J. Child Lang. **22** (1995), 531–556.

Weird past tense forms*

FEI XU AND STEVEN PINKER
Massachusetts Institute of Technology

(*Received*: 22 *December* 1993. *Revised*: 5 *August* 1994)

ABSTRACT

It is often assumed that children go through a stage in which they systematically overapply irregular past tense patterns to inappropriate verbs, as in *wipe-wope*, *bring-brang*, *trick-truck*, *walk-has walken*. Such errors have been interpreted both as reflecting over-use of minor grammatical rules (e.g. 'change *i* to *a*'), and as reflecting the operation of a connectionist pattern associator network that superimposes and blends patterns of various degrees of generality. But the actual rate, time course, and nature of these errors have never been documented. We analysed 20,000 past tense and participle usages from nine children in the CHILDES database, looking for overapplications of irregular vowel-change patterns, as in *brang*, blends, as in *branged*, productive suffixations of *-en*, as in *walken*, gross distortions, as in *mail-membled*, and double-suffixation, as in *walkeded*. These errors were collectively quite rare; children made them in about two tenths of one per cent of the opportunities, and with few stable patterns: the errors were not predominantly word-substitutions, did not occur predominantly with irregular stems, showed no consistency across verbs or ages, and showed no clear age trend. Most (though not all) of the errors were based closely on existing irregular verbs; gross distortions never occurred. We suggest that both rule-theories and connectionist theories have tended to overestimate the predominance of such errors. Children master irregular forms quite accurately, presumably because irregular forms are just a special case of the arbitrary sound-meaning pairings that define words, and because children are good at learning words.

[*] This research was supported by NIH Grant HD 18381 and NSF Grant BNS 91-09766 to the second author, and by the McDonnell–Pew Center for Cognitive Neuroscience at MIT. We thank Gary Marcus for comments, and Michael Ullman and Marie Coppola for technical assistance. Address for correspondence: Fei Xu, Institute for Research in Cognitive Science, University of Pennsylvania, 3401 Walnut Street, Suite 400C, Philadelphia, PA 19104–6228 fei@lincs.cis.upenn.edu or Steven Pinker E10-016, MIT, Cambridge, MA 02139 steve@psyche.mit.edu.

INTRODUCTION

In learning a language, children must generalize from the sentences they hear from their parents, pulling out grammatical patterns and using them to construct an unlimited number of words and sentences of their own. In doing so they must avoid overgeneralizing: producing forms that their language community does not recognize as part of the language. In the study of language development, children's grammatical overgeneralization errors have always taken on a special interest (see Pinker, 1984, 1989, and Marcus, Pinker, Ullman, Hollander, Rosen & Xu, 1992, for reviews), because they constitute clear cases of children creating forms they could not have heard from their parents, and simultaneously pose the challenge of how the errors are eventually unlearned given that children receive no reliable feedback from their parents signalling every kind of error they make (see Brown & Hanlon, 1970; Marcus, 1993).

The most-cited overgeneralization errors in children involve application of the past tense suffix *-ed* to irregular verbs, as in *breaked* and *comed*. After several months of using correct irregular past tense forms like *broke* and *came*, children begin producing these 'overregularized' forms as well, generally in the third year, and continue to use both correct and incorrect forms well into the school-age years, when most incorrect forms gradually drop out. This 'U-shaped' developmental sequence suggests that children first memorize the correct irregular forms (e.g. *broke*) and then acquire a rule, 'add *-d*', which they overapply.

But in understanding children's errors, it is crucial to have quantitative data on their prevalence, time course, and distribution. Typically in the study of child language, an investigator will report anecdotally that children make a certain kind of error, without indicating how often, relative to correct forms, the error is made (see Pinker, 1989, Stromswold, 1990, and Marcus *et al.* 1992 for documentation of examples). The problem with such reports is that they are equally consistent with a child making the error in 100% of the opportunities, indicating a system that is intent on finding and using generalizations to the point of forgetting or suppressing the exceptions, and a child making the error in 0·000001% of the opportunities, indicating a system that is extremely good at recording the exceptions but prone to sporadic malfunctioning.

Overregularization errors provide a telling example. Most textbooks and secondary sources assume that children make the errors between 50 and 100% of the time, and conclude that children suppress irregular forms in a radical reorganization of their grammatical systems and have to re-learn the irregulars later. But until recently there were few primary sources that actually reported error rates from representative speech samples, and the secondary sources rarely cited the figures that did exist. Marcus *et al.* (1992)

carried out the first detailed quantitative study, using spontaneous speech transcripts of 83 children involving more than 11,500 past tense forms. They found that contrary to common belief, overregularizations are rare in spontaneous speech: the errors occur with a mean of 4%, and a median 2·5%, of irregular past tense forms. Moreover, these low rates are not artefacts of averaging over children, verbs, or ages; the errors occur sporadically for all children throughout the preschool years and for all their commonly used verbs. Marcus *et al.* argued that children do not go through a stage where they ignore the exceptional nature of irregular verbs and apply the rule to them indiscriminately. In fact children are never even indifferent between correct and overregularized forms; at all times they prefer the irregular. The errors come from occasional failures in a system that is designed to suppress them. Though retrieving an irregular form from memory automatically blocks application of the regular rule to the stem (that is why adults do not say *breaked*), memory retrieval is not perfect, especially in children, and when retrieval fails in a child who has command of the regular rule, the rule is applied to the stem, and an overregularization error results. The low error rate allowed Marcus *et al.* to present a parsimonious theory in which no radical reorganizations were necessary to account for the onset or disappearance of the errors: as the child hears irregular past tense forms more and more often, the memory entry for them becomes stronger, retrieval more reliable, and overregularizations less and less likely. Their conclusions underscore the necessity of having extensive quantitative data while theorizing about children's errors; in the absence of such data, the errors can mistakenly be interpreted as the modal response and unnecessary theoretical complications ensue.

Though application of the regular suffix *-ed* to irregular forms is the most common kind of error that children make in their past tense forms, it is not the only kind, and in this paper we examine several rarer errors, which we collectively refer to as 'weird past tense forms'. The most prominent is the overapplication of *irregular* inflectional patterns to regular verbs or inappropriate irregular ones, such as *wipe-wope* (cf. *write-wrote*, *ride-rode*) and *bring-brang* (cf. *sing-sang*, *ring-rang*). These 'irregularizations' were first noted by Chamberlain (1906), who found examples such as *hide-hod* and *take-tooken* in his daughter's speech. Bowerman (1982), Bybee (1985), and Pinker & Prince (1988) also presented some examples, such as *bite-bote* and *break-brekked*, and in elicitation studies, Bybee & Slobin (1982) and Marchman (1989) found irregularization errors like *think-thunk* and *sting-stang* in children from four to nine.

Irregularization errors show that even though irregular verbs are by definition idiosyncratic and unpredictable, children are not simply memorizing them as an unrelated list of words, but must be sensitive to the patterns found among them, such as in *ring-rang*, *sing-sang*, and *drink-drank*, or in

speak-spoke, *steal-stole*, and *freeze-froze* (see Pinker & Prince, 1988, for a complete list). Applications of irregular patterns to new forms can also be seen in other language phenomena. One example is historical change in English; *kneel-knelt*, for example, is a relatively recent addition to English (about two centuries old), and *sneak-snuck*, common today in younger Americans, is even more recent. A related phenomenon is dialect divergence: nonstandard dialects have many local irregular forms like *drag-drug* and *slide-slood* (Mencken, 1936). Finally, in experiments in which adult subjects are asked to supply past tense forms for novel verbs like *spling*, they often suggest *splang* or *splung* (Bybee & Moder, 1983; Prasada & Pinker, 1993).

The existence of generalizable patterns in both regular and irregular verbs has led some linguists and psychologists to attribute both kinds of generalizations to grammatical rules. For example, Kiparsky & Menn (1977), Kiparsky (1982), and Halle & Mohanan (1985) have proposed that alongside the regular rule 'add *-d*' there are minor, subregular rules such as 'change *i* to *a*', which apply only to a small group of verbs such as *ring*, *sing*, *sit* and *spring*. Irregularizations would occur when children ignored the restriction, or had hypothesized a restriction that included a greater range of verbs than is appropriate for the adult language.

The hypothesis that irregularizations are the result of over-applied minor grammatical rules depends on the assumption that children actually go through a stage in which they apply such rules systematically. Indeed, several researchers have assumed that children go through a stage in which they irregularize all verbs, or irregularize some verbs all the time. For example, Haber (1975) writes:

> For example, considering the English past tense again, it is a well attested fact that some children learn the regular rule first, and apply it to all verbs productively. Later they learn that some verbs are irregular, and sort these out one by one. Other children, beginning with the regular rule, ABANDON IT FOR AN IRREGULAR RULE WHICH THEY INDISCRIMINATELY APPLY TO ALL VERBS, only later separating out the truly irregular ones and returning to the regular rule for the rest. Still other children progress in a different fashion, generalizing some irregular rule initially, and so forth ... [emphasis ours]

Similarly, Kiparsky & Menn (1977) wrote,

> Learning the past tense of *bring* might involve the following stages:
> I. *bring-brought* (present and past tense forms of verbs are learned separately in the lexicon).
> II. *bring-bringed* (the child learns the regular dental suffix, and overgeneralizes it).

> III. *bring-brought* (*brought* is relearned as an exception to the general rule).
> IV. *bring-brang* (the child learns the ablaut rule *i* → *a* and overgeneralizes it).
> V. *bring-brought* (*brought* relearned).
> (We do not claim, of course, that this whole sequence will occur for this particular word in each child; merely that such a pattern of bracketing out successively finer generalizations is typical.)

Irregularization errors are also discussed as if they were a common occurrence, indicating use of a linguistic rule, by Bowerman (1982) and Pinker & Prince (1988).

Irregularizations are also predicted to occur, though not at a consistently high rate, by an alternative to the theory that grammatical computation consists of symbol-processing rules. It has been proposed that children abstract out thousands of correlations between the sounds of stems and the sounds of their past tense forms, superimpose these correlations in a pattern associator memory, and generalize to new forms based on their overlap in sound with previously encountered ones. Pattern associator memories are usually implemented as connectionist models (also called Parallel Distributed Processing (PDP) models or Artificial Neural Networks), and PDP models of the past tense have been devised by Rumelhart & McClelland (1986), Plunkett & Marchman, (1991, 1993), MacWhinney & Leinbach (1991), Daugherty & Seidenberg (1992), and Sproat (1992). They have also been implemented as symbol-processing systems by Ling & Marinov (1993).

A key feature of these pattern associator models is that there is no architectural distinction between regular and irregular verbs; the devices are designed to pick up simultaneously on patterns of many degrees of generality. Irregularization errors, like overregularizations, occur when a correlation between a set of stem phonological features and a set of past phonological features has been strengthened across a set of verbs to such an extent that it overrides the correlations that have been learned for features of that verb alone. For example, experience with *ring-rang*, *spring-sprang*, and *sing-sang* might cause connections between *ing* units and *ang* units to be so strong that when the system is faced with inflecting *bring*, the *ing-ang* connection prevails over connections specific to *bring* (e.g. *bri-bro*, *ing-ought*, and so on), and a blend of correct and incorrect features is activated, resulting in *brang*. Thus irregularizations are natural outputs of most PDP past tense models, and indeed are sometimes cited as signs of the psychological reality of the models (e.g. Rumelhart & McClelland, 1986; Marchman, 1989). For example, Plunkett & Marchman's (1991) model irregularized the verbs it had been trained on between 3·2 and 5·2%; indeed, the version they describe in Plunkett & Marchman (1993), under many training conditions, produced

irregularization errors such as vowel-changes at a higher rate than over-application of the regular suffix (Plunkett & Marchman, 1993; Marcus, in press). Rumelhart & McClelland's (1986) model irregularized novel verbs (ones it had not been trained on) at a rate of 11·6%, and Egedi & Sproat's three-layer back-propagation enhancement of that model (see Sproat, 1992) irregularized novel verbs at a rate of 23·5%. Whether and how often such errors occur depends on a complex combination of the way the pattern associator is set up (e.g. how features are represented, how many 'hidden units' it has in the layer mediating between input and output), how many similar words there are in the training set and how similar and frequent they are, and the model's particular history of training. Therefore, one cannot simply use irregularization rates to test the psychological reality of PDP past tense models in general, though different models within that framework can be compared in terms of their ability to duplicate such rates. In addition other aspects of irregularization, such as their form and timing, can be brought to bear on the models.

Aside from rules and pattern associators, there is a third kind of representation that the mind might use for irregulars. A lexical entry is an abstract representation – an index, pointer, or address – of a word root, consisting of an arbitrary link between a sound, a meaning and a grammatical category. Irregulars could be stored as linked pairs of lexical entries, and in both rule theories and PDP theories such entries could be used to curtail the application of common irregular patterns (rules or connections) to inappropriate verbs like *bring* and *slay*: the entries would simply list the appropriate forms outright, linked to their stems. Lexical entries would prevent errors when they were properly retrieved. But even pure entries could give rise to a kind of irregularization error, if the retrieval pathways for the entries were in some way sensitive to the entries' sounds (see, e.g. Dell & Reich, 1980; Stemberger, 1983; Bybee, 1991): if the pathways were weak or momentarily degraded, they could point to an incorrect similar entry. Indeed, Bybee & Slobin (1982) found that 71 out of 85 (83·5%) of the novel change responses produced by their adult subjects in an elicitation task were real English verbs, such as *seat-sat*, *weed-wed*, *glide-glow*, and *drink-drunk*), and Bybee (1985) noted that in an informal elicitation experiment with her son, *slay* was given the past tense *slung*. Bybee & Slobin suggest that such errors could have resulted from misretrievals of whole lexical entries, thus providing a third hypothesis for the cause of children's irregularization errors.

Most discussions of irregularization and its causes assume that the errors occur reasonably frequently. Is there any evidence supporting these hypotheses? In Bybee & Slobin's (1982) elicitation experiment, third graders (eight- and nine-year-olds) irregularized existing verbs at a rate of 2·4%; in Marchman's (1988), four- to six-year-olds irregularized at a rate of 13 to 7%.

(Marchman's figures are difficult to interpret, however, because when the children simply repeated back the stem unaltered, she counted the response as an irregularization, as if they had been analogized from verbs like *hit-hit* and *cut-cut*. But the children could just as easily have been not trying to convert the verb to a past tense form at all.) In Berko's study (1958), four- to seven-year-olds irregularized nonce verbs at the rate of 0·78% (as a proportion of the total number of forms they produced in the past tense), and never irregularized the two existing English verbs in the study, *ring* and *melt*.

Elicitation tasks, however, can yield misleading estimates of error rates, as an examination of published rates of overregularization (e.g. *comed*) makes clear. Bybee & Slobin (1982) found third graders (eight- and nine-year-olds) overregularized between 2 and 55%, depending on the verb subclass; Kuczaj (1978) found that the overregularization rates for three- to four-year-olds, five- to six-year-olds, and seven- to eight-year-olds were 29, 49 and 1%, respectively. Yet the mean rate of overregularization in spontaneous speech, as reported by Marcus *et al.* (1992) was only 4%. Thus it is important to estimate the rate of irregularization in spontaneous speech, which so far has never been done.

The rate of irregularization is highly relevant to understanding the mental mechanisms underlying inflection. If children systematically irregularize (e.g. some children go through a stage in which they irregularize all *-ing* verbs to *-ung* 100% of the time), it would suggest that they have hypothesized an incorrect irregular rule. Similarly, if irregularizations are not 100% but nonetheless fairly common, that could support a pattern associator model lacking lexical entries or other representational resources dedicated to individual words, which would thus tend to favour PATTERNS over word-specific idiosyncratic information. But if irregularization is very rare, it would suggest that the child has very accurate rote memory abilities; irregularizations would be occasional slip-ups in an architecture designed to reproduce irregulars verbatim almost all the time. Any model of the child (connectionist, rule-based, or hybrid) would have to provide resources sufficient to reproduce irregular words accurately most of the time from the earliest stages.

Extensive data on the nature and time-course of irregularization errors are also important. For example, in most pattern associators, words are represented as distributed patterns of activation over a set of phonological feature units, and the inflected output form is a complex blend of appropriate and inappropriate features contributed from connections strengthened over hundreds of different verbs. As a result, all kinds of blended errors are possible, not just the vowel-changes and suffixations that would be the direct applications of minor rules to inappropriate stems. The Rumelhart–McClelland model (1986) produced systematic vowel changes (*cling-clung*), systematic blends (*weep-wept*, *brown-brawned*), and gross distortions (*mail-*

membled, *tour-toureder*); Prasada & Pinker (1993) found that the model was especially prone to the distortions when presented with unusual-sounding stems. The Egedi–Sproat model (Sproat, 1992) also produced such forms (e.g. *conflict-conflafted*, *wink-wok*, *quiver-quess*, *satisfy-sedderded*), as did the MacWhinney–Leinbach model until a special non-associative pathway was added to the model to overcome that tendency (MacWhinney, personal communication). Thus it is of interest to see how often children produce such distortions. Pinker & Prince (1988) discuss several other patterns of weird past tense forms produced by the Rumelhart–McClelland model, each of which raises the question of children's behaviour with such forms.

In this paper we report the first large-scale quantitative analysis of irregularizations in children's spontaneous speech, involving 20,000 past tense and participle usages from nine children (ages two–seven) in the CHILDES database (MacWhinney & Snow, 1985, 1990). In addition to simple irregularizations, irregularization blends and gross distortions, we looked at two other kinds of weird past tense forms. One consisted of generalizations of the irregular participle suffix *-en*, as in *walken*, *sangen*, or *tooken*; Fletcher, 1991, presents data on these errors from one child. Pinker & Prince (1988), Marcus *et al.* (1992, in press), and Kim *et al.* (1994) present linguistic evidence that would categorize these as irregularizations, but some rule theories (Kiparsky, 1982) treat *-en* as a limited affixation rule for participles (used, for example, in *take-took-taken* and *break-broke-broken*), which gets overgeneralized by children. Second, we looked at double-marked forms such as *raineded*, which are frequently made by the Rumelhart–McClelland model. We also document as far as possible the consistency, nature, and time course of all these errors.

METHOD

Subjects

Nine children from the CHILDES database (MacWhinney & Snow, 1990) were selected for the present project; they are listed in Table 1.

Procedure

Since an exhaustive hand-search of the transcripts was impractical, five methods were employed to find examples of possible irregularizations by a computer search. First, we generated a list of all possible irregularizations, by extracting all vowels found in existing English past tense forms of irregular verbs and applying them to all regular and irregular verbs found in CHILDES (both in the adults' and children's speech). For example, for the existing irregular verb *sit-sat*, we searched for *sate*, *set*, *sot*, *sought*, *sut* and so on

TABLE 1. *Transcripts sampled*

Name of child	Corpus	Age of child	No. of past tense & participle tokens
Abe	Kuczaj	2;4–5;0	4744
Adam	Brown	2;3–4;10	3125
Eve	Brown	1;6–2;3	365
Mark	MacWhinney	0;7–6;0	2353
Naomi	Sachs	1;1–5;1	833
Nathaniel	Snow	2;5–3;9	541
Peter	Bloom	1;9–3;2	1385
Ross	MacWhinney	2;6–8;0	4746
Sarah	Brown	2;3–3;9	2476
		Total	20,568

(analogous to *made*, *let*, *forgot*, *bought* and *shut*, respectively).[1] However, the resulting list is so long that an impractically large number of transcript lines contained at least one example from the list. Thus we trimmed the list in the following way: an irregularized form was searched for only if at least one child used the stem at least once. For example, the regular verb *to whet* has a possible irregularized form *what* based on *forget-forgot*, but none of the children ever used the stem *to whet*, so we never searched for *what*.

The second method of finding possible irregularizations was to check all non-words in the transcripts. We used the CLAN program *freq* (MacWhinney, 1990) to generate a list of all word-forms used by each child, and filtered the list through an on-line dictionary, resulting in a list of the non-words.

Third, in Adam, Eve, Sarah and Abe's transcripts, many phonological and morphological errors and other novel forms had been noted by the original investigators, and were marked in the transcripts by special symbols.

Fourth, we created a list of possible *-en* generalizations: all irregular stems plus *en* (e.g. *singen*), all irregular past forms plus *en* (e.g. *sangen*), and all regular verb stems in CHILDES plus *en* (e.g. *walken*), and searched the transcripts for instances of them.

Finally, we created a list of all doubly-suffixed versions of the irregular and regular verbs found in the children's transcripts, such as *brokeded*, *walkeded* and so on. The great majority of these would have turned up in the list of nonwords anyway, but occasionally one of the comment-lines in a transcript, which we did not search for nonwords, had noted an error that had not been transcribed in the child's speech-lines themselves.

Once the computer had created these lists of candidate errors, we utilized the *kwal* program in the CLAN package to extract the immediate context in

[1] We thank Michael Ullman for generating the list.

which these words were uttered. We looked through these parts of the transcripts by hand and recorded the ones that corresponded to incorrectly inflected past tense forms.

Though this search could in theory have missed some weird past tense forms, we are confident that it caught virtually all of them. First, it successfully found the errors that Cazden (1966) had listed in her study of Adam, Eve and Sarah's transcripts, (*beat-bate* and *beat-bet*; she also found *hit-heet* for Adam, but it is nowhere to be found in the CHILDES version of the transcripts). Second, it found all the errors that Stromswold (1990) listed during her very extensive hand-search for grammatical errors in the same transcripts. Third, it found the three examples given in Marcus *et al.* (1992): *bite-bat* (Adam), *sweep-sweepened* (Abe) and *trick-truck* (Adam).

We did not distinguish past tense and past participle usages in this data set: data from all children include forms pooled over past tense AND past participle forms, both correct forms and errors. This was because most verbs in English (even many irregulars) use the same form for past tense and past participle forms (e.g. *walked-has walked*, *fought-has fought*). Moreover, even for verbs with distinct past and participle forms, children's frequent omission of auxiliaries (e.g. *I gone*) makes it difficult, if not impossible, to determine when a child is confusing past tense with past participle forms and when he or she is using them correctly without an auxiliary.

No-change verbs like *hit* and *cut* present a problem: an alphanumeric search program like *kwal* cannot distinguish present and past tense uses. Fortunately, past tense uses of these verbs have been coded by hand for Adam, Eve, Sarah and Abe (data reported in Marcus *et al.* 1992). If we assume that the proportion of past tense tokens that consist of no-change verbs is similar across all children, then we can use that proportion from these four children (8·2 %) to adjust the total number of past tense tokens for each of the other five children accordingly. For example, if a child had 92 past tense/participle tokens, excluding no-change verbs, we estimated that the child actually had 100 past tense/participle tokens.

Errors were divided into two main categories: irregularizations, such as *trick-truck*, and double-marked forms, such as *sweep-sweepeded* and *make-makeded*. The category of irregularizations was further subdivided into incorrect vowel changes, such as *bring-brang*, correct present or past form plus *-en*, such as *take-tooken*, and blending errors, such as *bring-brunged*, where a vowel change was combined with the regular suffix or some other change found in past tense forms.

RESULTS

Because none of the CHILDES transcripts was transcribed phonetically, we cannot be sure of the details of the child's pronunciation. In this paper we

TABLE 2. *Irregularizations found in the transcripts*

Child	File	Age	Stem	Irregularized form	No. of irreg	No. of correct	No. of stem+*ed*	No. of past+*ed*
Abe	ABE048	2;10	*trip*	*trippen*	1	4	0	0
	ABE130	3;9	*bring*	*brunged*	1	2	5	1
	ABE190	4;8	*shoot*	*shotten*	2	10	2	1
	ABE190	4;8	*shoot*	*shotten*		(see *shoot* above)		
Adam	ADAM04	2;4	*fit*	*feet*	1	4	0	0
	ADAM15	2;10	*bite*	*bat*	1	12	0	0
	ADAM34	3;7	*beat*	*bate*	1	1	2	0
	ADAM37	3;8	*trick*	*truck*	1	6	0	0
Eve	EVE10	1;10	*sleep*	*slep*	2	0	0	0
	EVE10	1;10	*sleep*	*slep*		(see *sleep* above)		
Mark	ROSS60	3;5	*shoot*	*shooten*	1	6	0	0
	BOYS63	3;8	*say*	*set*	1	243	0	0
	BOYS66	3;10	*bring*	*brang*	8	4	0	0
	BOYS66	3;10	*jump*	*janged*	1	37	0	0
	BOYS71	4;2	*see*	*sawn*	2	80	0	0
	BOYS77	4;7	*bring*	*brang*		(see *bring* above)		
	BOYS77	4;7	*eat*	*aten*	1	54	3	0
	BOYS78	4;8	*bring*	*brung*		(see *bring* above)		
	BOYS78	4;8	*bring*	*brung*		(see *bring* above)		
	BOYS80	4;10	*bring*	*brang*		(see *bring* above)		
	BOYS80	4;10	*see*	*sawn*		(see *see* above)		
	BOYS82	4;11	*bring*	*brang*		(see *bring* above)		
	BOYS82	4;11	*swing*	*swang*	1	0	1	0
	BOYS92b	5;7	*bring*	*brang*		(see *bring* above)		
	BOYS92b	5;7	*bring*	*brung*		(see *bring* above)		
Nathaniel	NATH24	3;4	*bring*	*brang*	1	0	0	0
Peter	PETER19	2;10	*lift*	*left*	1	0	0	0
Ross	ROSS38	3;4	*crush*	*crooshed*	1	3	0	0
	ROSS44	3;8	*fight*	*fooed*	1	2	3	0
	ROSS45	3;10	*close*	*closen*	1	15	0	0
	ROSS53	4;7	*bring*	*brid*	2	11	4	0
	ROSS53	4;7	*sit*	*sought*	1	9	1	0
	ROSS55	4;9	*bring*	*brang*		(see *bring* above)		
	ROSS58	5;1	*drink*	*dranken*	1	10	1	0
	BOYS73	6;3	*swing*	*swang*	3	1	1	0
	BOYS73	6;3	*fling*	*flang*	1	0	0	0
	BOYS76	6;5	*swing*	*swang*		(see *swing* above)		
	BOYS92b	7;6	*swing*	*swang*		(see *swing* above)		
Sarah	SARAH035	2;10	*bite*	*bet*	1	11	0	0

reproduce the forms using the orthography with which the transcriber transcribed them.

In the lists in Tables 3 and 5, and in cases where the orthography is ambiguous as to pronunciation, we supply within parentheses a broad phonetic transcription representing our best guess of the child's pronunciation, based on the sound patterns common in American English and on the orthographic renderings in the rest of that child's transcripts.

We found 63 tokens of weird past tense forms, including 39 irregularizations of 14 verbs (e.g. *trick-truck*, *bring-brang*, *shoot-shooten*) and 24 double-marked forms of 17 verbs (e.g. *break-brokted* [broktid], *drown-drownded*). Thirty-three of the irregularizations were of irregular verbs (e.g. *bring-brung*); 6 were of regulars, (e.g. *crush-crooshed*) [kroʃt]). Results are summarized by child, in Table 2 (which also lists each child's over-regularization errors, for use in calculating error rates), and by verb, in Table 3.

Irregularizations

How often do children irregularize overall? The mean rate of past tense irregularizations was calculated for each child by dividing the number of errors by the number of opportunities for making such errors. The number of opportunities was defined as the number of times when the child was attempting to mark the past tense on a verb, which we operationalized as the number of times that the child succeeded in marking the past tense in any overt form whatsoever (correct forms, irregularizations, and over-regularizations, both those involving the stem such as *breaked*, and those involving the irregular past tense form such as *broked*; see Marcus *et al.* 1992). Thus the error rate was calculated as:

$$\frac{\text{Irregularized tokens}}{\text{Irregularized tokens} + \text{correct past and participle tokens} + \text{overregularized tokens}}$$

(For regular verbs, of course, there are no overregularized tokens by definition.)

Note that we did not count stem forms as examples of opportunities for the child to mismark the past tense. One practical reason was that the children's transcripts were not coded for stem forms used in past tense contexts (e.g. *Yesterday I break it*). But a more important reason is that even in a past tense context, one is never sure whether a non-marking error was caused by the child trying to mark the verb for tense and using an unmarked form as his or her guess, or whether the child was not trying to mark the verb for tense at all and just using the infinitive or stem (see Marcus *et al.* 1992, Chapter III, Section III, for extensive discussion). Of course, the effect of including stem forms in the denominator, as opportunities to make an error, would be to LOWER our estimates of the irregularization rate. Since one of our conclusions will be that such rates are lower than previously believed, excluding stem forms from the denominator makes this conclusion more robust.

Table 4 summarizes our estimates of the rates of irregularization. Children rarely irregularize. Collapsing across tokens, the mean error rate for irregular verbs was 0·0023, about two-tenths of one percent, ranging from 0 to 0·009

TABLE 3. *Summary of irregularizations*

Vowel change		
Irregular verbs	*bring*	*brang, brung, brunged, brid*
	bite	*bat, bet*
	beat	*bate*
	fight	*fooed* [fud]
	fit	*feet*
	fling	*flang*
	say	*set*
	sit	*sought*
	sleep	*slep*
	swing	*swang*
Regular verbs	*crush*	*crooshed* [kroʃt]
	jump	*janged* [ʤæŋd]
	lift	*left*
	trick	*truck*
Suffixation of *-en*		
Irregular verbs	*drink*	*dranken*
	eat	*aten*
	see	*sawn*
	shoot	*shooten, shotten*
Regular verbs	*close*	*closen*
	trip	*trippen*

TABLE 4. *Irregularization error rates*

	Irregular verbs			Regular verbs			Overall
Child	Correct & overregularized past & participle tokens	Irregularization tokens	Irregularization rate	Correct past & participle tokens	Irregularization tokens	Irregularization rate	Irregularization rate
Abe	2486	3	0·0012	2258	1	0·0004	0·0008
Adam	2649	3	0·0011	476	1	0·0021	0·0013
Eve	313	2	0·0063	52	0	0	0·0054
Mark	1558	14	0·0089	795	1	0·0013	0·0063
Naomi	527	0	0	306	0	0	0
Nathaniel	321	1	0·0031	220	0	0	0·0018
Peter	1038	0	0	347	1	0·0029	0·0007
Ross	3089	9	0·0029	1657	2	0·0012	0·0023
Sarah	2278	1	0·0004	198	0	0	00004
Total	14,259	33	0·0023	6309	6	0·0010	0·0019

across children. The mean error rate across tokens for regular verbs was 0·0010, one-tenth of one percent, ranging from 0 to 0·003 across children. The overall error rate, collapsing regular and irregular verbs, was 0·0019, less than two-tenths of one percent, ranging from 0 to 0·0063 across children.

verb roots ending in *-ing* are irregular (Pinker & Prince, 1988), but they embrace a variety of past tense forms, such as *bring-brought*, *cling-clung* and *ring-rang*. Some support for the assumption that irregularity fosters further irregularization comes from Mencken's (1936) list of dialectal variations in American English: the irregular forms are overwhelmingly substitutions of one irregular form for another (*bring-brung*, *fight-fit*, *keep-kep*), rather than substitutions of an irregular form for a regular one.

Children did not exclusively irregularize already-irregular verbs: 4 of the 14 verbs that underwent a vowel-change, and 2 of the 6 that were suffixed with *-en*, were regular. The data do contain a hint of a bias in favour of irregularizing irregulars. For each child, we compared the rate of irregularization for irregular verbs and regular verbs. The difference (0·0025 vs. 0·0009) is in the right direction, and is seen in 6 of the 8 children who irregularized, but it is not statistically significant; $t(7) = 1{\cdot}65$, $p = 0{\cdot}14$.

Are irregularizations real words? One might ask whether irregularizations result from word retrieval problems, given that many of them are real words which are irregular-sounding (e.g. *bite-bet*, *say-set*, *sit-sought*, *lift-left*). This pattern would be consistent with two assumptions about word storage: (1) each word, including irregular forms, has its own local representation, rather than being assembled out of parts as it is needed; (2) words are organized by sound, so that similar words activate each other and might get confused during retrieval. We calculated the percentage of real words among irregularizations. Only 19·4% of all irregularizations are real words. This shows that irregularizations are not exclusively, or even predominantly, word-retrieval errors. Indeed it is not even clear that the errors are biased towards being real words, because there are many ways in which a substitution of a vowel among the 180-odd irregular verbs of English could fortuitously result in some other verb. If one takes all possible vowel-changes among all the verbs that the children used in CHILDES, excluding ones for which no child ever used the stem form (see Method), one finds that 41% of them turn out to be real words. Thus real-word substitutions are, if anything, LESS frequent than one would expect by chance. Presumably the reason they are LOWER than chance would predict is that children favour some vowel substitutions over others (see Stemberger, 1993, for discussion), and this must make any comparison of substitution errors with base rates tentative. But for now, there is no support for the hypothesis that irregularizations even tend toward being whole-word retrieval errors, and such errors are by far not the most common kind.

Kinds of irregularization errors. Most discussions of productive extensions of irregular patterns assume that a pattern found in an existing family of irregulars is applied directly to a new stem. For example, the *i-a* ablaut

pattern found in *ring-rang*, *sing-sang*, *drink-drank* and so on, is applied to *bring*. This is often called a 'proportional analogy', and has played a prominent role in explanations in diachronic linguistics. Similarly, suffixation of *-en* to a stem or past form, as in *taken* or *broken*, might be applied in exactly that way to a new verb stem or past form, like *breaken*, *tooken*, or *walken*. In contrast, modern generative rule theories like that of Halle & Mohanan (1985) decompose these relations into sets of quasi-independent featural changes. For example, a single vowel-laxing rule might underlie *feed-fed*, *flee-fled*, *shoot-shot*, *say-said* and *hear-heard*, and the differences among them be effected by independent suffixation and other-vowel change rules, such as 'vowel-shift'. In PDP models such decomposition is even more extreme, and various unpredictable blends such as in *mail-membled* and *tour-tourder* can result. Thus it is of interest to see whether irregularization errors are precise proportional analogies, less direct but still lawful vowel changes, or haphazard blends and other odd forms.

There are a total of 25 different error forms in the sample (comprising 39 tokens and 20 different verbs). They break down into 18 errors involving an incorrect vowel-change (30 tokens), and 7 involving incorrect *-en* suffixation to the stem or past tense form (9 tokens). Another way of dividing them is according to their relation to existing irregular pairs. Sixteen out of the 25 error types (60%) are direct analogies of existing irregular pairs (that is, the errorful form is related to its stem exactly as some existing English past form is related to its stem (like *bring-brang* and *sing-sang*, *shoot-*(*has*) *shooten*, and *take-*(*has*) *taken*), or *drink-*(*has*) *dranken* and *break-*(*has*) *broken*). These direct analogies include (*a*) all 7 of the *-en*-suffixed forms, which invariably used the verb's stem or past form; (*b*) 7 over-applications of a simple vowel-change (*bring-brang*, *fling-flang*, *swing-swang*, *bring-brung*, *beat-bate* (cf. *eat-ate*), *sleep-slep* (cf. *feed-fed*), and *trick-truck* (cf. *stick-stuck*); (*c*) *bring-brid*, which over-applies the replacement of a final stem consonant with *d* seen in *have-had* and *make-made*; and (*d*) *sit-sought*, which replaces the rhyme of the stem with *-ought* exactly as is seen in *think*, *bring*, *fight*, *catch*, *buy*, *seek* and *teach*.[2]

Even the 9 forms that are not perfect analogies to existing irregular stem-past pairs are mostly quite close. *Say-set*, a devoiced version of the correct form *said*, presumably reflects the many irregular verbs in which the stem vowel is laxed to /ɛ/ and a /t/ is suffixed, such as *deal-dealt*, *feel-felt*, *mean-meant*, and so on. *Lift-left* is also moderately well related to these verbs, and to the *feed-fed* family where the vowel of a *t-* or *d*-final stem is laxed. *Bite-bat* and *bite-bet* also rhyme with existing irregulars (e.g. *sat* and *met*) even though their stems do not correspond to those verbs' stems. These four forms are consistent with the suggestions of Bybee & Slobin (1982) and

[2] It is possible that *sleep-slep* involves deletion of the *t* in *slept* rather than generalization of an irregular pattern.

Bybee & Moder (1983) that irregular patterns are product-oriented schemas, abstracted from the properties of the past tense forms themselves, not necessarily from an operation that would generate a past tense form from its stem in a predictable way (though recall that exact stem-past analogies were far more common).

In two more errors, a vowel underwent a change seen in an existing family and the *-d* suffix was added as well: *brunged* and *fight-fooed* [fud] (cf. *fly-flew*). These also have a ready analogical source: the various irregulars that involve a blend of a vowel change and a suffix, like *sleep-slept*, *say-said*, *flee-fled*, and *tell-told*. Slightly farther out, *jump-janged* [jæŋd] appears to blend the *-ang* rhyme seen in *sang*, *rang*, *sprang*, *sank*, *drank* and so on, with the regular suffix.

Another error, *crush-crooshed* [kroʃt], probably involves an analogical transformation of the stem itself, not an error in the past tense formation process. In colloquial English, *goosh* is a common alternative form of *gush*, and *squoosh* is an alternative to *squish*. Both in terms of sound and meaning (involving squeezing), these forms may have inspired *croosh* as an alternate version of *crush*, which then could simply have been fed into the regular rule, much as other quasi-onomatopoeic creations of children, such as Abe's *poonked*, are productively inflected with the regular pattern.

This leaves only one form that has no close counterpart among existing verbs: *fit-feet*. The inspiration is unclear, but it may have something to do with the many irregular STEMS that are similar (*meet*, *bleed*, *feed*, *lead*, *read*). Conceivably this confusion of stem and past patterns is similar to the one that led in the history of English to the backwards-ablaut seen in *fall*, *hold* and *come*.

In sum, irregularizations are generally quite closely analogized from existing irregular patterns, though the analogies are not always perfect 'proportional' ones. Both minor-rule theories, and PDP models which superimpose subregularities extracted across the entire set of verbs, are in principle compatible with this pattern, as long as their generalizations are kept close to the input patterns. Entirely absent from the data, however, are the wilder stem distortions occasionally produced by the various PDP models, such as *mail-membled*, *tour-toureder*, *smairf-sprurice*, *quiver-quess* and *satisfy-sedderded*.

The seven overapplications of the *-en* participle suffix raise the question of whether the children used *-en* systematically, as an incorrect version of the regular affix for participles. The plausibility of this hypothesis depends on how often children use participle forms correctly, a figure that has so far been hidden in our data, which pools past tense forms and participles. In order to calculate an error rate for *-en* as a proportion of participle forms, we need to estimate what proportion of our data for correct past and participle forms consists of participles alone. Adams (1938) and Smith (1935) sorted various

TABLE 5. *Doubly-suffixed verb forms*

Child		File	Age	*Stem*	*Form*	No. of doubly-suffixed	No. of correct	No. of stem+*ed*	No. of past+*ed*
Abe	Irreg	ABE045	2;10	*sweep*	*sweepted* [swiptid]	2	0	2	0
		ABE071	3;1	*break*	*brokted* [broktid]	1	38	4	12
	Reg	ABE141	3;10	*drown*	*drownded* [draondid]	1	1	0	0
		ABE176	4;4	*rain*	*rainded* [reɪndid]	2	0	0	0
Adam	Reg	ADAM34	3;7	*tie*	*tie+d+ed* [tadid]	3	7	0	0
		ADAM35	3;8	*stop*	*stop+t+ed* [stoptid]	1	14	0	0
		ADAM44	4;2	*rope*	*ropted* [roptid]	1	0	0	0
Naomi	Irreg	n20	1;10	*throw*	*throweded* [θrodid]	3	5	4	0
		n76	2;11	*tear*	*teareded* [terdid]	1	0	0	0
	Reg	n37	2;0	*crash*	*crasheded* [kræstid]	1	0	0	0
Ross	Irreg	ROSS54	4;8	*hit*	*hitdid* [hɪtdid]	1	27	1	0
	Reg	ROSS31	2;11	*need*	*neededed* [nididid]	1	11	0	0
Sarah	Irreg	SARAH129	4;10	*buy*	*buyeded* [badid]	1	18	0	0
		SARAH132	4;11	*make*	*make+t+ed* [mektid]	1	52	8	0
	Reg	SARAH130	4;11	*pick*	*pickeded* [pɪktid]	1	0	0	0
		SARAH127	4;10	*jump*	*jump+ed+ed* [ʤʌmptid]	2	1	0	0
		SARAH127	4;10	*like*	*like+ed+ed* [latktid]	1	3	0	0

verb forms by hand in the speech of children between the ages of two and five, the ages of the children in our sample. Their tables reveal that 19·8% of the total of past tense forms and past participle forms consisted of past participles. Applying this proportion to our own data, we find that the *-en* error rate was only 0·0017, less than two-tenths of one percent of the opportunities. This suggests that the children were unlikely to have been treating *-en*-suffixation as a productive rule.

Fletcher (1985) has reported data from a girl who passed through a stage in which she over-applied *-en* to irregular verbs about 15% of the time. (Zwicky, 1970, reports a similar phenomenon in his four-year-old daughter.) Fletcher suggests that the girl may have preferred *-en* for verbs with certain phonological endings, especially *t* and *d*, and though the differences among verbs were small and inconsistent, this is a hypothesis that warrants further exploration. It is possible that some children may temporarily misconstrue a common irregular suffix like *-en* as a regular one. For example, Clahsen, Rothweiler, Woest & Marcus (1992) adduce several kinds of evidence that some German children treat the nonregular plural suffix *-en* as a regular suffix. Whether this is the correct explanation for the children studied by Fletcher and Zwicky is unclear.

Doubly-suffixed past tense forms

The second major class of errors we looked for was doubly-suffixed forms, such as *tie+d+ed* and *sweepted*. The errors found are listed in Table 5. The error rates for the double-marked forms were calculated similarly to the error rate of irregularizations, and are summarized in Table 6. We found that the mean error rate was 0·0012 across all children, less than two-tenths of a percent, ranging from 0 to 0·0060. Note that this error rate is an order of magnitude lower than the rate of the Rumelhart–McClelland model (8·1%), though the fact that Rumelhart & McClelland presented data only from untrained verbs makes the two figures not directly comparable.

The Rumelhart–McClelland model produced these errors only with regular verbs ending in *p* and *k*. Pinker & Prince (1988) suggested the following explanation. The *p* and *k* verbs share most of their phonological features with *t*. Since the model's generalizations are based on overlapping phonological features, learning that the phonemes *t* and *d* should output [id] transfers to other phonemes that share a large number of features with *t* and *d*, namely *p*, *b*, *g* and *k*. The model also acquired a tendency to add *t* to *p* and *k*; thus it was liable to blend 'add *t*' with 'add [id]', resulting in double marking. However, irregular verbs never call-for [id], so the probability of blending the two patterns is much smaller.

Rumelhart & McClelland predicted that children might make such errors for these verbs too. Pinker & Prince (1988) presented some preliminary data,

TABLE 6. *Rates of double marking (DM) for individual children*

	Correct DM		Error rate	Correct DM		Error rate	Total error rate (collapsing
	irreg	irreg	irreg	reg	reg	reg	irreg & reg)
Abe	2486	3	0·0012	2258	3	0·0013	0·0013
Adam	2694	0	0	476	5	0·0104	0·0016
Eve	313	0	0	52	0	0	0
Mark	1558	0	0	795	0	0	0
Naomi	527	4	0·0075	306	1	0·0033	0·0060
Nath	321	0	0	220	0	0	0
Peter	1038	0	0	347	0	0	0
Ross	3089	1	0·0003	1657	1	0·0006	0·0004
Sarah	2278	2	0·0009	198	4	0·0198	0·0024
Total	14,259	10	0·0007	6309	14	0·0022	0·0012

but the current sample offers a better set of tests. First, we examined whether only regular verbs are doubly-suffixed. This was not so: 7 of the 17 verbs were irregular. Though regular verbs were more likely to be doubly-suffixed than irregular ones (0·0039 vs. 0·0011), the difference is not significant; $t(7) = -1{\cdot}19$, $p = 0{\cdot}27$. Four out of nine children had a higher double-suffixing rate for regulars, one child had a higher rate for irregulars, and four children did not make any double-marking errors.

Second, we determined whether regular verbs whose stems end in *p* or *k* are the only ones doubly-suffixed. Though the sample includes 8 errors with *p*/*k*-final verbs, the other 9 ended in other phonemes. Errors occurred at a slightly higher rate for *p*/*k*-final verbs (mean across children = 0·0068 vs. 0·0011), but the difference was too inconsistent across children to reach statistical significance ($t(7) = 0{\cdot}024$, $p = 0{\cdot}82$); two children out of nine had a higher error rate for regular verbs ending in *p* or *k*; three had a higher rate for other regular verbs, and four children never produced doubly-suffixed errors with regular verbs.

In sum, children rarely make double-suffixing errors in spontaneous speech, and the errors are not restricted to regular verbs ending in *p* or *k*, but occur in a variety of subclasses, including regulars ending in sonorants, fricatives, and vowels, irregulars, including no-change *t*-final ones (*hitdid*), and even a *d*-final regular resulting in the very odd form *neededed*.[3]

Why do double-marked forms occur at all? As we have seen, the blending phenomenon seen in the RM model can explain only some of the errors.

[3] *Hitdid* is unlikely to be a mis-ordering of the verb *hit* with respect to the auxiliary *did*. Stromswold (1990) analysed over 50,000 children's sentences containing auxiliaries and found no errors in which the auxiliary was improperly ordered with respect to other verbal elements.

There are two simple alternative possibilities. One is that children might occasionally misanalyse the syllabic version of the syllabic past tense form, which occurs in English exclusively with *t*/*d*-final verbs. Instead of analysing *patted* as '*pat* (stem)+*-ed* (past tense morpheme)', or (as Pinker & Prince, 1988, argue is appropriate for adults), '*pat* (stem)+[i] (inserted by phonological rule of epenthesis)+*-d* (past tense morpheme)', they analyse it as '*pat* (stem)+[-tid] (past tense morpheme)'. The suffix [-tid] is then occasionally applied productively, either as a temporary, weak alternative rule, or as a piece of a stored past tense form available for analogizing. (Similarly, *pretended* could be misanalysed as *pretend*+[dɪd], and the spurious suffix applied to form *hitdid*.) This would explain why the [tid] is applied so indiscriminately across verb stems. Another possibility is that before the child acquires the regular rule, he or she stores the regular pasts just like the irregular ones (see Marcus *et al.* 1992). Once the child acquires the regular rule, the stored form gets fed into the *-d* rule system. For instance, the child first memorizes the correct past tense form *picked* without analysing it as *pick*+*d*; when he or she acquires the regular rule, *picked* gets fed into the regular rule system, resulting in *pickeded*. (There are, however, no clear data uniquely supporting this explanation.)

DISCUSSION

We have systematically analysed a large sample of children's speech for irregularization and double-suffixing errors, and have tested a large number of hypotheses that have been proposed in the literature. There are two main discoveries.

First, children very rarely produce any kind of weird past tense forms, including over-analogized irregular patterns, addition of the participle suffix *-en*, systematic blends, or multiple applications of the regular suffix. The rates for all of these errors combined were in the order of tenths of a percentage point, and as far as we could tell from these admittedly limited data, were never stable preferences. This suggests that despite the salient patterning found among the irregulars, which both PDP and generative rule theories take such pains to capture, the predominant psychological mechanism by which children acquire the irregulars does an excellent job at simply memorizing each form verbatim and reproducing it the great majority of the time. As Marcus *et al.* (1992) point out, presumably this is simply a manifestation of children's well-documented ability to learn words (at a rate estimated to be about one every two waking hours in the preschool years; see Pinker, 1994), together with the assumption that irregulars simply ARE words, rather than the product of an on-line computation applied to the stem each time. The irregulars are idiosyncratic, but this should not especially challenge the child, because virtually all word roots are idiosyncratic in the

sense that they record a completely arbitrary pairing between a sound and a meaning, as has been obvious at least since the time of de Saussure. The textbook picture in which children forget or abandon irregulars in some kind of regularity-driven reorganization and have to re-learn them later has no motivation, either empirical or theoretical.

Second, though weird past tense forms do occur, probably in all children's speech, they are haphazard occurrences that defy most of the generalizations that have been proposed for them. They are not predominantly word-substitutions, do not occur predominantly with irregular stems, show no consistency across verbs or ages, show no obvious age trend, do not invariably conform to precise proportional analogies, and, in the case of doubly-suffixed forms like *like*+*ed*+*ed*, are not restricted to regular verbs or to verbs with any identifiable phonological pattern. This further suggests that the errors are sporadic malfunctions in a system designed to suppress them, not recurring products of the system. One can say, however, that the predominant pattern is for the errors to match closely an existing pattern exemplified among irregular pasts, indeed, usually a pattern exemplified by existing stem-past pairs. The farther a possible kind of error deviates from an existing pattern, the less likely a child is to use it, and truly bizarre distortions never occurred at all. Therefore the pattern-extracting mechanism responsible for irregularization errors is conservative in its analogizing.

The results speak against any model, rule-based or connectionist, that would allow irregular pattern-generalizing mechanisms to run free, uninhibited by the exact forms of irregular items. This would include any rule model that proposed that children are prone to overapplying minor irregular rules, and lexicon-free connectionist models like that of Plunkett & Marchman (1993), which irregularizes at the same rate that it overregularizes (unlike children, whose irregularization rate is an order of magnitude lower than their overregularization rate). It also suggests that the tendency of the Rumelhart–McClelland model (and the three-layer back-propagation enhanced version designed and tested by Sproat, 1992) to produce severe distortions like *mail-membled* or *satisfy-sedderded* is not psychologically realistic.

One theory that is compatible in a general way with the results is the rule/associative memory model discussed in Pinker & Prince (1988, 1992), Pinker (1991), and Marcus *et al.* (1992, in press). This model proposes that there is a rule of grammar, analogous to rules of syntax, that concatenates a suffix to a stem for regular past tense and participle forms. Irregular forms are stored as memorized linked pairs of lexical entries in the mental dictionary. The patterns shown across the irregulars are due to the associative nature of memory: when X is linked to Y, the properties of X are also linked to the properties of Y, so that new items similar to X (that is, sharing properties with X) have some probability of activating the properties of Y. In

other words, though *string* and *strung* are represented as separate, linked words, ensuring good verbatim retrieval, the mental representation of the pair's phonological properties overlaps in part with similar forms like *sling* and *think*, so that the learning of *slung* is rendered easier, and irregularizations like *brung* occur with nonzero probability as the result of noise or decay in the parts of the representation that code the identity of the lexical entries themselves.

Thus we do not interpret the results presented here as evidence against connectionist models in general, which may capture interesting aspects of the associative component of memory.[4] The issue raised by these data is that any model, connectionist or otherwise, will have to include resources that implement something close to the traditional linguist's notion of lexical entries for individual words. Some of the more extreme enthusiasts for connectionism deny that the mind has any representations for words (e.g. Daugherty & Seidenberg, 1992; Plunkett & Marchman, 1991, 1993), but because their models implement only the past-tense mapping, not semantics, syntax, or morphological structure, the denial is surely premature (see Kim, Marcus, Pinker, Hollander & Coppola, 1994, for discussion). Moreover, even for the past tense mapping itself, other connectionists have been forced to acknowledge the need for lexical entries, both explicitly and by their practice.

Both Rumelhart & McClelland (1986) and MacWhinney & Leinbach (1992) entertained the possibility of adding representations for lexical entries to their models, and in ways that they did not acknowledge, they had already snuck them into the models they did implement. Rumelhart & McClelland had an unusual 'blurring' scheme in which each word was represented not only by its phonological feature sequences, but also by an extra set of incorrect phonological feature sequences, unique to the word and activated for it every time the word was presented (that is, it was not stochastic noise). These 'blurred' feature sequences (many of which were phonologically impossible and thus not actually used to represent any other word's phonological composition) therefore served as surrogates for the unique identity of the particular word, just what a lexical entry is designed to capture (and helpful in distinguishing similar stems with different past forms like *spring*, *bring*, and regular *blink*). MacWhinney & Leinbach, also with little explanation, represented each word twice: as a left-justified representation of the full ordered phonological string, and as a right-justified representation of the word's rhyme. Since the patterns among irregulars are defined mainly over their rhymes, the right-justified representation is ideal for picking up on these subregularities to support similarity-based generalization, while the left-justified full representation is then available to code idiosyncratic lexical differences, a surrogate for the traditional notion of a verb's lexical entry as the stored locus of its idiosyncratic information.

[4] See the papers by Pinker, Prince, Marcus and their collaborators for critical discussion of such models; their arguments hinge exclusively on the circumstances of application of the REGULAR pattern, which we have not discussed here.

In sum, the challenge presented by irregularizations is to explain the occurrence of rare, and only quasi-systematic, pattern generalizations in the context of extremely accurate overall reproduction of memorized word-specific forms. We suspect that any model, whether it emphasizes rules or associations, will be able to handle this challenge by positing some kind of pattern-associator memory only if that mechanism is subordinate to representations specific to individual words.

REFERENCES

Adams, S. (1938). Analysis of verb forms in the speech of young children, and their relation to the language learning process. *Journal of Experimental Education* **7**, 141–4.

Berko, J. (1958). The child's learning of English morphology. *Word* **14**, 150–77.

Bowerman, M. (1982). Reorganizational process in lexical and syntactic development. In E. Wanner & L. R. Gleitman (eds), *Language acquisition: the state of the art*. New York: Cambridge University Press.

Brown, R. & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (ed.), *Cognition and the development of language*. New York: Wiley.

Bybee, J. L. (1985). *Morphology: a study of the relation between meaning and form*. Philadelphia: Benjamins.

Bybee, J. L. & Moder, C. L. (1983). Morphological classes as natural categories. *Language* **59**, 251–70.

Bybee, J. L. & Slobin, D. I. (1982). Rules and schemas in the development and use of the English past tense. *Language* **58**, 265–89.

Cazden, C. B. (1966). The acquisition of noun and verb inflections. *Child Development* **39**, 433–48.

Chamberlain, A. F. (1906). Preterite forms, etc., in the language of English-speaking children. *Modern Language Notes* **21**, 42–4.

Clahsen, H., Rothweiler, M., Woest, A. & Marcus, G. F. (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition* **45**, 225–55.

Daugherty, K. & Seidenberg, M. (1992). Rules or connections? The past tense revisited. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Dell, G. S. & Reich, P. (1980). Toward a unified model of slips of the tongue. In V. Fromkin (ed.), *Errors in linguistic performance: slips of the tongue, ear, pen, and hand*. New York: Academic Press.

Fletcher, P. (1985). *A child's learning of English*. Oxford: Blackwell.

Haber, L. (1975). Muzzy theory. *Proceedings of the 11th Regional Meeting of the Chicago Linguistics Society*. Chicago: University of Chicago.

Halle, M. & Mohanan, K. P. (1985). Segmental phonology of modern English. *Linguistic Inquiry* **16**, 57–116.

Kim, J. J., Marcus, G. F., Pinker, S., Hollander, M. & Coppola, M. (1994). Sensitivity of children's inflection to morphological structure. *Journal of Child Language* **21**, 173–209.

Kiparsky, P. (1982). Lexical phonology and morphology. In I. S. Yang (ed.), *Linguistics in the morning calm*. Seoul: Hansin.

Kiparsky, P. & Menn, L. (1977). On the acquisition of phonology. In Macnamara, J. (ed.), *Language learning and thought*. New York: Academic Press.

Kuczaj, S. A. (1978). Children's judgements of grammatical and ungrammatical irregular past tense verbs. *Child Development* **49**, 319–26.

Ling, C. X. & Marinov, M. S. (1993). Answering the connectionist challenge: a symbolic model of learning the past tenses of English verbs. *Cognition* **49**, 235–90.
MacWhinney, B. & Leinbach, J. (1991). Implementations are not conceptualizations: revising the verb learning model. *Cognition* **40**, 121–57.
MacWhinney, B. & Snow, C. (1985). The Child Language Data Exchange System. *Journal of Child Language* **12**, 271–96.
—— & —— (1990). The Child Language Data Exchange System: an update. *Journal of Child Language* **17**, 457–72.
Marchman, V. (1988). Rules and regularities in the acquisition of the English past tense. *Center for Research on Language Newsletter*. University of California, San Diego, 2(4).
Marcus, G. F. (1993). Negative evidence in language acquisition. *Cognition* **46**, 53–85.
—— (in press). U-shaped language development in children and connectionist networks. *Cognition*.
Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J. & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development* **57** (4, Serial No. 228).
Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R. & Pinker, S. (in press). German inflection: the exception that proves the rule. *Cognitive Psychology*
Mencken, H. L. (1936). *The American language*. New York: Knopf.
Pinker, S. (1984). *The language instinct*. New York: Morrow.
—— (1989). *Learnability and cognition*. Cambridge, MA: MIT Press.
—— (1991). Rules of language. *Science* **253**, 530–5.
—— (1994). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
Pinker, S. & Prince, A. (1988). On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition* **28**, 73–193.
Plunkett, K. & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: implications for child language acquisition. *Cognition* **38**, 43–102.
—— & —— (1993). From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition* **48**, 21–69.
Prasada, S. & Pinker, S. (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes* **8**, 1–56.
Rumelhart, D. E. & McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland, D. E. Rumelhart & The PDP Research Group, (eds), *Parallel distributed processing : explorations in the microstructure of cognition*. Vol. 2: *Psychological and biological models*. Cambridge, MA: Bradford Books/MIT Press.
Smith, M. E. (1935). A study of some factors influencing the development of the sentence in preschool children. *Journal of Genetic Psychology* **46**, 182–212.
Sproat, R. (1992). *Morphology and computation*. Cambridge, MA: MIT Press.
Stemberger, J. P. (1983). Inflectional malapropisms: form-based errors in English morphology. *Linguistics* **21**, 573–602.
—— (1993). Vowel dominance in overregularizations. *Journal of Child Language* **20**, 503–21.
Stromswold, K. (1990). Learnability and the acquisition of auxiliaries. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Department of Brain and Cognitive Sciences.
Zwicky, A. (1970). A double regularity in the acquisition of English verb morphology. *Papers in Linguistics* **3**, 411–18.