

# Restrictions on the Meaning of Determiners: Typological Generalisations and Learnability

Tim Hunter, Jeffrey Lidz, Alexis Wellwood and Anastasia Conroy  
*University of Maryland*

## 1. Introduction

It is standardly assumed that a close relationship holds between (i) the learnability of languages with some property  $P$ , and (ii) the existence of languages with property  $P$  as revealed by typological studies. If it is not possible for a human to acquire a language with  $P$ , then clearly no speakers will be found of any language with  $P$ . The inverse, though not logically necessary, is often also implicitly thought to be true: that if no natural language exists with  $P$ , then languages with  $P$  are unlearnable. This is not without reason, of course. Since the question of how children manage to acquire their native languages as quickly as they do is still largely unanswered, linguists are interested in discovering possible constraints on the learner's hypothesis space, for which typological generalisations would seem to make good candidates.<sup>1</sup> Assuming that we would like the formalisms used to describe natural language semantics to have the ability to express all and only the languages that human beings can naturally acquire, discoveries about constraints on the learner's hypothesis space in turn dictate which formalisms are too powerful and which are too weak.

In this paper we examine the relationship between learnability and typology in the area of determiner meanings. We begin with two generalisations about the meanings that determiners of the world's languages are found to have, and investigate the learnability of fictional determiners with unattested meanings. If participants in our experiments fail to learn such determiners, then this would suggest that they are unattested **because** they are unlearnable. If, on the other hand, participants are able to learn the determiners in question, then some other explanation for their absence in the languages of the world is necessary.

Specifically, the generalisations we consider are (i) that all natural language determiners are conservative, and (ii) that no natural language has a determiner analogous to *most* that expresses "less than half" rather than "more than half". The standard formal tools used to describe determiner meanings are powerful enough to express both kinds of unattested determiner meanings. This is undesirable if they are indeed unlearnable. We find that learnability correlates with (i), whereas it does not with (ii). Thus the lack of nonconservative determiners in the world's languages

---

Thanks to Bill Idsardi, Paul Pietroski and Alexander Williams for helpful advice; to Imogen Davidson White and Nathan Ycas for help organising and running many of our child subjects; and to the parents and children at the Center for Young Children at the University of Maryland.

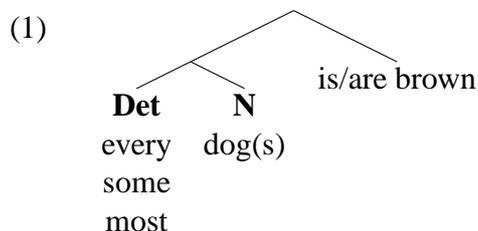
<sup>1</sup>See Chomsky (1965: p.55), Chomsky (1975: p.37), among many others.

can plausibly be put down to the fact that they are unlearnable, and we should strive for a formalisation of determiner meanings which precludes their occurrence. But the fact that no *less than half* determiner has been discovered can not be explained by the same logic, and so there is no reason to disprefer formalisms in which it can be expressed.

In Section 2 we briefly review the relevant theoretical background and present the two typological generalisations we will investigate, and then turn, in Section 3 and Section 4, to experiments testing the learnability of unattested determiners.

## 2. Formal Properties of Determiners

We take determiners to be those words that may occur in the syntactic frame illustrated in (1),<sup>2</sup> and assume that they express a relation between the two sets denoted by the determiner's two arguments (Mostowski 1957) as shown in (2).



- (2)
- $\llbracket \text{every} \rrbracket (X)(Y) \iff X \subseteq Y$
  - $\llbracket \text{some} \rrbracket (X)(Y) \iff X \cap Y \neq \emptyset$
  - $\llbracket \text{most} \rrbracket (X)(Y) \iff |X \cap Y| > |X - Y|$

When the determiners of the world's languages are analysed in this way — associating each with a binary relation on sets — a surprising generalisation emerges: every attested determiner expresses a relation that is conservative, as defined in (3) (Barwise and Cooper 1981, Higginbotham and May 1981, Keenan and Stavi 1986).<sup>3</sup>

- (3) A two-place relation on sets  $\mathcal{R}$  is **conservative** if and only if the following biconditional is true:  $\mathcal{R}(X)(Y) \iff \mathcal{R}(X)(X \cap Y)$

The English determiner *every* is conservative because *every dog is brown* has the same truth condition as *every dog is a brown dog*; or, more formally:

<sup>2</sup>We remain agnostic about many of the details of the syntax of these sentences, and for this reason limit our attention to quantifiers in subject positions. What is important for us is that “determiner” is defined distributionally.

<sup>3</sup>Two apparent counterexamples are *only* and *many*. Closer examination quickly shows that *only* is not a determiner, as defined distributionally. While at first *only dogs are brown* looks superficially like *some dogs are brown*, *only* can appear in many other positions where *some* and *every* cannot, eg. *dogs only/\*some/\*every are brown*, and *dogs are only/\*some/\*every brown*. The case of *many* is less clear, complicated by context-dependence, but can also plausibly be made to fit with the conservativity generalisation; see e.g. Keenan and Stavi (1986) and Herburger (1997).

$$(4) \quad \llbracket \text{every} \rrbracket(X)(Y) \iff X \subseteq Y \iff X \subseteq (X \cap Y) \iff \llbracket \text{every} \rrbracket(X)(X \cap Y)$$

It is a simple matter to define a determiner just like *every*, but with its arguments reversed — call it *grevery*, defined in (5), such that *grevery dog is brown* is equivalent to *every brown thing is a dog* — that is not conservative.

$$(5) \quad \llbracket \text{grevery} \rrbracket(X)(Y) \iff Y \subseteq X \iff \llbracket \text{every} \rrbracket(Y)(X)$$

Here, while it is the case that  $\llbracket \text{grevery} \rrbracket(X)(X \cap Y)$  is always true,  $\llbracket \text{grevery} \rrbracket(X)(Y)$  is not, so the biconditional in (3) does not hold. No attested determiner has this meaning.

A nonconservative determiner need not be the “reverse” of a conservative determiner. Another fictional nonconservative determiner that one can imagine is *equi*, defined in (6).<sup>4</sup>

$$(6) \quad \llbracket \text{equi} \rrbracket(X)(Y) \iff |X| = |Y|$$

Thus *equi dogs are brown* would mean that there are exactly as many dogs as there are brown things. Since this is certainly not equivalent to the statement that there are exactly as many dogs as there are brown dogs, *equi* is not conservative.

It should be clear that the standard formal tools used here are equally capable of representing both conservative and nonconservative determiners. Ideally, we would like determiners with all and only the meanings that the human language faculty allows to be statable in our semantic theory. If the absence of nonconservative determiners cross-linguistically is due to an inability of the language faculty to associate such meanings with sentences of the form in (1), a view we claim our experimental results support, then the standard toolkit “overgenerates” determiner meanings, and we should strive for an alternative which does not.<sup>5</sup> The *equi* example is a particularly striking instance of a relation that seems to be a “basic” or “simple” one that we would expect to find a determiner expressing, **if** the space of possible determiner meanings is precisely the space of binary relations on sets.

A second surprising absence from the world’s languages is a determiner — call it *fost* — that is analogous to *most* but expresses “less than (half)” rather than “more than (half)”. This is defined in (7); compare with *most* in (2).

$$(7) \quad \llbracket \text{fost} \rrbracket(X)(Y) \iff |X \cap Y| < |X - Y|$$

This determiner is conservative, since it depends on the same two sets as does *most*,<sup>6</sup> but is nonetheless unattested (Hackl 2009). This is rather striking when one considers that it is so minimally different from *most*, (an equivalent of) which, while not widespread, does appear in a number of natural languages (Hackl 2009).

<sup>4</sup>Thanks to Paul Pietroski (p.c.) for this example.

<sup>5</sup>Some alternatives that depart from the standard and eliminate this overgeneration have been proposed by Pietroski (2005), Bhatt and Pancheva (2007), and Fox (2002).

<sup>6</sup>Note that *fost* is not the “reverse” of *most* in the way that *grevery* is the reverse of *every*. That determiner, which one might call *grmost*, would be defined as:

$$\llbracket \text{grmost} \rrbracket(X)(Y) \iff |Y \cap X| > |Y - X| \iff \llbracket \text{most} \rrbracket(Y)(X)$$

This, in combination with the fact that *most* is the only attested determiner that is not expressible in first-order logic (Barwise and Cooper 1981), prompted Hackl (2009) to suggest that *most* may not be an atomic determiner, but rather a complex expression constructed from morphemes *more* and *-est*, in such a way that does not likewise permit *less* and *-est* to form a complex determiner with the meaning of *fost*. He explains the fact that *fost* is unattested by proposing a theory in which it is not expressible, while at the same time making it possible to restrict atomic determiner meanings to just those expressible in first-order logic — certainly two desirable consequences. The results of our experiments, however, indicate that *fost* is learnable, so Hackl’s theory appears to be too restrictive. An explanatorily adequate theory must be able to express the meaning of *fost*, even if no natural language makes use of this option.<sup>7</sup>

We present our experiments concerning the learnability of nonconservative determiners and the learnability of *fost* in Section 3 and Section 4, respectively.

### 3. Experiment 1: Conservativity

This experiment is also reported in Hunter and Conroy (2009).

---

<sup>7</sup>Of course, this leaves open the question of why it is that no natural language has a determiner with the meaning of *fost*, but we believe that its learnability is reason enough to prefer formalisms in which it is expressible. There are other conceivable approaches: if the absence of *fost* is for some pragmatic or functional reason, for example, one could choose to strive for a formalism that enforces the pragmatic or functional pressures which prevent *fost* from naturally occurring. But we follow convention in assuming that the goal of linguistic theory is to characterise possible states of the human mind, and our results reported below suggest that human minds can, albeit artificially, enter into a state where a word with the syntactic distribution of a determiner has the meaning of *fost*.

Horn (1972) argues for a lexicalisation constraint that rules out a class of quantifiers, possibly including those of the form *fost*  $X$ , on the basis of their pragmatic inferentiability from an extant paradigm, so that a language gets “only those lexical items it actually needs” (p. 251). His demonstration centers around the possibility of lexicalising quantifiers along with negation, observing that across domains (e.g. nominal and modal quantification, the binary connectives) there is lexicalisation of  $\neg\exists \equiv \forall\neg$  (e.g. *none*) but not of  $\exists\neg \equiv \neg\forall$ . This generalisation accounts for the fact that devices for expressing  $\neg\forall$  on a given scale come for free with any assertion containing that scale’s existential quantifier, via implicature. As is (by now) commonly assumed, an assertion of *Some girls called* implicates *Some girls didn’t call* or, equivalently, *Not all girls called*. Simplifying somewhat, Horn predicts that if it is true that  $\exists$  implicates  $\exists\neg$ , then  $\exists\neg \equiv \neg\forall$  will not lexicalise. This account correctly predicts the absence of the quantifier we label *nall* on pragmatic grounds, thus not precluding its learnability. It likewise predicts the absence of  $\neg\text{most}$ , but does not predict the absence of  $\text{most}\neg$ , although both are unattested in natural languages. Since  $\neg\llbracket\text{most}\rrbracket(X)(Y)$  and  $\llbracket\text{most}\rrbracket(X)(\neg Y)$  differ in truth value only if  $|Y \cap X| = |Y - X|$ , and we do not consider this case in our experiments on *fost*, nothing we have done determines whether by *fost* we mean  $\neg\text{most}$  or  $\text{most}\neg$ . So one can think of our experiment as testing whether **either**  $\neg\text{most}$  or  $\text{most}\neg$  is learnable; and it remains an open question whether the absence of the determiner we show to be learnable can be explained by Horn’s pragmatic restrictions.

### 3.1. Two novel determiners: *nall* and *grall*

The question we address is whether children permit structures like (1) to have non-conservative meanings. While it has been shown that children will sometimes accept non-adult-like interpretations of quantificational sentences (Inhelder and Piaget 1964), previous research is silent with respect to the specific question of conservativity.<sup>8</sup> To investigate this question, we attempted to teach children novel determiners. If children have no inherent restrictions on determiner meanings, then we would predict that they will be able to learn both novel conservative determiners and novel nonconservative determiners. However, if the typological generalisation that we observe reflects a restriction imposed by the language faculty, then we predict that children will succeed in learning novel conservative determiners, and will not succeed in learning novel nonconservative determiners.

To test these predictions we created two novel determiners, one conservative and one nonconservative. The conservative one, *nall*, is defined in (8).

$$(8) \quad \llbracket \text{nall} \rrbracket(X)(Y) \iff X \not\subseteq Y \iff \neg(X \subseteq Y)$$

So *nall girls are on the beach* is the negation of *every girl is on the beach*: we might paraphrase it as *not all girls are on the beach*. For example, it is true in the scene shown in Figure 1(a), but false in the scene shown in Figure 1(b). Since *nall* is the “negation” of the conservative determiner *every*, it is also conservative<sup>9</sup>.

The novel nonconservative determiner, *grall*, is defined in (9).

$$(9) \quad \llbracket \text{grall} \rrbracket(X)(Y) \iff Y \not\subseteq X \iff \neg(Y \subseteq X) \iff \llbracket \text{nall} \rrbracket(Y)(X)$$

So *grall girls are on the beach* is the “reverse” of *not all girls are on the beach*: it is true if and only if not all beach-goers are girls. For example, it is true in the scene shown in Figure 1(b), but false in the scene shown in Figure 1(a). Since the “lived on” set (the beach-goers) is not expressed as the **internal** argument of *grall*, *grall*

<sup>8</sup>The common finding of “symmetrical responses” in studies of universal quantification (Inhelder and Piaget 1964, Philip 1995, Drozd and van Loosbroek 1998) do not speak to a hypothesised conservativity constraint. This common finding is that some children judge a sentence like *every boy is riding an elephant* to be true if and only if (i) every boy is riding an elephant, and (ii) every elephant is ridden by a boy. On this reading, *every* is not expressing a two-place relation between the set of boys (the denotation of its internal argument) and the set of elephant riders (the denotation of its external argument, *is riding an elephant*); and if it is not expressing a two-place relation between these two sets, then it is certainly not expressing a nonconservative relation between them (like *equi* does). On the (questionable) assumption that *every* was indeed being analysed as a determiner in these sentences, this finding would be evidence against the standard restriction on determiner meanings that permits only relations between the sets denoted by their internal and external arguments — let alone the stronger restriction to only conservative relations — perhaps in favour of an account involving quantification over elephant-riding events as in Philip (1995) or over elephants as in Drozd and van Loosbroek (1998). But further experiments have suggested that these interpretations are a methodological artifact resulting from infelicitous contexts in any case (Crain et al. 1996).

<sup>9</sup>Suppose that  $\mathcal{R}$  is conservative, and that  $\mathcal{R}'(X)(Y) \equiv \neg\mathcal{R}(X)(Y)$ . Then

$$\mathcal{R}'(X)(Y) \iff \neg\mathcal{R}(X)(Y) \iff \neg\mathcal{R}(X)(X \cap Y) \iff \mathcal{R}'(X)(X \cap Y)$$

so  $\mathcal{R}'$  is also conservative.

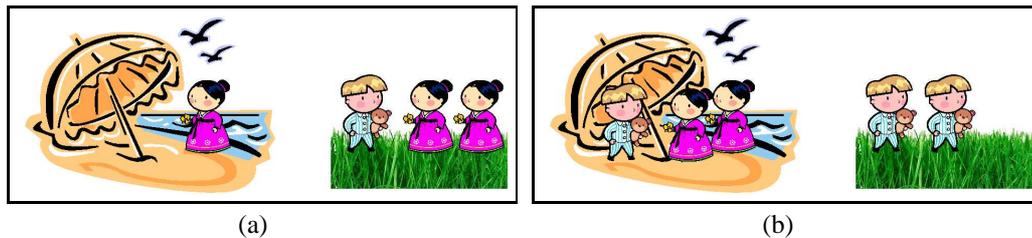


Figure 1: Two sample cards from Experiment 1. In the conservative condition, the puppet would like only the card in (a): *nall girls are on the beach* is true in (a), but false in (b). In the nonconservative condition, the puppet would like only the card in (b): *grall girls are on the beach* is false in (a), but true in (b).

is **not** conservative.<sup>10</sup>

Our experiment compares children’s ability to learn *nall* with their ability to learn *grall*, based on equivalent input. Note that since the conditions expressed by these two determiners are just the “mirror image” of each other, there is no reason to expect a difference in how easily they can be learnt — unless there are constraints on the semantic significance of being the internal or external argument of a determiner, since this is all that distinguishes *nall* from *grall*. A finding that children are able to learn *nall* but not *grall* would therefore be difficult to explain by any means other than such a restriction on the way internal and external arguments of determiners are interpreted.

### 3.2. Experiment design and methodology

Each participant was assigned randomly to one of two conditions: the conservative condition or the nonconservative condition. Participants in the conservative condition were trained on *nall*, and participants in the nonconservative condition were trained on *grall*; each participant was then tested on whether he/she had learnt the determiner he/she was exposed to.

To assess the participants’ success in learning, we used a variant of the “picky puppet task” (Waxman and Gelman 1986). The task involves two experimenters. One experimenter controls a “picky puppet”, who likes some cards but not others. The second experimenter places the cards that the puppet likes in one pile, and the cards that the puppet does not like in a second pile. The child’s task is to make a generalisation about what kinds of cards the puppet likes, and subsequently “help” the second experimenter by placing cards into the appropriate piles.

The experimental session was divided into two phases: warm-up and target. During the warm-up phase, the experimenter ensured that the child could sort cards into piles. For example, in one warm-up item the child would be told “The puppet only likes cards with yellow things on them”, and would be asked to sort a number of cards into “like” and “doesn’t like” piles. The warm-up phase contained three

<sup>10</sup>The fact that *grall* happens to live on its external argument makes it anticonservative — unlike *equi*, which is neither conservative nor anticonservative — but this is not relevant here.

items; the cards, and the criterion to be met for the puppet to like them, differed from each item to the next.

The target phase used cards like those shown in Figure 1, and was divided into a training period and a test period. The child was told that the puppet had revealed to the experimenter whether he liked or disliked some of the cards, but not all of them. The child was told that the experimenter would sort what they could, but that the child would then have to help by sorting the remaining cards that the puppet was silent about. During the training period the experimenter sorted five cards, according to the criterion appropriate for the condition: in the conservative condition, the puppet likes cards where *nall girls are on the beach* is true, and in the nonconservative condition, the puppet likes cards where *grall girls are on the beach* is true. The experimenter placed each card into the appropriate pile in front of the participant, providing either (10a) or (10b) as an explanation as appropriate. Note that whatever the determiner being tested, it was pronounced *gleeb*; the names *nall* and *grall* have been used solely for expository convenience in this paper.

- (10) a. The puppet told me that he likes this card because gleeb girls are on the beach.
- b. The puppet told me that he doesn't like this card because it's not true that gleeb girls are on the beach.<sup>11</sup>

Having placed all the training cards (those that “the puppet had told the experimenters about”) in the appropriate piles, the experimenter turned the task over to the child for the test period. The experimenter handed five new cards to the child, one at a time, and asked the child to put the card in the appropriate pile, depending on whether or not the child thought the puppet liked the card. The experimenters recorded which cards the child sorted correctly and incorrectly with respect to the criterion used during training. The cards that the experimenter had sorted during the training period remained visible throughout the testing period.

The same training cards and the same testing cards were used in both conditions, though whether the puppet liked or disliked the card varied from one condition to the other. Table 1 shows, for each card, the number of girls and boys on the beach and on the grass, and whether each condition's relevant criterion is met or not. These were designed to be as varied as possible, while maintaining the pragmatic felicity of the two crucial target statements. The total number of characters on each card was also kept as close to constant as possible: either five or six for each card. The number of training cards that the puppet likes is the same in each condition (three), so the situation that the participant is presented with during the training phase is analogous across conditions.

The participants were 20 children, aged 4;5 to 5;6 (mean 5;0). Each condition contained 10 children: ages of those in the conservative condition ranged from 4;5 to 5;5 (mean 4;11), and ages of those in the nonconservative condition ranged from 4;11 to 5;3 (mean 5;1). The two groups did not differ significantly in age ( $t = -1.4141$ ,  $df = 18$ ,  $p > 0.174$ ).

---

<sup>11</sup>Negation was always expressed in a separate clause to avoid any undesired scopal interactions.

Card	beach		grass		<i>nall girls are on the beach</i>	<i>grall girls are on the beach</i>
	boys	girls	boys	girls		
Train 1	2	0	1	2	true	true
Train 2	0	2	3	0	false	false
Train 3	0	1	2	3	true	false
Train 4	2	3	0	0	false	true
Train 5	2	1	1	2	true	true
Test 1	3	0	0	2	true	true
Test 2	0	3	3	0	false	false
Test 3	2	3	0	2	true	true
Test 4	1	2	2	0	false	true
Test 5	1	2	0	2	true	true

Table 1: The distribution of girls and boys on each card in Experiment 1

Condition	Conservative	Nonconservative
Cards correctly sorted (out of 5)	above chance ( $p < 0.0001$ , mean 4.1)	at chance ( $p > 0.2488$ , mean 3.1)
Subjects with “perfect” accuracy	50%	10%

Table 2: Summary of results from Experiment 1

### 3.3. Results

The results indicate that children exposed to the novel conservative determiner successfully learnt it, and that children exposed to the novel nonconservative determiner did not. The results are summarised in Table 2.

First we can consider how many cards children in the two conditions sorted correctly. If children never succeeded in learning the determiner’s meaning, we would expect performance to be at chance (namely 2.5 cards correctly sorted out of 5). Children in the conservative condition performed significantly better than chance ( $\chi^2 = 74.160$ ,  $df = 5$ ,  $p < 0.0001$ ), sorting an average of 4.1 cards correctly, whereas children in the nonconservative condition did not ( $\chi^2 = 6.640$ ,  $df = 5$ ,  $p > 0.2488$ ), sorting an average of 3.1 cards correctly.

Alternatively, we can consider how many children in each condition performed “perfectly”, sorting all five test cards correctly. Of the children in the conservative condition, five out of ten sorted all test cards correctly, whereas only one child out of ten in the nonconservative condition sorted all test cards correctly, indicating a correlation between conservativity of the determiner and success in learning ( $p = 0.07$ , Fisher’s exact test).

The results are even more telling when we look more closely at the responses of the one child who sorted all five test cards correctly in the nonconservative condition. This child told the experimenters that the puppet was confused about which characters on the cards were boys and which were girls. Recall that in this condition the true criterion for the puppet to like a card was *grall girls are on the*

*beach*, or equivalently *not all beach-goers are girls*. But another statement equivalent to these is *some boys are on the beach*. So if the child thought that the puppet intended the internal argument of the determiner in the crucial sentence to denote the set of boys, then she in fact learnt a **conservative** meaning for *gleeb*, with a meaning like *some* has. One might even be tempted to suggest that she was led to believe that the puppet was confusing boys with girls **because of** a requirement that *gleeb* be understood conservatively.

Of course, these results should only bear on the issue of determiner meanings to the extent that we are confident that the participants really did understand the relevant parts of the explanations in (10) to have the structure shown in (1). Had we found no difference between the conservative and nonconservative conditions, one might be hesitant to reject the hypothesis that determiners are restricted to conservative meanings, because of the possibility that participants were not analysing the crucial word as a determiner. But it is unlikely that we would have found results consistent with the independently motivated restriction to conservative determiner meanings if participants had not been using determiner structures.

## 4. Experiment 2: *fost*

### 4.1. Experiment design and methodology

In this experiment we investigate the learnability of the fictional determiner *fost*, introduced in Section 2. The methodology differs only slightly from that used in Experiment 1. These differences were generally due to the increased complexity of *fost* over *nall* and *grall*: it requires comparisons of non-zero cardinalities (or equivalent), rather than mere detection of the presence of girls/boys on the beach/grass. (Note that knowing only whether each number is zero or non-zero in Table 1 is sufficient to determine the truth or falsity of either of the sentences investigated in Experiment 1.)

First, the number of cards was increased: we now use six training cards instead of five, and twelve testing cards instead of five. Seeing a larger number of cards sorted is necessary to distinguish between candidate sorting procedures, both from the point of view of the participant figuring out the puppet’s criterion during the training phase, and from that of the experimenter figuring out what a participant thinks *gleeb* means on the basis of their sorting of testing cards.

Second, the pictures on the cards were modified. A sample card is shown in Figure 2. The layout was changed to make grouping of characters by gender or by location equally salient, with each card effectively showing a two-by-two “grid”.<sup>12</sup> On every card, no “cell” in this grid ( $\{\text{girls, boys}\} \times \{\text{beach, grass}\}$ ) was ever empty. As soon as one training card had been placed in the “like” pile and one in the “dislike” pile, the participant is able to determine that the novel determiner being tested must rely on comparisons between non-zero cardinalities.

---

<sup>12</sup>Note that the layout used in Experiment 1, if anything, biases the participant towards grouping

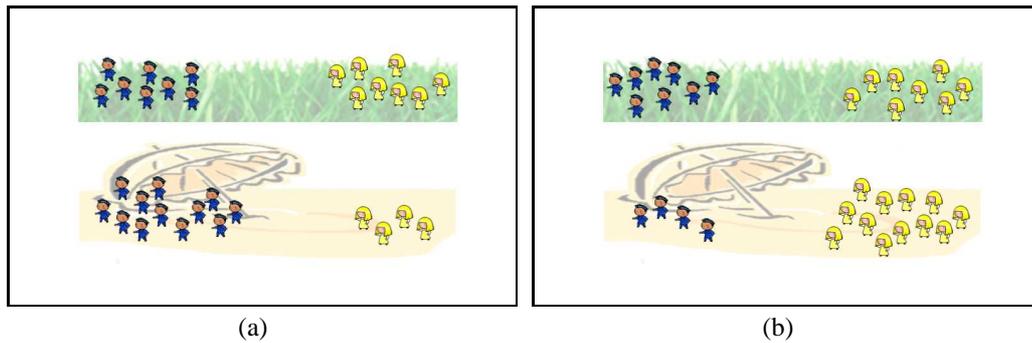


Figure 2: Two sample cards from Experiment 2. The relevant sentence, *fost girls are on the beach*, is true in (a) but false in (b); so in this experiment the puppet likes the card in (a) but not the card in (b).

Table 3 shows, for each card, the number of girls and boys on the beach and on the grass. In this experiment we only test one novel determiner, *fost* (always pronounced *gleeb*, as in Experiment 1), and the truth value of the relevant sentence is shown for each card.<sup>13</sup> The number of characters in each “cell” was either 4, 8 or 12. Numbers smaller than four were avoided so that none of the sets’ cardinalities could be compared within the subitisable range (Kaufman et al. 1949, Mandler and Shebo 1982, Trick and Pylyshyn 1994, Le Corre and Carey 2007).

In this experiment we tested both children and adults. In the case of children, the training cards sorted by the experimenter remained visible in their piles throughout the testing period as for Experiment 1, whereas for adults each training card was removed from sight before the next was presented. This was intended to make it difficult to detect the relevant generalisation just by examining all six training cards sorted, without taking any notice of the novel determiner.

We tested 12 children and 12 adults using the *gleeb* paradigm as described for Experiment 1 (i.e. with explanations as shown in (10)). In addition, we tested another 12 adults in a “no information” condition, and another 12 adults in a “full information” condition. In the “no information” condition, the explanations given by the experimenter while sorting the training cards were as shown in (11); in the “full information” condition the explanations given were as in (12).

- (11) a. The puppet told me that he likes this card.  
 b. The puppet told me that he doesn’t like this card.

---

of characters by location (rather than by gender), thus towards a nonconservative interpretation of the novel determiner. The new layout is intended to avoid any bias in either direction.

<sup>13</sup>In Experiment 1 the crucial comparison was between participants’ success learning *nall* and their success learning *grall*. There is no minimally different fictional neighbour to *fost*, so we must use different methods to evaluate participants responses when trained and tested on *fost*. See the discussion of results, below. (Using the non-fictional neighbour *most* as the meaning of *gleeb* for half the participants, so that analysis could proceed as for Experiment 1, would not necessarily be useful since participants trained on *most* may benefit from the existence of a determiner with this meaning in their lexicon.)

Card	beach		grass		<i>fost girls are on the beach</i>
	boys	girls	boys	girls	
Train 1	4	8	8	12	true
Train 2	12	8	8	4	false
Train 3	12	12	4	4	false
Train 4	12	4	4	12	true
Train 5	12	4	8	8	true
Train 6	4	12	8	8	false
Test 1	8	8	4	12	true
Test 2	4	12	12	4	false
Test 3	8	12	4	8	false
Test 4	4	4	12	12	true
Test 5	8	8	12	4	false
Test 6	4	4	12	8	true
Test 7	8	12	8	4	false
Test 8	8	4	12	8	true
Test 9	12	8	4	12	true
Test 10	4	8	12	4	false
Test 11	12	12	4	8	false
Test 12	8	4	8	12	true

Table 3: The distribution of girls and boys on each card in Experiment 2

- (12) a. The puppet told me that he likes this card because less than half of the girls are on the beach.
- b. The puppet told me that he doesn't like this card because it's not true that less than half of the girls are on the beach.

The participants were 12 children, aged 4;0 to 5;5 (mean 4;7), and 36 University of Maryland undergraduates. Adults received either course credit or payment for participation.

#### 4.2. Results

To begin we consider the three adult conditions, which vary only in the amount of information provided about the puppet's sorting criterion during the training phase. The crucial comparison will be between the "no information" condition, where participants are provided with no information beyond the distribution of training cards into "like" and "dislike" piles, and the *gleeb* condition, where they know in addition that the relevant criterion can be expressed as *gleeb girls are on the beach*. Participants in the *gleeb* condition therefore have a strictly smaller hypothesis space than do those in the "no information" condition, since they are limited to criteria that can be expressed by a natural language determiner. We would expect this reduction of the hypothesis space to be helpful if and only if the relevant criterion is in it — that is, if and only if *fost* is a possible natural language determiner.

To analyse the results we categorise participants according to the hypothesis about the puppet’s “liking criterion” which he/she most likely adopted, given his/her pattern of responses to the test cards. Four possible hypotheses we consider are in (13).

- (13) a. **most:** The puppet likes a card iff *most girls are on the beach*.  
 $\llbracket \text{most} \rrbracket (X)(Y) \iff |X \cap Y| > |X - Y|$
- b. **fost:** The puppet likes a card iff *fost girls are on the beach*.  
 $\llbracket \text{fost} \rrbracket (X)(Y) \iff |X \cap Y| < |X - Y|$
- c. **grmost:** The puppet likes a card iff *grmost girls are on the beach*.  
 $\llbracket \text{grmost} \rrbracket (X)(Y) \iff |Y \cap X| > |Y - X| \iff \llbracket \text{most} \rrbracket (Y)(X)$
- d. **grfost:** The puppet likes a card iff *grfost girls are on the beach*.  
 $\llbracket \text{grfost} \rrbracket (X)(Y) \iff |Y \cap X| < |Y - X| \iff \llbracket \text{fost} \rrbracket (Y)(X)$

In addition, we consider the possibility that a participant did not adopt any such specific hypothesis about the puppet (e.g. as a result of not being able to detect the relevant generalisation across the training cards), and sorted the test cards randomly. We therefore have five possible hypotheses in total. The choice of how to classify each participant was made on the basis of Bayes Factor model selection: we judge a participant to have adopted hypothesis  $H_i$  iff  $K_{ij} \geq 3$  for all other hypotheses  $H_j, j \neq i$ , where  $K_{ij}$  is the Bayes factor corresponding to the choice between  $H_i$  and  $H_j$ .<sup>14</sup> Note that some participants will remain unclassified on this criterion (but as we will see, this is rare in our results).

Figure 3 shows, for each of the three adult conditions, the number of participants classified according to each of the five hypotheses. Unsurprisingly, in the “full information” condition, participants overwhelmingly responded according to the *fost* hypothesis. In the “no information” condition, many participants (seven) responded according to the “random” hypothesis: these participants were apparently

---

<sup>14</sup>Intuitively, this means we took a participant to have adopted hypothesis  $H_i$  iff it is at least three times more likely that this participant adopted  $H_i$  than it is that he/she adopted any other hypothesis; the factor of 3 is the accepted threshold for the evidence to be considered “substantial” (Jeffreys 1961) (in much the same way as  $p < 0.05$  is the accepted threshold for “significance” in hypothesis testing). More specifically,

$$K_{ij} = \frac{Pr(D|H_i)}{Pr(D|H_j)}$$

where  $D$  is the participant’s sequence of responses (twelve responses, one for each testing card). This value is equivalent to  $\frac{Pr(H_i|D)}{Pr(H_j|D)}$  on the assumption of a uniform prior distribution over hypotheses. In the case of the four hypotheses corresponding to determiners, we assumed a 10 percent “error rate”: that is, if a participant adopts a determiner *det*, we assume that for each card the probability of sorting the card in accordance with *det* is  $\frac{9}{10}$ . (This is necessary to avoid the probabilities  $Pr(D|H)$  from going straight to zero whenever one response in  $D$  disagrees with  $H$ .) Thus if the sequence of responses  $D$  contains  $n$  responses in accordance with *det*, and therefore  $12 - n$  responses that disagree with *det*, then

$$Pr(D|H_{det}) = \left(\frac{9}{10}\right)^n \left(\frac{1}{10}\right)^{12-n}$$

The random hypothesis represents guessing randomly for each card, so  $Pr(D|H_{random}) = \left(\frac{1}{2}\right)^{12}$  for any sequence of responses  $D$ .

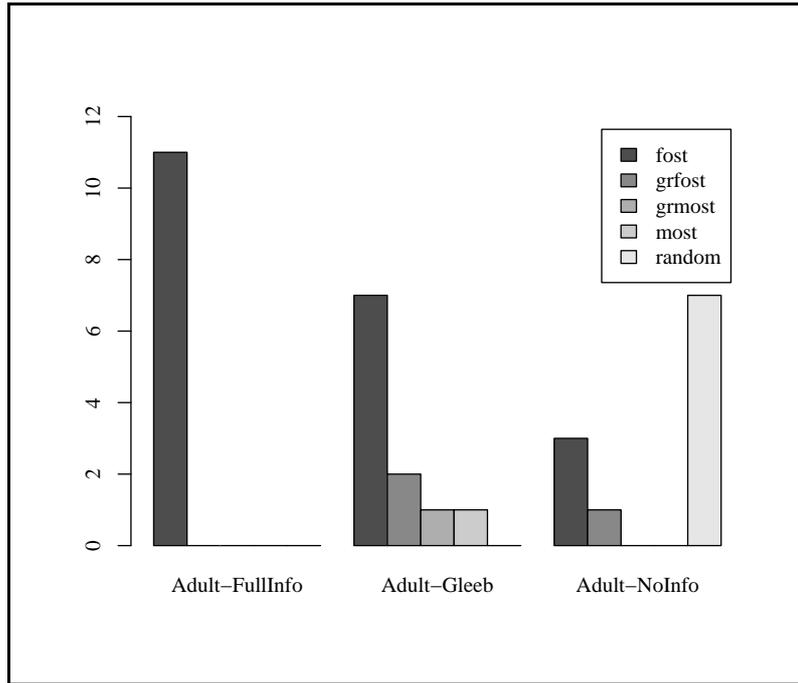


Figure 3: Adult participants classified according to the hypothesis we estimate each adopted

unable to extract any useful generalisation from the training phase. Some participants (three) in the “no information” condition, however, did manage to classify the test cards according to the correct *fost* criterion, purely on the basis of the distribution of training cards. When we turn to the *gleeb* condition, we see an increase in the number of participants who responded according to the correct *fost* criterion (and a reduction in the number who responded randomly, seemingly unable to detect a generalisation). When we distinguish only participants who adopted the *fost* hypothesis and those who adopted another, this increase is marginally significant by Fisher’s exact test; see Table 4.

As mentioned above, the only distinction between the “no information” condition and the *gleeb* condition is the additional linguistic information, in the latter case, that the correct underlying “liking criterion” is expressible as *gleeb girls are on the beach*. This increase in successful responses is only predicted if *gleeb* can be understood as *fost*. If *gleeb* could not be understood as *fost*, we would actually predict a **decrease** in correct responses if anything, since the additional information would rule out the puppet’s actual “liking criterion” and so participants should be led away from the correct hypothesis.

These results indicate that the fictional determiner *fost* lies inside the space of learnable natural language determiners, and therefore suggest that its typological absence (unlike that of *grall*) should not be put down to intrinsic constraints imposed by the language faculty. One would need to be cautious in interpreting the results of the adult participants in the *gleeb* condition alone, because of the possibility that participants adopt a strategy of searching for the relevant generalisation

	<i>fost</i>	other	random	other	nonconservative	other
Adult-NoInfo	3	8	7	4	1	10
Adult-Gleeb	7	4	0	11	3	8
Fisher’s exact test result	$p = 0.099$		$p = 0.002$		$p = 0.293$	

Table 4: The results of three different Fisher’s exact tests examining the distributions shown in Figure 3.

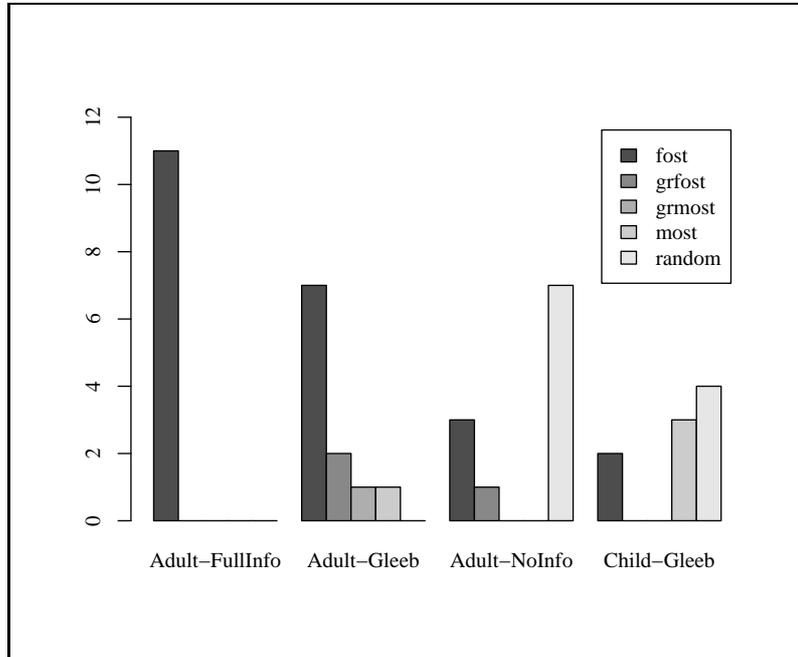


Figure 4: All participants classified according to the hypothesis we estimate each adopted

while ignoring the explanations provided using *gleeb*. But the “no information” condition provides a baseline measure of how successful this strategy can be, so the key point is the relative increase in the number of *fost* respondents in the *gleeb* condition over this baseline.

Table 4 shows two other analyses of the pattern in Figure 3. First, if we distinguish only the participants who adopted the random hypothesis from those who adopted another, we find strong evidence ( $p = 0.002$ ) that the extra information in the *gleeb* condition correlated with a decrease in the number of “random” response sets, indicating that there is at least some non-trivial restriction on the binary relations expressible by natural language determiners. Second, there is no evidence ( $p = 0.293$ ) that the extra information in the *gleeb* condition correlated with an increase in nonconservative (i.e. either *grmost* or *grfost*) response sets, despite the appearance of such an increase in Figure 3.

Given this finding for adults, we analysed the responses from child participants (who were only tested in the *gleeb* condition) in the same way. The results are shown alongside those of adults in Figure 4. The conclusions to draw from these

are less obvious but some interesting observations can be made. In particular, no child participant responded in accordance with either of the two nonconservative hypothesised determiners, *grmost* or *grfost*; this provides additional support for the conclusion we drew from Experiment 1, that children will not hypothesise nonconservative determiner meanings. Those who responded in accordance with any incorrect determiner we considered all chose *most*, despite the fact **every** training card provides evidence falsifying this possibility:  $\llbracket\text{fost}\rrbracket(X)(Y) \iff \neg\llbracket\text{most}\rrbracket(X)(Y)$ . In contrast, three of the six training cards were consistent with the *grmost* hypothesis and four with the *grfost* hypothesis, yet these were never adopted.

## 5. Conclusion

In this paper we have presented two experiments examining the relationship between learnability and typology in determiner meanings. A strong correlation is often thought to hold between learnability and typological generalisations, but recall that only in one direction is this connection logically necessary. While an unlearnable determiner will of course not be found in any natural language, determiners which are not found in any natural language need not — despite familiar reasons to suspect so — necessarily be unlearnable. Our experiments have investigated two typological generalisations in this light, only one of which, results indicate, constitutes an instance of the conventionally suspected correlation. Specifically, we found no evidence of participants successfully learning (unattested) nonconservative determiner meanings, but did find evidence of participants successfully learning an (unattested) determiner meaning “less than half” (which we call *fost*). This suggests that the typological absence of nonconservative determiners can be put down to unlearnability, but that the same cannot be said for the absence of *fost*; and in turn gives us reason to prefer theories of natural language semantics that allow “less than half”, but do not allow nonconservative relations, to be expressed as a determiner meaning.

## References

- Barwise, J. and Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4:159–219.
- Bhatt, R. and Pancheva, R. (2007). Degree quantifiers, position of merger effects with their restrictors, and conservativity. In Barker, C. and Jacobson, P., editors, *Direct Compositionality*, pages 102–131. Oxford University Press, Oxford.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.
- Chomsky, N. (1975). *The Logical Structure of Linguistic Theory*. Plenum Press, New York, NY.
- Crain, S., Thornton, R., Boster, C., Conway, L., Lillo-Martin, D., and Woodams,

- E. (1996). Quantification without qualification. *Language Acquisition*, 5(2):83–153.
- Drozd, K. and van Loosbroek, E. (1998). Weak quantification, plausible dissent, and the development of children's pragmatic competence. In *Proceedings of the 23rd Boston University Conference on Language Development*, pages 184–195, Somerville, MA. Cascadilla Press.
- Fox, D. (2002). Antecedent contained deletion and the copy theory of movement. *Linguistic Inquiry*, 33(1):63–96.
- Hackl, M. (2009). On the grammar and processing of proportional quantifiers: *most* versus *more than half*. *Natural Language Semantics*, 17:63–98.
- Herburger, E. (1997). Focus and weak noun phrases. *Natural Language Semantics*, 5:53–78.
- Higginbotham, J. and May, R. (1981). Questions, quantifiers and crossing. *The Linguistic Review*, 1(1):41–80.
- Horn, L. R. (1972). *On the Semantic Properties of Logical Operators in English*. PhD thesis, UCLA.
- Hunter, T. and Conroy, A. (2009). Children's restrictions on the meanings of novel determiners: An investigation of conservativity. In Chandlee, J., Franchini, M., Lord, S., and Rheiner, G.-M., editors, *BUCLD 33 Proceedings*, pages 245–255, Somerville, MA. Cascadilla Press.
- Inhelder, B. and Piaget, J. (1964). *The Early Growth of Logic in the Child*. Routledge & Kegan Paul, London.
- Jeffreys, H. (1961). *The Theory of Probability*. Oxford University Press, New York.
- Kaufman, E. L., Lord, M. W., Reese, T. W., and Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology*, 62:498–525.
- Keenan, E. and Stavi, J. (1986). A semantic characterization of natural language determiners. *Linguistics and Philosophy*, 9:253–326.
- Le Corre, M. and Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, 105:395–438.
- Mandler, G. and Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, 29:224–236.
- Mostowski, A. (1957). On a generalization of quantifiers. *Fundamenta mathematicae*, 44:12–36.
- Philip, W. (1995). *Event quantification in the acquisition of universal quantification*. PhD thesis, University of Massachusetts, Amherst.
- Pietroski, P. M. (2005). *Events and Semantic Architecture*. Oxford, New York, NY.
- Trick, L. M. and Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, 101(1):80–102.
- Waxman, S. R. and Gelman, R. (1986). Preschoolers' use of superordinate relations in classification and language. *Cognitive Development*, 1:139–156.