

Marcus, G. F. (1996) Why do children say “brea**k**ed”? *Current Directions in Psychological Science*, **5**, 81-85.

Why do children say “break**ed”?**

Gary F. Marcus

Revised: January 30, 1996

CD 95-10

Errors can yield special insight into learning mechanisms. In language development, perhaps the most notorious error is the past tense overregularization. Most English verbs form their past tense regularly, by adding the suffix *-ed*, e.g., *walk-walked*. About 180 verbs, though, form their past tense idiosyncratically, e.g., *sing-sang*. Overregularizations result when the regular *-ed* suffix is applied to an irregular verb, e.g., *singed*.

Because parents almost never overregularize, these errors demonstrate that the language learning involves more than mere imitation. Instead, children must possess mechanisms that detect and extend linguistic generalizations.

Although the production of overregularizations has typically been ascribed to the application of a mental rule, the mere fact that the regular pattern has been overextended doesn't guarantee that overregularization errors are produced by a rule. Instead, as Rumelhart and McClelland showed, a single uniform neural network that contains no explicit rules and makes no explicit distinction between regular and irregular words can produce overregularizations.¹

In 1988, Pinker and Prince pointed out several limitations to Rumelhart and McClelland's model.² Since then characterizing the mechanisms responsible for overregularization has become a central focus of detailed empirical comparisons between symbolic, rule-based models and connectionist neural network models that explicitly forsake rules in favor of networks of connections between nodes.

THE RULE AND MEMORY MODEL

A model of overregularization that my colleagues and I have proposed depends on the existence of mental rules. On this model, children's grammars and adult's grammars are structured as similarly as possible. The model has three simple components.

First, speakers have access to a symbolic, default rule, saying roughly, 'to form the past tense, add *-ed* to any word carrying the symbol Verb.' What makes this a rule is that it can apply to any verb, regardless of its resemblance to stored examples. What makes it a default is that it applies anytime access to the lexicon fails, that is, to any word that lacks a stored past tense form, including low-frequency words (*sнарfed*), unusual sounding words (*ploamphed*), and complex words that are not treated as roots, such verbs derived from nouns (*The soldiers ringed the city*). Children readily generalize the *-ed* inflection to nonsense words like *wug* and to novel denominal verbs (e.g., *ring* meaning to put a ring on a finger) with *-ed*.³

Second, past tense forms of irregular roots are stored in memory. Since all memory is fallible, memory for irregulars is imperfect.

Third, the irregular lexicon always takes precedence over the rule, hence *sang* blocks *singed*. The default applies if and only if no irregular inflected form can be found.

When memory for irregulars is taxed, even adults may overregularize. For instance, many adults would overregularize *strive* as *strived*, because the irregular form *strove* is rare and hence difficult to retrieve. Upon retrieval failure, the default rule steps into the breach, yielding an overregularization.

The rule-and-memory model holds that children's grammars are similarly structured. But because children have had less exposure than adult to correct forms, their memories for irregulars are weaker. Whenever access to an irregular past tense form fails, the child adds *-ed*, producing an overregularization.

To test this model, using data from a publicly available archive of children's spontaneous speech, CHILDES⁴, my colleagues and I conducted a systematic, quantitative analysis of children's overregularizations, extracting and analyzing 11,521 past tense utterances from the spontaneous speech of 83 children.⁵

Several observations support the rule-and-memory model. First -- contrary to popular opinion reported everywhere from *Newsweek* to the primary literature -- there is no stage in which children completely replace correct forms with overregularizations. Instead, we discovered that children overregularize in only about 4% of their opportunities, demonstrating a systematic preference for correct irregulars. Thus, errors appear to be a consequence of performance limitations rather than a qualitatively inaccurate grammar. Figure 1, representative longitudinal plots of three children, show that low rates of overregularization are not artifacts of averaging over time; these low rates are also consistent across children and hold for most individual verbs.

< Please Insert Figure 1 >

If overregularizations are the consequence of retrieval failure, verbs that are more difficult to remember should be more likely to be overregularized. To test this, we used parental frequencies of irregular verbs as an index of retrievability. As predicted, frequency was negatively correlated with overregularization - the more often a parent used a verb, the less likely the child was to overregularize it. The correlation is in the predicted direction for 18 of 19 children's overregularization rates⁶, mean = -0.34.

Verbs that are reinforced by families of similar-sounding irregulars should be easier to remember. Thus *sing*'s past tense *sang* should be easier to remember because of the presence of similar-sounding verbs that take similar patterns, such as *ring-rang*, and *drink-drank*. Indeed, verbs with greater irregular cluster strength (as measured by the number and frequency of similar neighbors) were less prone to overregularization. The correlation is in the predicted direction for 16 of 19 children's overregularization rates, mean = -0.08, even when frequency is partialled out.

If these errors disappear as a consequence of improving facility at retrieving correct forms, rather than qualitative grammatical change, then they should disappear gradually. Indeed, the sample of preschoolers that we studied overregularized at a rate of 4.2%; a sample of first graders at 2.5%,

fourth-graders 1%, and even adults overregularize occasionally, about once in every 25,000 opportunities.

The “add -ed” rule, because it is English-specific, cannot be innate. Before learning it, retrieval failures should be left unmarked (*I sing yesterday*). These errors should be replaced by overregularizations only once the English-specific default -ed is acquired. As predicted, before children reliably inflect regular verbs for past tense in contexts that require it (e.g, *I walked yesterday*), irregular verbs are used correctly or left unmarked, but never overregularized. Figure 1 shows that the period of overregularization appears to coincide with the development of reliable regular past tense marking in contexts which demand regular past tense inflection. Rates of regular tense marking are significantly greater during the period of overregularization than before (Adam: 73% vs. 8%; Sarah 85% vs. 44%; Eve 66% vs. 11%), suggesting that overregularization is tied to the acquisition of a rule.⁷

PARALLEL DISTRIBUTED PROCESSING

Consider now the connectionist alternative. In single uniform networks such as those proposed by Rumelhart and McClelland and Plunkett and Marchman, words are represented as distributed patterns of activation over a network of nodes and connections. Simplifying somewhat, *sing* would be represented as a set of units representing the initial consonant, the vowel nucleus, and the final consonant. The network is presented with pairs of stem and past tense forms. Learning the mapping from *sing* to *sang* involves strengthening connections that run from the input nodes that represent the sound *ing* to the output nodes that represent the sound *ang*. Similar-sounding words overlap in their representations, yielding an explanation for why adults sometimes inflect the novel word *spling* as *splang*.⁸

Regular inflection is treated identically: exposure to *walk-walked* strengthens the connections between input nodes representing *alk* and output nodes representing *alked*, generalizing to similar words like *talk-talked*.

Overregularization are produced when regular patterns exert too strong an attraction on irregular verbs. For instance *grow* might be overregularized to *grewed*, because words like *glow-glowed* increase the strength of the connection between nodes representing the input *ow* and output *owed*. As the network hears more such pairs, the chance of overregularization increases. Thus, “the level of generalizations ... [is] closely related to the total number [and proportion] of regular verbs in the vocabulary.”⁹ This relationship might be dubbed “the regular attraction hypothesis” and it is central to most connectionist models of inflection.¹⁰

But while the network’s overregularization is driven primarily by attraction to stored regular stem-past pairs, four types of evidence suggest that children’s overregularizations may not be.

Longitudinal test

In network models, overregularization occurs during sudden shifts in the proportion of vocabulary that is regular.¹¹ Contrary to the network prediction, the child's overregularization appears to be independent of changes in the composition of the vocabulary. Quantitatively, there is no positive correlation between increases in the number or proportion of regular verbs acquired (whether measured in types or tokens) by the child and the child's rate of overregularization. Instead, Figure 2 shows that the proportion of regular verbs in a child's vocabulary increases less rapidly during the period of overregularization than before -- precisely the opposite of the regular attraction hypothesis.¹² Thus, the onset of overregularization thus does not depend on dramatic increases in the proportion of vocabulary that is regular.

< Please Insert Figure 2 >

Lexical test

The regular attraction hypothesis predicts that irregular verbs will be more frequently overregularized if they are similar to regular verbs; thus *feel* should be drawn to *feeled* by similar neighbors such as *heal-healed* and *peel-peeled*. To test this hypothesis, we calculated a verbwise measure of regular cluster strength, based on each verb's number (and frequency) of similar-sounding regular neighbors. There was no correlation between regular cluster strength and overregularization: verbs with many regular neighbors were no more likely to be overregularized than verbs with few regular neighbors, suggesting that overregularizations are not the result of an attraction to regular verbs.

Comparison between noun plurals and verbs

The English noun plural system contains a greater proportion of regular words (excluding learned Latinate plurals like *bacterium-bacteria*, over 98% of noun plural types are irregular) than the past tense (only 86% of the 1000 most frequent types are regular). But, like *-ed*, the *-s* plural is generalized whenever access to memory for irregular roots fails, including plurals derived from names (*We had a fabulous meal at the Childs*), onomatopoeia (*The cartoon climax had seven POWs*), and unusual-sounding words (*Kiosks*). Further evidence that regular inflection is generalized independently of type frequency comes from the fact that individual children's rates of overregularizing the plural do not differ significantly from their rates of overregularizing the past tense.¹³

Cross-linguistic test

If generalization of regular inflection is truly independent of type frequency, speakers should generalize regular inflection in the same way even when there are far fewer regular words. One test case is the German *-s* plural, which applies to fewer than 10% of German nouns (whether

measured by types or tokens) and is just one of five different suffixes, $-\emptyset$, $-e$, $-er$, $-(e)n$, $-s$. Yet the infrequent $-s$ plural behaves as the default. Just as the English plural $-s$ applies when access to irregulars is blocked, as with names, borrowings, and onomatopoeia, so too does the German $-s$ plural, applying to names (*Thomas Manns*), borrowings (*Kiosks*), onomatopoeia (*Wau-waus*).¹⁴

German-speaking children often overregularize with $-s$, more often than might be predicted by a frequency-based approach¹⁵, and children are more likely to generalize the $-s$ pattern than irregular patterns to novel words that are unusual sounding or presented as names.¹⁶ These facts strengthen the conclusion that generalization of the default does not require high regular type frequency.

CONCLUSIONS

Overregularization errors do not correspond with increases in the proportion of vocabulary that is regular, are not more common for irregular words which resemble other regular verbs, are not more common in inflectional systems that have a higher proportion of regular words, and occur even in languages in which the regular pattern is infrequent. Rather than generalizing regular inflection by analogy to stored exemplars, children generalize default inflection by a rule which applies whenever access to irregulars fails.

Default inflection seems to be a natural property of many linguistic systems, readily learned by children. Although the input to children varies widely, English and German-speaking children rapidly converge on the same sorts of inflectional systems. The English past tense, English noun plural (almost entirely regular) and German noun plural system (almost entirely irregular) appear to develop similarly. Because single uniform neural networks tend to be closely tied to the input, they have difficulty explaining why children develop similar linguistic systems despite differing input conditions. Instead, generalization of regular inflection seems to be best explained in terms of a rule.

The key property of a rule appears to be its ability to treat all instances of a class equally, regardless of their degree of resemblance to stored forms. Rules don't generalize in a gradient of similarity but rather apply in an all-or-none fashion applying to any item carrying the appropriate symbol. For example, the add $-ed$ rule applies just as readily to any novel word carrying the symbol [verb]. This mode of generalization, which is driven not by resemblance but by the presence of a symbol, may be central to other domains of cognition.

For example, this ability to suppress resemblance and treat all tokens of a class equally is to be central to word recognition. The phenomenon of categorical speech perception reflects a mechanism that allows us to treat many different examples of a given phoneme equally, ignoring whether a given token of the phoneme is more or less similar to previously heard examples (because of variability in pitch, accent, and voice-onset timing).

Similarly, we can identify someone as a grandmother regardless of whether she has gray hair or cooks chicken soup; likewise, we can override perceptual similarity to identify a raccoon as a raccoon, even if it has been surgically transformed to look like a skunk.¹⁷ There is little doubt that the mind can track information about similarity and resemblance (and calculate that Priscilla Presley is not a typical grandmother), but the ability to override information about resemblance and treat all members of a class *identically* also appears to be an essential property of cognition -- precisely the work done by rules that manipulates symbols.

A simple model, that depends on the existence of such symbolic rules, explains a wide range of data: Irregular forms are retrieved from memory and block the application of the default regular rule (add *-ed*); if a child fails to retrieve the irregular form, the regular rule applies by default, and the child produces an overregularization.

Acknowledgments

I thank my collaborators S. Bartke, U. Brinkmann, H. Clahsen, M. Coppola, M. Hollander, J. J. Kim, T. J. Rosen., M. Ullman, R. Wiese, F. Xu, and especially, S. Pinker. For helpful comments on an earlier draft I thank Paul Bloom. The research described in this article was supported by an NDSE Graduate Fellowship and University of Massachusetts Faculty Research Grant to Marcus, NIH 18381 and NSF BNS 91-09766 to Steven Pinker, and an ACLS/DAAD collaborative research grant to Clahsen, Pinker, and Marcus.

Gary F. Marcus is an Assistant Professor of Psychology at the University of Massachusetts, Amherst. Address correspondence to Gary Marcus, Department of Psychology, University of Massachusetts, Amherst MA 01003, or by e-mail: marcus@psych.umass.edu.

¹D. Rumelhart and J. McClelland, On learning the past tenses of English verbs. Implicit rules or parallel distributed processing? In: J. McClelland, D. Rumelhart and the PDP research group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. (MIT Press: Cambridge, MA, 1986); for a more recent model, see K. Plunkett and V. Marchman, From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition*, **48**, 21-69 (1993).

²S. Pinker, and A. Prince, On language and connectionism: Analysis of a Parallel Distributed Processing model of language acquisition. *Cognition* **28**, 73-193 (1988).

³Wug: J. Berko, The child's learning of English morphology. *Word 14*: 150-177 (1958).

Denominal verbs: J. J. Kim, G. F. Marcus, S. Pinker, M. Hollander, and M. Coppola, Sensitivity of children's inflection to grammatical structure, *Journal of Child Language*, *21*, 173-209 (1994).

Distinctions between regular and irregular morphology: S. Pinker, Rules of Language, *Science*, *253*, 530-555 (1991).

⁴B. MacWhinney and C. Snow, The Child Language Data Exchange System, *Journal of Child Language*, *12*, 271-296 (1985).

⁵G. F. Marcus, S. Pinker, M. Ullman, M. Hollander, T. J.. Rosen, and F. Xu, *Overregularization in Language Acquisition*. (*Monographs of the Society for Research in Child Development*), *57* (4, Serial No. 228) (1992).

⁶Lexical correlations could be tested for only 19 children, since the remainder lacked individual databases or never overregularized in our samples.

⁷Marcus, et al., 1992, p 104.

⁸S. Prasada, and S. Pinker, Similarity-based and rule-based generalizations in inflectional morphology. *Language and Cognitive Processes*. **8**, 1-56 (1993).

⁹Plunkett & Marchman, note 1.

¹⁰G. F. Marcus, U. Brinkmann, H., Clahsen, R. Wiese, and S. Pinker, German inflection: The exception that proves the rule, *Cognitive Psychology* **29**, 189-256 (1995).

¹¹The model of Plunkett & Marchman (1993) depends on a similar, abrupt external change, in which the onset of overregularization coincides with a sudden shift in training regime from a slow pace that allows the network to learn each irregular verb individually to a rapid pace that forces the network to generalize. See G. F. Marcus, The acquisition of inflection in children and multilayered connectionist networks, *Cognition.*, **56**, 271-279 (1995).

¹²Marcus, et al. (1992), Chapter V, using novel sampling techniques from biostatistics, have shown that this is not an artifact of a cumulative vocabulary measure that necessarily decelerates over time.

¹³G. F. Marcus, Children's overregularization of English plurals: a quantitative analysis, *Journal of Child Language* **22**, 447-459 (1995).

¹⁴See Marcus, et al., note 10.

¹⁵The only systematic study of how input frequency affects network generalization is a series of simulations conducted by Plunkett and Marchman, 1993. In this study, when the initial proportion of regular vocabulary was less than 50%, their simulation failed to generalize the regular pattern to "novel indeterminate stems".

¹⁶S. Bartke, G. F. Marcus, and H. Clahsen, Acquiring German Noun Plurals. In D. MacLaughlin & S. McEwan (eds.) *Proceedings of the 19th Annual Boston University Conference on Language Development*. Boston: Cascadilla Press (1995).

¹⁷F. C. Keil (1989), *Concepts, Kinds, and Cognitive Development.*, Cambridge, MA: Bradford Books. G. F. Marcus, Children's Overregularization and Its Implications for Cognition, in P. Broeder and J. M. J. Murre (Eds.), *Models of Language Acquisition: Inductive and Deductive Approaches*. Cambridge, MA: MIT Press, 1996.

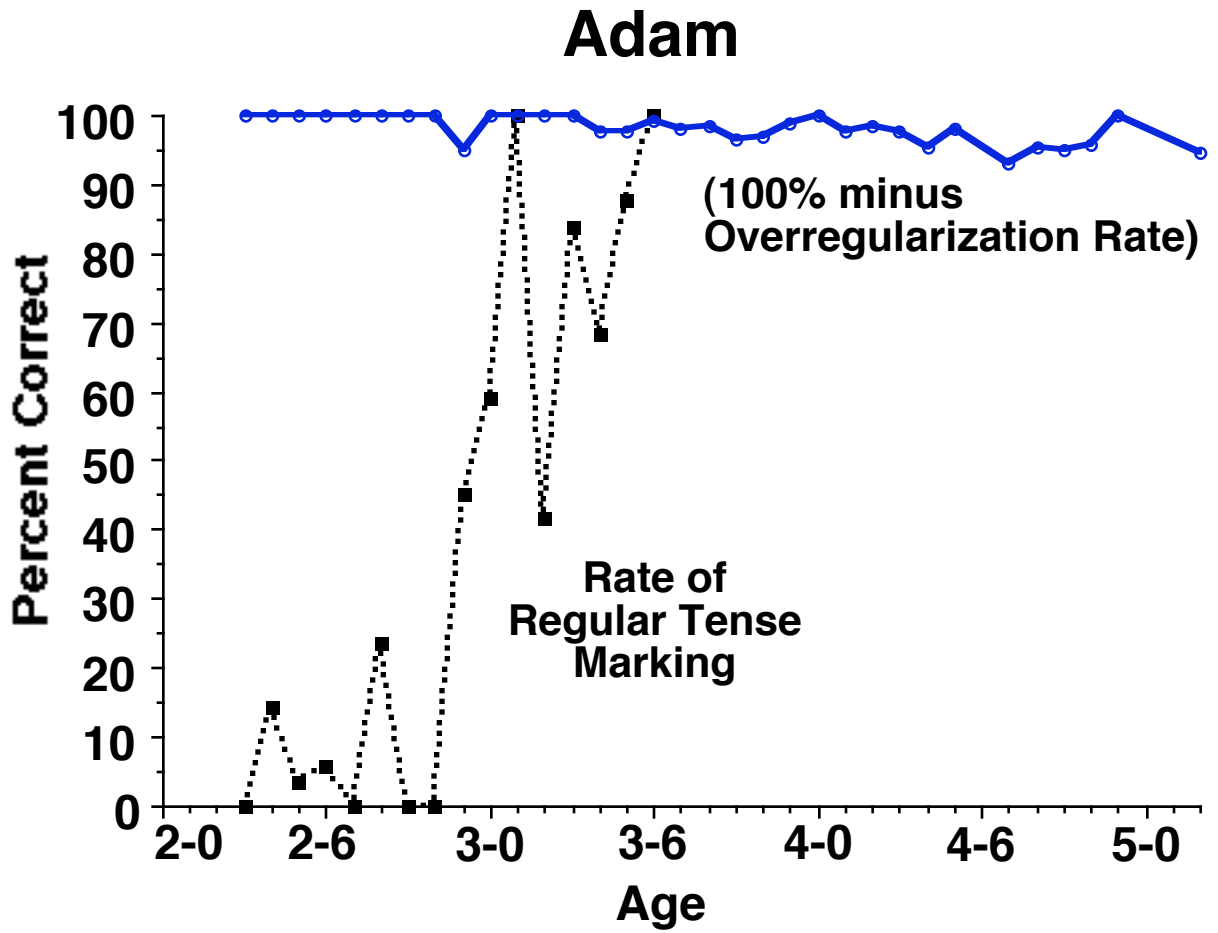


Figure 1a

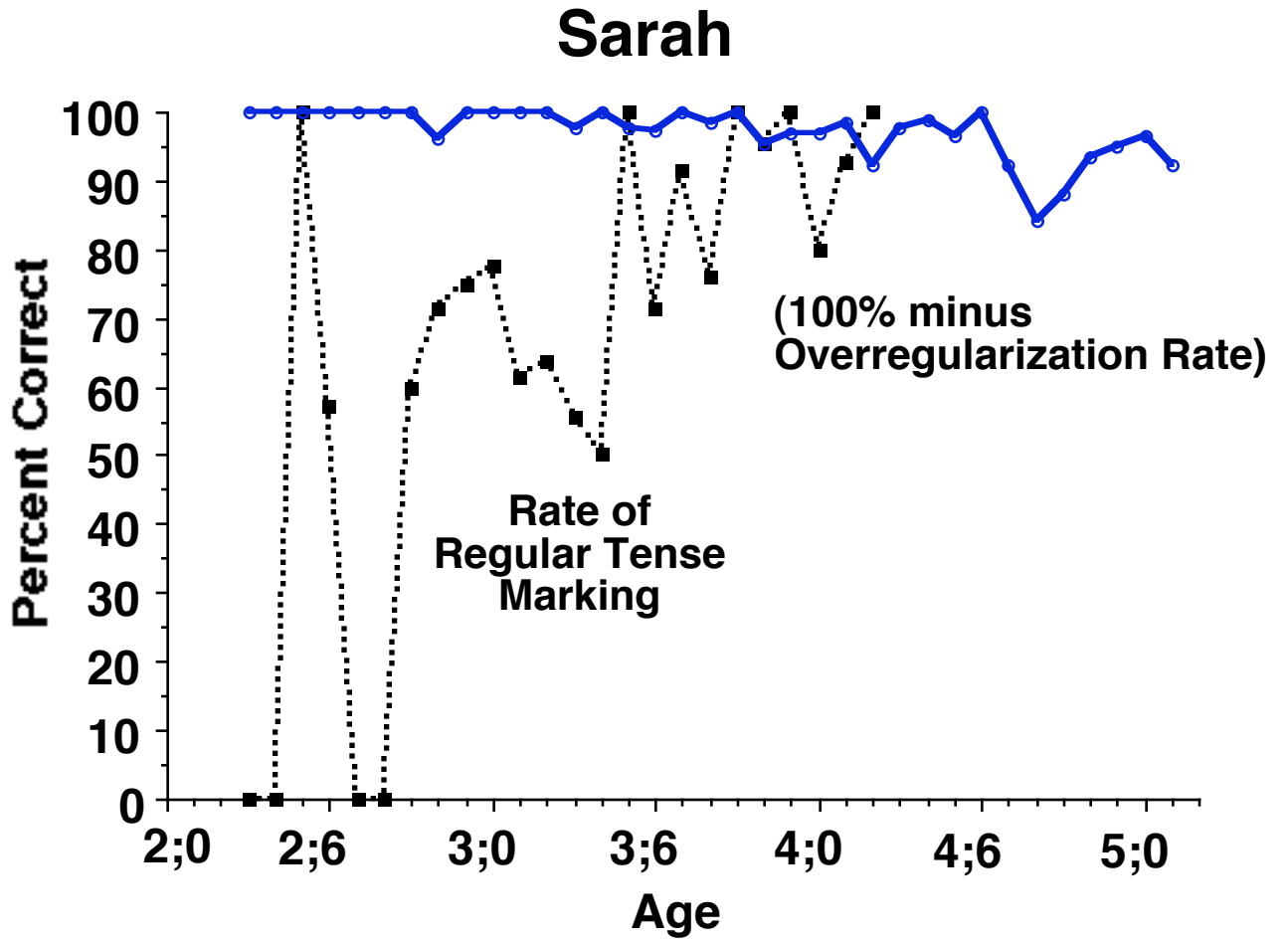


Figure 1b

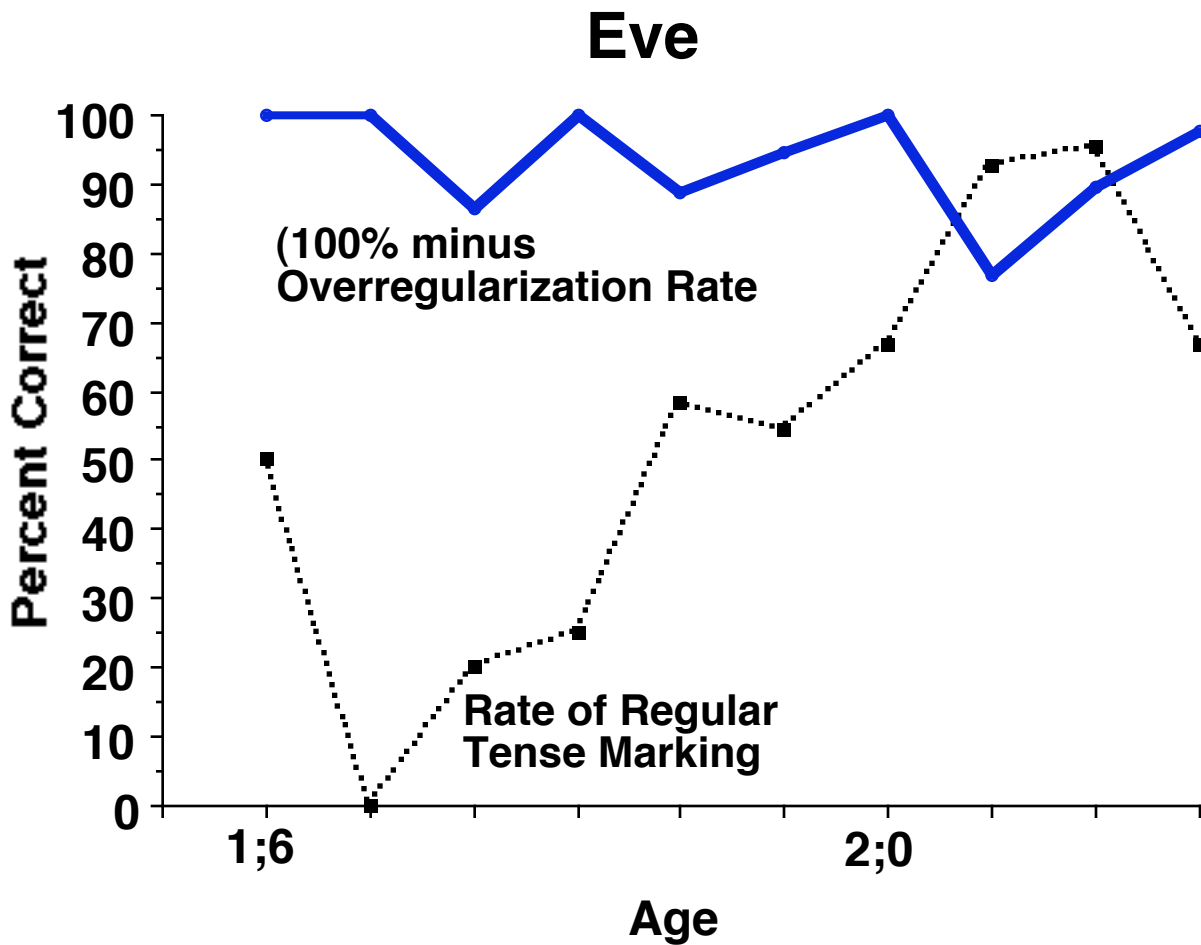


Figure 1c

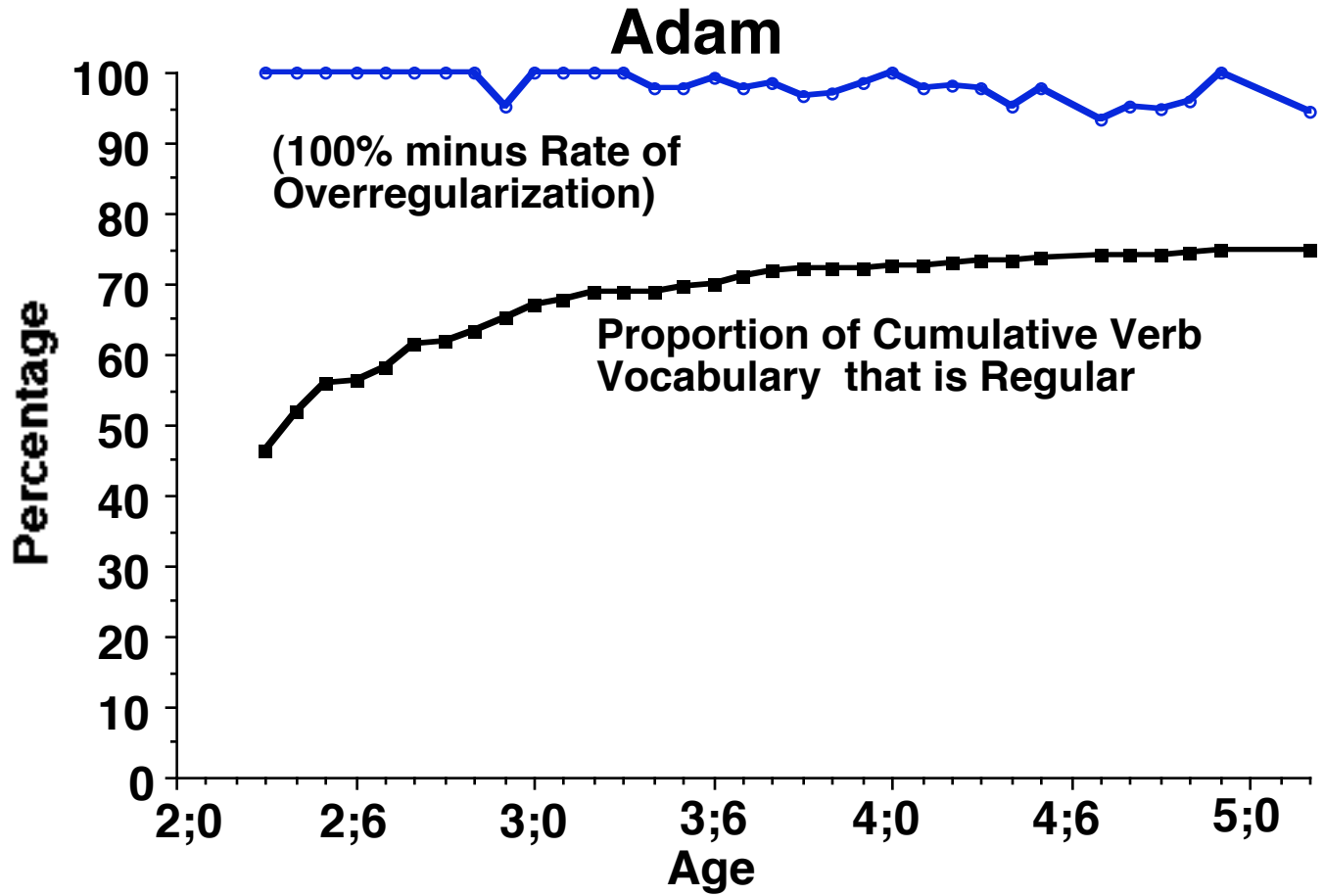


Figure 2a

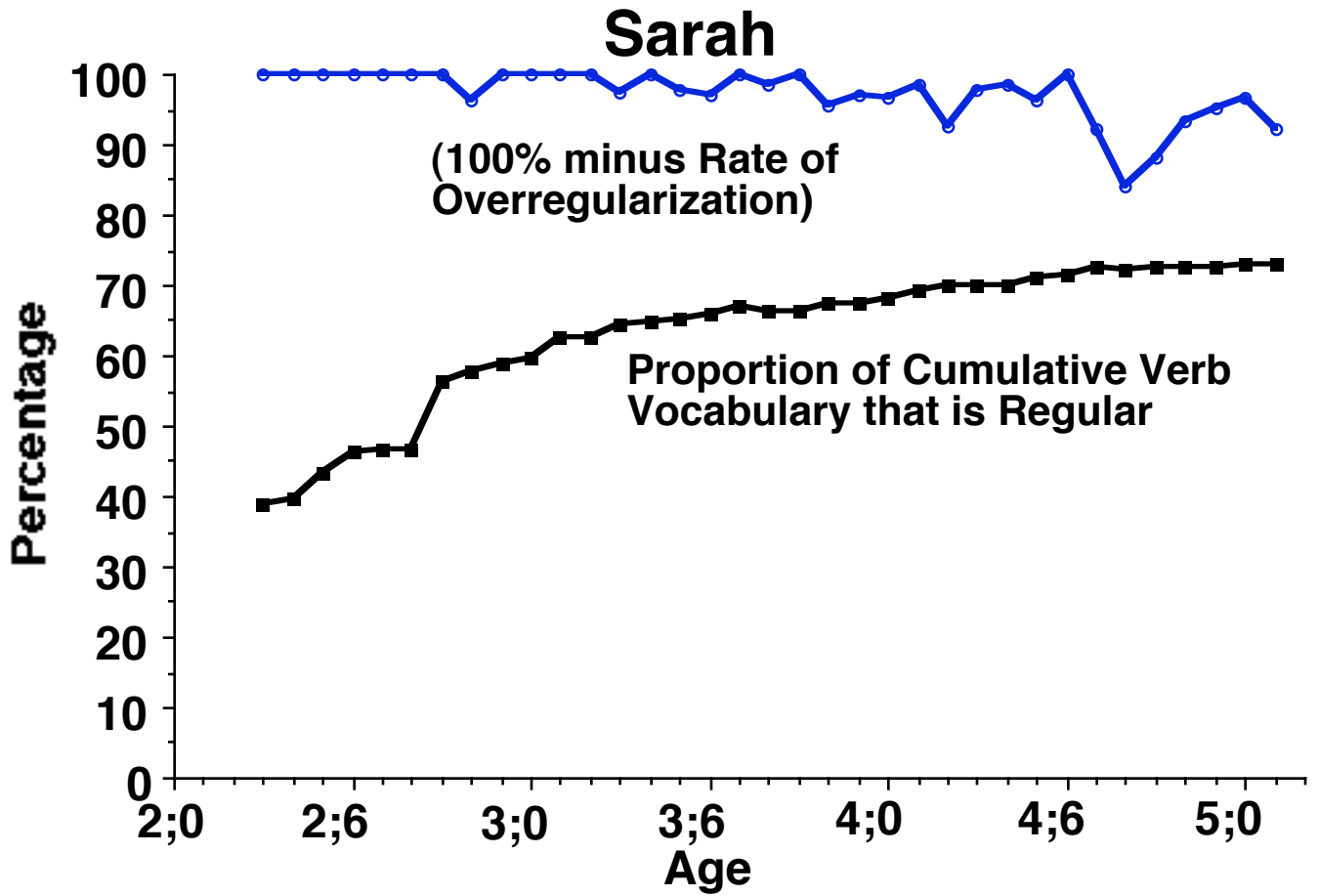


Figure 2b

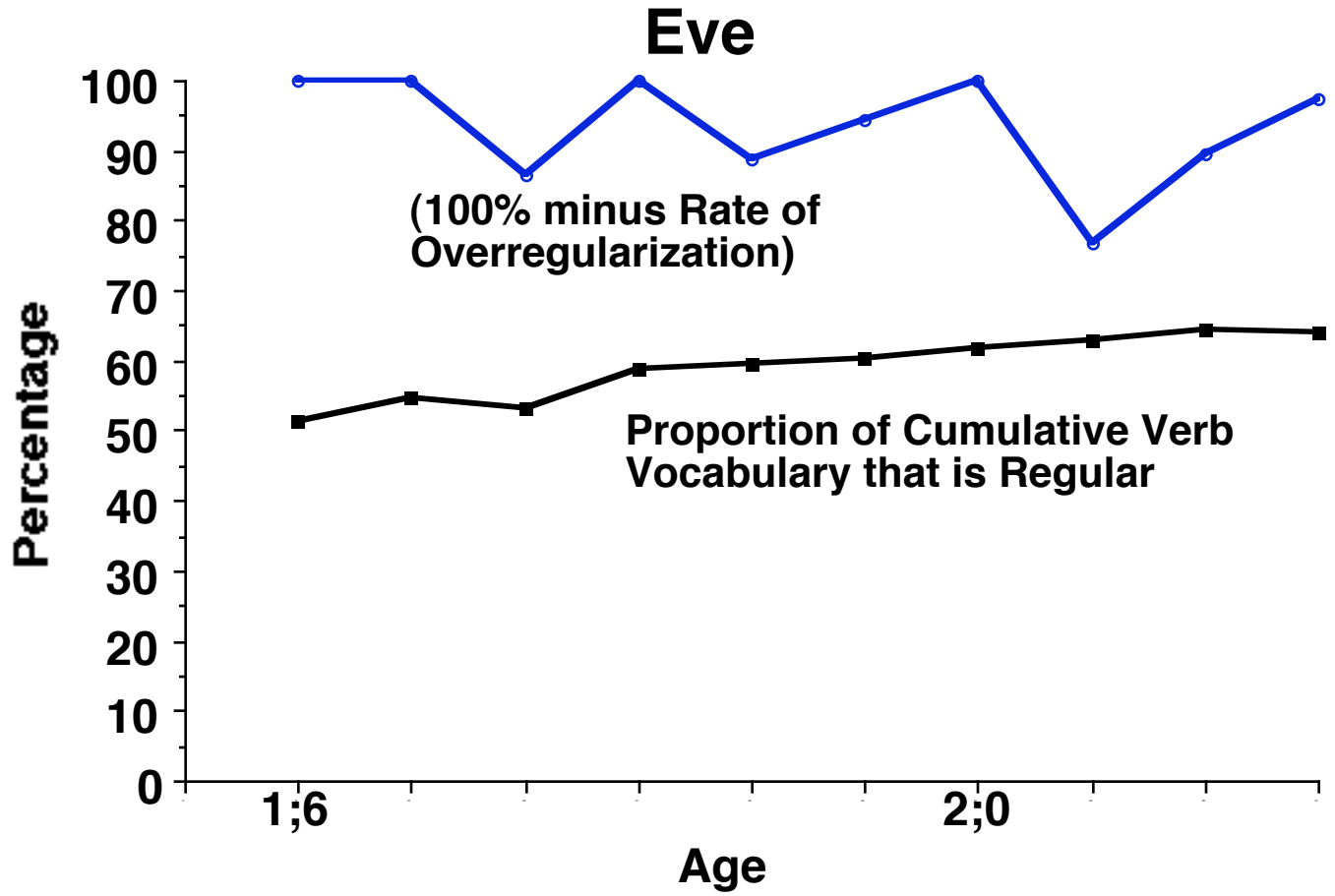


Figure 2c

Caption for Figure 1(a-c): A child's rate of overregularization subtracted from one, and the child's rate of supplying the regular past tense in contexts which required past tense marking.

Caption for Figure 2(a-c): A child's rate of overregularization subtracted from one, and the proportion of the child's cumulative verb vocabulary that is regular.

Possible layout suggestion on the following two pages.