers, whose role in an utterance is more about grammatical well-formedness than a strong contribution to meaning.

Keywords: classifiers; word order; noun phrase; artificial language learning; silent gesture; linguistic universals

Introduction

It has been known since at least Greenberg (1963) that some typological patterns are more common than others, and in fact some are completely absent from the world's languages. For example, within the Noun Phrase (NP), when demonstratives, numerals and adjectives precede the noun, they tend to occur in that order, whereas if they follow the noun, the order tends to either remain the same, or completely reverse (Universal 20).

Other typological tendencies regarding word order within the Noun Phrase concern the positioning of *classifiers*. In some of the world's languages, numerals cannot occur with only a noun (the equivalent of *"three dogs"*), and instead require the inclusion of a classifier (Greenberg 1990 [1972], Cheng & Sybesma 1998; Chierchia 1998; Aikhenvald 2000). This is demonstrated in (1) with examples in Mandarin (classifiers are henceforth glossed as CL).

(1) a. *yi zhī māo*
one CL cat
'one cat'

b. *liang tiáo sheng-zi*
two CL rope
'two ropes'

c. *sān píng guo zhī*
three CL fruit juice
'three bottles of juice'

d. *si duī shù-zhī*
four CL tree-branch
'four piles of branches'

The *sortal* classifiers in (1a-b) reflect properties of the referent: *cat* occurs with a classifier for animate entities, whereas *rope* occurs with a classifier often used for long and thin objects. The *mensural* classifiers in (1c-d) reflect how the referent is individuated or "packaged": *juice* occurs with a classifier that measures out bottle (which is the same as the noun for "*bottle*"), whereas *tree branch* occurs with a classifier that denotes a referent of a collective nature (i.e., not individual tree branches). More generally, while classifiers often reflect physical properties of the referent, such as size, shape, and length, this is not always the case, and so classifier-noun parings require rote memorization (Erbaugh, 1996).

Here we focus on the *linear order* of nouns, numerals and classifiers within the noun phrase, which could logically occur in six different orders. Based on a survey of one hundred languages, Greenberg (1972) first pointed out that the two orders where the numeral and the classifier occur on *different* sides of the noun are not found in the world's languages (Table 1, highlighted in red: Medial1 and Medial2). A second generalization concerns the order of the numeral and classifiers in languages where both occur either before or after the noun: the orders where the numeral *precedes* the classifier (green: Frequent-Final and Frequent-Initial) are much more common compared to the two orders where the numeral *follows* the classifier (yellow: Rare-Final and Rare-Initial). Additional evidence for this generalization comes from Her et al. (2015, 2019) who found the order Num-CL in 86% of 194 languages from the SMATTI families (Sinitic, Maio-Yao, Austroasiatic, Tibeto-Burman, Tai-Kadai, and Indo-Aryan).

Table 1: The six logical orders of Noun (N), Numeral (Num) and classifier (CL) and their typological status.

Order	Examples	Label
Num CL **N**	Mandarin, Japanese	Frequent-Final
N Num CL	Thai, Shan (Kra-Dai)	Frequent-Initial
CL Num **N**	Ibibio (Niger-Congo)	Rare-Final
N CL Num	Jingpho (Tibeto-Burman)	Rare-Initial
CL **N** Num	No languages attested	Medial1
Num **N** CL	No languages attested	Medial2

An interesting hypothesis for why certain patterns are more common across the world's language is that they arise from cognitive biases at the level of individual speakers. In other words, when crafting a message, individuals are biased to use certain word orders (e.g., ordering subjects before objects in sentences, or nouns before adjectives the noun phrase level) and these preferences get encoded as linguistic conventions over time (Typological Prevalence Hypothesis: Gentner & Bowerman, 2009). If this hypothesis is correct, then

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

typological patterns should be observed in cases where individuals improvise language (i.e., in thew absence of existing linguistic conventions).

One methodology used to examine cognitive biases at the level of individuals is artificial language learning (Hudson Kam & Newport, 2005, 2009). Focusing on word order in the Noun Phrase, Culbertson and Adger (2014) pioneered an 'extrapolation' paradigm: English speakers were first exposed to NPs with a numeral *or* an adjective, and later asked to produce NPs with both a numeral *and* an adjective. Participants spontaneously produce the typologically more common patterns, even when that meant that producing orders that do not exist in their native language.

Another methodology used to test the Typological Prevalence Hypothesis is improvised gesture (Goldin-Meadow et al., 2008, et seq.). For example, speakers of various languages produced Subject-Object-Verb (SOV) by default, even if this was not the order in their native language (Goldin-Meadow et al., 2008, Gibson et al., 2013), echoing the dominance of SOV word order across the world's languages (41%; Dryer, 2013). Similarly, Culbertson, Schouwstra and Kirby (2020) found that, when speakers spontaneously gesture complex noun phrases (i.e., determiner, numeral, adjective and noun as in '*these four striped triangles*'), they tended to favor orders that are typologically more common (see also Jaffan, et al., 2020).

Our goal here is to examine whether the typological preference for certain word orders concerning classifiers is also observed at the level of individuals. This question has been examined by Wang (2024) who conducted three artificial language learning experiments and did *not* find a preference for the typologically-preferred orders. We consider the possibility that this is due to methodological choices. First and most importantly, participants were directly instructed about the meaning of classifiers, and were explicitly told that the classifiers were required for counting. As such, participants did not spontaneously conceptualize classifier meanings in relation to the noun and the numeral. In addition, the paradigm was such that participants were presented with the noun, and had to add the numeral and the classifiers, meaning that only N-initial orders were considered. It is possible that these methodological choices mask any cognitive biases that drive word order.

To avoid explicit instruction and instead allow participants to spontaneously learn classifier meanings, we chose to focus on group-like classifiers (e.g., 1c and 1d), which, unlike sortal classifiers (e.g., 1a and 1b), have a clear semantic meaning. In noun phrases like *three bottles of juice* or *four piles of branches*, classifiers like *bottle* and *pile* do not simply echo physical properties of the referent, but instead they specify a certain grouping of the object (see Lehrer, 1986). Because these meanings exist in all languages, this may lead to a situation where participants simply use the word order in their native language; however, previous studies have clearly demonstrated that speakers can prefer an order that is not attested in their native language (e.g., Goldin-Meadow et a., 2008). We first conducted an artificial language learning experiment (Exp. 1); then, because our experiment did not replicate typological patterns, we also conducted an improvised gesture experiment (Exp. 2).

Experiment 1: Artificial language learning

We created a miniature artificial language, and used an exposure-then-test paradigm to address how learners linearize these elements. Participants first learned the words for nouns (seven objects), classifiers (two group classifiers) and numerals (four numerals) *separately* – see Figure 1. This order of exposure was chosen so it would translate into a rare typological ordering (N-CL-Num, Rare-Initial in Table 1). During the test phase, participants had to produce combinations of the three categories of the learned lexical items, as in *"four piles/bunches of barsu"* (cf. Culbertson & Adger, 2014). We used a production task, where participants were shown the lexical items in a "word bank", and were asked to produce a phrase by clicking on the relevant lexical items in order (adopted from Fedzechkina et al., 2020). This paradigm is meant to elicit spontaneous production without explicit instruction.

Nouns	Classifiers		Numerals
barsu	benu	zimi	kup (1)
bliffen			sat (2)
doakla			vin (3)
nagid			yod (4)
rizbi			
slagum			
velmik			

Figure 1: The lexicon in Exp 1 The classifier *benu* was used for a neat bunch, and *zimi* for a disorganized pile.

If the typological patterns shown in Table 1 arise from cognitive biases at the level of individual speakers, then the word orders produced in this task should reflect typological preferences. Specifically, patterns should be dominated by orders where the numeral precedes the classifier, whether they both precede or follow the noun. In addition, orders where the numeral and the classifier occur on the two sides of the noun are expected to be dispreferred (Medial1: Num-N-CL and Medial2: CL-N-Num, both unattested typologically).

Alternatively, participants may simply repeat the order in which they learned the elements, which would yield a N-CL-Num ordering. This may be broadly compatible with the idea that learners prefer to match input distributions (Hudson Kam & Newport, 2005, 2009; Fedzechkina & Jaeger, 2020). Importantly, although this order is indeed attested cross-linguistically, it is less common than orders where the numeral precedes the classifier (Rare-Initial, in Table 1). Of course, another possibility is that, in the absence of instruction on how to order elements, participants will impose the word order of their native language (English),

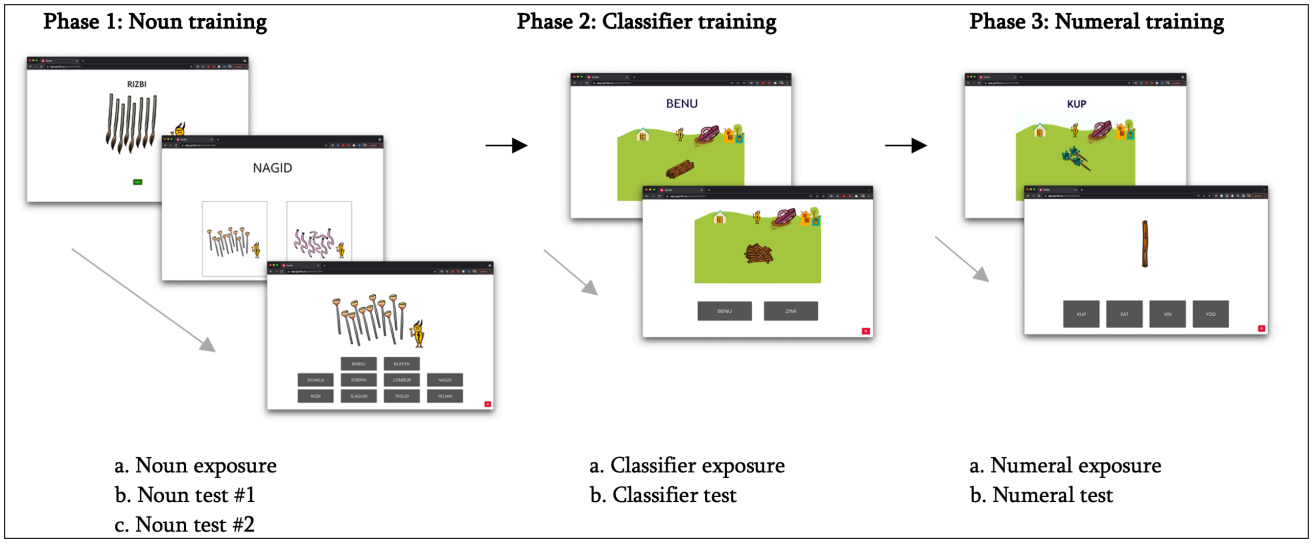

Figure 2. Experimental procedure illustrating the three learning phases. Images are screenshots showing sample trials for all steps in the task. Arrows indicate the progression through the task.

producing Num-CL-N, as in *three piles of sticks*. Because this is a common order typologically (Frequent-Final), if we find this order alone, the results would not be informative

Method

Participants. Fifty-six English speakers were recruited from the University of Toronto participant pool, and through an open call. Participants received course credit or $5 for completing the experiment. A language background questionnaire was used to ensure participants were English-dominant, and that they had at most little knowledge of any classifier language.

Nine participants were excluded, due to language background (N=2) or because their accuracy in training as below 70% (N=7), resulting in data from 47 participants.

Materials. The artificial language consisted of 13 novel words (phonotactically legal in English), adapted from Fedzechkina et al., (2012): seven bi-syllabic nouns referring to physical entities, four monosyllabic numerals (CVC), and two group classifiers (with a CVCV syllable structure) – see again Figure 2.

The visual stimuli were created using the Procreate and Notability apps on the iPad. The 'alien objects' (more below) resembled real-world artifacts (e.g., ladder, scarf, poles) and natural kinds (e.g., a flower with unconventional colors). The two classifiers were associated with two types of groups: a neat bunch (*benu*) and a disorganized pile (*zimi*).

Procedure. The experiment was created and hosted on Gorilla Experiment Builder (www.gorilla.sc, Anwyl-Irvine et al, 2019), and was run on a computer or a tablet.

At the beginning of the experiment, participants were told that they would be learning a novel 'alien' language to communicate with the alien Electrobud-1 who was going to sort the inventory of their spaceship. Participants first learned words in the alien language, and them produced instructions to tell the alien which items are going back into the spaceship. Participants went through all phases in one sitting, which took 10 to 35 minutes to complete.

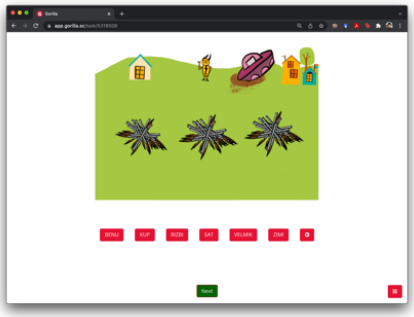

Figure 3: Test phase for Experiment 1. Participants chose words from a word bank to describe the target object.

Phase I: Learning

1. NOUN. Participants were first presented with the seven objects along with their labels, one at a time. Noun exposure was followed by two tasks. First, participants saw one word and two objects, and had to choose the object that corresponded to the noun; feedback was provided. There were seven trials in this block. Regardless of their performance on the first task, participants advanced to the second noun task. In the second task, participants saw a picture of one object, with a list of all seven nouns (i.e., the full lexicon), and had to choose the appropriate noun. Participants received feedback on each trial and were required to correctly label at least six nouns (out of seven; 86% accuracy) in order to advance to the next phase. If this

threshold was not met, they repeated this task until they reached criterion.

2. CLASSIFIER. Participants were first presented with an image depicting one group classifier at a time across 6 trials, with the respective word displayed above. The set of images showed two kinds of collectives, '*benu'* (neat bunches) and '*zimi'* (disorganized piles).

After exposure, participants saw an image depicting a pile or a bunch of objects, and had to choose the appropriate classifier. Feedback was provided on each of 14 trials (seven for each classifier). At least 12 correctly labelled classifiers (80% accuracy) were required to advance to the next stage. Participants who did not achieve threshold repeated the exposure and test phase until they reached criterion.

3. NUMERAL. Participants were first presented with pictures of one to four objects, alongside the respective numeral in the artificial language: exposure included 12 trials of this type.

After exposure, participants saw a picture depicting one to four objects, and had to choose the correct word. Ten trials were presented, and feedback was provided. They advanced to the next phase if they achieved 80% accuracy (or 8/10 trials); otherwise, they repeated the exposure and test phases until they reached criterion.

Phase II: Test

Participants were asked to produce a string of words to "help the alien check the inventory". On critical trials (n=14), the image depicted two or three collections of objects (bunches or piles), which were meant to elicit all three elements (noun, classifier, numeral): seven for the neat group classifier '*benu'*; seven for the disorganized classifier '*zimi'*. We also included baseline trials (n=7), where the image depicted individual objects, meant to elicit just noun and numeral.

Participants had to form a string of words by selecting from a word bank that contained all the words participants learned, arranged in alphabetical order. (cf. Fedzechkina & Jaeger, 2020) – see Figure 3. A 'backspace' button was available for editing their response before moving to the next trial. If participants chose less than two lexical items (the minimum number of elements required on certain trials), a message appeared telling them to provide more information ("You need to be more specific!").

Results

We first assessed successful learning of the lexical elements. All three nominal elements were acquired with a high degree of accuracy. Acquisition of the nouns was at 91%, classifiers at 88%, and numerals at 90%. This gave rise to an overall mean accuracy of 90%. To reach this accuracy, participants needed at least one block to learn the nouns (mean: 1.38 blocks), one block for classifiers (mean: 1 block), and one block for numerals (mean: 1 block).

To assess how participants ordered the elements, we coded the order used on critical trials (653 out of 987 trials). We then excluded 51 strings that had multiple lexical errors or any missing element. Thus, the order analysis includes 602 trials (92.2% of the data).

The distribution of participants' word orders involving classifiers is plotted in Figure 4. The Frequent-Final word order (Num-CL-N) is the most common. Because this word order mimics English (e.g., *three bunches of sticks*), we are unable to conclude whether this is a result of typological prevalence or native language influence. The second most common order is Medial2 (Num-N-CL). This is surprising because this order is unattested in the world's languages. In our data, it is more common than both Frequent-Initial and both Rare orders.

We further note that the order of exposure did not seem to have an effect on the orders produced. Specifically, the training exposed participants to nouns, classifiers, and then numerals. Nevertheless, participants generally maintained the preference for the numeral to precede the other elements, as seen in the top four produced orders (Frequent-Final, Medial2, Frequent-Initial and Rare-Final). In fact, there were only 24 instances of a participant producing the order reflecting the blocked learning of lexical items during training (Rare-Initial: N-CL-NUM, 4% of all 602 included coded strings). When participants chose to produce nouns initially, there was little preference for NUM-CL and CL-NUM (40 versus 24 counts respectively). Even when participants separated the numeral and classifier (in red), there was a much higher preference for the numeral to precede the classifier, as opposed to having the numeral after the classifier (111 versus 17 counts). These results are also not consistent with the possibility that participants ordered the elements with ease of lexical access. Since there were two classifiers, four numerals and seven nouns in the lexicon, this would predict CL-Num-N (Rare Final) orders to be produced, but this was not overwhelmingly chosen (65 counts).

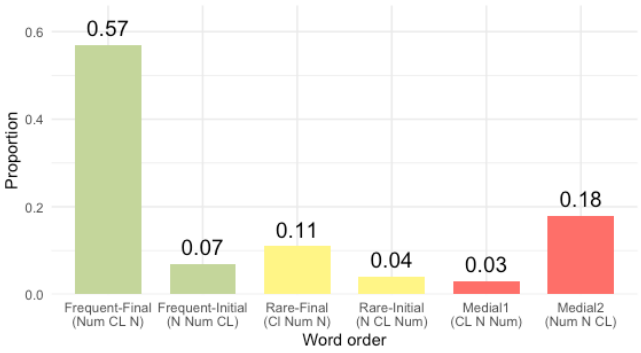

Figure 4: Word orders produced in Experiment 1

Our statistical analysis addressed two questions. First, we asked whether participants avoid orders that are unattested typologically. Using an intercept-only logistic regression model (with an offset for a 1/3 baseline, and crossed random effects for participants and items), we asked whether N-medial orders were produced less than chance (i.e., CL-N-Num: .03 and Num-N-CL: .18, or .21, is different from chance, which is 1/3). The observed N-medial responses were significantly less likely to be produced than chance (β = -2.68, SE = 0.82, z = -3.25, p = .001). Second, we asked whether numerals were overall more likely to be ordered

before the classifier, which is the more prevalent order typologically. Indeed, participants ordered the numeral before the classifier on 525 trials (or 83%). Here an intercept-only logistic regression model (with crossed random effects for participants and items) showed that numerals preceded the classifier significantly more than chance (β = 7.19, SE = 1.63, z = 4.40, p < .001; chance here is 1/2). Thus, while participants preferred the typologically prevalent order of numerals and classifiers, we cannot rule out the possibility that thus is driven by the order in their native language. In that sense, our results align with Wang (2024) in that we do not find evidence for the Typological Prevalence Hypothesis.

Experiment 2: Improvised gesture

We considered the possibility that artificial language learning always involves *some* explicit instruction, and so we chose to examine the same question using a fully spontaneous task, namely improvised gesture. Specifically, we had participants gesture noun phrases that included a noun, a numeral and a classifier by asking participants them to refer to a target picture that appeared in an array of four – see Figure 5. The reason for including three "context" pictures in addition to the target picture, is to encourage participants to gesture *all three* "elements": these are required in order to differentiate the target image from the other images. For example, the (red-framed) target image in Figure 5 cannot be distinguished from the others by gesturing only one or two of three elements: pencil (N), number (Num) or bunch (CL).

Method

Participants. We collected data from English speakers (n=21). Seventeen could also speak a classifier language as a first language or a heritage language (e.g., Mandarin, Cantonese, Japanese and Korean).

Materials. Each display contained four photographs (Figure 5), where the target was highlighted in red. On critical trials, the target showed objects that involved three elements (noun, numeral and classifier). Across 12 critical trials, we used two numerals (*2* and *3*), six classifiers (*bunch, pack, stack, bag, scoop, box*), and 12 nouns (*pencil, straw, toothbrush, razor, book, bowl, apple, ball, clip, marble, crayon, match*). The three nontarget images in each display were designed to ensure that all three elements would be gestured. For example, *two bunches of pencils* occurred alongside images of *two pencils*, *three pencils* and *two bunches of sticks.* We also included six fillers trials (*saltshaker, ruler, egg, banana, whistle, comb*). These fillers were included to discourage a situation where participants would repeat the same pattern.

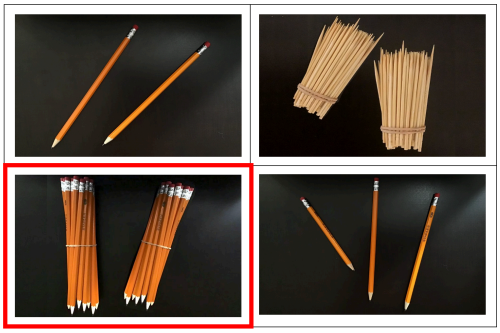

Figure 5: Example display for Exp. 2.
The target ('two bunches of pencils') is marked in red.

Procedure. The experiment was conducted via Zoom (during the pandemic). Participants were asked to use their hands and upper body to silently gesture the target such that a partner in a future study could use their gesture to identify the correct picture in the same array. Participants completed one practice trial before the 18 experimental trials. Each session lasted about 30-45 minutes.

Results

Gestures were transcribed for word order from the video recordings. Nouns and classifiers were typically gestured using their shape or function (e.g., a scribbling action for '*pencil*'), whereas numerals were either gestured by raising fingers or repeating the same gesture several times in different locations. We excluded trials where: (i) participants mouthed words (7 trials); (ii) two elements were co-gestured, preventing an assessment of linear order (e.g., gesturing scoop with their left hand and marble with their right hand: 14 trials), (iii) participants produced gestures for "quantifiers" rather than numerals (e.g., 'many'; 17 trials), or (iv) trials with disfluencies (i.e., absent or uninterpretable gestures: 4 trials). A further 38 trials were excluded because they only contained two elements. We thus analyze 172 (or 68.3%) of 252 trials. We note that this rate of exclusion is similar to other gesture paradigms (e.g., 65.1% in Christensen, Fusaroli & Tylén, 2016). We coded these for order, by logging the first gesture sequence with all noun phrase elements. For example, if a numeral was expressed twice, the later occurrence was not recorded.

Figure 6 plots the word orders of the gestured produced, using the labels and shading from Table 1. These results show that two of the orders were used the most, specifically Frequent-Final (Num-CL-N: 32%) and Rare-Initial (N-CL-Num: 32%). The remaining four orders were distributed about evenly (7-10%). Thus, the orders preferred in improvised gesture did not reflect typological preferences.

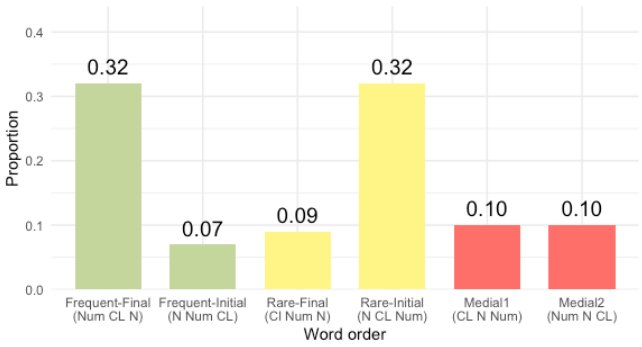

Figure 6: The word orders produced in Exp. 2

Notably, the Rare-Initial (N-CL-Num) order is not found in English or in any of the classifier languages spoken by participants.

Again, our statistical analysis asked two questions. First, whether participants avoid the typologically-unattested orders. Using an intercept-only logistic regression model (with an offset for a 1/3 baseline and crossed random effects for participants and items), we asked whether N-medial orders were gestured less than chance. Indeed, these N-medial word orders (i.e., CL-N-Num: .10 and Num-N-CL: .10, or .20) were gestured significantly less than chance (β = -0.87, SE = 0.33, z = -2.65, p = .008). Our second question was whether numerals were more likely to be gestured *before* classifiers. Participants gestured the numeral before the classifier on 84 out of the 172 trials (or 48.8%). An intercept-only logistic regression model, with random intercepts for participants and items, showed that this is not different from chance (β = 0.030, SE = 0.555, z = 0.054, p = .96). In other words, participants avoided the typologically unattested noun-medial orders, but there is no evidence for a preference for Num-CL over CL-Num orders. Thus, unlike previous gestural studies replicating word order patterns both in the noun phrase (e.g., Culbertson et al., 2020; Jaffan et al., 2020) and on the sentential level (e.g., Goldin-Meadow et al. 2008), Exp. 2 did not reflect typological preferences for classifier word order through a gestural, non-linguistic paradigm.

General Discussion

We used artificial language learning (Exp. 1) and improvised gesture (Exp. 2) to examine the hypothesis that word order patterns across the world's languages are shaped, at least in part, by cognitive biases at the level of individuals, and therefore should be reflected in tasks where there is not prior linguistic convention. Our first question was whether speakers avoid the typologically-unattested N-medial orders. Both tasks revealed a pattern where these are avoided, suggesting some support for the Typological Prevalence Hypothesis. Our second question was whether speakers prefer the typologically-preferred order, Num-CL. This was the case in Exp. 1 (artificial language); however, because this was driven by native language order alone, it is not evidence for the Typological Prevalence Hypothesis. In Exp. 2 (improvised gesture), there was no evidence that participants preferred the typologically-prevalent order.

Our results are surprising in the context of prior work on word orders within the noun phrase. First, typological preferences regarding the order of nouns, adjective, numerals and demonstratives were mirrored in artificial language learning (Culbertson & Adger, 2014). Second, gestural paradigms also demonstrated that speakers prefer typologically-prevalent orders, both when gesturing nouns, adjectives and numerals (Jaffan et al., 2020) or when also including demonstratives (Culbertson et al., 2020). However, the absence of support for the Typological Prevalence Hypothesis in our results is actually similar to Wang (2024), the only prior study to examine the ordering of classifiers and numerals.

One possibility is that our results actually reveal that cognitive biases might not play a meaningful role in **"finer grained" ordering phenomena**, such as those involving numerals and classifiers within noun phrases. This contrasts with more coarse-grained phenomena such as the ordering of sentential elements (cf. Goldin-Meadow et al., 2008: Subject-Object-Verb) and the ordering of more distinctively "meaning-bearing" noun phrase elements (cf. Culbertson & Adger, 2014; Culbertson et al., 2020; Jaffan et al., 2020).

Relatedly, we note that in Exp. 2 the dominant orders were ones where classifiers were *closer* to the noun, namely Num-CL-N (i.e., Frequent-Final) and N-CL-Num (i.e., Rare-Initial). This points to the possibility that what was intended to express classifiers was perceived by speakers as *adjectives*. Typologically, adjectives tend to occur closer to the noun than numerals (Dryer, 2013), a preference that has been previously replicated using silent gesture tasks (Schouwstra, Kirby & Culbertson 2017; Culbertson et al., 2020; Jaffan et al., 2020). Thus, if speakers interpreted these elements as adjectives, these are the orders that would be expected, namely Num-ADJ-N and N-ADJ-Num, the latter being the most frequent word order cross-linguistically (Dryer, 2013). In fact, this converges with Schouwstra et al., (2017) who found that English-speaking participants gestured adjectives after the noun, unlike the word order found in English.

Classifiers are grammatical units that sit on the cline between adjectives and numerals, given that they might be described as "modifying" nouns while being associated with counting. At an abstract level, classifiers and adjectives are similar in that they encode properties about the referent. Recall that the sortal classifier in (1a) highlights the animate property of the referent, while the sortal classifier in (1b) highlights the long and thin shape of the referent. Similarly, the mensural classifier in (1c: *ping* 'bottle') highlights the quantity of the referent, whereas the one in (1d: *dui* 'pile') highlights the configuration of the referent, similar to the group-like classifiers in our experiments (e.g., *bunch* and *pile*) (we note that both exist independently as nouns). This comparison raises two interesting questions for future research. First, what drives the difference in linguistic convention between adjectives and classifiers? While it seems intuitive that adjectives are sensitive to the position of the noun and classifiers are sensitive to the position of the numeral, we would need to explain why adjectives occur on either side of the noun, whereas classifiers have a strong preference to follow the numeral. A second question concerns the why only certain semantic properties (e.g., animacy or shape) become grammaticalized as classifiers, while other properties (e.g., color, size or state) do not.

In sum, although future research is necessary to address these questions, the present study demonstrates that typological patterns – particularly "finer grained" ordering phenomena, such as those involving numerals and classifiers within noun phrases – are not always straightforwardly reflected in cognitive biases at the level of individuals.

References

Aikhenvald, A. Y. (2000). *Classifiers: A typology of noun categorization devices*. OUP Oxford.

Allan, K. (1977). Classifiers. *Language*, *53(2),* 285-311.

Anwyl-Irvine, A.L., Massonié J., Flitton, A., Kirkham, N.Z., Evershed, J.K. (2019). Gorilla in our midst: an online behavioural experiment builder. *Behavior Research Methods.*

Cheng, L. L. S., & Sybesma, R. (1998). Yi-wan tang, yi-ge tang: Classifiers and massifiers. *Tsing Hua Journal of Chinese Studies, 28(3),* 385-412.

Chierchia, Gennaro. (1998). Reference to kinds across language. *Natural Language Semantics 6,* 39–405.

Christensen, P., Fusaroli, R., & Tylén, K. (2016). Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition*, *146*, 67-80.

Culbertson, J., & Wilson, C. (2013). Artificial grammar learning of shape-based noun classification. *Proceedings of the Annual Meeting of the Cognitive Science Society*.

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

Culbertson, J, Schouwstra, M. & Kirby, S. (2020). From the world to word order: Deriving biases in noun phrase order from statistical properties of the world. *Language*, *96(3),* 696-717.

Dryer, M. S. (2013). Order of subject, object, and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology.

Erbaugh, Mary S. (1986). Taking stock: The development of Chinese noun classifiers historically and in young children. In *Noun classes and categorization*, ed. C. Craig. Amsterdam: John Benjamins.

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, *44*, (pp. 17897–17902).

Fedzechkina, M., & Jaeger, T. F. (2020). Production efficiency can cause grammatical change: Learners deviate from the input to better balance efficiency against robust message transmission. *Cognition*, *196*, 104-115.

Gentner, D., & Bowerman, M. (2009). Why some spatial semantic categories are harder to learn than others: The typological prevalence hypothesis. In *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (Vol. 465, p. 480).

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological Science, 24(7),* (pp. 1079-1088).

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, *105(27),* (pp.9163-9168).

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of language* (pp. 73–113). MIT press.

Greenberg, J. H. (1990[1972]). Numerical classifiers and substantival number: Problems in the genesis of a linguistic type. In K. Denning & Kemmer (eds.), *On language. Selected writings of Joseph H. Greenberg*, 166-193. Stanford, CA: Stanford University Press. [First published 1972 in Working Papers on Language Universals 9. 1-39. Stanford, CA: Department of Linguistics, Stanford University.]

Her, One-Soon, Hui-Chin Tsai, Kun-Han Lin, Marc Tang, & Meng-Chang Lee. (2015). Unification of numeral bases and numeral classifiers: Evidence from SMATTI. Paper presented at the *Workshop on the syntax and semantics of numerals, the 48th Annual Meeting of the Societas Linguistica Europaea*.

Her, O-S. (2017). Deriving classifier word order typology, or Greenberg's Universal 20A and Universal 20. *Linguistics*, *55(2),* 265-303.

Her, O.-S., Tang, M., & Li, B.-T. (2019). Word order of numeral classifiers and numeral bases: Harmonization by multiplication. *Language Typology and Universals*, 72(3), 421-452.

Jaffan, J., Klassen, G., Yang, J., & Heller, D. (2020). Semantic influences on emergent preferences of word order: Evidence from silent gesture. *Proceedings of the 42nd Meeting of the Cognitive Science Society.*

Kam, C. L. H., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1(2),* 151-195.

Kam, C. L. H., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, 59(1), 30-66.

Lehrer, A. (1986). English classifier constructions. *Lingua*, *68(2-3),* 109-148.

Schouwstra, M., Kirby, S., & Culbertson, J. (2017). Silent gesture and noun phrase universals. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 39).

Wang, F. (2024). *From cognition to word order universals: an artificial language learning approach*. Doctoral dissertation, University of Edinburgh.