Of mouses and mans: the role of production and feedback in language learning

Megan Waller (meganwal@andrew.cmu.edu)
Department of Psychology, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213

Nazbanou Nozari (bnozari@andrew.cmu.edu)
Department of Psychology, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213

Daniel Yurovsky (yurovsky@cmu.edu)
Department of Psychology, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213

Abstract

Do children learn language from the words that they produce themselves? Because children know that they have imperfect knowledge of language, they could simply ignore their own productions. However, children could also learn from their productions – using what they say and how their caregivers respond to update their language models. Using irregular plurals as a case study, we conducted a large-scale corpus analysis and two experimental studies to understand the role of children's productions and caregivers' responses in language learning. We demonstrate that children do learn from their own production, with errorful utterances leading to more errors. However, at least in some contexts, children can use implicit corrections from parents to offset the negative effects of their errors. Children thus appear to learn not only from their caregivers' productions, but also from their own productions and from the relationship between the two.

Keywords: development; language acquisition; language production; corpus studies

Introduction

How do children learn their native language? For decades, research in cognitive science has conceptualized this question as a problem of statistical inference. Children's caregivers provide sample utterances of their native language, and from these children must infer the underlying system of linguistic rules that explain why they heard these utterances and not others (e.g. Pinker, 1979). Over time, this framework has led researchers to ask questions like (1) what information is available to children in these utterances? (Gleitman, 1990), (2) what are the constraints that children impose on their inferences? (Markman & Wachtel, 1988), and (3) what machinery do children use to perform these inferences? (Smith & Yu, 2008).

But are the utterances that children hear from caregivers and peers the only input to language learning? Could children's *own* utterances also play a role? Because children know that they are not yet language experts, they could ignore their own productions. Alternatively, children might learn from their own utterances in two different ways. One possibility is that children may treat these utterances the same way they treat those produced by their caregivers. When children produce correct utterances of their native language, these utterances may may act as additional instances from which the rules of language can be inferred. Conversely, when children produce incorrect utterances, these errors could impede learning because they are learned in the same way as correct utterances (Middleton & Schwartz, 2012). Another possibility is that if children learn language using error-based mechanisms, then their own errors could be an especially powerful and beneficial vehicle for learning (Metcalfe, 2017; Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010). In error-based learning accounts, the amount of learning is proportional to the discrepancy between the learner's prediction and the observed outcome. If parents provide a corrective signal when children make errors, error-based learning accounts would predict the most effective learning happens in these moments.

To understand the role that children's utterances play in their learning, and to distinguish between errorless and error-based accounts of language acquisition, we focus on a case study of learning plural morphology. In English, plural nouns are generally marked by the addition of an '-s' to their base form – for example, 'cat' becomes 'cats.' However, some words are exceptions to this rule: 'mouse' becomes 'mice.' When learning irregular nouns, children sometimes go through a phase of overgeneralizing grammatical rules, thus producing 'mouses.' These overregularizations have been an important target of explanation for theories of language learning because children do not tend to hear the word 'mouses' in their caregivers' productions (Marcus et al., 1992; McClelland, Rumelhart, & PDP Research Group, 1987).

How do children learn to stop saying 'mouses?' An influential set of accounts have argued that the key lies not in the linguistic input that children hear, but in the relationship between their own productions and their parents' responses. Although parents rarely explicitly correct their children's errors (e.g. "no, you say mice"), they do provide implicit negative evidence in the form of reformulations – replaying the child's utterance, but changing the error to the correct form ("yes, those are mice!" Chouinard & Clark, 2003; Hirsh-Pasek, Treiman, & Schneiderman, 1984). While this correction is less direct than explicit feedback, reformulations provide a direct contrast to the child's error, supporting error-based learning.

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

Reformulations as an error-signal have received empirical support in both corpus and empirical studies. For instance, in a seminal study, Chouinard & Clark (2003) showed that parents are more likely to reformulate children's errors than to repeat correct utterances. In addition, children have been shown to repeat their parents' utterances more after reformulations than non-corrective repetitions (Farrar, 1992). Finally, grammatical forms of irregular words are more common in children's speech after receiving reformulations than other non-corrective responses (Nelson, Carskaddon, & Bonvillian, 1973; Saxton, 2000). The key experimental support comes from a set of studies by Saxton, showing that while learning novel irregular conjugations (e.g. 'pell'–'pold'), children were much more likely to learn the target form after receiving reformulations than when the experimenter provided the form first (Saxton, 1997; Saxton, Kulcsar, Marshall, & Rupra, 1998).

However, these corpus studies and the key experiments possess critical limitations that prevent us from drawing strong inferences about the available language input or the learning mechanisms that children apply to it. Although corpus evidence has been provided by a number of different research groups, all of the extant studies investigate a very small sample of children. Chouinard & Clark (2003), for example, describes learning in only 5 children. The experiment results, although large in effect size, investigate contexts that are quite different from those in which children typically receive reformulations. In particular, they focus explicitly on word learning – children knew that they did not know the correct form and their goal was to learn it (Saxton, 1997; Saxton et al., 1998). When children are not explicitly aware of the goal of word learning, especially when they are already familiar with the nouns, implicit corrections may not provide the same corrective benefit.

In a corpus study and two experiments, we investigated the contribution of children's productions and conversational partners' reformulations to learning irregular plurals like 'mice.' Using a larger sample than previous research, we assessed whether parents replays and reformulations can be a significant learning signal for the child. In a set of experiments, we evaluated if the benefits of reformulations found in prior work generalizes to a more naturalistic learning game. These studies together will provide valuable insight on how children's own production experience and parents' responses shape word learning.

Study 1

Building on a paradigm developed by Chouinard & Clark (2003), we investigated the frequency of reformulations in parent's speech. Reformulation rates estimated in prior work vary widely, from 10% to 60% of errors (Chouinard & Clark, 2003; Hirsh-Pasek et al., 1984). To get a more precise estimate, we investigated all children available in the North-American and British English corpora of CHILDES (MacWhinney, 2000). We also measured children's uptake of corrections provided in reformulations. Together, these analyses provide a clearer picture of how helpful implicit negative evidence in the form of reformulations could be to an ideal learner. We focused specifically on the production of commonly encountered irregular plural nouns like 'mice,' as they will be the stimuli used in our experimental Studies 2 and 3.

Method

From within the CHILDES database, we sampled from both the English-United Kingdom and the English-North American corpora any instance in which a child said a plural form of our target irregular plurals (mice, geese, snowmen, men, women, children, teeth, feet), along with the two utterances that followed. We searched for the correct form and two common forms of overregularization errors: adding an '-s' on either the singluar form ('mouses'), or the plural form ('mices'). While this may not capture every time the child attempts a plural form in this corpora, this sample likely covers the most common cases.

We selected any interactions in which anyone other than the child responded to the initial utterance. This sample contained a total of 2394 utterances across 223 children. We then detected whether the parent repeated the form the child used (replay), or switched to a different form (reformulation). We evaluated how often parents replay or reformulate in their responses, and whether these rates differ depending on whether or not the child made a mistake. We then examined whether children replied with the correct form after receiving a reformulation to evaluate the effectiveness of this error signal.

Results & Discussion

When children produced the correct irregular plural (e.g. 'mice'), parents did often repeat their correct productions. However, when children produced an error like 'mouses,' parents were *also* likely to replay their errorful overregularizations (Figure 1A). Parents were reliably more likely to replay correct productions (34.1%) than incorrect ones (16.1%). However, because the difference in base rates was so large – children were overwhelmingly more likely to produce a correct irregular plural – replays are a weak signal even for an ideal learner.

In contrast, reformulations were a clear signal about the correctness of the utterance – parents only provided reformulations when children made an error. We found the prevalence of reformulations – 27.3% – low relative to some other estimates (e.g. Chouinard & Clark, 2003). Nonetheless, because parents never reformulated children's correct productions, this prevalence is sufficient to be a powerful learning signal for an error-based learner. Did children take up these reformulations? When we looked at children's responses to parents' reformulations, we did not find much evidence of uptake. Children were actually twice as likely to repeat their incorrect overregularizations as they were to take up parents' correct reformulations (Figure 1B).

This corpus analysis thus provides support for the hypothesis that the relationship between children's productions and

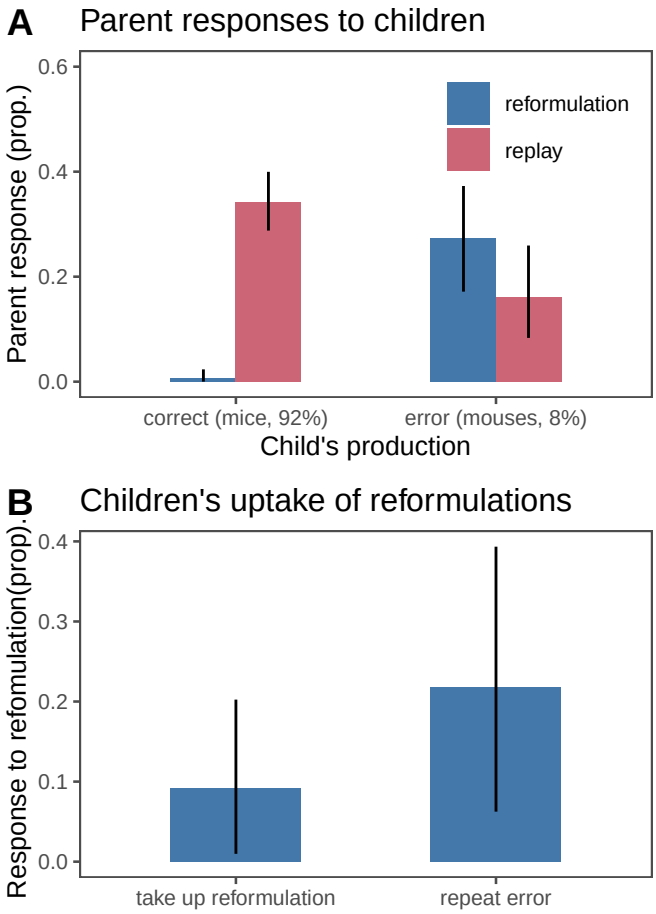

Figure 1: A) Parents' rates of replay and reformulation of children's irregular plural. B) Children's response to parents' reformulations. Error bars indicate 95% confidence intervals estimated by non-parametric bootstrapping.

parents' responses could be a source of information for children learning language. Although parents frequently repeat both correct and overregularized plurals, they do reformulate a significant fraction of errors, and these reformulations could provide an important learning signal for error-based learners. However, at least in their immediate responses, children did not appear to take up parents' reformulations.

In prior experimental work, reformulations rapidly accelerated children's learning of novel irregulars. However, these experiments explicitly tell children that their goal is to learn new words. Given that reformulations are not an explicit correction, the emphasis on word learning may make these reformulations more effective than they are in children's real-life interactions. In Study 2, we developed an experimental paradigm that makes reformulations incidental to the primary goal, more akin to the interactions in our corpus analysis where children are talking to parents with familiar words. We then asked, experimentally, whether reformulations support language learning.

Study 2

In contrast to our corpus analysis, prior experimental work shows strong evidence for children's learning from reformulations. For example, Saxton (1997) taught children novel irregular words either through repetition (in which the experimenter would produce the word and then prompt the child to repeat it), or through reformulation (in which children would attempt the word first, and the experimenter would reformulate their errors). In this study, children learned much better from reformulation than from repetition. In Study 2, we adapted Saxton's paradigm to be more analogous to the real-world interactions in which word learning is backgrounded relative to proximal goals. Children were asked to produce irregular plurals in the context of a reference game with familiar words, and their propensity to make errors was measured. Across conditions, children either repeated the experimenter's correct irregular words, or they were asked to make the first reference and their errors were reformulated. If children learn through error-based learning mechanisms, and reformulations are a strong error signal that children can interpret, then children who received reformulations should produce fewer overregularizations than children who repeated after the experimenter.

Method

Participants Sixty-six three and four-year-olds were recruited from a Psychology Department database and through word of mouth to participate. Families were given $10 Amazon gift cards for their participation. Five children did not complete the game, and one child was removed for experimenter error to yield our preregistered sample of 60 children (OSF registration: bit.ly/3ueVKTb). Another child was removed post-hoc due to parental interference. Our final sample consisted of 59 children (mean age = 4.00 years).

Procedure On Zoom, children were told they would play a counting game to help the experimenter fix their dream machine. Each session had four phases: a pre-test, two blocks of training trials, and a post-test (Figure 2). In the pre-test, children were instructed to greet each set of plural nouns by saying "hi geese!" to serve as an evaluation of the child's prior irregular word knowledge. If the child did not know the name or used the wrong word (e.g. saying "girl" instead of "woman"), the experimenter would provide the singular form of the noun and ask the child to try again. Half of the pictures were control regular nouns, the other half were target irregular nouns (mice, geese, leaves, teeth, feet, men, women, snowmen).

In blocks 1 and 2, children were told that the dream machine had trouble telling which of two dreams was correct, and the child could help the experimenter by counting and naming the objects that distinguished the experimenter's dream from the similar dream (Figure 2). All of the irregular plural nouns occurred once within each block. In each trial of the game, the child counted the pictures in the dream that blinked and had a line of bubbles leading to it. Children

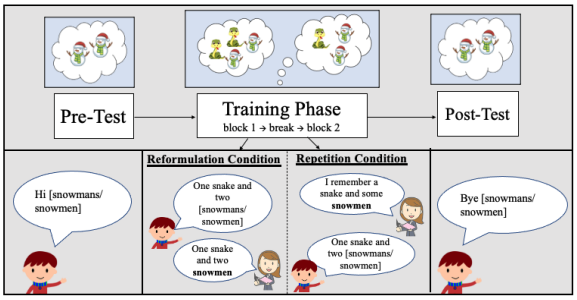

Figure 2: Diagram of experiment flow for Study 2.

responded with a compound phrase counting the two sets of pictures, such as "one snake and two snowmen." If the child named the wrong dream or counted incorrectly, the experimenter would remind the child to follow the bubbles and ask them to count again. There were two fixed orders of trials counterbalanced across children.

For children in the Reformulation condition, children counted first and then the experimenter responded with the correct noun forms. Therefore, if the child made a mistake such as saying "snowmans" instead of "snowmen", the experimenter's response of "snowmen" provided a reformulation to their error. If the child did not make a mistake, the experimenter replayed the correct response.

For children in the Repetition condition, the experimenter said each correct from before the child's turn (e.g. "I remember a snake and some snowmen). The experimenter did not give any feedback after children's productions. This condition controlled for exposure to the correct noun forms. While in both conditions children heard the correct form twice, the children only received negative evidence to their mistakes in the Reformulation condition. During the post-test, the children say goodbyes to each of the noun pairs again (e.g. "bye-bye geese"), to test for any change in children's irregular noun production after the intervention.

Data Preprocessing Zoom sessions were transcribed by hand in Microsoft Excel by two coders. The coders then marked whether the child provided the right number (numerical error), whether the child used the correct root noun to refer to the object (lexical error), and whether the child used the correct plural form (plural error). If the child did not provide any number or word during a trial, it was coded as a numerical or lexical error respectively.

The coders both individually coded six of the same videos to detect the reliability of their coding. Coder judgements were reliably similar across all 3 error types (numerical: Cohen's $\kappa = 0.92$, lexical: $\kappa = 0.84$, and plural: $\kappa = 0.85$). The two coders then met to reconcile any differences they had on those six participants.

After calculating error rates for individual children, we removed 8 children who made counting errors for over half the trials, as it was clear they did not understand the directions.

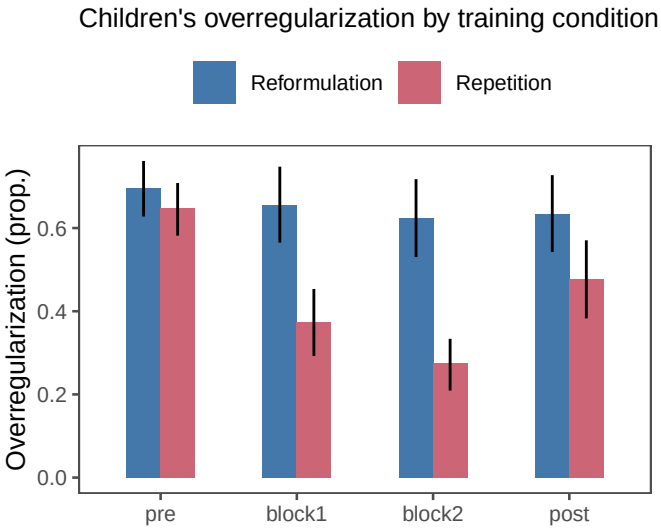

Figure 3: Overregularization error rates with 95% confidence interval from non-parametric bootstrapping.

Any trials in which the child used the wrong lexical word were removed from analyses. If the child counted in a way that changed the noun from singular to plural or from plural to singular, those trials were also removed (e.g. counting one mouse instead of two mice).

Results and Discussion

Children in both the Repetition and Reformulation condition overregularized the irregular nouns at similar rates at pre-test ($M = 65\%$, $\beta_{Repetition} = -0.34, p = .199$; mixed-effects logistic regression). In the Repetition condition, compared to their error rates at pre-test, children produced reliably fewer errors during the game and at post-test ($\beta_{block1} = -1.44$, $\beta_{block2} = -1.97$, $\beta_{post} = -0.88$, $ps < .001$). In contrast, children in the Reformulation condition did not produce fewer errors during the game or at post-test compared to pre-test ($\beta_{block}1 = -0.24$, $\beta_{block2} = -0.42$, $\beta_{post} = -0.38$, $ps > .1$). Thus, in contrast to Saxton (1997), children benefited from the Repetition condition, but not the Reformulation condition (Figure 3).

To further compare these two conditions, we fitted a mixed effects logistic regression, predicting whether or not the child made a grammatical error with fixed effects for section (pre- or post-test), and condition (Repetition or Reformulation), and random effects for both subject and target word. There was not a significant effect of the interaction between section and condition ($\beta = 0.46$, $p = .189$).

Although children in the repetition condition did show some improvement, there was no evidence that production improved in the reformulation condition, and certainly not to a greater extent than in the repetition condition. This finding is radically opposed to the previous report by Saxton (1997), who demonstrated much better learning after reformulations than repetitions. One could argue that the current results are better representative of incidental learning, which characterizes most learning situations from reformulation in everyday

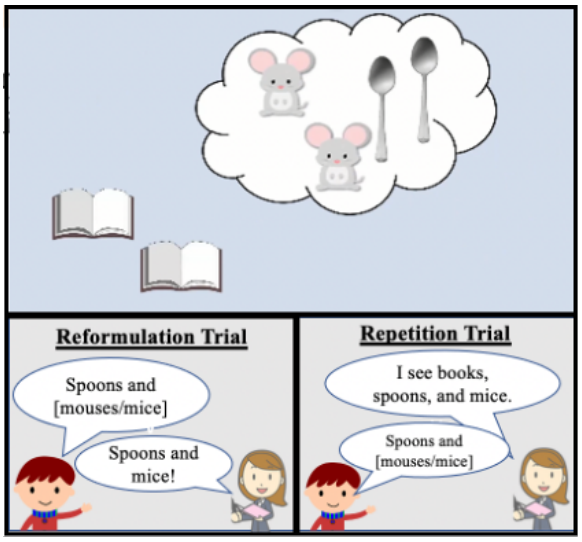

Figure 4: An example of Study 3 training phase for both the Repetition and Reformulation trials.

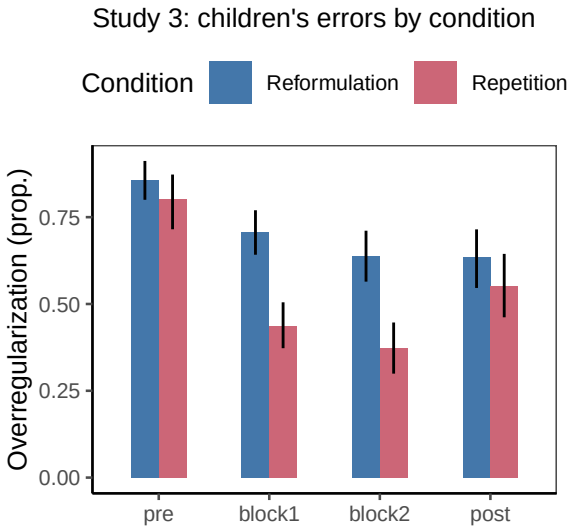

Figure 5: Study 3 overregularization error rates with 95% confidence interval from non-parametric bootstrapping.

life. However, before we conclude that reformulations are not a very effective way of vocabulary learning in children, we must rule out two confounds: (a) grammatical knowledge is quite variable in this age range. It is thus possible that the between-group differences in children's morphological knowledge were stronger than our experimental manipulation. (b) The counting task can be demanding in this age group. When made the central goal of the game, this demanding task may distract the child from processing other information not directly relevant to the task goal. Study 3 was designed to address these two issues. In a within-subject design, we removed the counting demand from the task in Study 2, and made the experimental items the focus of the communicative task.

Study 3

Method

Participants Sixty-one children of our preregistered sample of 100 3-and-4 year olds have been collected (OSF registration: bit.ly/3KYvAKe). One child did not complete the game. Families were recruited in the same manner as Study 2, and offered the same $10 gift card compensation.

Procedure As in Study 2, children were told they would help fix the experimenter's dream machine. Rather than immediately entering the pretest as in Study 2, children first labeled the images in singular form to establish the root word, and provide the experimenter the opportunity to correct the lexical item without using a plural. The game then proceeded to the the pre-test as before: children counted the pairs of pictures (e.g. "two mice"). The target plurals were all the same, except "leaves" was replaced with "children" because of the difficulty of distinguishing children's correct production of "leaves" from the overregularized "leafs".

During blocks 1 and 2, three pairs of pictures appeared outside the dream cloud, and then two of the three pairs moved into the cloud. The child was instructed to tell the experimenter which pictures went into the cloud, saying something like "spoons and [mouses/mice]" (Figure 4). Rather than a between-subjects design as in Study 2, we shifted to a within-subjects design in which children each participated in both conditions for different words. The assignment of words to Reformulation and Repetition conditions was counterbalanced across participants. If the word was assigned to Repetition trials, the experimenter labeled all the pairs first, saying "I see trees, dogs, and children, but which go into my dream?" and then the pictures moved and the child replied. If the word was assigned to the Reformulation condition, the experimenter prompted without labels (e.g. "what was in my dream?"), and then the child replied, and finally the experimenter provided the correct response. Then at post-test, children counted the pairs again to assess whether there was any difference in learning between the words they repeated compared to the words the experimenter reformed.

Data Preprocessing Zoom sessions were transcribed by hand in Microsoft Excel by a team of four coders. The coders then marked whether the child made a lexical or plural error during each trial. The coders separately coded a small set of videos in pairs to detect the reliability of their coding. Coder judgements were on average reliably similar across error types (lexical average Cohen's $\kappa = 0.91$ and plural: $\kappa = 0.89$).

Results and Discussion

Children again produced fewer errors after Repetition compared to pre-test ($\beta_{block1} = -2.32$, $\beta_{block2} = -2.54$, $\beta_{post} = -1.9$, $ps < .001$); however, this time after receiving Reformulations, children showed improvement during the second

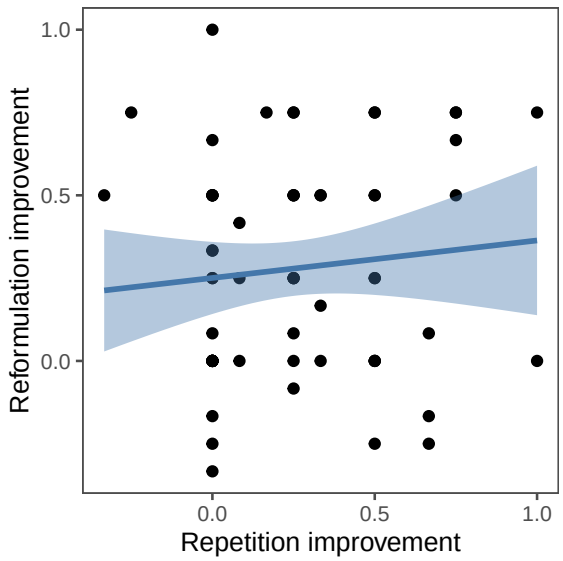

Figure 6: The relationship between improvement for Repetition words and Reformulation words. Points show individual participants. The line shows the line of best fit and a 95% confidence interval.

and post-test trials (β_{block2} = -1.26, β_{post} = -1.55, $ps < .001$). This result indicates that the focus on counting rather than labeling in Study 2 may have hindered learning, but in neither study were Reformulations more helpful than Repetition (c.f. Saxton (1997), Figure 5).

Study 3 was designed to make reformulations as salient as possible, while still maintaining a naturalistic, interactive context. While in Study 2 reformulations produced no learning benefit, in Study 3 children did improve after reformulations. The simplifications in the design intended to make reformulations more salient may have allowed children to learn from this feedback. Alternatively, the benefit for reformulated words could instead have come indirectly from the learning benefit of the Repetition trials. That is, the children could have learned not to overregularize the irregular plurals which they repeated after the experimenter, and generalized this knowledge to words that appeared in the Reformulation condition (a spillover effect). If so, we should expect to see a correlation between the amount that individual children benefited from Repetition and the amount they benefited from Reformulation. Figure 6 shows the relationship between the amount that children reduced overregularization in the two within-subjects conditions. Contrary to the spillover hypothesis, there was no reliable relationship between the two conditions ($r(\ 55\ ) = 0.06$, $p = .646$).

While reformulations were helpful in reducing overregularization in this study, reformulations still were not more helpful than repeating after an experimenter. Together with the results of Study 2, these data suggest that children benefit from receiving corrections, but that they pay a cost for making errors. Further, when corrections are harder to identify and process, as in Study 2, the cost paid for making an error can outweigh the benefit of corrections.

General Discussion

Using irregular plural morphology, we investigated the role of children's productions and parents' responses in children's language learning. In a large-scale corpus analysis, we confirmed prior research suggesting that parents' reformulations could provide an important corrective signal for language learning. However, in two subsequent experiments, we showed that the practical effect of reformulations on learning may be smaller than estimated in prior experimental work (Saxton, 1997, 2000). In Study 2, when reformulations were less salient for the child, children showed no improvement, indicating the feedback was not sufficient to correct their mistakes. On the other hand, children in the repetition condition improved, indicating that children's own production of the correct forms improved their learning. In Study 3, when the task demands were simpler, children were able to learn from the reformulations, but this learning was not more beneficial that repetition. Considering these results together, it seems that the positive impact of reformulations at best offsets the negative learning-impact of making errors in the first place. Together, these studies suggest that children can use error-based learning as part of a suite of many language-learning mechanisms, but perhaps not as their main tool.

Overall, children appear to learn not only from their caregivers' productions, but also from their own productions and from the interaction between these two inputs. Even though the experimenter provided the correct forms the same number of times, in Study 2, children's own production experience led to different learning outcomes. These results suggest that children's errors can have a negative impact on later production, which brings into question the generalizability of error-based accounts as a key mechanism for language learning. In our experimental paradigm, making errors and receiving feedback was not more beneficial than learning through repetition. However, these two experiments only test the short-term effects on learning familiar nouns, so future research should investigate the effects of production and feedback on the learning of novel nouns as well as long-term retention. In the short term, it appears that producing the correct word form, rather than making errors and receiving feedback, is reliably more helpful for learning. However, when feedback is clear, reformulations can counteract the negative impacts of producing errors.

Acknowledgements

This research was funded by James S. McDonnell Foundation Scholar Award in Understanding Human Cognition #220020506 to DY. We thank Angela Wei, Riya Chaudhry, Lily Madojemu, and Christina Choi for their work transcribing and coding videos for this project.

References

Chouinard, M. M., & Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child*

Language, *30*(3), 637–670.
Farrar, M. J. (1992). Negative evidence and grammatical morpheme acquisition. *Developmental Psychology*, *28*(1), 90.
Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.
Hirsh-Pasek, K., Treiman, R., & Schneiderman, M. (1984). Brown & hanlon revisited: Mothers' sensitivity to ungrammatical forms. *Journal of Child Language*, *11*(1), 81–88.
MacWhinney, B. (2000). *The childes project: Tools for analyzing talk. Transcription format and programs* (Vol. 1). Psychology Press.
Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, i–178.
Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157.
McClelland, J. L., Rumelhart, D. E., & PDP Research Group. (1987). *Parallel distributed processing, volume 2: Explorations in the microstructure of cognition: Psychological and biological models* (Vol. 2). MIT press.
Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology*, *68*, 465–489.
Middleton, E. L., & Schwartz, M. F. (2012). Errorless learning in cognitive rehabilitation: A critical review. *Neuropsychological Rehabilitation*, *22*(2), 138–168.
Nelson, K. E., Carskaddon, G., & Bonvillian, J. D. (1973). Syntax acquisition: Impact of experimental variation in adult verbal interaction with the child. *Child Development*, 497–504.
Pinker, S. (1979). Formal models of language learning. *Cognition*, *7*(3), 217–283.
Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, *34*(6), 909–957.
Saxton, M. (1997). The contrast theory of negative input. *Journal of Child Language*, *24*(1), 139–161.
Saxton, M. (2000). Negative evidence and negative feedback: Immediate effects on the grammaticality of child speech. *First Language*, *20*(60), 221–252.
Saxton, M., Kulcsar, B., Marshall, G., & Rupra, M. (1998). Longer-term effects of corrective input: An experimental approach. *Journal of Child Language*, *25*(3), 701–721.
Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.