The mechanism underlying the learning of rules and exceptions in 14-month-old infants

Elena Koulaguina and Rushen Shi
Université du Québec à Montréal

1. **Introduction**

A crucial ability in language acquisition is the abstraction of syntactic rules from exemplars in the input and the generalization of the rules to novel instances. In addition to rule learning and generalization, infants also need to learn particular cases which do not apply to the rules, that is, exceptions. The learning system must be able to accomplish both linguistic needs. In this study we examine the exact nature of input distributions that lead to rule generalization and exception learning.

With regards to rule generalization, Marcus et al. (1999) showed that by seven months of age infants can already extract rules and, more importantly, generalize them to novel instances. Infant were trained with strings of monosyllabic elements that were combined according to a rule (either ABB or ABA pattern). Infants showed evidence of abstracting the rule and generalizing it to novel items. In Marcus et al. (1999) all the training input conformed to the rule, containing no exceptions (i.e., no noise instances). Gomez and LaKusta (2004) went beyond the ideal noise-free input conditions. They examined how 12-month-old infants' generalization of rules may be affected by input noise levels. Across several experiments, they gradually added noise instances that did not follow the dominant rule in the training input. The instances for the dominant rule went from 100% to 83% and to 67% while noise instances increased. The noise instances followed a different rule. Gomez and LaKusta showed that infants in the 100% and 83% rule-conforming input conditions learned and generalized the dominant rule to novel instances. They failed to do so in the 67% condition.

The results from Gomez and LaKusta (2004) suggest that infants' successful rule learning and generalization require the input to be consistent with the rule by a dominant percentage. A certain amount of noise in the input does not impede rule generalization. The focus of Gomez and LaKusta (2004) was on the effects of overall token frequency of rule instances, i.e. the overall occurrence of a dominant rule relatively to another rule. Other factors such as type frequency (i.e., the number of distinct exemplars) and token-per-type frequency (i.e., the number of occurrences for each distinct exemplar) were not specifically examined. In fact, when they manipulated the overall token frequency of rule instances (hereafter "overall frequency") from 100%, 83%, to 67%, the type frequency co-varied at the same percentages. On the other hand, since the number of occurrences for each rule utterance type remained the same across experiments, the token-per-type frequency was controlled. Therefore, infants' failure in learning in the 67% experiment could be resulted from low type frequency of rule instances, or low overall frequency of rule instances, or both.

Wonnacott et al. (2008) attempted to specifically manipulate type and token-per-type frequencies in a study with adults. Using an artificial language, they designed two training conditions in which verbs were used in distinct verb-argument structures (Verb_Agent_Patient versus Verb_Patient_Agent_ka). The exemplars representing the structures were different in type, token-per-type, and overall frequencies. In one condition, the dominant structure was represented by seven different verbs, whereas the other structure (i.e., noise) appeared exclusively with one verb. This condition is similar to the manipulation of the 83% experiment in Gomez and LaKusta (2004). Hence, the first structure was dominant by both type and overall frequency. The token-per-type frequency (i.e., the number of occurrences for each distinct verb type) was held equal for the two structures. In the other training condition, both structures were represented by all eight verbs, but each verb occurred seven times with one structure and only one time with the other structure. The two structures were therefore equal in verb type frequency, while one of the structures was dominant by token-per-type frequency and by overall frequency. Thus, Wonnacott et al. (2008) dissociated type and token-per-type frequencies across the two input conditions, so that in one condition, a particular verb-argument structure was dominant by type, and in another, dominant by token-per-type frequencies. Following training, participants in both conditions displayed a bias for using novel verbs in the dominant structure. These results may be interpreted as evidence that any dominant frequency, be it either type, or token-per-type can guide rule learning and generalization.

However, the role of token-per-type frequency for rule generalization is not clear because the type frequency in the second condition of Wonnacott et al. (2008) was not strictly controlled. Although the two structures had an equal frequency in terms of verb type (i.e., both structures occurred for each verb type), each verb across the seven repetitions occurred with different noun arguments, such that the dominant structure with each verb was realized as seven different sentence types. In other words, if we consider the type variability due to the combinations of verbs with nouns in the sentences, rather than the verbs alone, one structure becomes dominant by type frequency. The variability of nouns cannot be ignored here since the structures to be learned in Wonnacott et al. (2008) concerned verb-argument relations. Therefore, the rule generalization results shown in this condition may be due to the impact of dominant type frequency, token-per-type frequency and/or overall frequency.

It should be noted that the observed rule learning in both conditions in Wonnacott et al. (2008) could be attributed to the dominance in the overall frequency, as in the study of Gomez & Lakusta (2004). It is possible that participants were simply acquiring a bias towards the structure which they heard more often in general.

One study that did show the differential role of type and token in learning while controlling for the overall frequency is that of Gerken (2006). Nine-month-olds were trained with exemplars of a grammar, either AAB or ABA. In one condition, A's and B's were all variable word types. In another condition, A's were variable as in the first condition, whereas a single B-word was used in all training sequences. When tested on new utterances containing A and B novel words, only infants in the B-variable condition showed evidence of generalizing the trained rule to these cases. Infants from the B-single-word condition did not achieve this more abstract generalization. Instead, they learned that the trained rule (e.g., AAB) must contain the specific B-single-word that had occurred during training. That is, these infants only generalized the trained rule to novel A words, but confined the rule to the specific B-single-word.

The difference between the B-variable and B-single-word input conditions in Gerken (2006) is that in the former, B is higher by type and lower by token frequencies, whereas in the latter, B is higher by token, and lower by type frequencies. Infants learned a grammatical pattern and generalized it to novel utterances containing new B words only after being exposed to the high-type, low-token B-variable condition. Since the overall frequency of B's was the same across both conditions, it could not have been responsible for the differential results of the two conditions. Therefore, it may be the high type frequency that determines generalization to novel cases. The low-type, high-token training, on the contrary, led to the more restrictive learning of specific items (single B-word).

Unlike the studies of Gomez and Lakusta (2004) and Wonnacott (2008), Gerken's experiments (2006) were not designed to test rule generalization when noise cases are present in the input. Our research interest is to understand how type, token-per-type and overall frequencies may contribute differentially to rule generalization and to the learning of rule exceptions when input contains noise. Specifically, in the present study we tested two hypotheses. First, we hypothesized that high type frequency of rule instances, relatively to noise types, leads to rule abstraction. Second, high token-per-type frequency of noise instances should lead to exception learning. That is, noise instances should resist being overgeneralized to a learned rule if their token-per-type frequency is high.

To test these hypotheses, we kept the overall frequency of the rule and noise instances equal within the same training input set. Rule instances were high in type frequency and low in token-per-type frequency. Correspondingly, noise instances were low in type frequency and high in token-per-type frequency. We predicted that after being exposed to this input, which dissociated type and token-per-type frequencies, infants should generalize the type-dominant rule to novel instances (Experiment 2) but should resist the generalization of the rule to the noise items (Experiment 3).

The rules in our design involved word order movement. Rule-governed instances were three-word utterances each followed immediately by the same utterance in which a word order movement occurred. Noise instances, also three-word utterances, were cases with no movement.

Unlike previous studies, which all used artificial languages, we used a natural language, Russian. In Russian, different word orders are all allowed, thus allowing us to produce our stimuli with equal naturalness. Because our infants never heard Russian prior to the study, no semantic information was available in the stimuli to them. As mentioned above, in the training phase we manipulated the type and token-per-type frequencies of sentences conforming to one word order movement rule, as well as those of non-movement sentences (noise), while keeping the overall frequency equal for the movement versus non-movement sentences. In the experiment on rule learning and generalization to novel instances, two different word order movement rules (one of which being the trained rule) were applied in the test phase to novel items (Experiment 2). Experiment 3 tested the learning of exceptions. In this experiment the two word order movement rules were applied in the test phase to the noise sentences from the training phase.

2. Experiment 1

Before examining the effect of rule instances versus noise on learning, it is important to determine whether in the absence of any noise, learning syntactic movement rules from natural language materials would at all be possible. In Marcus et al. (1999) and Gerken (2006) stimuli were simple, monosyllabic CV words combined into three-word strings by rules, and each string included a reduplication of one of two elements (e.g., AAB). Infants as young as seven months of age learned the rules in Marcus et al. (1999), although Gerken (2006) did not fully replicate this finding with seven-month-olds. Gerken obtained robust results only with nine-month-old infants. Our stimuli were of greater complexity. Multisyllabic Russian words were arranged in three-word sentences with no identical word forms. The word order movement required the infants to register each sentence in memory and track its moved version. Given that our task was more demanding in that of Marcus et al. (1999) and Gerken (2006), we chose an older age group, 14-month-old infants.

2.1.1. Participants and Materials

Sixteen infants aged 14 months from various linguistic backgrounds completed the experiment. None of the infants had any prior exposure to Russian.

Materials were 12 Russian sentences recorded by a female Russian native speaker in the child-directed speech style. The speaker clearly separated the words when producing each sentence. Eight of those sentences werse used as training stimuli, and four as novel instances in the test phase. All original sentences had a Subject-Verb-Subordinate structure (i.e., S-V-Sbd). The subordinate part of sentence was variable. It could be an adverbial, a direct object or an indirect object. This design served to diversify the morphological features. All the words were highly variable in phonotactic and morphological properties, and in the number of syllables.

Training sentences presented a word order movement rule applied immediately to each sentence. For example, S-V-Sbd sentence was followed immediately by the same sentence transformed into V-S-Sbd (for one training condition: Rule 1) or into S-Sbd-V (for the other condition: Rule 2). For each of the two training conditions, the eight sentence pairs (S-V-Sbd and its moved version) were presented randomly four times. Sentences within each sentence pairs (i.e., the original and its moved version) were separated by 700 msec. Sentence pairs were separated from other sentence pairs by 1200 msec.

Test materials included four novel sentences not included in the training set. Two of the four test sentences went through the "S-V-Sbd to V-S-Sbd" transformation (Rule 1), and the other two novel sentences went through the "S-V-Sbd to S-Sbd-V" (Rule 2) transformation. These stimuli were separated by the same inter-stimulus intervals as the training materials.

The visual stimulus for all trials was an animation with many multi-colored circles, changing sizes. A water wave sound was used for contingency training and the post-experiment trial. An animation of blue bubbles accompanied by a cricket sound served as the attention getter.

2.1.2. Procedure and Design

The experiment contained the following steps:

1. Training: passive listening phase, when each infant was exposed to either the Rule 1 or Rule 2 training set (i.e., "S-V-Sbd to V-S-Sbd" or "S-V-Sbd to S-Sbd-V" sentences); the duration for the total of Rule 1 training materials was 314 sec, and the duration for Rule 2 training materials was 305 sec.

2. Pre-test: each infant heard two novel test sentences going through S-V-Sbd to V-S-Sbd transformation, and two other novel sentences going through S-V-Sbd to S-Sbd-V transformation. The order of these two trials was counter-balanced across infants. The trials were identical to the test trials in Step 4 (see below). This stage allowed infants to hear one full version of each test stimulus before it could be interrupted by a fully infant-controlled procedure at the later steps of the experiment. It served as a basis for the potential recognition of particular sentences associated with one of the two rules after hearing the very first word in the test phase (Step 4).

3. Contingency training: two contingency training trials were designed to teach the infants that they could control the duration of trials. A trial would terminate if the infant looked away from the screen. Minimum look-away for terminating a trial was 2 sec. The maximum duration of each trial was 9 sec if looking lasted till the end of a trial. Trials starting from this stage were all fully infant-control.

4. Test phase: 10 test trials. This phase repeated exactly what infants had heard during the Pre-test trials (Step 2), except that trials were fully infant-controlled. Maximum trial duration was 21 sec if the infant looked till the end of a trial. Each test trial presented two sentence pairs going through a rule. The order of the two sentences within a trial was fixed.

5. Post-experimental phase: one trial; identical to contingency training trials, except that the maximum trial length was 21 sec. This trial enabled us to determine if the infant was on task throughout the experiment. If so, looking time should increase during this post-trial, which presented auditory stimuli that were distinct from the 10 test trials.

In the passive listening training phase (Step 1), infants and their parents were invited tp a sound chamber. There was a TV screen and a sofa in the room, and the speakers were hidden behind curtains. Infants were given soft toys, a way to keep them from boredom. Parents were instructed to keep silent. During the presentation of the sentences, infants could see an animation on the screen, with bright multi-colored circles slowly changing sizes.

After the passive listening phase, parents and infants left the toys behind and moved to another acoustic chamber for Step 2-5 of the experiment, which were executed by an experimental program (Cohen, Atkinson, & Chaput, 2000). The infant sat on the parent’s lap. The parent wore headphones to hear masking music. She or he was asked to not interfere with infant’s reactions. The experimenter, who was blind to the audio-visual stimuli, observed the infant’s eye movement from a closed-circuit TV in an adjacent room, and pressed down a computer key whenever the infant looked at the screen. The experimental software presented the stimuli and automatically recorded all looking times. Each trial in Steps 2-5 were initiated by the infant’s look to the screen. In Step 2 infants were presented with the test sentences once in two trials with a fixed length of 21 sec.

Half of the infants were trained with Rule 1 (Step 1), and the other half with Rule 2. In the test phase (Step 4), all the infants were presented with two Rule 1 and two Rule 2 novel sentences, the same stimuli as in Step 2, in separate trials. The test trials were characterized by two trial types. In one type the two sentences confirmed to the trained rule, whereas in another type the other two sentences conformed to the non-trained rule. These two trial types were presented alternatingly for 10 test trials in total (5 for each type). As part of the counter-balancing, half of the infants were presented with novel test sentences going through the trained rule as the first test trial, and the other half heard sentences going through the non-trained rule as the first test trial.

Within all trials auditory stimuli were presented simultaneously with visual stimuli of moving circles on the screen. In between trials, an attention getter was presented, an animation of blue bubbles accompanied by cricket sound.

2.2. Results

Each infant's looking times during the two trial types in the test phase (Step 4) were calculated, i.e., the trials presenting sentences conforming to the trained rule and those presenting sentences conforming to the second rule that infants did not hear during training. A paired sample T-test revealed that infants showed a significant discrimination of these two rules. In particular, they looked longer while listening to sentences complying with the non-trained movement rule than to sentences complying with the trained rule, $t(15) = 2.44$, $p < 0.05$, 2-tailed.

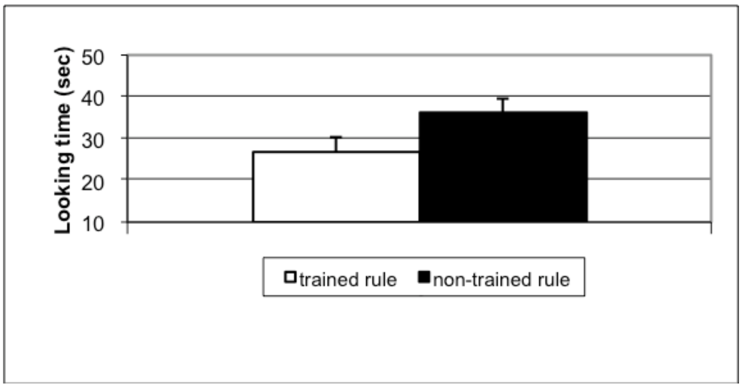

Fig. 1. Looking time results during the test phase of Experiment 1. After the training phase of 100% rule-consistent input, infants' looking time while listening to novel instances conforming to the non-trained rule was significantly longer than their looking time while listening to novel instances conforming to the trained rule.

The results suggest that after a brief exposure to an unfamiliar natural language, 14–month-olds can learn movement rules and generalize the rules to novel instances, in the absence of any phonological, morphological and semantic cues.

3. Experiment 2

In Experiment 2 we inquired how type frequency of rule exemplars and noise cases in the input can affect infants' capacities to extract syntactic regularities and apply them to novel instances. In order to test the role of type frequency, we dissociated type and token frequencies of rule and noise instances in the training sample. High type frequency of 80% was assigned to rule instances, whereas the exceptions were presented with a low type frequency of 20%. Token frequency per type was arranged in the opposite direction: 20% for rule instances, and 80% for noise. The overall frequency of both rule instances and rule-evading noise instances was held equal, at 50% level, a control to assure that it could not affect learning.

If dominant type frequency is the determining factor for syntactic rule learning, infants should succeed in generalizing the trained rule to novel instances despite the fact that token frequency for rule exemplars was not dominant. But if token-per-type frequency is determinant, infants should fail to learn the rule.

3.1.1. Participants and Materials

Participants were 16 infants aged 14 months from various linguistic backgrounds, with no prior exposure to Russian. Materials were 12 new Russian sentences recorded by the same Russian native speaker in the same way as in Experiment 1. Ten of these sentences were used as training stimuli, and two as novel instances in the test phase. They had a Subject-Verb-Object structure, with consistent morphological markings and a distinct phonotactic profile for each part of speech. All words contained two syllables. Within a sentence pair the original and the moved version were separated by 700 msec. The pause between “rule” and “noise” types, between any pairs, and between any two noise sentences was 1200 msec.

Out of the ten training exemplars, eight sentences presented a word order movement rule. For example, an SVO sentence was followed immediately by the same sentence transformed into VSO (for one of the training conditions: Rule 1) or into SOV (for the other training condition: Rule 2). Two “noise” sentences went through no movement and appeared in both training conditions. Infants were randomly assigned to one of the two conditions.

For the training stimuli, the eight alternating sentence pairs (either Rule 1 or Rule 2) were presented four times each, while being intermixed randomly with two “noise” sentences presented sixteen times each. This way, the total number of alternating sentence pairs and that of “noise” sentences, i.e., overall frequency, was kept equal (8x4=32 rule instances, and 2x16= 32 noise instances). That is, the overall frequency for both rule instances and exceptions was at 50%. The type frequency was 80% for rule instances and 20% for exceptions. On the other hand, the token-per-type frequency was 20% (1:4) for rule instances but 80% (1:16) for exceptions. This design enabled us to tease apart the factors of type frequency and token-per-type frequency while controlling for overall frequency. The total duration of the training phase was 340.26 sec for the Rule 1 training condition and 339.67 sec for the Rule 2 training condition.

The test stimuli were two novel SVO sentences and their moved version (SVO-VSO and SVO-SOV).

3.1.2. Design and Procedure

Design and procedure were nearly identical to the ones in Experiment 1, except that Step 2 and Step differed in maximum trial length. At Step 2 infants heard the two novel test sentences and their moved version once, whereas at Step 4 the same stimuli within a trial would be repeated up to three times if the infant looked till the end of the trial. In one trial one of the two novel test sentences went through the SVO-VSO rule. In the other trial the second sentence went through the SVO-SOV rule. The sentence and rule application were counter-balanced across infants, e.g., one group of infants heard one sentence being applied to SVO-VSO while another group heard the same sentence in SVO-SOV transformation. The trial duration in Step 2 was 6 sec. The order of the first test trial was also counter-balanced, such that half of the infants heard Rule 1 as the first test trial whereas the other half heard Rule 2 as the first trial. Sentences in the test phase were separated by the same inter-stimulus intervals, as in the training phase.

The maximum trial length of the test phase (Step 4) was 20 sec. As in Experiment 1, the test phase was characterized by two test trial types, one for the trained rule, and the other for the non-trained rule. If type frequency was the determining factor for rule learning and generalization, we should obtain significant looking time difference for the two trial types in the test phase. This prediction was based on the fact that only type frequency was dominant for rules (80%) in our design, with token-per-type frequency and overall frequency both non-dominant for the rule instances (20% and 50% respectively).

3.2. Results

As in Experiment 1, each infant's looking times during the two test trial types were calculated, i.e., the sentence conforming to the trained rule versus that conforming to the other rule that infants never heard before. A paired sample T-test revealed that infants looked longer while listening to the trial type presenting the non-trained movement rule, than to the trial type presenting the trained rule $t(15) = 2.65$, $p < 0.05$, 2-tailed.

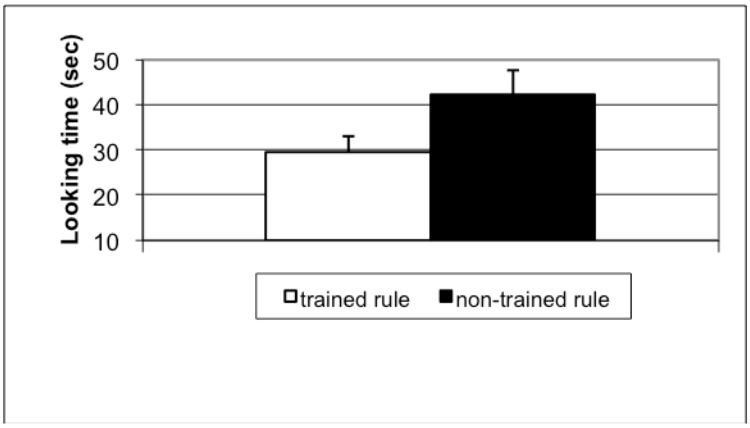

Fig. 2. Looking time results during the test phase of Experiment 2. Training Phase: input consisted of rule-consistent instances at 80% type frequency and 20% token-per-type frequency, and noise instances at 20% type frequency and 80% token-per-type frequency. Test phase: infants' looking time while listening to novel instances conforming to the non-trained rule was significantly longer than their looking time while listening to novel instances conforming to the type-dominant rule in the training input.

The results suggest that after a brief exposure to an unfamiliar natural language, 14–month-olds can learn movement rules and generalize the rules to novel instances. Crucially, our results show that the type frequency of rule instances, but not their overall frequency, was important for rule learning.

4. Experiment 3

Experiment 3 concerns a different question, one that is related to the issue of rule generalization to novel instances. Our idea is that when a rule is learned despite a certain degree of noise, infants also learn some aspects of the noise instances. We hypothesized that high token-per-type frequency of noise instances leads to learning exceptions, i.e. cases for which the rule should not be applied (that is, resisting overgeneralization). Our reasoning was that if indeed those "noise" instances resist overgeneralizations, it would be because they were better consolidated in memory due to a greater number of repetitions per type.

Logically, learners could derive two possible interpretations when encountering the kind of noise in Experiment 2, i.e., sentences that did not go through any movement:

-- True exceptions: those particular instances were not subject to the rule;

-- False exceptions: those instances have not had a chance to apply the rule, but they should be able to.

Experiment 2 showed that high type frequency of rule exemplars relative to "noise" can lead to infants' learning of syntactic regularities and their generalization to novel instances. In Experiment 3 we asked whether under the condition of exposure identical to that of Experiment 2 infants could also learn exceptions (the noise) and resist applying the learned rule to them.

4.1. Participants, Materials, Design and Procedure

Sixteen 14-month-old infants from various linguistic backgrounds, with no prior exposure to Russian completed this experiment. Materials were identical to the training materials of Experiment 2. The only difference between Experiments 2 and 3 was that in the test phase of Experiment 3, the movement rules were applied to the "noise" sentences from the training phase, but not to any novel sentences. In one test trial type the trained rule was applied to one noise sentence. In the other trial type the non-trained rule was applied to the second noise sentence. All other aspects of the design and procedure were identical to those in Experiment 2.

4.2. Results

looking times to the trained rule versus non-trained rule during the test trials were calculated for each infant, i.e., trials presenting a noise sentence being applied to the trained (and learned) rule versus those presenting a noise sentence being applied to the non-trained rule. A paired sample T-test showed no difference in looking times to the two trial types, $t(15) = 0.19$, $p = 0.85$, 2-tailed.

Fig. 3. Looking time results during the test phase of Experiment 3. Training phase: same as that of Experiment 2; input consisted of rule-consistent instances at 80% type frequency and 20% token-per-type frequency, and noise instances at 20% type frequency and 80% token-per-type frequency. Test phase: the specific noise instances from the training input were presented with the trained rule versus non-trained rule. Infants showed no looking difference towards the two rules.

These results show that under the conditions favorable for rule generalization to novel instances (as shown in Experiment 2), infants resisted applying the acquired rule to noise instances that were high in token frequencies. This suggests that such high token-per-type frequency determines the learning of exceptions. This may be because high token-per-type frequency of exceptions leads to memory consolidation of exceptions, inhibiting overgeneralization of such cases to the learned rule.

5. Discussion

Experiment 1 demonstrates that 14-month-olds can learn syntactic movement rules of a natural language and generalize them to novel instances. Infants were able to do so even in the absence of any phonotactic, morphological and semantic cues. These results are consistent with those of Marcus et al. (1999) and Gerken (2006), who showed with an artificial language that infants can generalize rules to novel instances. Our results show that infants can perform rule abstraction and generalization with complex speech stimuli from an unknown natural language.

The main goal of our study was to understand the distributional nature of the input that derives rule learning and abstraction. Experiment 1 showed that learning was successful when infants heard exemplars that were 100% consistent with a rule. The training phase of Experiments 2, on the other hand, contained rule instances mixed with non-alternating noise sentences. We inquired if this kind of noise sentences can potentially impede learning. In the training sample overall frequency was controlled by being kept equal, 50% for both rule and noise sentences. Type and token-per-type frequencies of rule and noise instances were dissociated: there were more variable rule sentences (80% type frequency) with each occurring few times (20% token-per-type frequency), whereas noise sentences were very scarce (20% type frequency) with each occurring highly frequently (80% token-per-type frequency).

We found that infants could learn rules from this input and generalize them to novel instances (Experiment 2). This suggests that learning was driven by high type frequency of the rule instances.

The purpose of Experiment 3 was to study the factor underlying infants' learning of exceptions to a rule. Infants heard the same training input as in Experiment 2, but were tested on whether they would apply the trained rule to the noise sentences. Those sentences appeared as non-alternating instances (i.e., rule-evading noise cases) during training. In principle, the non-alternating noise sentences could either be interpreted as true exceptions, or, on the contrary, as false exceptions (i.e., sentences that just had not had a chance to be heard with the rule but could logically be). Infants did not show evidence of applying the rule to the noise sentence during the test phase. We interpret this result as suggesting that they treated those "noise" instances as true exceptions and resisted overgeneralizing them to the dominant movement rule in the training input.

Taken together, the results confirm our hypotheses that high type frequency of rule instances, relatively to noise type, leads to rule abstraction, whereas high token frequency of noise instances leads to exception learning. Further controlled experiments are needed to fully assess the potential contribution of token-per-type frequency and overall frequency for rule abstraction. The results of Experiment 3 motivate the need for examining the conditions under which noise instances are not learned as exceptions, but rather treated as being eligible for rule applications, i.e., overgeneralization. We suggest that when token-per-type frequency is low for noise instances, they tend to be overgeneralized to the learned rule. This idea is coherent with our findings in Experiment 3, i.e., high token-per-type frequency favoring exception learning.

Unlike the previous studies (Gomez & Lakusta, 2004; Wonnacott, 2008), the nature of the noise in our training stimuli was a different kind – we used non-alternating noise sentences, instead of a movement rule that differed from the dominant rule in the training input. Both kinds of noise in fact exist in natural linguistic environment. Our particular design, i.e., rule-evading noise instances, had the advantage of allowing us to examine the circumstances under which infants would treat "noise" as true or false exceptions, a logically more interesting question about learnability. Future research should examine whether the mechanism which we showed here for exception learning applies to the other, overt kind of noise instances. This is necessary given that overgeneralizations are observed for both kinds of noise instances in language acquisition.

Acknowledgment

This research was funded by grants from NSERC, SSHRC AND CFI to the second author. We thank all parents and infants who participated in this study.

References

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

Gerken, L. (2006). Decisions, decisions: infant language learning when multiple generalizations are possible. *Cognition*, *98*, B67-B74.

Gómez, R. L., & LaKusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science, 7*(5), 567–580.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*, 77–80.

Wonnacott, E., Newport, E.L., & Tanenhaus, M.K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, *56*, 165-209.