Language Acquisition

ISSN: 1048-9223 (Print) 1532-7817 (Online) Journal homepage: https://www.tandfonline.com/loi/hlac20

Rule Generalization from Inconsistent Input in Early Infancy

Elena Koulaguina & Rushen Shi

To cite this article: Elena Koulaguina & Rushen Shi (2019) Rule Generalization from Inconsistent Input in Early Infancy, Language Acquisition, 26:4, 416-435, DOI: 10.1080/10489223.2019.1572148

To link to this article: https://doi.org/10.1080/10489223.2019.1572148

Published online: 08 Apr 2019.

LANGUAGE ACQUISITION
2019, VOL. 26, NO. 4, 416–435
https://doi.org/10.1080/10489223.2019.1572148

Routledge
Taylor & Francis Group

Rule Generalization from Inconsistent Input in Early Infancy

Elena Koulaguina[a,b] and Rushen Shi[a]

[a]Département de Psychologie, Université du Québec à Montréal; [b]Cognition and Brain Plasticity Unit, University of Barcelona

ABSTRACT

Children begin to learn abstract rules at an early age, in an implicit way, without access to rule descriptions. They rely on specific rule instances that they encounter. However, rule instances often co-occur with rule-inconsistent instances. One kind of inconsistent input, non-application instances, constitutes a learnability problem. For example, a child might hear many instances of the dative shift rule in English, such as *Mary gave the book to me* → *Mary gave me the book*, and partial cases such as *He will bring toys to you* or *He donated books to John*, without being told which of the latter sentences (i.e., non-application instances) would be rule-possible or exceptions. We examined whether and how non-application instances may impact rule learning per se. Fourteen-month-old infants were passively exposed to an unfamiliar natural language. In Experiment 1 half of the training input supported an artificial word-order shift rule, and the other half were non-application singletons without shifting. During test, infants failed to generalize the rule to new instances, suggesting that the non-application cases in training were treated as non-rule cases and might have impeded rule learning. In Experiment 2 rule instances were dominant in type frequency relative to non-rule instances in the training input, and infants showed rule generalization, confirming that the non-rule instances in Experiment 1 indeed impeded learning. The token frequency of individual instances did not affect rule generalization. Experiments 3 and 4 replicated the same findings (of Experiments 1 and 2 respectively) with stimuli containing no morphological cues, demonstrating that the mechanism underlying abstract rule learning is robust, likely being one of the earliest learning mechanisms available to humans.

ARTICLE HISTORY
Received 09 December 2017
Accepted 18 December 2018

1. Introduction

A crucial characteristic of abstract rules is their productivity. The productive knowledge gives rise to the ability to recognize and produce an infinite number of novel instances that can be generated by an abstract rule. A rule can be taught, with the rule directly given, as is often done in educational settings. Infants and young children, however, typically encounter specific instances only, from which they derive an abstract rule. For example, linguistic input to children contains only specific instances. Caregivers do not present rules per se when interacting with children. Nevertheless, language development in the first few years of life shows remarkable rule learning at various linguistic levels. Even more impressive is the fact that children's rule learning often occurs under input conditions where rule instances are copresent with rule-violation cases—for example, the regular *–ed* past tense verbs and irregular forms such as *went* and *ran* in English. In this study we inquire how infants at the earliest stage of linguistic development learn rules implicitly from input that contains both rule-consistent and inconsistent instances.

Our particular interest concerns the role of non-application instances in the input for rule learning. Children's input contains a subset of the output that can be generated by an abstract

CONTACT Rushen Shi shi.rushen@uqam.ca; Elena Koulaguina elena.koulaguina@gmail.com Département de Psychologie, Université du Québec à Montréal, C.P. 8888, Succursale Centre-Ville, Montréal, Québec H3C 3P8 Canada.

rule. For example, a child may hear many present-past-tense verb alternations (e.g., *touch – touched, move – moved, play – played*, etc.), but also non-application singletons (e.g., *walk, hit*, etc). Such non-application singleton cases have been the center of the learnability literature for the past few decades (since Baker 1979). As it is known that the child receives no negative evidence indicating which forms are exceptions to a rule, it seems paradoxical how a child can learn that the past-tense rule should be applied productively if he happens to only hear *walk* and not yet *walked*, but that he should inhibit from applying the rule to *hit* - **hitted* upon just hearing *hit*. Past research focused on how non-application exceptions of a productive rule can be learned. The question that is equally important but has not been studied concerns the status of non-application input for infants' rule learning per se. In this article we report four experiments that tested the following specific questions: (i) do infants consider non-application singleton cases as possible targets for rule application? and (ii) how do infants learn general rules from input containing both rule instances and non-application singleton instances?

Rule learning has been demonstrated in previous experimental studies using 100% rule-consistent input, i.e., all instances fully demonstrating a rule with no rule-deviant instances. Infants as young as 7 to 9 months of age can learn algebraic-like abstract rules (e.g., Marcus et al. 1999; Gerken 2006). In those studies, one group of infants was exposed to stimuli such as *le di di* and *wi je je*, which conformed to the ABB rule. Another group was trained with instances of a different rule (e.g., ABA) such as *le di le* and *wi je wi*. In the test phase, infants discriminated novel stimuli such as *ba po po* versus *ba po ba*. There is also evidence that infants can learn this kind of rules (e.g., ABB, ABA, AAB) when exposed to nonspeech stimuli (including nonspeech sounds, gestures and visual patterns) under some experimental conditions (Marcus, Fernandes & Johnson, 2007; Rabagliati et al. 2012; Frank et al. 2009; Johnson et al. 2009; Saffran et al. 2007). For example, in Saffran et al. (2007), 7-month-old babies learned ABA and ABB rules and applied them to novel instances after being trained with rule-consistent animal pictures (dogs and cats) that were presented simultaneously on the screen. The rules used in those studies involved identical repeating elements (e.g., ABB, ABA). Such rules are perceptually salient to infants (Gerken et al. 2014). Even the newborn brain detects speech stimuli containing such repetition patterns, as shown in studies using near-infrared spectroscopy measures (Gervain et al. 2008; Gervain, Berent & Werker 2012).

Infants' learning of more complex rules has been shown in a few studies (e.g., Gerken & Knight 2015; Gómez & Lakusta 2004; Koulaguina & Shi 2013). Gómez & Lakusta (2004) trained 12-month-olds with instances of an aXbY artificial grammar, where a and b pseudo-words resembled grammatical words in natural language (e.g., determiners, auxiliaries), and X and Y pseudo-words resembled open-class words (e.g., nouns), which were monosyllabic and disyllabic respectively. In one experiment, 'a' words *alt* and *ush* always occurred with disyllabic words (e.g., *alt coomo*, etc.) and 'b' words *ong* and *erd* always with monosyllabic words (e.g., *ong deech*, etc.). During the test phase, infants discriminated novel rule strings (i.e., new X/Y words with the appropriate a/b words) and novel rule-violation strings (the new X/Y words violating the trained aXbY dependencies), suggesting that they categorized the novel X/Y words by abstracting their phonological structure (number of syllables) and linking them with appropriate a/b words.

Infants can also learn abstract word-order rules. In Koulaguina & Shi (2013), infants were passively exposed to three-word sentences created from a natural language (Russian), a language unknown to the participants. The sentences were each shifted in word order, ABC to BAC for one group of infants (e.g., *Staya letit klinom* → *Letit staya klinom*), and ABC to ACB for the other group (e.g., *Staya letit klinom* → *Staya klinom letit*). After the passive training phase, 14-month-olds distinguished novel sentences (containing all novel words) being shifted to the trained rule from those being shifted to the other untrained rule. The two groups of infants showed opposite-looking patterns consistent with their respective training input, revealing that they learned the word-order rules and generalized them to novel instances.

These findings concerned infants' rule generalization when training input was 100% rule instances. Limited experimental work has been done on rule learning under inconsistent input conditions, nearly all with adults and older children. In artificial language-learning studies with adults (Wonnacott, Newport & Tanenhaus 2008; Wonnacott & Newport 2005) and children aged 5 to 7 years (Wonnacott 2011; Hudson Kam & Newport 2005), rule generalization was observed after participants were trained with input containing rule instances mixed with a small proportion of inconsistent cases. For instance, in the input condition where most verbs were in the "VOS particle" structure (e.g., *Glim*-V *blergen*-O *tombat*-S *ka*-particle 'The giraffe hit the lion') and one verb in the "VSO" structure (e.g., *Frag*-V *flugat*-S *nagid*-O 'The bee tickled the elephant'), learners used novel verbs more in the dominant "VOS particle" structure (Wonnacott, Newport & Tanenhaus 2008). In those studies adults tended to show probability matching (i.e., reproducing the proportion of the trained dominant structure in the test phase), whereas children regularized more (i.e., producing more cases with the dominant structure beyond its training proportion).

The study of Gómez & Lakusta (2004) was the only one that tested infants' rule generalization when input was inconsistent. Using the aXbY artificial language, the authors manipulated the proportions of rule cases versus violations of the training input across experiments, with 100% rule instances in one experiment (as discussed previously), and reduced proportions of rule instances in other experiments. In one condition, for example, rule cases were 'a' words (*alt, ush*) occurring with disyllabic words (e.g., *alt coomo*, etc.) and 'b' words (*ong, erd*) with monosyllabic words (e.g., *erd deech*, etc.). The violations had the opposite pairings (e.g., *alt* and *ush* with monosyllabic words such as *alt deech; ong* and *erd* with bisyllabic words such as *erd coomo*). The experiments showed that 12-month-olds tolerated 17% of violation cases in the learning input and generalized the dominant rule to novel test words. However, when violation cases reached 33% (and rule cases dropped to 67%), rule generalization failed. The results suggest that the proportion of rule instances relative to violation cases determines infants' rule generalization.

Overall, rule productivity shown in adults, children, and infants under inconsistent input conditions in the aforementioned studies is in line with models of rule learning in similar input conditions (e.g., Rumelhart & McClelland 1986; Yang 2005, 2016). The connectionist model by Rumelhart & McClelland (1986) advanced the importance of high type frequency of rule instances for rule learning, using the case of English past tense (*-ed*). In one learning condition they exposed their model to one regular verb and one irregular verb. This led to the learning of the specific trained verbs but no productive rule learning. In another condition they exposed their model to many different regular verbs and only one exception, resembling the frequency distribution of regular and irregular verbs in English: Regulars are higher in type and lower in token frequencies, whereas irregulars are lower in type but higher in token frequencies. The model learned the regular rule after 10 exposures to 17 regular verbs and 1 irregular verb. However, the model was not able to generalize the rule when exposed to two regular and eight irregular verbs. Subsequently, the model could achieve rule generalization after being exposed to 334 regular and 76 irregular verbs in addition to the first batch.

A mathematical model of rule learning from inconsistent input was proposed recently: the Tolerance Principle (Yang 2005, 2016), according to which a general rule will be formed when the number of non-rule cases does not exceed the total number of cases divided by the natural log of the total number (i.e., $e \le N/\ln N$, where e being exceptions and N being total number of cases). The N is supposedly type frequency. In Gómez and Lakusta (2004), rule generalization succeeded when the proportion of non-rule instances was below the threshold of tolerance but failed when the level of non-rule instances was above the threshold.

The model of Rumelhart and McClelland (1986) and the Tolerance Principle of Yang (2005, 2016) only considered non-rule cases that were overtly different from the forms of the productive rule. It is not clear how non-application cases, which are not overt rule violations, may affect rule learning, since such cases could logically yield two kinds of output: (i) rule forms that did not yet have a chance to occur, and (ii) true non-application exceptions.

The non-rule cases in the input of the previous studies were similar to overt violations in natural languages (e.g., *go* → *went*). Such violations seem to be salient, as they stand out as a clear indication that they are different from rule instances. Non-application singleton cases, however, are a subtle and theoretically more interesting kind of input. As mentioned earlier, the verb *hit*, which has no overt irregular form for its past tense, is an exception to the past-tense rule. It is unlike regular verbs, which might also happen to appear as singletons in a child' input (e.g., *walk*).

Consider also the well-known example of dative shift in English, as discussed in Baker's Paradox (Baker 1979), e.g., *give the book to John → give John the book*. Dative shift is widely attested in many ditransitive verbs, allowing children to learn the rule and apply it productively to new ditransitive verbs in both constructions without having heard both. However, certain verbs are only possible in one construction, e.g., *donate the book to John*, but not in the shifted construction, e.g., **donate John the book*. The key point is that children receive only positive input but no negative evidence indicating that the unattested *donate John the book* is unacceptable. The system cannot have a general mechanism to block such unattested cases, as this would be against the infinite use of abstract rules, which is a property of human language. The child faces the paradoxical needs for the rule to be productively applied and for certain unattested exceptions to be inhibited from the productive use without being told which ones. Unattested cases such as dative shift thus constitute a challenging learnability problem. Amazingly, children do succeed in managing both needs, learning both kinds of knowledge. How they achieve the learning, however, remains to be fully understood.

A rich body of literature considered how children may resolve the paradox (e.g., see the reviews in Pinker 1989), and the emphasis was on possible ways to learn unattested exceptions, with the assumption that the general rule was already learned. In our study we focused on the learning of the general rule per se. We tested how the occurrence of non-application instances, either rule possible or not, may affect rule learning. That is, how do learners interpret non-application cases, and what impact do these cases have on rule learning? Experimental work with systematic manipulation of attested and non-application input is sparse and is absent in research with infants. By studying infants, we aimed at revealing rule-learning mechanisms available to humans from the earliest stage of life.

We used the same kind of artificial grammar as in Koulaguina and Shi (2013), but we manipulated the distribution of rule instances and non-application instances in the training input. The rule instances exhibited a word-order shift rule, and the non-application ones did not exhibit any word-order shift. We tested how such input might affect rule learning in 1-year-old infants. In particular, we examined rule generalization to new sentences that were not part of the training input, neither among the rule instances nor among the non-application instances.

In Experiment 1 we tested whether infants treated non-application exemplars as possible targets for rule application or as non-rule instances. The results showed that the input in this experiment did not lead to rule learning, indicating that the non-application cases were treated as non-rule instances. Experiment 2 tested the hypothesis that rule generalization is determined by the type frequency, rather than the overall frequency, of rule instances, relative to non-rule instances. *Type frequency* refers to the number of different instances, and *overall frequency* includes the number of different instances as well as the number of repetitions of these instances (i.e., their token frequencies). For example, eight different instances each occurring four times would yield eight for the type frequency and 32 for the overall frequency. As discussed previously, Rumelhart and McClelland (1986) advanced the importance of type frequency for the learning of regular rules. Yang's Tolerance Principle (Yang 2005, 2016) also operates in terms of type frequency. As expected, infants showed rule generalization in Experiment 2, by relying on the high type frequency of rule instances relative to non-rule instances. The findings of the two experiments were replicated in Experiments 3 and 4 with sentences containing no morphological cues to the word-order shifts.

2. Experiment 1

In Experiment 1 we asked whether infants' rule generalization would be affected by non-application cases. We constructed artificial rules using a natural language (i.e., Russian), which was unknown to our participants. We exposed infants to an input set in which half (i.e., 50%) of the three-word sentences underwent a word-order shift rule (ABC–BAC or ABC–ACB), and half (i.e., 50%) were non-application singleton cases (ABC) that did not go through the shift. Given that our infants were very young and may have general difficulty in engaging in the task, we used consistent morphological markings for A, B, and C words to make the rules easier to learn. The markings were present for both rule utterances and non-application cases. Infants were then tested on their learning of the trained rule with entirely new sentences (i.e., not the non-application sentences).

We predicted that if infants treated the non-application cases as possible targets for rule application, then these instances might not impede rule generalization. In this case, the input could be considered 100% rule-consistent, and infants should learn the rule and apply it to entirely new instances. If, alternatively, the non-application cases were treated as rule violations, the proportion of rule instances would be only 50%, lower than the level possible for rule learning based on previous experiments with infants (Gómez & Lakusta 2004) and lower than the threshold for rule learning according to the Tolerance Principle proposed by Yang (2005, 2016). In such a case, infants should fail to learn the rule.

2.1. *Participants*

Sixteen infants aged 14 months (9 boys and 7 girls) from various linguistic backgrounds completed the experiment. The age ranged from 14 months and 3 days to 14 months and 24 days (M = 14 months and 15 days). None of the infants had any prior exposure to Russian. Fourteen other infants were tested, but their data were not analyzed for various reasons such as fussiness (5), getting out of camera range during test trials (1), crying (1), lack of interest (1), parental interference (3), and looking time too short—for 2 seconds or less in six or more test trials (3). One other infant did not complete the experiment.

2.2. *Materials*

Materials were 18 Russian sentences (see Table 1) recorded by a female Russian native speaker in the child-directed speech style. We decided to use Russian because different word orders of the same sentence strings are grammatical in this language, thus allowing us to avoid any possible speech disfluency related to ungrammaticality. Sixteen of these sentences were used as the training stimuli and two as the novel instances in the test phase. Sentences had a Subject-Verb-Object structure. All words were bisyllabic. We selected Russian words that had consistent morphological markings for each position within the base-order sentences (ABC). All A words ended with –*a*, B words ended with –*it*, and C words ended with –*ku*. The words in the shifted sentences (BAC or ACB) kept their original morphological markings. For instance, for the sentences that went through the ABC–BAC rule application (e.g., *Vika darit murku → Darit Vika murku*), the A words from the base ABC order appeared as the second word in BAC, always with the -*a* suffix.

Table 1. Sentences used as stimuli in Experiment 1.

Experimental phase	Base ABC sentences for the word-order shift rules (either ABC–BAC or ABC–ACB; each rule instance consisted of a base sentence and its shifted version)	Non-application sentences (non-shift singletons)
Training phase	*Vika darit murku*	*Gena vidit lavku*
	Dima gonit galku	*Lida zharit manku*
	Lera nosit dochku	*Kesha penit rechku*
	Njuta varit kashku	*Ljova kopit lesku*
	Gosha lovit rybku	*Nina lechit sivku*
	Vova katit lodku	*Goga vozit zhuchku*
	Rada manit koshku	*Sasha parit repku*
	Zhora lepit belku	*Seva kosit gorku*
Test phase	*Njura topit pechku*	
	Tjoma rubit lipku	

Of the 16 training exemplars, 8 sentences went through a word-order shift rule. For example, an ABC sentence also appeared in the BAC order (for one of the training conditions, Rule 1) or in the ACB order (for the other training condition, Rule 2). Since for infants the unknown Russian words had no syntactic categories, we will refer to these rules as ABC–BAC (Rule 1) and ABC–ACB (Rule 2). Eight other non-application sentences only appeared in the base order (i.e., ABC singletons) and did not go through any shift.

We created two input strings, each containing eight alternating sentence pairs (either Rule 1 or Rule 2). Each pair, which formed one rule instance, consisted of a base sentence immediately followed by its alternating version (e.g., ABC–BAC: *Vika darit murku* → *Darit Vika murku* in one string; ABC–ACB: *Vika darit murku* → *Vika murku darit* in the other string). The eight pairs in each string were conjoined randomly with the eight non-application sentences. The total number of alternating pairs (i.e., rule instances) and of non-application singleton instances was kept equal, both at 50%. These stimuli repeated four times in the final file of each string (8 x 4 times =32 rule sentence pairs, and 8 x 4 times = 32 non-application singleton instances). The total durations of the training strings were 341 s for the "Rule 1 + non-application" condition and 340 s for the "Rule 2 + non-application" condition.

The test stimuli were two other novel ABC sentences and their shifted version (ABC–BAC and ABC–ACB). In both training and test, the original and the shifted versions within a sentence pair were separated by about 700 msec. The pause between "rule" and "non-application" types, between any pairs, and between any two non-application cases was longer, approximately 1,200 msec.

In the training strings, average sentence duration was 2.49 s (*SD* = 0.2) for Rule 1, 2.51 s (*SD* = 0.21) for Rule 2, and 2.39 s (*SD* = 0.15) for non-application sentences. In the test strings, average sentence duration was 2.53 s (*SD* = 0.096) for Rule 1 and 2.54 s (*SD* = 0.05) for Rule 2.

The visual stimulus for all trials was an animation of colorful circles gradually changing sizes on a white background of a screen. In addition, an animation of blue bubbles accompanied by cricket calls was used as the attention-getter. These stimuli were the same across all experiments in this study.

2.3. *Design and procedure*

The present experiment consisted of the same steps (see Table 2) as those of Koulaguina and Shi (2013), in which infants showed the capacity to learn and generalize similar word-order shift rules from fully rule-consistent input. Our sentences, however, were all different from those in that study.

2.3.1. *Training (passive listening)*

Each infant was exposed to one of the training conditions, either the "Rule 1 + non-application" (i.e., ABC–BAC pairs intermixed with ABC singleton sentences), or the "Rule 2 + non-application" training set (i.e., ABC–ACB pairs intermixed with ABC singleton sentences). Half of the infants were randomly assigned to one condition, the other half to the other condition.

Table 2. Experimental phases and speech examples in Experiment 1.

Experimental phase	Auditory stimuli
1. Training phase	Rule input (8 sentence pairs, 4 times each), e.g., ABC–BAC: *Vika darit murku – Darit Vika murku* Non-application singletons (8 sentences, 4 times each), e.g., ABC: *Gena vidit lavku*
2. Pre-test	Trial 1 (or 2): one novel sentence & its shifted version in the trained rule, e.g., ABC–BAC: *Njura topit pechku –Topit Njura pechku* Trial 2 (or 1): the other novel sentence & its shifted version in the non-trained rule, e.g., ABC–ACB: *Tjoma rubit lipku –Tjoma lipku rubit*
3. Contingency training	Sinewave sound
4. Test phase	Same sentences as in the pre-test; Trained rule versus non-trained rule
5. Post-test	Sinewave sound

2.3.2. *Pre-test*

All infants heard one trial with one new test sentence undergoing the ABC to BAC shift (i.e., Rule 1) and another trial with another novel sentence undergoing the ABC to ACB shift (i.e., Rule 2). The sentences were identical to the sentences used as test stimuli in Step 4 (see the following). Each trial was initiated when the infant looked at the screen at about one meter in front and terminated when the sentence pair was played once until the end regardless of whether the child continued to look at the screen during the trial. This pre-test phase allowed infants to hear one full version of each test stimulus pair. This served as a basis for the potential recognition of the particular sentences, each associated with one of the two rules, after infants began hearing the early part of the stimuli in a trial in the test phase (Step 4). The status of the first pre-test trial was counterbalanced across infants such that half heard the Rule 1 pre-test trial first, and the other half heard the Rule 2 pre-test trial first. The duration of each pre-test trial was of 6 s (Step 2). The particular sentences and rule application were counterbalanced across infants—that is, one group of infants heard Sentence 1 in Rule 1 and Sentence 2 in Rule 2, while another group heard the Sentence 1 in Rule 2 and Sentence 2 in Rule 1.

2.3.3. *Contingency training*

Two contingency training trials were designed to teach the infants that they could fully control the duration of trials by their eyes. Auditory stimuli were sinewave sound. A trial was initiated when the infant looked at the screen, and it terminated if the infant looked away from the screen. Minimum look-away for terminating a trial was 2 s. The maximum duration of each trial was 9 sec if looking lasted till the end of a trial. Trials starting from this step were all fully infant controlled.

2.3.4. *Test phase*

The test stimuli were the same two novel sentence pairs as in Step 2, one pair for Rule 1, and the other for Rule 2. The test stimuli were characterized by two trial types, the trained rule versus the non-trained rule. Infants from both training conditions all heard the same test stimulus types. The trained rule for one group was the non-trained rule for the other group and vice versa. The two types of trials alternated for a total of 10 test trials in Step 4.

This phase differed from Step 2 only in two aspects: The trials were fully infant controlled, and the sentence pair within each test trial was presented three times if the infant looked until the end of the trial. The maximal trial duration was 21 s if the infant did not look away for more than 2 s during a trial. The counterbalancing of the first test trial as well as the particular sentence and rule application followed that of Step 2. For example, if Rule 1 was presented first in Step 2, it was also presented first in Step 4.

2.3.5. *Post-experimental phase*

One trial (21 s) presented the same auditory and visual stimuli as those in the contingency training trials. This trial marked the end of the experiment.

Each infant was tested individually. For the passive listening phase (Step 1, training phase), the infant and the parent were invited to a sound chamber. There was a TV screen and a sofa in the room, and speakers for auditory presentation were adjacent to the left and right sides of the TV screen. The infant was given toys, a way to forestall boredom. The parent was instructed to keep silent. She or he could play silently with the child. The child could move around freely in the room. During the presentation of the sentences, an animation of bright multicolored circles that slowly changed sizes was shown on the screen.

After the passive listening phase, the parent and infant left the toys behind and moved to another acoustic chamber for Steps 2–5 of the experiment, which were executed by an experimental program (Cohen, Atkinson & Chaput 2000). The change of environment between the training and the test was intended to maintain the infant's attention for the test phase. The infant sat on the parent's lap facing a TV monitor. The parent wore headphones to hear masking music. She or he was asked not to interact with the infant. The experimenter, who was blind to the audiovisual stimuli, observed the infant's eye movement from a closed-circuit TV in an adjacent room. The experimental software presented the stimuli and automatically recorded all looking times. Each trial in Steps 2–5 was initiated by the infant's looking toward the screen.

For all trials, speech stimuli were presented simultaneously with the visual stimuli of circles growing and reducing in size on the screen. Between trials, the attention-getter (blue bubbles accompanied by a cricket sound) appeared automatically on the screen.

The dependent measure was looking times toward the screen during test trials that presented speech conforming to the trained rule and during test trials that presented speech confirming to the non-trained rule, i.e., Step 4. If infants learned and generalized the rule, they should yield a significant looking-time difference for the two trial types in the test phase.

2.4. Results and discussion

Each infant's looking times were calculated for the two test trial types in Step 4, i.e., trials for the trained rule versus those for the other rule that had not been present in the training. The average looking time per trial was taken for each of the two trial types.

To assess if the specific rules in the training input for counterbalancing purposes (Rule 1 ABC–BAC or Rule 2 ABC–ACB) yielded differences in rule learning, we first ran a mixed ANOVA (within-subject factor: Test Rule [trained vs. non-trained]; between-subject factor: Training Group [Rule 1 vs. Rule 2]). There was no main effect of Test Rule, $F(1,14) = 0.031$, $p = .862$, nor main effect of Training Group, $F(1,14) = 0.003$, $p = .96$, nor an interaction between the two, $F(1,14) = 0.025$, $p = .875$. Since the interaction was not significant, we collapsed the data of the two groups in subsequent analyses.

A paired samples *t*-test revealed that infants showed no discrimination between the trained and non-trained rules, $t(15) = 0.18$, $p = .858$, *two-tailed*, *Cohen's d* = 0.046. Average looking time per trial was 7.4 s ($SE = 1.06$) for the trained rule and 7.19 s ($SE = 0.91$) for the non-trained one (see Figure 1).

The results show no evidence of rule learning, suggesting that the non-application cases were treated as rule-deviant instances. Had infants treated the singletons as possible targets for rule application, the training input would have been fully consistent with the rule, and infants should have learned the rule and generalized it to the new instances. Instead, infants perceived the non-application cases as distinct from the rule exemplars. In this input condition (i.e., 50% non-application cases), rule generalization did not occur.

Having established that the non-application cases were perceived by infants as non-rule instances, we then conducted Experiment 2 to examine the distribution of rule cases that might lead to rule generalization. In Gómez & Lakusta (2004), in which rule generalization was observed, the type

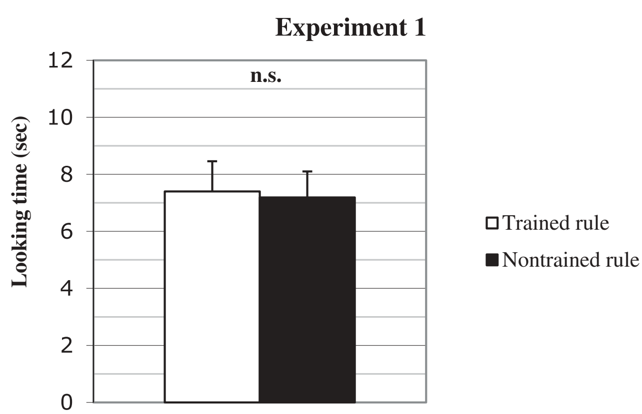

Figure 1. Means and standard errors of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 1.

frequency and the overall frequency of rule instances were kept equal, both high relative to rule-violation cases. The two kinds of frequencies were not separated. Thus, it remains unclear which kind was the determining factor. In our Experiment 1 the non-rule cases (i.e., the non-application cases) were equally high in type frequency and overall frequency. Thus, it was uncertain which of the two factors might have affected rule learning. We suggest that it was the high type frequency of non-rule cases, but not their overall frequency, that impeded rule generalization in Experiment 1. In Experiment 2 we teased apart the two kinds of frequency: type frequency and overall frequency. We hypothesized that low type frequency of non-rule cases and high type frequency of rule cases should lead to rule generalization, even if the overall frequency of non-rule cases may be high due to many repetitions (i.e., tokens) of the few instances. Our manipulation of high token frequency for non-rule cases in fact resembles what occurs commonly in natural languages: Rule exceptions tend to be frequent items, e.g., past-tense irregulars such as *went* in English. Based on our hypothesis, the high token frequency of non-rule items should not be problematic for rule learning as long as they are low in type frequency relative to rule instances.

3. Experiment 2

The purpose of this experiment was to test the hypothesis that high type frequency of rule instances relative to non-rule instances influences rule learning. To do so, we modified the relative type ratio between rule instances and non-rule instances of Experiment 1. The proportion of the type frequency of rule instances was now much higher, whereas their overall frequency was kept the same as that for non-rule instances. That is, relative to the non-rule cases, the rule instances were nondominant in overall frequency; the type frequency of non-rule cases was made very few, but their overall frequency was made relatively high by increasing their repetitions (i.e., token frequency).

Specifically, we used the same rule instances of the training input in Experiment 1 but reduced the type frequency of non-application cases from eight to two, while increasing the number of the repetitions (i.e., token frequency) of each of the two from 4 to 16 times. This assured that the overall frequency of rule and non-rule instances stayed equal, both 32 (for rule, 8 types x 4 repetitions = 32, and for non-rule, 2 types x 16 repetitions = 32). In contrast, as a result of the manipulation, type frequency differed for the two kinds of instances, 80% for rule and 20% for non-rule. The manipulations of type frequency and overall frequency distributions across experiments are summarized in Table 3.

If, as we hypothesized, infants rely on the relative high type frequency of rule instances to learn rules, they should succeed in Experiment 2. If, alternatively, it is more important to hear dominant overall frequency of rule instances, rule generalization should fail.

3.1. *Participants*

Sixteen infants (eight boys) aged 14 months from various linguistic backgrounds completed the experiment. Their ages ranged from 14 months and 15 days to 15 months and 5 days (M = 14 months and 26 days). None had any prior exposure to Russian. Eight other infants were tested, but their data were not included in the analysis for various reasons, including fussiness (1), crying (2), parental interference (2), experimenter error (2), and looking toward the screen for 2 seconds or less on six or more test trials (1).

Table 3. Summary of the training input distributions in Experiments 1–4.

	Experiments 1 & 3	Experiments 2 & 4
Number of rule cases	8 sentence pairs x 4 times	8 sentence pairs x 4 times
Type frequency	50%	80%
Overall frequency	50%	50%
Number of non-rule cases	8 singletons x 4 times	2 singletons x 16 times
Type frequency	50%	20%
Overall frequency	50%	50%

Table 4. Sentences used as stimuli in Experiment 2.

Experimental phase	Base ABC sentences for the word order shift rules (either ABC–BAC or ABC–ACB; each rule instance consisted of a base sentence and its shifted version)	Non-application sentences (non-shift singletons)
Training phase	*Vika darit murku*	*Gena vidit lavku*
	Dima gonit galku	*Lida zharit manku*
	Lera nosit dochku	
	Njuta varit kashku	
	Gosha lovit rybku	
	Vova katit lodku	
	Rada manit koshku	
	Zhora lepit belku	
Test phase	*Njura topit pechku*	
	Tjoma rubit lipku	

3.2. *Materials and design*

Materials were 12 Russian sentences, a subset from Experiment 1 (see Table 4). Of the 10 sentences, 8 were used for creating the instances of the word-order shift rules (Rule 1: ABC–BAC, for Condition 1; Rule 2: ABC–ACB, for Condition 2). Two non-rule sentences were non-application cases, appearing in the base order (ABC) without any alternation. These base-order sentences appeared in both training conditions. Thus, by type frequency distribution the rule instances (80%) were dominant over non-rule instances (20%).

For each of the eight rule instances (either Rule 1 or Rule 2) in the training input, the sentence pair consisted of the base sentence immediately followed by its alternating version, and the eight instances were intermixed randomly with two singleton sentences. In the sound file for each of the two training strings, one for each condition, the eight sentence pairs for the rule each occurred four times, and the two singleton sentences each occurred 16 times. This way, the total number of occurrences (i.e., the overall frequency) of the pairs and that of the singletons was kept equal (8 x 4 = 32 rule sentence pairs, and 2 x 16 = 32 non-rule instances), both at 50%. The token-per-type frequency was thus four times higher for non-rule instances than for rule instances. The total durations of the training strings were 341 s for the “Rule 1 + non-rule” training condition, and 340 s for the “Rule 2 + non-rule” training condition. In the training strings, average sentence duration was 2.49 s (SD = 0.2) for Rule 1, 2.51 s (SD = 0.21) for Rule 2, and 2.32 s (SD = 0.11) for non-rule sentences.

The test stimuli were those of Experiment 1, i.e., the two novel ABC sentences and their moved version (ABC–BAC and ABC–ACB). In both training and test, the original and the shifted versions within a sentence pair were separated by about 700 msec, and the pause between “rule” and “non-rule” types, between any pairs, and between any two non-rule sentences was longer, approximately 1,200 msec, as in Experiment 1.

Thus, Experiment 2 differed from Experiment 1 only in the training input. All the stimuli from Step 2 to Step 5 (see Table 2) were identical for the two experiments.

3.3. *Procedure*

The procedure was the same as in Experiment 1.

3.4. *Results and discussion*

As in Experiment 1, each infant’s looking times during the two types of test trials (i.e., trained rule versus non-trained rule) were calculated. To assess if the specific rules in the training input for counterbalancing purposes (Rule 1 ABC–BAC or Rule 2 ABC–ACB) yielded differences in rule learning, we first ran a mixed ANOVA—within-subject factor: Test Rule (trained vs. non-trained); between-subject factor: Training Group (Rule 1 vs. Rule 2). The main effect of Test Rule,

$F(1,14) = 7.303$, $p = .017$, was significant, but not the main effect of Training Group, $F(1,14) = 0.313$, $p = .585$, nor the interaction between the two, $F(1,14) = 1.626$, $p = .223$. Since the interaction was not significant, we collapsed the data of the two groups in subsequent analyses.

A paired samples *t*-test revealed that looking times for the trained and non-trained rules were significantly different, $t(15) = -2.65$, $p = .018$, *two-tailed*, *Cohen's d* = –0.662. Across the two training conditions, average looking time per trial was 5.89 s (*SE* = 0.73) for the trained rule and 8.47 s (*SE* = 1.07) for the non-trained one (see Figure 2).

Thus, infants discriminated between the trained and non-trained rule applied to novel sentences. The overall frequency of non-rule instances (50%) was kept the same for both Experiments 1 and 2. The crucial difference was the proportion of type frequency of rule instances. In Experiment 2 the type frequency of rule instances was dominant (80%), and infants learned the rule. In contrast, the rule instances were nondominant in type frequency (50%) in Experiment 1, leading to no rule generalization. Crucially, the type frequency of rule cases was identical in absolute number across the two experiments. However, the relative type frequency of rule cases was higher in Experiment 2 due to the lowered type frequency of non-rule cases. Therefore, the relative high type frequency of rule instances was the determining factor for the generalization of word-order shift rules.

The A, B, and C words in the Experiments 1 and 2 were marked consistently by distinct morphological endings (*-a* for A, *-it* for B, *-ku* for C). There is evidence in the literature that infants as young as 11 months of age segment and encode consistent morphological endings from variable stems and can generalize this analysis to new words containing novel stems (Marquis & Shi 2012). Is it possible that infants in Experiment 2 simply encoded the alternation patterns of these specific endings, rather than forming the abstract rule involving the whole test words (stems plus endings)? Specifically, are the results in Experiment 2 simply about the tracking of specific endings that were heard during training, i.e., "*-a* … *-it* … *-ku*" shifted to "… *-it* … *-a* … *-ku*", or shifted to "… *-a* … *-ku* … *-it*"? We argue that this was not the case. Importantly, the morphological endings and their shifts in the training input were identical across the two experiments, occurring with exactly the same number of repetitions. Hence, if the successful learning in Experiment 2 was only about tracking those specific items, they should have also shown discrimination in Experiment 1. They should have recognized the alternations of endings even better in Experiment 1 than in Experiment 2, as the number of stems was greater in the input of Experiment 1, favorable for tracking the endings in light of Marquis and Shi (2012). However, infants showed no learning in Experiment 1. The combined results suggest that the learning in Experiment 2 was abstract: Infants formed the alternation rules for the whole words and generalized the rules to novel instances containing new stems.

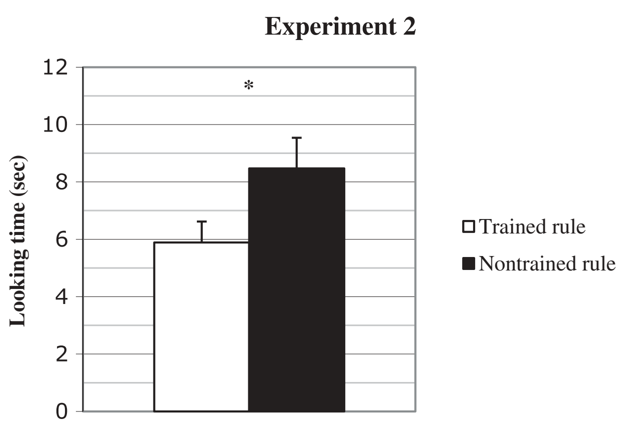

Figure 2. Means and standard errors of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 2.

To further examine the role of type frequency and the abstractness of the rule learning, we conducted a 2 x 2 mixed ANOVA for Experiments 1 and 2. The dependent variable was looking time during test trials. The within-subject factor was Test Rule (trained vs. non-trained), whereas the between-subject factor was Type Frequency of Rule Instances (80% in Experiment 2 vs. 50% in Experiment 1). No significant effect of Test Rule was found, $F(1, 30) = 2.506$, $p = .124$. The between-subject effect of Type Frequency was not significant, $F(1, 30) = 0.011$, $p = .919$. The interaction between Test Rule and Type Frequency approached significance, $F(1, 30) = 3.461$, $p = .073$, suggesting that different type frequencies indeed affected the results in the two experiments. However, the fact that the interaction between Test Rule and Type Frequency was only marginally significant suggests that morphological marking may have led infants to be somewhat uncommitted to the trained word order shift rule in Experiment 2. This agrees with the common observation that languages with rich morphological markings (e.g., Russian) are freer in word order than languages with less or no morphology (e.g., Chinese).

To control for the influence of the morphological markings on infants' learning of word-order shift rules and to definitively test abstract rule generalization, we conducted Experiments 3 and 4 using the same design as in Experiments 1 and 2 respectively, except that there were no consistent morphological markings and the test words had no shared elements with the training input.

4. Experiment 3

Experiment 3 was identical to Experiment 1 in all respects except that A, B, and C words in the training and test stimuli contained no consistent morphological markings. The words contained variable endings, since Russian is a heavily morphological language. However, for infants unfamiliar with Russian, the variable endings did not constitute any cue that could assist the learning of the word order shift rules. Therefore, in all the following description, the absence of consistent morphological markings will be referred to as 'no markings.'

4.1. *Participants*

Sixteen infants (nine boys and seven girls) aged 14 months from various linguistic backgrounds completed the experiment. Their age ranged from 14 months and 13 days to 15 months and 14 days (M = 15 months and 0 days). None of the infants had any prior exposure to Russian. Seven other infants were tested, but their data were not included in the analysis for various reasons such as fussiness (1), crying (1), experimenter error (1), and looking to the screen for 2 seconds and less during six or more test trials (4). Five other infants did not complete the experiment.

4.2. *Materials*

Eighteen new sentences with new words were created and produced by the same Russian native speaker as in Experiments 1 and 2. Of those sentences, 16 were used as the training stimuli (see Table 5), and two as novel instances in the test phase. All words contained two syllables. To obtain the unmarked learning input, we created the original sentences (ABC) with variable parts of speech for words in A, B, and C positions. For example, words in the A position could be a noun, a verb, or an adverb.

As in Experiment 1, eight base sentences were used to form rule instances in the training. Each of the rule instances had two alternating counterparts (Rule 1: ABC → BAC; Rule 2: ABC → ACB). Eight other sentences were non-rule singletons (i.e., ABC) with no alternation.

As in the other experiments, the original and the shifted versions within any sentence pair were separated by approximately 700 ms. The pause between rule and non-rule types, between any pairs, and between any two non-rule sentences was approximately 1,200 ms.

Table 5. Sentences used as stimuli in Experiment 3.

Experimental phase	Base ABC sentences for the word order shift rules (either ABC–BAC or ABC–ACB; each rule instance consisted of a base sentence and its shifted version)	Non-application sentences (non-shift singletons)
Training phase	*Machty gnutsja lukom*	*Stanut reki polny*
	Zina gladit plat'e	*Otzvuk smekha sladok*
	Pojte pesnju druzhno	*Kozam travok ssyplju*
	Veter vybil okna	*Seno pahnet volej*
	Dimke snilos' pole	*Skrojut tuchi solntse*
	Chistim tufli vaksoj	*Tanets veren bubnu*
	Budesh vilkoj kushat'	*Pishem bukvy krasnym*
	Flagi utrom snjali	*Obuv' skinul rezvo*
Test phase	*Snova milyj vesel*	
	Vizhu nosik belki	

In the training strings, average sentence duration was 2.63 s (SD = 0.19) for Rule 1, 2.65 s (SD = 0.18) for Rule 2. The average sentence duration for non-rule sentences was 2.46 s (SD = 0.18). In the test strings, average sentence duration was 2.55 s (SD = 0.09) for Rule 1 and 2.55 s (SD = 0.09) for Rule 2.

4.3. *Design and procedure*

The design and procedure were identical to those in Experiment 1. The training input consisted of eight rule instances. For each rule instance the base sentence immediately went through a word-order shift. There were eight singleton non-alternating ABC sentences. Hence, the non-rule cases were 50% both by overall frequency and by type frequency relative to the rule instances. We expected that rule generalization should be impeded in Experiment 3, similarly to Experiment 1.

The total duration of the new training stimuli was 343.7 s for the "Rule 1 + non-rule" training condition and 344.4 s for the "Rule 2 + non-rule" condition. The maximal duration of pre-test, contingency training, test trials, and post-test were identical to those in Experiments 1 and 2.

4.4. *Results and discussion*

Each infant's looking times during the two test trial types (trained rule versus non-trained rule) were calculated. As with Experiments 1 and 2, we first ran a mixed ANOVA—within-subject factor: Test Rule (trained vs. non-trained); between-subject factor: Training Group (Rule 1 vs. Rule 2). The main effect of Test Rule, $F(1,14) = 0.316$, $p = .583$, was nonsignificant, and the main effect of Training Group, $F(1,14) = 3.568$, $p = .08$, did not reach significance. The interaction between the two was also not significant, $F(1,14) = 1.258$, $p = .281$. Since the interaction was not significant, we collapsed the data of the two groups in subsequent analyses.

A paired samples t-test revealed no discrimination between the trained and non-trained rules, $t(15) = 0.56$, $p = .585$, *two-tailed*, *Cohen's* $d = 0.14$. Average looking time per trial was 7.98 s (SE = 1.33) for the trained rule and 7.51 s (SE = 1.04) for the non-trained rule (see Figure 3).

We further compared infants' performance with and without consistent morphological markings, i.e., Experiments 1 versus 3, under the conditions of the identical distribution of rule and non-rule instances. Both Experiments 1 and 3 had high ratio of non-rule instances in type frequency, and both experiments had nondominant rule instances. We conducted a 2 x 2 mixed ANOVA. The dependent variable was looking time during test trials. The within-subject factor was Test Rule (trained vs. non-trained), and the between-subject factor was Morphological Cues (present in Experiment 1, absent in Experiment 3). There was no effect of Test Rule, $F(1, 30) = 0.228$, $p = .636$. The between-subject effect of Morphological Cues was also not significant, $F(1, 30) = 0.107$, $p = .745$. Importantly, there was no interaction between Test Rule and Morphological Cues, $F(1,30) = 0.034$, p = .855, suggesting that looking responses in Experiments 1 and 3 were the same. Both experiments showed no evidence of rule generalization.

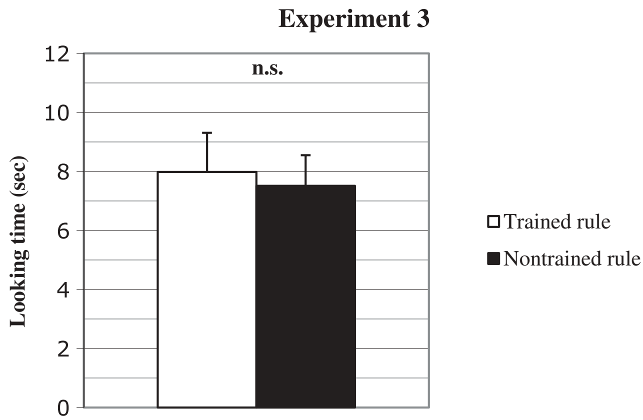

Figure 3. Means and standard errors of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 3.

The results of Experiment 3 suggest that in the absence of any morphological cues, infants interpreted the non-application singleton cases as non-rule instances, as in Experiment 1. Given the successful learning in Experiment 2, the lack of generalization in Experiments 1 and 3 must have been due to the high type frequency of non-rule instances (50%).

In Experiment 4 we again tested the hypothesis that lowering the type frequency of non-rule instances (thus in effect increasing the relative type frequency of rule instances) would lead to rule generalization, as in Experiment 2. The only difference between Experiments 2 and 4 was that the stimuli in the latter experiment contained no morphological cues, which enabled us to clearly examine the role of type frequency for abstract rule learning.

5. Experiment 4

5.1. *Participants*

Sixteen infants (10 boys and 6 girls) aged 14 months from various linguistic backgrounds completed the experiment. Their age ranged from 14 months and 9 days to 14 months and 29 days (M = 14 months and 18 days). None of the infants had any prior exposure to Russian. Eight other infants were tested, but their data were not included in the analysis for various reasons such as fussiness (1), crying (2), parental interference (3), experimenter error (1), and looking to the screen for 2 seconds and less during six or more test trials (1). Four other infants did not complete the experiment.

5.2. *Materials*

In Experiment 4 we used the same training stimuli as in Experiment 3, except that for the non-application instances, we kept only two of the eight sentences (see Table 6).

Specifically, eight sentences were used to form rule instances in the training input, and two other sentences were non-application instances. Each of the rule instances had two alternating counterparts (Rule 1: ABC → BAC; Rule 2: ABC → ACB). The two non-rule sentences were singletons (i.e., ABC) with no alternation. The interstimulus intervals (ISIs) were the same as in Experiments 1–3. For rule sentences in the training, the average sentence duration was the same as in Experiment 3. For non-rule sentences, the average sentence duration was 2.59 s (SD = 0.16). The test sentences were those of Experiment 3.

Table 6. Sentences used as stimuli in Experiment 4.

Experimental phase	Base ABC sentences for the word order shift rules (either ABC–BAC or ABC–ACB; each rule instance consisted of a base sentence and its shifted version)	Non-application sentences (non-shift singletons)
Training phase	*Machty gnutsja lukom* *Zina gladit plat'e* *Pojte pesnju druzhno* *Veter vybil okna* *Dimke snilos' pole* *Chistim tufli vaksoj* *Budesh vilkoj kushat'* *Flagi utrom snjali*	*Stanut reki polny* *Otzvuk smekha sladok*
Test phase	*Snova milyj vesel* *Vizhu nosik belki*	

5.3. Design and procedure

The design and procedure were identical to Experiment 2. The training input consisted of eight rule instances and two singleton instances. The overall frequency of rule and singleton instances was kept the same (both at 50%), while the type frequency for rule instances was dominant (80%). This was achieved by increasing the repetition (i.e., token frequency) of the singleton utterances more than the rule utterances. The crucial difference from Experiment 2 was the absence of morphological cues in the stimuli of Experiment 4, allowing us to test unambiguously if infants could learn the abstract rules based on the dominant type frequency of rule instances and generalize to novel instances.

The total duration of the training phase was 349.7 s for the "Rule 1 + non-rule" training condition and 350.3 s for "Rule 2 + non-rule." The test stimuli as well as all the other stimuli starting from Step 2 (pre-test) were the same as in Experiment 3.

5.4. Results and discussion

Each infant's looking times during the two test trial types (trained rule versus non-trained rule) were calculated. As was done in previous experiments, we first ran a mixed ANOVA—within-subject factor: Test Rule (trained vs. non-trained); between-subject factor: Training Group (Rule 1 vs. Rule 2). The main effect of Test Rule, $F(1,14) = 7.588$, $p = .016$, was significant, but not the main effect of Training Group, $F(1,14) = 0.963$, $p = .343$, nor the interaction between the two, $F(1,14) = 0.311$, $p = .586$. Given the nonsignificant interaction, we collapsed the data of the two groups in subsequent analyses.

A paired samples *t*-test showed a significant discrimination between the trained and non-trained rules, $t(15) = -2.82$, $p = .013$, *two-tailed*, *Cohen's d* = −0.705. Average looking time per trial was 8.24 s ($SE = 1.2$) for the trained rule and 9.88 s ($SE = 1.27$) for the non-trained rule (see Figure 4).

To further examine infants' generalization in the absence of morphological cues, we conducted a mixed ANOVA to compare infants' responses across Experiments 3 and 4. The within-subject factor was Test Rule (trained vs. non-trained), and the between-factor was the Type Frequency of Rule Instances (80% in Experiment 4 vs. 50% in Experiment 3). There was no effect of Test Rule, $F(1, 30) = 1.324$, $p = .259$. The between-subject effect of Type Frequency was also not significant, $F(1, 30) = 0.645$, $p = .428$. Crucially, the interaction between Test Rule and Type Frequency was significant, $F(1, 30) = 4.271$, $p = .048$, indicating that infants in Experiments 3 and 4 responded differently to the trained and the non-trained rules, depending on the Type Frequency of rule instances in their training input. The significant interaction here contrasts with the ANOVA results for morphologically marked Experiments 1 and 2, where the interaction between Test Rule and Type Frequency did not reach significance (see the Results section of Experiment 2). This suggests that the morphological cues did not assist infants' learning of the word-order shift rules in those experiments. On the contrary, the morphological markings might have made infants

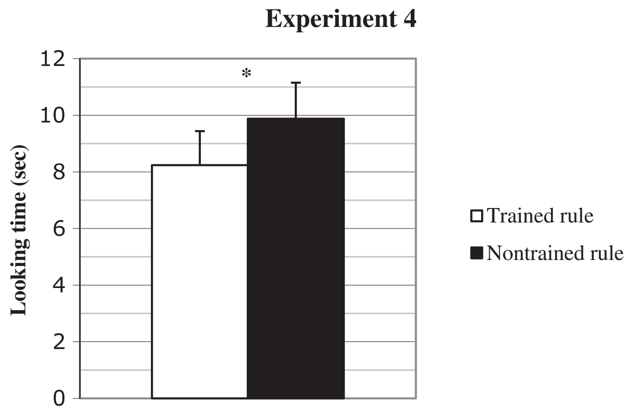

Figure 4. Means and standard errors of the average looking time per trial for test trials conforming to the trained rule vs. for the non-trained rule in Experiment 4.

slightly less committed to the trained word-order shift rules in Experiments 2 than in Experiment 4. Nevertheless, given the near significant Test Rule x Type Frequency interaction in Experiments 1 and 2, caution would be needed in interpreting the role of morphological markings for the learning of word-order shift rules.

To directly compare infants' performance with and without morphological cues under the conditions of the identical distribution of rule and non-rule instances, we conducted a 2 x 2 mixed ANOVA for Experiments 2 and 4. For both experiments the overall frequency of rule instances was relatively low (50% of all training input), but their type frequency was dominant for rule (80%) over the non-rule instances (20%). The only difference was the presence (Experiment 2) versus absence (Experiment 4) of morphological cues. The within-subject factor was Test Rule (trained vs. non-trained), and the between-subject factor was Morphological Cues (present in Experiment 2, absent in Experiment 4). The analysis showed a highly significant main effect of Test Rule, $F(1, 30) = 13.842$, $p = .001$. The between-subject effect of Morphological Cues was not significant, $F(1, 30) = 1.728$, $p = .199$. There was no interaction between Test Rule and Morphological Cues, $F(1, 30) = .677$, $p = .417$. This analysis suggests that with or without the morphological cues, infants generalized the rule to the new test instances, showing that the significant effect of the dominant type frequency of rule instances was comparable for Experiments 2 and 4.

These results suggest that infants successfully learned the rules and generalized them to novel instances based on the high type frequency of rule instances in the training input in Experiment 4, as in Experiment 2. In contrast, learning in Experiment 3 was impeded when the type frequency of rule instances was nondominant in comparison to that of non-rule instances. Furthermore, since infants had no morphological cues to rely on, the effect shown in Experiment 4 was a strong demonstration that infants learned the rules at an abstract level.

6. General discussion

In our study we considered several aspects of productive rule learning in early infancy. It is known that infants receive no explicit description of rules from caregivers. Their rule learning is implicit. They rely exclusively on exemplars in the input. However, exemplars supporting a rule often co-occur with violation cases, as is commonly seen for linguistic rules. The task is especially complicated by the fact that there is no negative evidence telling the child which are violation cases.

We examined a particular kind of input exemplars that has been the center of a learnability issue (since Baker 1979): non-application cases (related to "unattested cases" in the literature). Non-application cases (e.g., if a child heard only *walk* in his input, but not *walked*) can be exemplars for which a rule may apply, consistent with the productive nature of abstract rules. However, certain non-application cases are exceptions for which the rule should not apply—for example, the prohibition of the dative shift with certain verbs (e.g., *donate*). Past research focused on the question of how children can resolve this paradox and learn the non-application cases as exceptions even though no negative evidence is available. Our study asked a closely related question: How do non-application cases impact rule learning per se? To address this question, we tested infants' rule generalization to entirely novel cases rather than to non-application cases.

In Experiments 1 and 3, half of our training instances that each infant heard were sentence pairs supporting a word-order shift rule, and the remaining half were non-application singleton sentences that showed no shift. If infants treated the non-application sentences as possible targets for rule application, then the larger input including both shift and non-shift instances can be viewed as fully consistent for rule learning. However, our results showed that infants did not learn the rule. They failed to generalize the rules to completely new sentences. This suggests that the non-application cases in the training set were treated as non-rule instances. Relative to Experiments 2 and 4, the type frequency of non-application cases was too high in Experiments 1 and 3, impeding rule learning. When we lowered the type frequency of non-application cases in Experiments 2 and 4 (which in effect made the rule instances dominant in type frequency), rule learning was successful: Infants generalized the rule to new non-trained sentences. The results of the four experiments are consistent with the way the connectionist model of Rumelhart and McClelland (1986) behaves with respect to the English past tense and consistent with the general characteristics of the Tolerance Principle of Yang (2005, 2016). The results can only be explained if infants perceived the non-application exemplars as rule-deviant cases.

Is this finding paradoxical with the productive nature of rule generalization? We suggest that it is not. The evidence that infants treat non-application cases as deviant cases from rule instances does not mean that they cannot apply the learned rule to these non-application exemplars. As long as rule instances are sufficiently high in relative type frequency, infants can learn the general rule, and moreover, it has been shown in the literature that children have a tendency to even regularize inconsistent input (Hudson Kam & Newport 2005; Wonnacott 2011). Hence, treating non-application cases as violations does not have to prevent them from being generalized to a rule once the rule has already been learned. However, certain non-application cases are true exceptions and need to be inhibited from being overgeneralized. This can be resolved by independent means. In our previous study (Koulaguina & Shi, 2010) with the same training input (i.e., including the same non-application cases) as in Experiment 2 of the present study, 14-month-old infants avoided overgeneralizing these singleton cases when their token frequency was high. Similarly, Wonnacott, Newport & Tanenhous (2008) showed that high token frequency of specific sentences led adults to use those sentences in their trained specific forms and to avoid using them with the dominant rule. The learning system seems to be equipped with mechanisms to manage between productive use and exceptions, despite the seemingly contradictory non-application cases. We therefore suggest that Baker's Paradox is not paradoxical after all.

The second aspect of rule learning that we examined concerned the distributional properties of the training input underlying rule learning. In Experiments 2 and 4 we teased apart type frequency versus overall frequency of rule exemplars, making their type frequency dominant and overall frequency nondominant. Thus, infants were expected to show successful learning if they relied on the high type frequency of rule instances. However, they should show no learning if they simply relied on their overall frequency. Our results demonstrated that infants learned the rules and generalized them to new sentences that never appeared in the training set, indicating that high type frequency was the determining factor. This finding is consistent with the connectionist model of Rumelhart & McClelland (1986) and the Tolerance Principle of Yang (2005, 2016), which also assume the pertinence of type frequency. Furthermore, we showed that increasing the token frequency of non-rule instances, which led to the increased overall frequency of these instances, did not impede learning in Experiments 2 and 4. In fact, high token frequency of non-rule instances in Experiments 2 and 4 might have contributed to the rule learning. The repetitions of non-rule tokens could have freed up cognitive resources by reducing processing demands, thus making the type frequency of rule instances more salient.

The input conditions of type and token frequencies for rule learning in our study resemble that of natural languages. Dative alternation in English, for example, shows high type variability. The type frequency of shifting instances is high, whereas the type frequency of non-shifting instances is low (Yang & Montrul 2017), consistent with the input condition of our Experiments 2 and 4. In our Experiments 1 and 3, generalization failed because of high type and low token frequencies of non-rule cases. This situation is similar, for example, to grammatical gender in French, where there is a large type variability of nouns, which have no regular marking for gender. Therefore, it is impossible for a child to form a general rule; he has to learn the gender of each noun individually. In Experiments 2 and 4, on the contrary, the low type frequency of non-rule cases did not impede rule generalization, which was supported by the high type frequency of rule cases. This can be compared, for example, to the situation of grammatical gender in Spanish and Russian, where a large number of nouns of the same gender have the same ending, enabling the learner to form a general rule (e.g., Afonso et al. 2014). A small number of nouns are exceptions to the rule—for example, *mano* 'hand' is feminine and an exception to the general rule in Spanish that the *-o* ending occurs in masculine nouns. Similarly, *papá* 'dad' is masculine and an exception to the general rule in Spanish that the *-a* ending occurs in feminine nouns. Likewise, in Russian, masculine nouns *papa* 'dad' and *dedushka* 'grandfather' are exceptions to the general rule that the *-a* ending occurs in feminine nouns. A small number of such exceptions do not impede the learning of the general rule. Such exceptions have a high token frequency—this is possibly how they were spared from being regularized to the general rule and survived historic linguistic changes.

On the other hand, it should be noted that during language acquisition children's sensitivity to statistical regularities may interact with certain kinds of linguistic representations, which may be UG-based or specific to the child's native language (Lidz & Gagliardi 2015). Their rule learning does not always rely on statistical properties of the input. For example, in Gagliardi & Lidz (2014), Tsez-learning children privileged phonological cues over semantic cues in noun classification despite the fact that semantic cues were statistically dominant in the input. In our study, the training language was novel to infants, and no semantic information was accessible to them. Infants at an early age showed a fundamental sensitivity to type and token frequencies in rule learning.

The third aspect of our experiments was about morphological marking and its interaction with the word-order shift rules. In Experiments 1 and 2 all words were consistently marked by morphological endings, which could serve as the cues to the word positions in the sentences, whereas Experiments 3 and 4 had no consistent markings. We reasoned that there could be two opposing effects of morphological cues on rule learning. The cues could potentially highlight the word-order shift rules and allow infants to perform better in Experiments 1 and 2 than in Experiments 3 and 4. Alternatively, morphological markings could have a reverse effect—they might make the word-order rule harder to learn by making infants tolerate freer word orders, thus accepting both trained and non-trained shifts. This would resemble the fact that natural languages with rich inflectional morphology tend to have more flexible word orders (e.g., Russian) than languages with less or no morphology (e.g., Chinese). Our experiments did not yield supporting evidence for either an impedance effect or a facilitation effect. Overall, results with or without morphological cues were similar: Rule learning succeeded in Experiments 2 and 4 but failed in Experiments 1 and 3. Nevertheless, the interaction between Test Rule and Type Frequency in the ANOVA was significant for the nonmarked Experiments 3 and 4 but did not reach significance for the morphologically marked Experiments 1 and 2, suggesting that morphological markings might have made infants slightly uncommitted to the trained word-order shift rule in Experiment 2.

The consistent morphological markings in Experiments 1 and 2 allowed two possible kinds of learning: (i) abstract word-order shift rules with consistent endings (e.g., A-*a* B-*it* C-*ku* → B-*it* A-*a* C-*ku*; A-*a* B-*it* C-*ku* → A-*a* C-*ku* B-*it*), or (ii) alternations of specific endings (e.g., …-*a* …-*it* …-

ku → *…-it …-a …-ku; …-a …-it …-ku* → *…-a …-ku …-it*), which was not an abstract rule. Had infants only learned (ii), they should have recognized this specific alternation in both experiments, as both experiments presented exactly the same distribution of the specific endings and their alternations; moreover, Experiment 1, which presented more variable stems, could have yielded better recognition of the specific alternations, given that variability of stems can help infants recognize morphological endings (Marquis & Shi 2012). Our results, however, showed that infants went beyond tracking the alternations of the specific endings (as in ii) and were focusing on the abstract sentential structure between words. In Experiment 2 they learned the rules as in (i) and generalized the rules to sentences containing new stems that had never appeared in the training input. Our results demonstrate that the mechanism for abstract rule learning is available and automatic in preverbal infants.

We note a few other important findings of our study. Our results show that infants' rule learning was completely form-based. The use of a foreign language in our study ensured that there was no semantics in the stimuli that was known to any of our infants. Moreover, our infants were only 14 months old and had limited knowledge about the specific syntactic structures of their own language. They would unlikely impose any word-order shift bias from their native language. The rule learning demonstrated in our infants was thus language-general and likely to be the earliest and most basic learning mechanism available to humans. It is also interesting that infants succeeded in rule learning under the condition of passive exposure. Infants were playing toys with the parent while the stimuli were presented in the background. This suggests that abstract and productive rule learning is powerful and can occur unconsciously and automatically.

Taken together, our experiments provide empirical evidence in support of abstract rule learning in early infancy. In our experiments, as well as in our previous study (Koulaguina & Shi, 2010), non-application cases, i.e., those linked to "unattested" cases in the learnability literature, are treated as non-rule instances, the same as overt violation instances (such as in *go* → *went*). The distribution of non-application cases relative to rule instances directly impacts learning. Rule abstraction succeeds when the type frequency of rule input is dominant over that of non-rule cases but fails when the type frequency of rule input is nondominant relative to non-rule cases. Furthermore, our experiments demonstrate that infants learn abstract rules with and without morphological cues.

Acknowledgments

Material in this article formed part of the Ph.D. dissertation of the first author. This research was supported by NSERC, SSHRC and CFI grants to the second author. Partial reports of this work were presented at the BUCLD 34 meeting. We thank all parents and infants who participated in this study.

References

Afonso, Olivia, Domínguez, Alberto, Álvarez, Carlos J. & Morales, David. 2014. Sublexical and lexico-syntactic factors in gender access in Spanish. *Journal of Psycholinguistic Research* 43. 13–25.

Baker, Carl Lee 1979. Syntactic theory and the projection problem. *Linguistic Inquiry* 10(4). 533–581.

Cohen, Leslie B., Atkinson, Dan J. & Chaput, Harold H. 2000. *Habit 2000: A new program for testing infant perception and cognition* (Version 2.2.5c) [Computer software]. Austin: University of Texas.

Frank, Michael C., Slemmer, Jonathan A., Marcus, Gary F. & Johnson, Scott P. 2009. Information from multiple modalities helps 5-month-olds learn abstract rules. *Developmental Science* 12. 504–509.

Gagliardi, Annie & J. Lidz, Jeffrey. 2014. Statistical insensitivity in the acquisition of Tsez noun classes. *Language* 90(1). 58–89.

Gerken, LouAnn, Dawson,Colin,Chatila, Razanne & Tenenbaum, Josh. 2014. Surprise! Infants consider possible bases of generalization for a single input example. *Developmental Science* 18. 80–89.

Gerken, LouAnn & Knight, Sara. 2015. Infants generalize from just (the right) four words. *Cognition* 143. 187–192.

Gerken, LouAnn. 2006. Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition* 98. B67–B74.

Gervain, Judit, Macagno, Francesco, Cogoi, S., Peña, Marcela & Mehler, Jacques. 2008. The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences of the United States of America* 105(37). 14222–14227.

Gervain, Judit, Berent, Iris & Werker, Janet F.. 2012. Binding at birth: The newborn brain detects identity relations and sequential position in speech. Journal of Cognitive Neuroscience 24(3). 564–574.

Gómez, Rebecca L. & Lakusta, Laura. 2004. A first step in form-based category abstraction by 12-month-old infants. *Developmental Science* 7. 567–580.

Hudson Kam, Carla L. & Newport, Elissa L. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1. 151–195.

Johnson, Scott P., Fernandes, Keith J., Frank, Michael C., Kirkham, Natasha Z., Marcus, Gary F., Rabagliati, Hugh & Slemmer, Jonathan A. 2009. Abstract rule learning for visual sequences in 8- and 11-month-olds. *Infancy* 14. 2–18.

Koulaguina, Elena & Shi, Rushen. 2010. The mechanism underlying the learning of rules and exceptions in 14month-old infants. In Katie Franich, Kate Mesh Iserman & Lauren Keil (Eds.), A Supplement to the Proceedings of the 34th Boston University Conference on Language Development [BUCLD34], 1–13. Somerville, MA: Cascadilla Press. Retrieved from http://www.bu.edu/bucld/files/2011/05/34-Koulaguina-and-Shi-.pdf.

Koulaguina, Elena & Shi, Rushen. 2013. Abstract rule learning in 11- and 14-month-old infants. *Journal of Psycholinguistic Research* 42(1). 71–80.

Lidz, Jeffrey & Gagliardi, Annie. 2015. How nature meets nurture: Universal Grammar and statistical learning. *Annual Review of Linguistics* 1. 333–353.

Marcus, Gary F., Fernandes, Keith J. & Johnson, Scott P. 2007. Infant rule learning facilitated by speech. *Psychological Science* 18. 387–391.

Marcus, Gary F., Vijayan, Subramanian, Bandi Rao, Shoba & Vishton, Peter M. 1999. Rule-learning in seven-month-old infants. *Science* 283. 77–80.

Marquis, Alexandra & Shi, Rushen. 2012. Initial morphological learning in preverbal infants. *Cognition* 122. 61–66.

Pinker, Steven. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: The MIT Press.

Rabagliati, Hugh, Senghas, Ann, Johnson, Scott & Marcus, Gary F. 2012. Infant rule learning: Advantage language, or advantage speech? *PLoS ONE* 7(7). e40517.

Rumelhart, David E. & McClelland, James L. 1986. On learning the past tenses of English verbs. In David E. Rumelhart & James L. McClelland (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 2, 216–271. Cambridge, MA: The MIT Press.

Saffran, Jenny R., Pollak, Seth D., Seibel, Rebecca L. & Shkolnik, Anna. 2007. Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition* 105. 669–680.

Wonnacott, Elizabeth & Newport, Elissa L. 2005. Novelty and regularization: The effect of novel instances on rule formation. In Alejna Brugos, Manuella R. Clark-Cotton & Seungwan Ha, (eds.), *Proceedings of the 29th Annual Boston University Conference on Language Development [BUCLD]*, 663–673. Somerville, MA: Cascadilla Press.

Wonnacott, Elizabeth. 2011. Balancing generalization and lexical conservatism: An artificial language study with child learners. Journal of Memory and Language 65. 1–14.

Wonnacott, Elizabeth, Newport, Elissa L. & Tanenhaus, Michael K. 2008. Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology* 56. 165–209.

Yang, Charles & Montrul, Silvina. 2017. Learning datives: The Tolerance Principle in first and second language acquisition. *Second Language Research* 33(1). 119–144.

Yang, Charles. 2005. On productivity. *Linguistic Variation Yearbook* 5(1). 265–302.

Yang, Charles. 2016. *The price of linguistic productivity*. Cambridge, MA: The MIT Press.