1. Introduction

In a recent issue of this journal, Marcus (1995) criticizes Plunkett and Marchman's (Plunkett and Marchman, 1993, henceforth PM) connectionist account of the acquisition of verb morphology for employing "misleading graphing practices" (p. 273), "including arbitrary, unexplained changes in how graphs were plotted" (p. 277) that exaggerate "the apparent very close similarity between the learning of the past tense by Adam and the ... model". Furthermore, Marcus suggests that the onset of overregularization errors in the model is "triggered ... by an externally imposed discontinuity" (p. 276). He argues that "children show a U-shaped sequence of development which does not depend on abrupt changes in input" (p. 278) and points out that "U-shaped development in the simulation occurs only after an abrupt change in training regimen". Finally, Marcus argues that the pattern of errors in the model do not conform to the pattern observed in young children - "Children overregularize vowel-change verbs more than no-change verbs; the simulation overregularizes vowel change verbs less often than no-change verbs" (p.278) and "children, including Adam, overregularize more than they irregularize; the simulation overregularized less than it irregularized".

In this response, we briefly review the goals of the original model and reject Marcus' criticisms of our original comparisons between the model's profile of development and that observed in young children. We will demonstrate, following Plunkett and Marchman (1990), Plunkett and Marchman (1991), Plunkett and
Marchman (1993), that externally imposed discontinuities in the training regime constitute neither necessary nor sufficient conditions for the onset of overregularizations in connectionist models of English past tense acquisition. Yet, this in no way undermines the claim made in PM and elsewhere that small vocabulary size allows young children to correctly produce both regular and irregular past tense forms, and that non-linearities in vocabulary growth are a contributing factor to the subsequent onset of overregularizations (Marchman and Bates, 1994). We also reject the claim that the errors produced by the PM model are substantively different from those of real children.

2. The scope of the model

The primary goal of the PM model was "to determine whether gradual quantitative and structural changes in the verb vocabulary can lead to qualitative shifts in the manner in which a network organizes the mapping relationship between verbs stems and their past tense forms" (p. 28). Following Rumelhart and McClelland (1986), Plunkett and Marchman (1991) and MacWhinney and Leinbach (1991), PM explored the extent to which the profile of mastery and pattern of children's past tense errors could be explained in terms of an hypothesized underlying process that systematizes relationships between phonological representations of verb stems and past tense forms. One key component was an incremental program of vocabulary growth in which new verbs are added to the training regimen at varying rates. Initially, the network was trained on a small vocabulary (10 regular and 10 irregular verbs), subsequently incremented by one verb every 5 training epochs. When 100 verbs had been introduced, the rate of the addition of new verbs was increased to one verb every training epoch. Token frequency depended on when the verbs were introduced to the training set. Thus, the incremental character of the training environment, a small initial vocabulary size, several verb types, variable token frequencies, assumptions about the phonological definitions of stem and past tense forms, and the learning architecture, constituted PM's identification of the factors that determine the profile of development and pattern of errors produced by children learning the English past tense.

Of course, several additional simplifying assumptions fell outside the overall goals of the model. The training vocabulary was an artificial language limited to triphonemic strings that conformed to the phonotactics of English. Second, the PM network cannot produce alternative forms for homophonic stems (Pinker and Prince, 1988) which are presumably differentiated by access to some representation of the verb's meaning - though see Cottrell and Plunkett (1994). Finally, PM utilizes phonological representations that do not interact with other components of the linguistic system. Insofar as these factors impinge upon the child's acquisition - and we don't pretend to have supplied an exhaustive listing - PM is bound to fall short of capturing all aspects of past tense acquisition. Nevertheless, this observation does not undermine PM's potential to contribute to our understanding
of the extent to which children's performance can be explained with respect to a specific, tightly-constrained set of initial assumptions. We remain convinced that these assumptions were reasonable and valid, the model offered important insights into the factors determining language acquisition, and the field gained substantive and testable hypotheses for empirical studies in real children.

3. Developmental sequence

Much of the substance of the Marcus (1995) critique centres on PM's comparisons of the profile of overregularizations produced by the PM model and those produced by Adam (derived from data in Marcus et al., 1992). One major comparison involved Adam's rate of overregularization in tokens versus the model's rate as measured in types as a function of size of vocabulary. In Marcus (1995), Adam's overregularization rate is replotted as types against vocabulary size as measured in types (Figure 2(a), p. 273). Not surprisingly, the rate at which Adam is estimated to overregularize increases. Marcus argues that the similarity between Adam and the model is reduced, and more substantively, that the claims of general similarity between PM and children are not justified. It may at first appear that types-versus-types is the appropriate basis for comparing model and child. However, apart from the fact that type overregularization rates were not available in the original monograph, further reflection confirms that our original strategy is the most appropriate given an understanding of the computational mechanisms underlying learning in these models.

Since Marcus does not provide a clear explanation of his procedure, we must assume type overregularization rate is determined by observing whether any overregularization errors occurred on an irregular verb, summing the number of verb types on which an error occurred and dividing by the total number of irregular past tense types produced on each session. For example, if the child produced 10 tokens each of two verb types and made 8 overregularization errors on one verb and 1 overregularization on the other, then type overregularization rate would be 100% (versus 45% measured in tokens). Since children can oscillate in their accuracy of production, type-based rates tend to inflate error profiles compared to calculations of tokens.

In contrast, the PM model will give a constant response to a given verb stem at a particular point in training. In connectionist models of this type, the weights are fixed during testing and activations are calculated in a non-probabilistic, deterministic fashion. The network cannot vary its response to different tokens of the same verb and will produce the same answer given all presentations of a single verb token. Of course, the model could have been given the opportunity to respond in a probabilistic fashion to different verb tokens. Yet, PM made no claims as to the source of this variability in children, and so we adopted the measure which was most appropriate given the context of the modelling endeavour. We chose to compare network performance with child performance at a more general level,
refraining from drawing direct comparisons of absolute levels of overregularizations when they were clearly inappropriate – see PM (pp.45–46).

Marcus goes on to note that “Adam continues overregularizing at vocabulary sizes of nearly 400” when “the PM model has long since stopped altogether” (p. 273). This difference is not reflective of key components of the model’s behaviour, but is a consequence of simplifying assumptions regarding the training regime and vocabulary structure. Specifically, while Adam appears to be acquiring new irregular verbs throughout the period of development studied (see Figures 21 and 27 in Marcus et al., 1992, pp. 86 and 92), the PM model, in contrast, is introduced to its last irregular verb when vocabulary size reaches 269 (see PM, Table 2, p. 34). It is no surprise that errors stop shortly after the last irregular verb is introduced. Once the network has learnt to accommodate these final irregulars, the predominant response characteristic is to add a suffix to all new remaining regular forms (see PM, Figure 4a, p. 48). Addition of further irregular verbs would have resulted in further overregularizations. Note also that the proportion of irregular verb types at the end of training is 8.4% for PM versus 25% for Adam (estimated from Marcus et al., 1992, Figure 27, p.92). Adam has considerably more opportunities for error than the network.

Marcus criticizes PM for “truncating the x-axis range to only 320 verbs” (p. 273), prior to when Adam reaches the model’s final vocabulary size of 500. However, it is at this point when Adam’s overregularization rate also reduces temporarily to zero (see Marcus et al., 1992, Figure 33, p. 108). Thus, PM truncated the comparison at the point at which overregularization errors are not observed in either the model or Adam.

The kind of detailed comparison of the profile of errors in the model and Adam suggested by Marcus (1995) is unwarranted and potentially misleading. Clearly, the training data used in the model are not the same as those available to Adam. The graphing practices and analyses performed in the original PM model reflect justifiable assumptions about the nature of the task the network is required to perform and conditions for comparison with real children. Within the context of the goals established by PM, the mode of comparison offers a conservative and, we would argue, more appropriate assessment of the model’s general capacity to capture the developmental sequence of overregularization errors observed in young children. The discrepancies that Marcus (1995) points out have no bearing on these issues.

4. Input

In PM, we plotted the simulation’s overregularization rate in types and proportion of input vocabulary that is regular in tokens against vocabulary size.
(Figure 6, p. 61), with the goal of evaluating the role that vocabulary composition plays in the onset of overregularization errors. In the same figure, we also replotted the most comparable available data in Marcus et al. (1992), that is, Figure 28 which plots Adam’s overregularization rate in tokens and proportion of input vocabulary that is regular in types against age. PM draw the following conclusion: “In general, then, the data for Adam as well as the simulations reported here do not demonstrate a relationship between changes in vocabulary composition and overgeneralization rate after vocabulary size has achieved a particular level” (original italics, p. 61).

Marcus (1995) again critiques this comparison, and replots, Adam’s overregularization rate in types and proportion of input vocabulary that is regular in tokens against vocabulary size (Figure 4a, p. 274). Based on this re-analysis, Marcus (1995) shows that the proportion of parental input to Adam that is regular measured in tokens is considerably less than when this proportion is measured in types. This is to be expected since irregular verbs tend to have a high token frequency. The resulting low proportions of regular verbs in the parental input to Adam ( < 36%) appears to undermine PM’s claim that “generalization is virtually absent when regulars contribute less than 50% of the items overall” (p. 55).

It is useful here to recall the theoretical assumptions of the model, namely that children’s overregularization errors can be explained in terms of their attempt to systematize the relationship between phonological representations of verbs stems known to them and phonological representations of the past tense forms known to them. Plunkett and Marchman (1991) argued for the importance of distinguishing between the input to the child and the uptake by the child. We assume that children do not learn simultaneously all the verbs they hear in the input, just those that are salient to them. Of course, a verb’s saliency may change over developmental time, although we suppose that token frequency plays an important role throughout learning. Connectionist models use a filtered version of the verb’s raw token frequency to guarantee that low frequency verbs will be processed by the network in finite time. In essence, the modeller specifies both the “uptake” and the “input” environment in the assessment of the degree to which absolute token frequencies influence the saliency of the training item. As a result, the incidence of low frequency forms in the “uptake” environment are inflated relative to the hypothesized “input” environment. In plotting Adam’s parental input in token frequencies, Marcus fails to take account of the distinction between input and uptake that is crucial for assessing the impact of frequency in studies of children.

Even though the proportion of regulars in the input to Adam (measured in tokens) never exceeds 36%, throughout the period of development reported in Marcus et al. (1992), the number of different irregular verbs (i.e., types) that Adam knows is substantially less than the number of regular verbs that he knows (see Figure 27 in Marcus et al., 1992). These are precisely the conditions that PM predict for generalization in the network and the eventual onset of overregularization errors: “once the proportion of regular verbs exceeds 50%, the absolute number of regulars in the training set are allowed to act in consortium to produce generalization effects” (p. 55, italics added). Here training set refers to the uptake training environment.
PM concluded "We can identify two periods in the training of the network which are consistent with the hypothesis that vocabulary size plays a substantive role in the learning of inflectional verb morphology in young children. An early period of training results in a rapid transition in generalization characteristics ... no overregularization errors are observed ... In the second period (signalled by a flattening off in the rate of increase in proportion of regular verbs in the vocabulary), the generalization tendency is consolidated and the first overregularization errors are observed" (p. 63). The period of development reported by Marcus et al. (1992) and Marcus (1995) for Adam corresponds to this second period of development.

A stronger test of the PM model would be to show that overregularization errors can occur in children with small verb vocabularies where irregular verbs outnumber the regulars. Yet, Marcus et al. (1992) has already led us to reject this possibility by demonstrating the robustness of the developmental phenomenon that children pass through an initial period of error-free performance, that is, overregularizations do not occur when vocabularies are small. This pattern is also consistent with the non-linear relationship between vocabulary size and overgeneralizations observed in more than one thousand children using parental report (Marchman and Bates, 1994).

5. U-shaped development

Marcus (1995) astutely notices that the onset of overregularizations in the PM network coincides with an increase in the rate of expansion of verb vocabulary after 100 verbs. If the implication of this observation is that this discontinuity is likely to be a contributing factor to the timing of the onset of overregularization errors, then we would agree. However, it is clear that Marcus would like us to believe that this “externally imposed” discontinuity is the primary cause of the onset of overregularization errors in these models. We disagree completely.

We should first remind readers that Plunkett and Marchman (1991) demonstrated that overregularization errors and U-shaped learning can be observed in the absence of any discontinuities in vocabulary configuration, that is, when vocabulary size is held constant. It is odd that Marcus would propose that discontinuity must play a strong causal role in the PM model. It is further puzzling that Marcus (1995), in his footnote 2, cites Kruschke’s (Kruschke, 1990) erroneous claim that “Plunkett and Marchman (1991) does not exhibit U-shaped learning” (p. 61). While it is correct to claim that network performance monotonically improved on average throughout training, this improvement was achieved at the price of temporary local decrements in performance on specific verbs, that is, micro U-shaped learning. The claim that Plunkett and Marchman (1991) did not exhibit U-shaped learning is correct only if children’s U-shaped behaviour results from macro, system-wide decrements in performance. Given that Plunkett and Marchman (1991) clearly predicted a micro U-shaped pattern and Marcus et al. (1992)
confirmed this prediction, it is curious that Marcus (1995) reintroduces this misinterpretation.

Of course, one goal of the PM model was to evaluate the impact of vocabulary growth on past tense acquisition. Plunkett and Marchman (1990) had already concluded that expansion rate discontinuities can impact the type and timing of overgeneralizations. However, we also concluded that this factor was neither a necessary nor a sufficient condition for the onset of errors. To drive the point home, we present data from 10 new simulations (identical to the PM model except for their random seeds) in which expansion rate discontinuities are either eliminated ($N = 5$) or shifted forward (i.e., introduced earlier in training, $N = 5$).

Fig. 1 plots the results (1 minus overregularization rate) for simulations in which verbs are added to the training set at a constant rate across training. Three (of five) individual simulations are presented here. Note that the point of onset of overgeneralizations varies considerably: Fig. 1(a) at vocabulary size of 80; Fig. 1(b) at 110 (the same point as PM); Fig. 1(c) at 260 verbs. Clearly, a discontinuity in expansion rate is not a necessary condition for the onset of overregularization errors ($M = 176$ across all five simulations).

Fig. 2 plots the results for three (of five) simulations in which the epoch expansion discontinuity is introduced early in training after 50 verbs (vs. 100 verbs). In Fig. 2(a) and (b), the onset of overregularizations occurs after 80 verbs;
while in Fig. 2(c), the first error does not occur until 250! (overall $M = 152$). A comparison of the no-discontinuity versus the early-discontinuity results (paired on initial random seed) revealed no significant difference in point of error onset ($t(8) = -0.42$, $p = 0.68$), although it is likely that early discontinuities would consistently produce earlier errors in a larger sample of simulations. However, the onset of overregularization errors is rarely coincidental with the epoch increment discontinuity. Thus, as Plunkett and Marchman (1990) predicted and PM demonstrated (pp. 51–55), these data indicate that an epoch increment discontinuity is not a sufficient trigger for the onset of overregularization errors. Rather, the number of regular verbs needs to exceed a critical level before overregularizations occur (i.e., reaches a “critical mass”).

It has been demonstrated repeatedly that discontinuities in the training environment to which connectionist networks are exposed do impact upon the profile of mastery and pattern of errors observed. It has also been demonstrated repeatedly that the presence of such discontinuities are neither a necessary nor a sufficient condition for errors to occur. These results predict that we should expect to find overregularization errors by children who exhibit non-linearities in the growth of their verb vocabularies as well as by children who do not exhibit such non-linearities. In PM, we chose to report on a simulation that underwent a particular schedule of vocabulary growth incorporating “non-linearities in rate of vocabulary growth that are sometimes observed in longitudinal studies of young children (Dromi, 1987)” (PM, pp. 33–34). This assumption of non-linear growth can be further justified by noting that data from the MacArthur Communicative Development Inventory (CDI: Toddler) (Fenson et al., 1993) suggest that the number of verbs in children’s reported production vocabularies increases at a slow pace early in vocabulary growth, yet increases dramatically after vocabulary size exceeds about 100 items ($F(1, 5) = 258.4$, $p < .001$). Thus, while not the primary determinant of error onset in these networks, an acceleration in rate of vocabulary growth in children is nevertheless a compelling feature of acquisition that we chose to incorporate into our model.

6. Types of errors

Evaluating errors on the 10 simulations reported above, we can confirm Marcus’ observation that no-change verbs ($M$ typewise = 8.1%) are more susceptible to overregularization than vowel change verbs ($M$ typewise = 5.8%) ($t(9) = 2.4$, $p < 0.04$). This can be traced to two factors. First, in PM’s artificial language, no-change verbs all ended in an alveolar consonant; whereas, vowel change verbs shared both a vowel and consonant with the other members of their family resemblance clusters. Thus, more phonological cues to class membership were available for vowel change verbs, an important determinant of accuracy of performance (Plunkett and Marchman, 1991). Interestingly, Bybee (1995) has recently pointed out that in English, “the set of verbs that undergo no change in the past tense ... all end in /t/ or /d/ and have lax vowels (with the exception of
beat ...") (p. 431). If this constraint were incorporated into PM, errors on no-change forms would be reduced.

Second, no-change overregularizations in children may be rare due to their phonological similarity to fully inflected regular forms (i.e., they already end in an alveolar consonant). That is, the language processor might be fooled into believing that a suffix has already been applied (Bybee and Slobin, 1982). This explanation is not appropriate to the PM model given that it does not code temporal information and that phonological information is not available in its 2-bit suffix representation. It is noteworthy that connectionist models that attempt to capture the full phonological coherence of no-change verbs and the phonological similarity of their endings to suffixed forms do not overregularize no-change more than vowel change verbs (MacWhinney and Leinbach, 1991; Rumelhart and McClelland, 1986).

The model reported in PM produced a total of 7 overregularization errors on vowel change verbs, only one of which was "stem + ed"; however, this does not reflect a general trend. Across the 10 new simulations, rates of "stem + ed" errors ($M = 2.4\%$) and "past + ed" errors ($M = 1.65\%$) on vowel change verbs are considerably higher than those observed in the original PM model (0.1% and 0.7%, respectively). However, we find no evidence to suggest that "past + ed" errors tend to be more prevalent overall ($t(9) = 0.982, p < 0.36$). On the contrary, the results suggest that further simulations would reveal a significant tendency in the opposite direction. Finally, contrary to Marcus’ claim, overregularization rates ($M = 6.1\%$) were reliably higher than irregularization rates ($M = 4.3\%$) ($t(9) = 2.3, p < 0.05$). These data are consistent with Marcus’ expectations about the rates of various types of errors.

At the same time, just as individual simulations vary considerably in the point of onset of errors, the precise patterns of errors and error types can differ across simulations. We consider this variability to be one of the strengths of this approach as it offers the opportunity to address issues associated with individual variation in language development. This variability also highlights the dangers of comparing the performance of any given simulation with the performance of specific children such as Adam. The connectionist approach predicts that the pattern of mastery of the English past tense will be highly sensitive to the specifics of each child’s lexical history. Clearly, detailed comparisons between the performance of a particular simulation and a particular child, when the simulation is not designed to reflect that child’s specific inventory of verbs, is likely to be misleading.

7. Concluding remarks

What have we learnt from this re-evaluation of the analyses conducted in the PM modelling endeavour? In our view, we have been reminded of the wealth of potential insights and empirically testable hypotheses that can derive from tightly constrained attempts to model significant characteristics of children’s acquisition of the English past tense (see PM, pp. 58–59 for a fuller discussion). However, it is also
clear that those insights must be filtered by an understanding of the computational mechanisms underlying these types of models, as well as the simplifying assumptions that are built in by the modellers. Of course, a number of major issues remain to be addressed. All in all, however, we continue to be surprised just how much can be explained by appealing to such a small part of the complex apparatus that must be involved in children's acquisition of the English past tense.

Acknowledgments

Thanks to Patrick Juola and Donna Thal for comments on an earlier draft of the manuscript. The research was supported by grants to Kim Plunkett from the Economic and Social Sciences Research Council, UK and to Virginia Marchman from the National Institutes of Health–National Institute on Deafness and Other Communicative Disorders (5 R29 DC 02292).

References