Un-learning *Un*-prefixation Errors

Ben Ambridge (Ben.Ambridge@Liverpool.ac.uk)
Daniel Freudenthal (D.Freudenthal@Liverpool.ac.uk)
Julian M. Pine (Jpine@Liverpool.ac.uk)
Rebecca Mills (md0u419e@student.Liverpool.ac.uk)
Victoria Clark (ps0u51a2@liverpool.ac.uk)
Caroline F. Rowland (Crowland@Liverpool.ac.uk)
School of Psychology, University of Liverpool, Eleanor Rathbone Building
Bedford St South, Liverpool L69 7ZA, UK

Abstract

A simple three-layer feed-forward network was trained to classify verbs as reversible with *un-* (e.g., *unpack*) reversible with *dis-* (e.g., *disassemble*) or non-reversible (e.g., *squeeze*), on the basis of their semantic features. The aim was to model a well-known phenomenon whereby children produce, then subsequently retreat from, overgeneralization errors (e.g., **unsqueeze*). The model learned to correctly classify both the verbs in the training set and verbs held back during training (demonstrating generalization). The model demonstrated overgeneralization (e.g., predicting *unsqueeze* for *squeeze*) and subsequent retreat, and was able to predict adult acceptability judgments of the different *un-* forms.

Keywords: Un-prefixation, overgeneralization; language acquisition; no negative evidence problem

Overgeneralization in Language Acquisition

A central question in the cognitive sciences is that of how children acquire their native language. Since speakers do not simply repeat whole utterances verbatim, the key question is how children are able to form the generalizations that allow for the production of novel utterances whilst avoiding *over*-generalizations (i.e., utterances that adults would consider ungrammatical).

One generalization that English-speaking children must acquire (presumably on the basis of hearing such forms as *unpack*, *unhook* and *unfold*) is that it is possible to add the prefix *un-* to a verb to specify the reversal of an action (i.e., they must acquire an *un-VERB* construction). This allows a child who hears (for example) the verb *fasten* to produce *unfasten,* even if she has never previously heard this form. Evidence that speakers do acquire a productive *un-VERB* construction (as opposed to simply learning all *un-* forms by rote) comes from overgeneralization errors attested in children's speech (e.g., *I'm gonna *unhang it*; *How do you *unsqueeze it*?; Bowerman, 1988).

Given that children do produce such errors, the challenge for acquisition researchers is to explain how children "un-learn" these overgeneralizations, whilst retaining the capacity for productive forms. Because children do not seem to receive feedback from caregivers when they produce overgeneralization errors, this has become known as the 'no-negative-evidence' problem (Bowerman, 1988).

One proposed solution is the 'entrenchment' hypothesis. This was originally developed for verb argument structure overgeneralization errors where a verb (e.g., the intransitive verb *disappear*) is overgeneralized into a construction (e.g., the transitive causative *SUBJECT VERB OBJECT* construction as in **The magician disappeared the rabbit*). The entrenchment hypothesis states that repeated presentation of a verb (e.g., *disappear*) in one (or more) attested construction (such as the intransitive construction; e.g., *The rabbit disappeared*) causes the learner to gradually form a probabilistic inference that adult speakers do not use that particular verb in non-attested constructions. A number of studies (e.g., Ambridge et al, 2008; in press; submitted) have shown that, as predicted by this hypothesis, speakers rate argument structure overgeneralization errors as less acceptable for high- than low frequency verbs (e.g., *disappear* vs *vanish*).

Whilst this proposal appears to work well for argument-structure overgeneralization errors, it is less clear that the account can be applied to morphological overgeneralization errors such as *un*-prefixation (Bowerman, 1988). A learning mechanism that deems *un-* forms ungrammatical when the observed frequency of the bare form (or the ratio of the bare to the *un*-prefixed[1] form) reaches a certain threshold would seem likely to make errors. For example, based on frequencies in the *British National Corpus*, a learner would have to hear around 500 occurrences of *twist* before encountering the (perfectly acceptable) form *untwist*. On the other hand, the non-reversible forms *embarrass* and *detach* each occur only around 500 times in the entire corpus. Worse still, some verbs are far more frequent in *un-* than bare form (e.g., *unleash* = 365; *leash* = 9).

An alternative proposal is that children use verb semantics to determine the syntactic and morphological constructions in which particular verbs can and cannot appear (e.g., Pinker, 1989; Ambridge et al, 2008; in press; submitted). For example, with reference to the intransitive/transitive causative alternation, Pinker (1989) proposed that children form narrow-range semantic classes of verbs that are restricted to the intransitive construction (e.g., verbs of ''coming into or going out of existence'' such as *disappear* and *vanish*). In support of this proposal, Ambridge et al (2008) found that when taught *novel* verbs of "coming into

[1] Here and throughout, '*un*-prefixed' means 'prefixed with *un-*' not 'with no prefix'

or going out of existence", both children and adults rejected (i.e., judged as ungrammatical) transitive causative uses.

Li and MacWhinney (1996) sought to extend this verb-semantics account to the domain of *un*-prefixation errors. Again, verbs that may appear in this construction appear to share certain meaning components such as *covering, enclosing, surface-attachment, circular motion* and *hand-movements*. Whorf (1956) argued that it is not possible to specify which verbs may and may not appear in the *un*-construction with reference to a list of necessary and sufficient semantic features (as Pinker, 1989, argued for verb argument structure constructions). Rather, these meaning components seem to combine interactively in a manner that is not straightforwardly predictable.

Li and MacWhinney (1996) developed a computational model designed to test Whorf's (1956) speculation that the *un-* construction constitutes a semantic "cryptotype". These authors trained a standard three-layer backpropagation network (with six hidden units) to produce an output of *un-dis-* or *zero-* (the three output units) for each of 160 English verbs (49 of which take *un-*, 19 *–dis* and 92 no prefix [termed "zero" verbs]). The model had 20 input units, each representing a particular semantic feature (e.g., *circular movement*; *change of state*). For each verb, the input to the model was a 20-bit vector representing the extent to which the verb was deemed to instantiate each of the semantic features (as rated by 15 adult participants). Verbs were presented to their model in proportion to their type and token frequency in a corpus of adult speech. The model's task was to learn to categorize each verb as (a) reversible with *un-*, (b) reversible with *dis-* or (c) non-reversible. The model performed reasonably well under a variety of different training regimes, correctly classifying between 50% and 75% of *un-* taking verbs (depending on the simulation).

It is important to note at the outset that Li and MacWhinney's (1996) model (like the model presented in the current paper) does not solve the no-negative-evidence problem. The pre-classification of verbs as *un-*, *dis-* or *zero* means that the model is given exactly the information that the child would need but does not receive (i.e., which verbs can and cannot be reversed). However, the model is valuable in that it demonstrates that, in principle, (a reasonable approximation of) the input available to children contains sufficient information to allow for the formation of a semantic "cryptotype" for the construction. For example, one strength of the model is that it uses this cryptotype to produce "overgeneralization errors" similar to those produced by children (e.g., **unhold*, **unpress*, **unfill*, **uncapture*, **unsqueeze*, **unfreeze*, **untighten*).

Nevertheless, Li and MacWhinney's (1996) model does exhibit a number of shortcomings. First, this model actually has great difficulty learning some forms. In the first simulation, the model learned to correctly classify (defined as an RMSE < .25) only 15% of the *dis-* verbs. In a second simulation, where *dis-* verbs were entered into the training set early in training, performance on *dis-* verbs improved. However, this was at the expense of the model's performance on the *zero* verbs (25% correct, vs 74% in Simulation 1) and *un-* verbs (51% correct, vs 76% in Simulation 1).

Second, this finding suggests that the particulars of the training regime may have been instrumental in shaping the particular pattern of results obtained. An incremental training regime was used such that the model was pre-trained on a set of 20 high frequency zero-verbs with verbs gradually added to the training set based on their type (*un-*, *dis-* or *zero*) and token frequency. The rate at which items were added furthermore changed during training. This incremental training regime was aimed at reflecting the realities of acquisition. While it has been shown that such manipulations may be crucial for successfully simulating developmental data (e.g. Elman, 1993), the very fact that they can influence results suggests that caution may be required when developing incremental training regimes.

A third shortcoming of Li and MacWhinney's (1996) model is that it actually lacks an important source of information that is available to children; namely, the distribution of surface forms. Reversible and non-reversible verbs differ not only in their semantics (information which is available to the model) but also their distribution: The former sometimes occur with *un-/dis-*, whilst the latter do not. Because the input to the model is simply a set of semantic vectors, this information is not available.

The final shortcoming of Li and MacWhinney's (1996) model is that it has great difficulty in retreating from overgeneralization errors. This would seem to be a consequence of the fact that the model produces overgeneralization errors in a way that is quite different to children. The model's overgeneralization errors result from mis-classification of items (e.g., *squeeze* is incorrectly classified as an *un-* verb, presumably because it shares a number of semantic features with genuine *un-* verbs). The model has great difficulty in re-classifying such verbs correctly (presumably because much of the semantic overlap that caused the erroneous classification remains even after learning has reached asymptote). Intuitively, it would seem that at least some of children's overgeneralizations are caused not by misclassification, but by *functional pressure*: Presumably, children produce forms such as **unsqueeze* because they want to denote the reversal of (in this case) a squeezing action, have learned that the *un-* prefix serves this function and do not have an alternative form that expresses the required meaning. Later in development, children are able to avoid producing *un-* forms for verbs such as *squeeze*, even when they are under functional pressure to do so (note, however, that even adults occasionally produce forms that they would probably regard as "overgeneralizations" in such circumstances; as in the form **unlearn*, which appears in the title of this paper). Li and MacWhinney's (1996) model does not simulate this situation as it is never 'asked' to produce a reversed (or non-reversed) form of a particular verb, as required for the discourse context; verbs are simply probabilistically assigned to one of three categories.

Our goal in the present study was to address these shortcomings with a new version of the *un*-prefixation model. This model differs from that of Li and MacWhinney in a number of important ways. First, the model was trained using a regime that more accurately reflects the frequency of individual forms in the input. This allows us to achieve more accurate classifications, whilst avoiding the need for discontinuities in the training regime.

Second, we aimed to determine whether a model trained on the semantic features of a subset of the verbs is able to successfully generalize its acquired structure to novel items when presented with their semantic features. Although the ability to generalize will be a crucial feature of any model of this phenomenon, no such test was conducted by Li and MacWhiney (1996). This test is crucial in determining whether a semantics-based model can account not only for the retreat from overgeneralization errors, but also for the formation of the generalizations that allow for such errors (and correctly produced novel forms) in the first place.

Third, the new model was designed to simulate not only overgeneralization - which was observed in Li and MacWhinney's study - but also, crucially, the *retreat* from overgeneralization, which was not. This was achieved by including in the input signal a 'reversative feature', which was switched on for reversed forms and off for base (non-reversed) forms. The model was trained on reversible items in both their base (e.g., *pack*, *appear*) and reversed forms (e.g., *unpack*, *disappear*). For example, the set of semantic vectors representing the verb *pack* was trained with the reversative feature off (corresponding to presentation of *pack*) for some trials and on (corresponding to presentation of *unpack*) for others. This feature makes it possible to explicitly 'ask' the model to produce a reversed form for verbs that were never presented in this form during training. This maps closely onto the scenario where children produce overgeneralization errors (e.g., to denote the reversal of a squeezing action) and hence allows us to model both overgeneralization and the *retreat* from overgeneralization in a realistic way. The inclusion of this feature has two further advantages that would seem likely to facilitate learning and generalization. First, it makes it possible to present reversible verbs to the model with the relative frequencies of the reversed and non-reversed forms in speech to children. Second, the information that a verb has occurred in reversed form constitutes a powerful cue that the verb (or collection of semantic features) is indeed reversible.

The final advantage of the new model is that it allows us to simulate adult acceptability judgment data. The inclusion of the reversative feature means that the output (i.e., the activation of the *un-/dis-* units) of the model when asked to produce a reversative form for a verb never presented in this form during training (e.g., *squeeze*) can be taken as analogous to an "acceptability judgment" for the reversed form (e.g., *unsqueeze*). This makes it possible to evaluate the model's performance in a very fine-grained way, by investigating whether its "acceptability ratings" of the various verbs in *un-* form correlate with adults' judgments.

Method

Our learning task was designed to more closely mirror that faced by real learners. In particular, our models were trained on both the base form and the reversed form of reversible verbs. The simulation used the same set of 160 verbs used by Li and Macwhinney (1996), pre-classified as *un*-taking (*N*=40), *dis*-taking (*N*=19) or *zero* (*N*=92). The input to the model consisted of the 20-bit semantic vector employed by Li and Macwhinney (whom we thank for making these data available to us) as well as a one bit 'reversative' feature. The reversative feature was set to 0 when a verb was presented in its base form, and to 1 when a verb was presented in its reversed form (either *un-* or *dis-*). The model had three output units, one for each of the three prefixes 'zero' 'un' and 'dis', and six hidden units. The task of the model (during training) was to predict whether each verb was a *zero* verb, an *un-* verb or *dis-* verb. Training items were presented in their base- (i.e., with the reversative feature off) and reversed forms (i.e., with the reversative feature on) relative to their (log) frequency in the British National Corpus (BNC). For example, the model was presented with *fasten* (BNC frequency 667) both in its base form (i.e., with the reversative feature off) and in its reversed form (i.e., with the reversative feature on; BNC frequency of *unfasten* = 97). In both cases, the "correct" activation pattern of the *un-*, *dis-* and *zero* output units (for the purposes of backpropogation) was 1 0 0 (i.e., activation of the *un-* unit only). Likewise, zero verbs (which take neither *un-* nor *dis-*) were never presented with the reversative feature during training. The formal classification of items as *zero, un-* or *dis* was the same as that used by Li and MacWhinney (which was determined by adult raters). Whilst this classification can on occasion clash with BNC usage, this often represents cases where a prefixed form does not in fact represent the reversal of an action (e.g., *disapprove* has a meaning that is opposite to that of *approve*, but does not denote the reversal of this action). Thus we decided to respect the classifications of the adult raters, rather than determining classifications on the basis of corpus usage.

During testing, the model was presented with the training set with the reversative feature switched on for all items. The activations of each of the three output units were then read off. For *un-* and *dis-* verbs, the reversative feature had occasionally been switched on during training (and was always associated with a target of *un-* or *dis-*). For zero verbs, which had never been paired with the reversative feature this was a novel situation. This corresponds to a scenario in which a human learner is attempting to produce a reversative form of a verb never encountered in this form (e.g., *squeeze*) or judge the acceptability of a reversed form offered by an experimenter (e.g., **unsqueeze*). Early in development, children are quite willing to produce overgeneral forms like **unsqueeze*, before learning to reject them later on. In these simulations, the relative activation of the *un-* and *dis-* output units was taken to reflect the model's acceptability rating of these forms.

The model was implemented using LENS, with all parameters set to their default values. The model was trained for a total of 100,000 trials (with one verb presented each trial) and tested after every 5,000 trials. Individual forms were included in the training relative to their log frequency in the *British National Corpus*. The order of presentation of items was randomized.

Results

Classifying verbs in the training set

The first simulation was designed to investigate the model's ability to correctly classify the training items. In this simulation, an item was considered correctly classified if the activation of the target output node exceeded 0.7. The results for this simulation (averaged over 5 runs of the model) are depicted in Fig. 1.

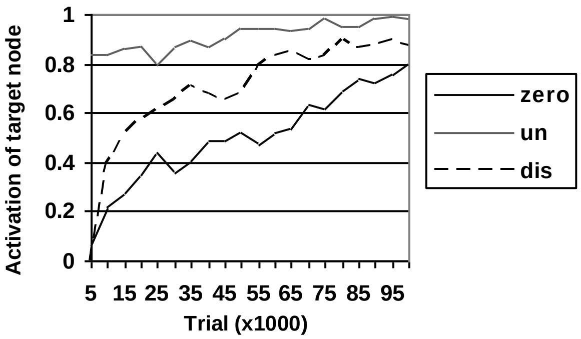

Fig.1: Proportion of correctly classified forms

As can be seen in Fig.1, the model is capable of correctly classifying an increasing number of items with increased training. Learning is particularly fast for the *un-* verbs, followed by the *dis-* verbs, and is slowest for *zero* verbs. However, even for the *zero* verbs, the model learns to ignore the fact that the reversative feature has been switched on (recall that the reversative feature is always switched on during testing). Thus despite the fact that the reversative feature was *always* associated with activation of either the *un-* or *dis-* output unit (and never the *zero* unit) during training, the model learns to correctly map 80% of *zero* verbs to the *zero* output unit when the reversative feature is switched on at test. This can be thought of as analogous to a child refusing to produce a form such as **unsqueeze* despite being under functional pressure to do so (or rating such a form as ungrammatical).

Generalization

Generalization – the ability to apply previously acquired "rules" or patterns to new items – is a key aspect of human linguistic competence. Given the semantics of novel verbs, both adults and children are able to determine whether or not this verb can be used in a particular construction (Ambridge et al, 2008; in press; submitted). (It is worth noting in passing that such findings are problematic for a purely statistical entrenchment account). Although we are aware of no studies that have investigated this phenomenon with regard to *un*-prefixation, it is reasonable to suppose that adults would be able to generalize in this way.

The second simulation was therefore designed to investigate the model's ability to generalize the knowledge it has extracted from the training set to novel items. This was done by removing 25% of the items from the training set (a different random set was held out for each of five runs). Testing then took place only on the items that were held out during training. Fig 2 shows the performance (average activation of the correct output node) of the model for these items, averaged over the five runs. As with the previous simulations, the model was trained for 100,000 trials. As can be seen from Fig. 2, the model is successful in generalizing its acquired knowledge to all three classes.

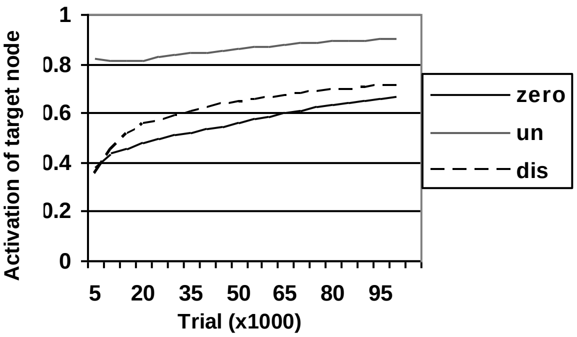

Fig. 2: Performance of the model on novel items.

Retreat from Overgeneralization

While the data presented thus far demonstrate that the model is capable of generalization, they do not demonstrate that the model – like children – produces, and then retreats from – overgeneralizations.

The data presented in Fig 1 suggest that the model produces overgeneralization errors, in that many *zero* verbs are incorrectly classified as *un-* or *dis-* verbs until relatively late in training. Nevertheless, this pattern is not necessarily indicative of overgeneralization behaviour. Even if a large percentage of *zero* verbs are not classified as such by a .70 criterion, it does not necessarily follow that the model is willing to overgeneralize on these items. For example, a verb activating the *zero* unit at 0.6 and the *un-* and *dis-* units each at 0.2 would be said to have failed in classifying the verb as a *zero* verb, but it would be odd to claim that the model was overgeneralizing the verb to *un-/dis-*. In order to more closely determine the model's willingness to over-generalize, we determined which output node showed the highest activation level for each of the *zero* verbs (for the simulation in which no verbs were held out). The results of this analysis are shown in Fig. 3. Early in training the *zero* node is most active for about 45% of *zero* verbs. Thus, when the reversative feature is switched on, the most active node is the *un-* or *dis-* node for 55% of zero-verbs (i.e, the model can be said to overgeneralize 55% of *zero* verbs to either *un-* or *dis-* when under functional pressure to do so). This decreases to around 10% at the end of training. Thus the model can be said to show the retreat from overgeneralization that is characteristic of children's learning.

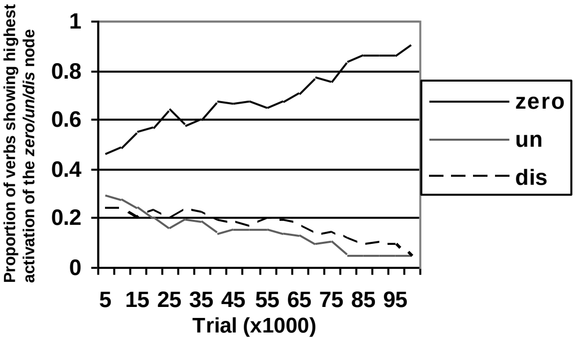

Fig. 3: Most active nodes for presentations of *zero* forms.

There are two possible reasons for the model's successful simulation of the pattern of overgeneralization on *zero* verbs. The theoretically interesting possibility is that this is caused by the presence of the reversative feature at test. On this explanation, it is the functional pressure of 'wanting' to reverse a verb (rather than an incorrect classification) that causes the model to overgeneralize (as we would argue is the case for children). A less interesting possibility, however, is that the class of *zero* verbs may simply be difficult for the model to learn (for example, it may be that *zero* verbs form a class that is less semantically cohesive than either *un*- or *dis*- verbs). This may cause the model to incorrectly classify *zero* verbs as either *un*- or *dis*- verbs. Indeed, misclassifications were the major cause of overgeneralization errors in Li And MacWhinney's (1996) simulations.

This possibility was tested by re-running our first simulation (with no items held out), with the modification that the reversative feature was switched on (when relevant) during training, but not at test, thus providing a baseline measure of the model's ability to classify items into the correct category. As in the first simulation, an item was considered correctly classified when the activation on the target node exceeded 0.7. The results of this analysis are shown in Fig. 4. As this figure demonstrates, the model is actually very successful in learning the zero-class. Thus, after a mere 5,000 trials, the model correctly classifies 75% of the zero-verbs.

These data suggest that the cause of the model's overgeneralizations is not the fact that the model incorrectly classifies many of the zero-verbs (though it may incorrectly classify some). Rather (as we would argue happens with children) the functional pressure to produce a reversative form (as instantiated in the model with the reversative feature) overrides the semantics of the *zero* class. With increased training the model (like children) learns to ignore this pressure and retreats from overgeneralization.

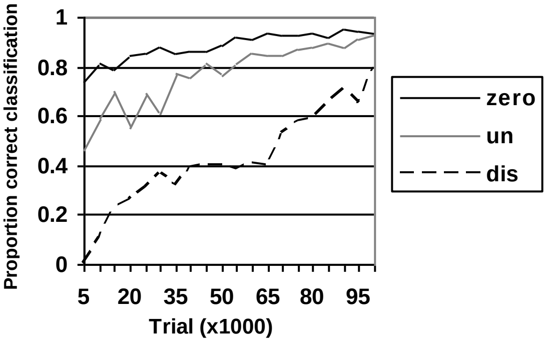

Fig. 4 Proportion of correctly classified items in the absence of the reversative feature.

Modeling Adult Acceptability Judgments

The data presented thus far show that the model displays a pattern of learning, generalization, overgeneralization and retreat from overgeneralization that is generally similar to that shown by children. In order to determine if the behaviour displayed by the model matches human behaviour more closely, we assessed the extent to which the model can simulate adult acceptability judgments.

Acceptability judgments of the base and *un*- form of each of the 160 verbs were obtained from 20 adult speakers of (British) English. Forms were presented in sentences with two different versions counterbalanced across participants. For example, 10 participants rated *Lisa bandaged her arm* and *Lisa unbandaged her arm* whilst 10 rated *Marge bandaged her friend's leg* and *Marge unbandaged her friend's leg*. Looking across all verbs, the correlation between the two sets was r=0.76 for the *un*-prefixed forms and r=0.55 for the base forms (both significant at p<0.001). This represents a reasonable upper-bound when assessing the model's ability to predict the human acceptability judgments.

In order to determine how well the simulation modeled the adult acceptability judgments, the mean adult judgments of (sentences containing) the *un*- forms were correlated with the model's activation of the *un*- node in the output bank (after 25,000 trials). Across all 160 verbs the correlations ranged from .68 to .73 for the five different runs of the model (all highly significant at p<0.001). Thus the model simulates to an impressive extent adults' ratings of the relative (un)acceptability of different *un*- forms.

This high correlation reflects the fact that adult judgments of overgeneralization errors are not binary but graded. Verbs that are highly incompatible with the semantic cryptotype for the construction (e.g., *talk*) are rated as extremely ungrammatical with *un*- (all raters gave **untalk* the lowest possible rating of 1/5). Non-reversible verbs that are, nevertheless, less semantically incompatible with the cryptoptype receive higher acceptability ratings (e.g., **unturn* = 1.67/5), whilst still being rated as unacceptable.

Indeed, even when looking only at the non-reversible (i.e,, *zero*) verbs, the model was able to predict the *extent* to which adults would consider the *un-* forms to be acceptable (notwithstanding the fact that all were, to some degree, unacceptable). Although the correlations were low (range .20-30) they were statistically significant for four of the five runs (*p*'s 0.01 – 0.05) and borderline for one (*p*=0.053). This is an important finding as the correlations calculated across all verbs will be somewhat inflated by the fact that verbs naturally cluster into two types: verbs that are reversible with *un-* and those that are not. Thus the adults and model could show a high correlation simply by rating the *un-*forms of all *un*-verbs as maximally acceptable (5/5 for adults, 1.0 *un-* unit activation for the model) and the *un-*forms of all *zero* and *dis-* verbs as maximally unacceptable (1/5 and 0.0). The fact that significant correlations between the predicted and actual acceptability of *un-* forms was observed, *when looking only at verbs that are not reversible*, demonstrates that the correlation observed was not simply an artefact of the fact that the verbs can be divided into two classes (*un*-taking and not-*un*-taking).

No significant model-adult correlations were found for acceptability ratings of the *un-* form of verbs that do take *un-* (i.e., where the *un-* form *is* acceptable, the model cannot predict the relative acceptability of the different un- forms). However, this is probably simply due to the fact that there is little relative acceptability (i.e., little variance) to explain, with most forms being rated as close to 5/5 (*M*=4.41, *SD*=0.76). The only *un*-taking verbs that received *un-* form ratings lower than 4/5 were *unarm, undelete, unmask* and *unscramble* (with the first two probably representing misclassifications). In any case, this issue is irrelevant to the question of the retreat from overgeneralization, as all these *un-* forms were acceptable (indeed, all had been encountered by the model and, presumably, the adults).

Discussion

The aim of the present study was to replicate and extend Li and MacWhinney's (1996) simulation of children's learning of *un-* prefixation. Specifically, we sought to implement a more plausible training regime in which both non-reversed and (where appropriate) reversed *un-/dis-*forms were presented in proportion to their frequency in a representative corpus. Another innovation was the introduction of a functional 'probe' for the reversative form which allowed us to investigate children's overgeneralization errors, and the retreat from such errors, in a more plausible way.

The first point to note is that the present model actually displayed better learning of the training set than Li and MacWhinney's (1996) original model. Thus we can be confident that the success of the previous model did not depend on unrealistic assumptions concerning the input or learning task, as a version of the simulation with (we would argue) more realistic assumptions actually performed better. The two key improvements would seem to be the more realistic training regime (including presentation of both reversed and non-reversed forms) and the presence of the reversative feature, which helps the model distinguish between reversible and non-reversible forms.

In addition to improved learning of the training set, the model was able to demonstrate generalization, overgeneralization and subsequent retreat from overgeneralization in a way that maps onto reports of children's performance. More impressively, the model was able to predict the relative (un)acceptability of the different *un*-prefixed forms as determined by adult raters.

With regard to theories of acquisition, the model adds to a growing body of evidence which suggests that pure statistical learning cannot explain how children form and retreat from grammatical (over)generalizations. Instead, what seems to be required is an account in which probabilistic learning of the semantics of particular verbs and constructions plays a key role (e.g., the *ILVACS* account of Ambridge et al, in press).

Of course, this model as it currently stands does not solve the 'no-negative-evidence' problem. To do so a model would need to determine which verbs are non-reversible or reversible with *un-* or *dis-*, without being given this information in the form of the correct output activation pattern. Such a model would likely need a more complex architecture than the simple feed-forward network used here. Nevertheless, the present set of simulations has demonstrated that a model that uses verb semantics to probabilistically learn verbs' argument-structure and morphological privileges is on the right tack with regards to solving the 'no-negative-evidence' problem.

References

Ambridge, B., Pine, J. M., Rowland, C. F., Jones, R., & Clark, V. (in press). A semantics-based approach to the 'no-negative-evidence'problem *Cognitive Science*.

Ambridge, B., Pine, J. M., Rowland, C. F., & Clark, V. (submitted). Restricting dative argument-structure overgeneralizations: A grammaticality-judgment study with adults and children. *Language*

Ambridge, B., Pine, J. M., Rowland, C. F., & Young, C. R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgments of argument-structure overgeneralization errors. *Cognition, 106*(1), 87-129.

Bowerman, M. (1988). The "no negative evidence" problem: how do children avoid constructing an overly general grammar? In J. A. Hawkins (Ed.), *Explaining language universals* (pp. 73-101). Oxford: Blackwell.

Li, P., & Macwhinney, B. (1996). Cryptotype, overgeneralization, and competition: a connectionist model of the learning of English reversive prefixes. *Connection Science, 8*, 3-30.

Pinker, S. (1989). *Learnability and Cognition: The Acquisition of Argument Structure.* Cambridge, MA: MIT

Whorf. B. K., (1956) *Language, Thought, and Reality. Selected Writings of Benjamin Lee Whorf*. John B. Carroll, ed. New York: Wiley