Are conservative quantifiers easier to learn? Evidence from novel quantifier experiments *

Jennifer Spenader[1] and Jill de Villiers[2]

[1] University of Groningen, The Netherlands
j.k.spenader@rug.nl
[2] Smith College, Massachusetts, U.S.A.
jdevilli@smith.edu

Abstract

Natural language quantificational determiners seem to always be conservative. Some researchers suggest that one explanation for this link between syntactic form and semantic interpretation may be due to conservative quantifiers being easier to learn. Previous experimental research showed that children learned a conservative novel quantifier (i.e. *not all* more easily than a novel non-conservative quantifier (i.e. *not only*) [Hunter and Lidz, 2013].

In a series of four experiments we further investigated this learnability claim. Experiment 1 replicated the original study, but with adult participants. We found no evidence that the conservative novel quantifier was easier to learn. Experiment 2 replicated and extended the original study, testing children. Again, we found no difference between the quantifiers. Experiments 3 and 4 introduce a new training and testing method, situation verification with correction. In Experiment 3, we tested this new method with known quantifier meanings, conservative *all* and non-conservative *only*, using a novel quantifier expression. Children were highly successful, and there was no difference between the two quantifier meanings. In Experiment 4, we used the same method to teach children the novel conservative and non-conservative meanings *not all* and *not only*. Most children were unable to learn either quantifier, and we again found no difference between the conservative and non-conservative quantifier in terms of learnability. In conclusion, in four experiments, we failed to replicate findings that conservative quantifiers are easier to learn. These results suggest that a learnability advantage is not a plausible explanation for conservativity's universality.

1 Introduction

It is a linguistic universal that quantificational determiners like *all* and *some* are always conservative (Keenan and Stavi, 1986; Barwise and Cooper, 1981). This means interpreting (1) or (2) only requires examining the restrictor set (pirates) and the intersection between the restrictor and scopal set (pirates with treasure chests). Other individuals with treasure chests are irrelevant. Syntactically, quantificational determiners also form a syntactic constituent with their restrictor set. In contrast, non-conservative quantifiers, like *always* and *only* (3), are adverbs, and do not form a syntactic constituent with their restrictor. To verify (3), we must first identify relevant non-pirates and then check that none of them possess treasure chests.

(1) All pirates have treasure chests.

(2) Some pirates have treasure chests.

*Thanks to Anna de Koster and the members of the Language Acquisition Lab in Groningen for useful comments on the poster associated with this paper.

(3) Only pirates have treasure chests.

Verifying the non-conservative quantifier may be more complex because syntactic information does not help in identifying the restrictor set, and the interpretation requires looking for the complement of the mentioned set, rather than the explicitly mentioned set. Since languages include both conservative and non-conservative quantifiers, where does this syntactic restriction that quantificational determiners are always conservative come from? Its typological universality has led to speculation that conservativity makes quantificational determiners easier to learn.

Hunter and Lidz [2013] investigated this claim experimentally by teaching young children novel quantifier meanings. To create two comparable novel quantifiers, they used the negations of two known quantifier meanings, the novel conservative *not all* and the novel non-conservative quantifier *not only*. Using a picture presentation method for training and testing, children were taught that a novel quantifier expression, *gleeb*, had one of these meanings. They found better performance for the conservative quantifier than for the non-conservative quantifier, suggesting that conservative quantifiers are easier to learn.

These experimental results suggest that conservative quantifiers are easier to learn. However, using Recurrent Neural Network computational models and semantic features as input, Steinert-Threlkeld and Szymanik [2019] were unable to find a learnability advantage for conservative quantifiers. They point out that this is unsurprising; without syntactic information there was actually no difference between the conservative/non-conservative quantifiers. They note that their results are consistent with a proposal by Romoli [2015] that determiner conservativity is a consequence of the syntax-semantic interface. Simplified, if quantifier raising is obligatory, the way in which sentences are constructed naturally leads to non-conservative quantifier meanings that, regardless of definition, will either be contradictory, tautological, or completely equivalent to existing quantifier meanings. This work suggests that conservativity is not universal because of a conceptual difference in inherent complexity, and therefore learnability, but because the syntax-semantics interface works similarly across languages in a way that makes non-conservative determiner meanings trivial.

We focus on experimentally investigating the learnability issue in four experiments. In Experiment 1, we replicated Hunter and Lidz [2013] study with adults. In Experiment 2, we replicated and extended it with children. Based on these results, we conclude that the method used in these and the original experiment were unlikely to have taught quantifier meanings, which means the results cannot support the conclusion that conservative quantifiers are easier to learn. We then present a new training and testing method, situation verification with correction, which solves a number of problems with the picture-presentation method. In Experiment 3, we show that this new method succeeds in teaching known quantifier meanings with novel expressions with just ten training items. This experiment serves as a control for teaching children novel quantifiers. Experiment 4 then used the new method to teach novel quantifier meanings with novel expressions. In both Experiment 3 and Experiment 4 we found no difference between the quantifiers.

2 Previous experiments teaching conservative and non-conservative quantifiers

Novel word learning experiments have been long used as a method to show that adults and children are sensitive to certain syntactic features when learning novel words (Naigles and Hoff-Ginsberg, 1995, Fisher et al., 2010), so to show a learnability difference between conservative

and non-conservative quantifiers, a novel word task is an obvious choice. Hunter and Lidz [2013] choose *not all* and *not only* as their novel quantifier meanings. Neither meaning exists as a single word quantifier in English or Dutch,[1] and while negation often adds complexity to interpretation, since they both incorporate negation they are still comparable. The set-theoretic definition that they use for the interpretation of the novel quantifiers is presented below:

(4) a. conservative: gleeb (A,B) is true iff A $\not\subseteq$ B.
b. non-conservative: gleeb' (A,B) is true iff B $\not\subseteq$ A.

Hunter and Lidz [2013] found better performance for the conservative quantifier than for the non-conservative quantifier. After five training instances, the mean accuracy for the conservative determiner on five test items was 4.1, significantly better than chance (χ^2 =74.160, df=5, p<0.0001) while for the non-conservative determiner the mean accuracy was only 3.1, not significantly different from chance (χ^2 = 6.640, df=5, p>0.25). Even stronger evidence comes from their finding that half the children exposed to the conservative quantifier had perfect responses to the five test instances. For the non-conservative quantifier, only one child gave perfect responses.

The original study done by Hunter and Lidz [2013] aimed to show that children will be more easily able to learn a conservative quantifier, either because it is easier to verify or because it is conceptually easier. They chose to work with children between the ages of 4-5. At this age, children are still in the process of acquiring the full meaning of existing natural language quantifiers, but should be able to easily interpret the basic meaning of the non-negated counterparts *all* and *only*.

They used a method that teaches a contrast in an indirect way, the "picky puppet" paradigm. This method links a contrast to what a puppet likes and does not like. We outline the method here since we also use it in both our Experiment 1 and 2. Large cards, printed in color, displayed various scenes on a beach and a grassy area next to the beach, with boys and girls (See Figures 1 and 2). A target sentence is either a correct or incorrect description of the cards. The same target sentence was used for all items, (5). Children were introduced to a puppet who likes some cards but dislikes others. The puppet is shy, and only whispers their likes and dislikes to the experimenter. For each card the children were told whether the puppet likes the card or not, with an explicit statement that was framed around the target sentences, e.g. ((6-a) or (6-b)):

(5) Gleeb girls are on the beach.

(6) a. Puppet told me that she likes this card because gleeb girls are on the beach.
b. Puppet told me that she doesn't like this card because it is not the case that gleeb girls are on the beach.

Liked cards were placed in a pile under a green smiley face, and disliked cards were placed under a red smiley face. After this training, participants were presented with five new cards to sort.All training and testing cards were presented in the same fixed order.

With conservative *gleeb* meaning *not all*, sentence (5) presented with Figure 1 will be true, because some of the girls are on the grassy area, but the same sentences will be false with Figure 2, because all the girls are on the beach. With non-conservative *gleeb* meaning *not only*, sentence (5) will be false with Figure 1, because only girls are on the beach, but it will be true with Figure 2 because some of the boys are also on the beach.

[1] Or any other language, as far as we know

Figure 1:
Gleeb=*Not all*: True
Gleeb=*Not only*: False

Figure 2:
Gleeb=*Not all*: False
Gleeb=*Not only*: True

3 Experiment 1: Replication with Adults

18 English speaking adults participated in Experiment 1 (M_{age}=21;1, Range=18;6-26;4, 17 female). In a between subjects design, 9 were taught conservative *not all*, and 9 were taught non-conservative *not only*, using the same materials, method and order of presentation used in Hunter and Lidz [2013].[2]

3.1 Results and Discussion Experiment 1

Unlike the children in the experiment by Hunter and Lidz [2013], our adults were not very successful at learning novel quantifier meanings with this method, with a mean accuracy of 56% with the conservative quantifier and 69% with the non-conservative. A mixed-effects model analysis also found no significant difference between the quantifier types (p>0.11).

We did however find a difference between the quantifiers in terms of the number of subjects that showed perfect accuracy. 4 subjects in the non-conservative condition (taught *not only*) were perfect (and one had 80% correct), but only 1 participant in the conservative condition had a perfect score. In debriefing the participants that gave perfect responses to the *not only* condition stated that they had noticed that if there are boys are on the beach, the card was 'liked' by the puppet. They then simply sorted cards with boys on the beach as 'liked'. Following this strategy doesn't require any reference to the target sentences. Another explanation given by some participants in the *not only* condition was that *gleeb* meant *non-*. These participants then also followed a strategy of checking if *gleeb girls=non-girls* were on the beach, which is in essence the same strategy.

There is also a potential simple non-linguistic strategy for *not all*, which is to look for girls on the grassy area. However, no participants reported using such a strategy, perhaps in part because few participants did well with the conservative quantifier.

The strategies mentioned by successful participants in the *not only* condition highlights how correct responses can be achieved for this task without any reference to language or without interpreting *gleeb* as a quantifier. These strategies are in part possible because all items used the same target sentence. Additional target sentences that refer to another character, e.g. *Gleeb boys are on the beach* would make the first strategy impossible.

Our results do not support the conclusion that conservative quantifiers are easier to learn, and in fact seems to support the opposite conclusion. But the adults are still much worse than the children. It is certainly the case that sometimes children are able to show performance

[2]Participants were not permitted to examine already sorted cards to make their decision, though many asked if they could.

Figure 3: Experiment 2, Test item 11. This items was designed to distinguish a strategy of just looking for boys on the beach from a quantifier interpretation for *not only* items.

# of cards	Quantifier meaning	# sorted correctly
5	Conservative	3.0
	Non-conservative	3.4
10	Conservative	5.6
	Non-conservative	6.7
11	Conservative	6.0
	Non-conservative	7.2

Figure 4: Experiment 2 results: Number of cards correctly sorted for each quantifier, out of 45, 10 and 11 test items. No differences significantly better than chance (χ^2 tests.)

that adults are not able to achieve, for various other reasons (e.g. overthinking). Given the abstract nature of the task, which we replicated exactly the original study used with children, may have encouraged adults to look for a non-linguistic strategy. Remember, the task simply stated that the puppet liked or did not like certain situations because of a linguistically quite complex reason (e.g. see (6-a) and (6-b)). This makes the target sentence easy to ignore. Is it possible that the children were attending to the linguistic information but that adults instead focused on non-linguistic strategies because the training method de-emphasized meaning? To investigate this further we replicated and extended the original experiment by Hunter and Lidz [2013] in Experiment 2

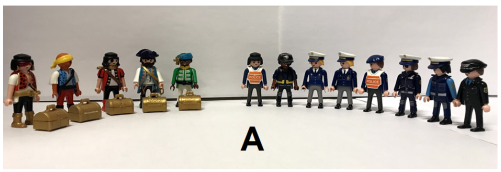
4 Experiment 2: Replication and Extension with Children

In Experiment 2 we had two goals, first to confirm the original results, and second, to make sure that responses given in the testing are reliable. The original study had 5 training items and 5 test items. In order to get a stronger confirmation of children's knowledge, we extended the testing part of the task to include 6 new test items. Each of these 5 test items was a minimal variation on the original 5 test items. The 6th new item differed by adding monkeys on the beach (and boys on the grassy area, see Figure 3). This picture should be true with *not only*, but if children followed the successful strategy of some of the adults in Experiment 1 and simply looked for boys on the beach, then they will incorrectly reject this item.

Using a between subjects design, 21 English-speaking children (M_{age}=5;0, 7 girls) participated. 10 children were taught *gleeb* with the novel meaning *not all* and 10 children were taught *gleeb* with the novel meaning *not only*. Experiment 2 duplicated the materials, procedure and order of presentation of Hunter and Lidz's original work and Experiment 1 for the training items and the first five testing items. Immediately following the five original test items, testing items 6-11 were presented, in a different random order for each participant.

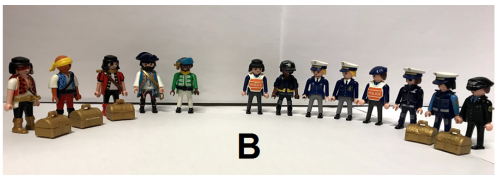
4.1 Results and Discussion Experiment 2

Children were equally bad at either quantifier. For the first 5 test items which were an exact replica of the Hunter and Lidz [2013] study, participants had an accuracy with conservative



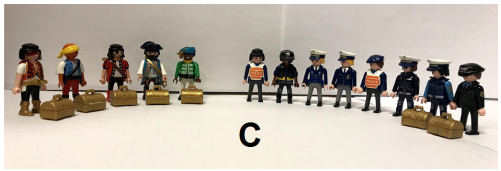
Flep = All	TRUE	No action
Flep = Only	TRUE	No action
Flep = Not all	FALSE	Remove from pirates
Flep = Not only	FALSE	Give to police

Figure 5: Type A Situation



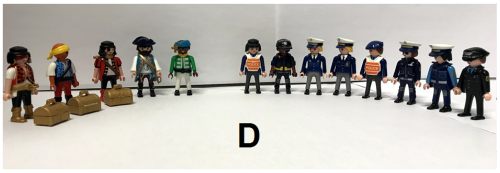
Flep = All	FALSE	Give to pirates
Flep = Only	FALSE	Remove from police
Flep = Not all	TRUE	No action
Flep = Not only	TRUE	No action

Figure 6: Type B Situation



Flep = All	TRUE	No action
Flep = Only	FALSE	Remove from police
Flep = Not all	FALSE	Remove from pirates
Flep = Not only	TRUE	No action

Figure 7: Type C Situation



Flep = All	FALSE	Give to pirates
Flep = Only	TRUE	No action
Flep = Not all	TRUE	No action
Flep = Not only	FALSE	Give to police

Figure 8: Type D Situation

gleeb of 60% and for non-conservative *gleeb* 68%, sorting 3.0 and 3.4 cards correctly out of five respectively.A mixed-effects model analysis found no significant difference between the quantifiers ($p>0.43$). The number of cards correctly sorted out of 10 and 11 are also given in Table 4. Also using the analysis from Hunter and Lidz [2013] comparing the cards sorted to chance with χ^2, there was no significant difference between the quantifiers.

Subjects were barely above chance in both Experiment 1 and 2, suggesting the picture presentation method fails to teach quantifier meanings. Instead, we developed an act-out alternative where true and false situations were presented for judgment to a puppet. Crucially, false descriptions were corrected by the puppet by manipulating objects, allowing children to immediately compare true and false situations.

5 Experiment 3: Known quantifier meanings using Situation Verification with Correction

To be sure that participants in testing actually understand quantifier meaning, we needed more information than simple picture verification can provide. For this reason we developed a novel training and testing paradigm: situation verification with correction.

In Experiment 3 we used this new method for training and testing with the known quantifiers *all* and *only*, presented as the nonce word *flep*[3] using five pirates and eight police officer figurines.

[3]We replaced *gleeb* with *flep*, as a more natural novel word for Dutch.

(7) a. Flep piraten hebben een schatkist. Klopt dat?
Flep pirates have a treasure chest. Is that true?
b. Hebben flep piraten een schatkist?
Have flep pirates a treasure chest? (*Do flep pirates have a treasure chest?*)

The training used a puppet and proceeded as follows. First, the experimenter distributed objects to the two different sets of figurines (*police and pirates*). Then the experimenter asked the puppet about how the situation related to the meaning of *flep*, using *flep* in a statement and in a question, e.g. (7-a), (7-b). The puppet judged the descriptive accuracy. If the description was incorrect, the experimenter asked the puppet to correct the situation by either adding or removing objects. When *flep* was used to mean *all*, the puppet gave an object to any mentioned characters that did not have one. When *flep* was used to mean *only* the puppet removed any objects that the non-mentioned characters had. There was always a surplus of objects for giving to figurines or for parking removed objects, as necessary, placed near the figurines for easy access.

16 Dutch native speaking children participated in Experiment 3. Seven children were taught *all* and nine children were taught *only*. Participants were first trained by being asked to observe the interaction between the experimenter and the puppet with 10 training items. Four different situations types were used, see Figures 3-6. Participants were presented with 3 Type A items, 3 Type B items, 2 Type C items and 2 Type D item. This resulted in 5 true items and 5 false items. After training, participants were presented with 10 test items, in the same distribution of types as the training items. Training and testing took approximately 10 minutes per child.

5.1 Results and Discussion Experiment 3

The children were highly successful in learning both *all* (98.5% correct) and *only* (96.6% correct), with most children giving perfect responses for all 10 test items. Many spontaneously used *flep* correctly even before the training was completed. There was no difference between the conservative and non-conservative quantifier.

These results showed that the new method, situation verification with correction, was successful in teaching known quantifier meanings with a novel quantifier expression with only 10 training items. It also showed that using 8 and 5 figurines was sufficient for encourage quantifier interpretations. Using this same method, we next attempted to teach children novel quantifier meanings with a novel quantifier expression in Experiment 4.

6 Experiment 4: Novel Quantifier Meaning using Situation Verification with Correction

In Experiment 4 we applied our new training method to the novel quantifiers *not all* and *not only* with the novel expression *flep*.

37 Dutch children (M_{age}=5;6, Range =5;0-6;6, 15 girls) participated in a between subjects design. 18 were taught *not all* and 19 were taught *not only* using the novel expression *flep* and 10 training and 10 test items. The method was the same as Experiment 3 except for the quantifier meaning.

When *flep* was used to mean *not all*, the puppet removed one object fromr the mentioned character. When *flep* was used to mean *not only* the puppet corrected false situations by giving objects to the non-mentioned character. The same four situation types were used as in Experiment 3. Participants were presented with 3 Type A items, 3 Type B items, 2 Type C

items and 2 Type D items for a total of 10 items each in training and testing. After training, children were asked to take over the role of the puppet, first verifying whether or not the situation presented was true or false, given the sentence and question, and then, if the sentence was false, children were asked to change the situation by manipulating the objects to make it true. Training and testing took approximately 10 minutes per child.

6.1 Results and Discussion Experiment 4

Few children were able to learn the novel quantifiers. The mean accuracy for *flep* with the novel quantifier meaning *not all* was 63% while the mean accuracy with the meaning *not only* was 54%. However, these means are not really informative. Examining individual accuracy, for children taught that *flep* meant *not all*, only five children out of 18 were successful, but these children actually had nearly perfect scores. For *flep* meaning *not only*, four children out of 19 were successful, again with nearly perfect scores. Similar to the results of Experiments 1 and 2, there was no difference between quantifiers.

Looking more closely at the children who made errors, a large subset of children picked up on the manipulation necessary for correcting false items, e.g. whether or not objects needed to be removed or added. But these children applied this action indiscriminately, showing they did not understand the intended meaning and also leading to low accuracy scores. The remaining children showed no identifiable pattern.

7 General Discussion

In four experiments, we failed to replicate findings that conservative quantifiers are easier to learn, instead finding no difference between the novel quantifiers and children and adult's success in the experiments. While our experiments were designed to test the learnability explanation directly, our results may also be relevant for evaluating other explanations for conservativity's universality.

If conservativity is instead a consequence of the way the syntax-semantic interface constructs meaning, then children's success in recognizing the non-conservative determiner meaning *only* distributionally presented as a determiner in Experiment 3 is surprising. These results show that children could comprehend and produce, even after only a handful of examples, a meaning with a syntactically dubious derivation. This suggests that at least for the ages and tasks tested, syntactic information was not necessary to learn the quantifier meaning. More research, is needed.

Why were children not able to learn the novel quantifier meanings *not all* and *not only* even with the new training method? This was unexpected given the ease with which the children were able to learn the known quantifier meanings in Experiment 3. One possible reason is that the negation makes the quantifier meaning much more complex. It is also well-known that children can have difficulties processing negation.

Our own impression is that the children who learned the quantifiers were children who were particularly good at the kind of puzzle the task represents. If this is the case, then the difficulty most children showed in the current study can be attributed to the conceptual novelty of the the quantifier meanings being taught. If we provide chidren with more training items, we might see a greater proportion of successful children. In that case, any difference found in the success rates between the two quantifiers would be more meaningful. We plan to investigate this further.

8 Conclusion

In four experiments we investigated the claim that natural language quantificational determiners are universally semantically conservative because conservativity makes them easier to learn. Contrary to previous experimental findings, we were unable to find a difference in ease of learnability between conservative and non-conservative quantifiers, both in experiments with adults and with children. We conclude that there is reason to believe that a learnability advantage is not a plausible explanation for conservativity's universality.

References

Jon Barwise and Robin Cooper. Generalized quantifiers and natural language. In *Philosophy, language, and artificial intelligence*, pages 241–301. Springer, 1981.

Cynthia Fisher, Yael Gertner, Rose M Scott, and Sylvia Yuan. Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(2):143–149, 2010.

Tim Hunter and Jeffrey Lidz. Conservativity and learnability of determiners. *Journal of Semantics*, 30(3):315–334, 2013.

Edward L Keenan and Jonathan Stavi. A semantic characterization of natural language determiners. *Linguistics and Philosophy*, 9(3):253–326, 1986.

Letitia R Naigles and Erika Hoff-Ginsberg. Input to verb learning: Evidence for the plausibility of syntactic bootstrapping. *Developmental Psychology*, 31(5):827, 1995.

Jacopo Romoli. A structural account of conservativity. *Semantics-Syntax Interface*, 2(1):28–57, 2015.

Shane Steinert-Threlkeld and Jakub Szymanik. Learnability and semantic universals. *Semantics and Pragmatics*, 12:4, 2019.