

Rules or Connections in Past-Tense Inflections: What does the Evidence Rule Out?

James L. McClelland
Center for the Neural Basis of Cognition
and Department of Psychology
Carnegie Mellon University
Pittsburgh, PA

Karalyn Patterson
Medical Research Council
Cognition and Brain Sciences Unit
Cambridge, UK

Abstract

Pinker and colleagues propose two mechanisms—a rule system and a lexical memory—to form past tenses and other inflections. They predict that children’s acquisition of the regular inflection is sudden; that the regular inflection applies uniformly regardless of phonological, semantic, or other factors; and that the rule system is separably vulnerable to disruption. A connectionist account makes the opposite predictions. Pinker has taken existing evidence as support for his theory, but the review presented here contradicts this assessment. Instead, the evidence supports all three connectionist predictions: gradual acquisition of the past tense inflection; graded sensitivity to phonological and semantic content; and a single, integrated mechanism for regular and irregular forms dependent jointly on phonology and semantics.

Teaser: A dual mechanism account of language processing make several predictions about the development, use, and disintegration of inflectional morphology; however, the evidence supports a single-system connectionist approach.

Key Words: Rules, Inflectional Morphology, Past Tense, Connectionist Models, Language Processing, Parallel-Distributed Processing

One view of language, originating with Chomsky [1,2], championed by Fodor and Pylyshyn [3] and widely pursued by Pinker [4-7], holds that abstract symbolic rules play a central role in the human language processing. This claim is part of a broader view that human cognitive mechanisms are symbolic, modular, innate, and domain-specific[4]. An alternative view, from Rumelhart and McClelland [8—See Box], challenges the need for the use of rules. This view arises within the Parallel Distributed Processing (PDP) or connectionist framework [9], in which cognitive processes are seen as graded, probabilistic, interactive, context-sensitive, and domain-general. Acquisition of language and other abilities occurs via gradual adjustment of the connections among simple processing units. Characterizations of performance as ‘rule-governed’ are viewed as approximate descriptions of patterns of language use; no actual rules operate in the processing of language.

These perspectives apply to many aspects of language, and, as Pinker and Ullman [10] suggest, to many other domains as well, but here we focus on inflectional morphology, especially the English past tense. The idea of a past tense rule arose from noting that young children sometimes regularize irregular verbs, producing ‘goed’ or ‘felled’ [11], and from the finding that children (and adults) typically produce regular forms for nonce (novel) words in a past-tense elicitation task [12]. Given a picture of a man said to be ‘ricking’ and a request to complete ‘Yesterday he ___’, the response is usually ‘ricked’. Since the child would never have heard ‘goed’ or ‘ricked’, such responses were thought to show use of a rule.

We address a specific notion of rules held by Pinker and his collaborators, in which rules are discrete, categorical, and symbolic objects used in a specialized, innate language module. For the English past tense, the rule takes as its argument any item identified only as a verb stem, and produces as its output its regular past tense. In English the output is stem + [d] (subsequent machinery realizes [d] as /d/, /t/ or /ɪd/, as in ‘loved’, ‘liked’ or ‘hated’, depending only on the stem-final phoneme). The rule is said to be uniform in its application and independent of the meaning, phonology, frequency of occurrence, or any other attribute of the verb stem to which it applies. A further characteristic often attributed to such rules is that their acquisition is sud-

Preparation of this article was supported by MH 47566, MH 64445 and by the Medical Research Council. We thank David Plaut, Michael Ramscar, and Mark Seidenberg for discussion and comments. Correspondence regarding this article may be sent to James McClelland (jlm@cnbc.cmu.edu), CNBC, 15 Mellon Institute, 4400 Forbes Avenue, Pittsburgh, PA, 15213 or to Karalyn Patterson (karalyn.patterson@mrc-cbu.cam.ac.uk), MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, UK, CB2 2EF.

Table 1
Predicted and Observed Aspects of Regular Inflection

Aspect	Prediction from		Observed
	Symbolic Rules	PDP Models	
Acquisition	sudden	gradual	gradual
Sensitivity:			
to phonology	no	yes	yes
to semantics	no	yes	yes
in development	no	yes	yes
in German plu.	no	yes	yes
Separability:			
genetically	yes	no	no
neurologically	yes	no	no

den. Thus Pinker [5] suggests that the child “deduces” the rule (p. 193), calling this an “epiphany” (p. 194), and a “‘Eureka’ moment” (p. 202). When we refer to symbolic rules, we mean rules with the characteristics just described.

Exceptions like ‘went’, ‘rang’ and ‘slept’ cannot be generated by the ‘add [d]’ rule. Pinker’s theory proposes that they are dealt with by a lexical mechanism that is sensitive to frequency and similarity and entirely distinct from symbolic rules. When planning to produce the past tense of a verb, the speaker first checks to see if an exceptional form can be retrieved from lexical memory. To account for the occasional occurrence of forms like ‘brang’ (as the past tense of bring) or ‘splung’ (as the past of the nonce verb spling), Pinker proposes that lexical memory has associative properties like PDP networks, and thus sometimes produces novel exception forms for inputs similar to known exceptions. In any case, if lexical memory offers up a form, it is produced; if not, the symbolic rule is used as a default. The theory encompassing the rule and the lexicon is called the dual mechanism account.

Pinker and his colleagues, having examined several predictions of their account, conclude that the available evidence provides convincing support. The predictions are strong enough that confirmation would indeed support the idea of the symbolic rule mechanism. Furthermore, clear evidence for the purported properties of the symbolic rule mechanism would contradict basic tenets of the PDP alternative. The PDP account denies that language and other cognitive processes are characterised by the discreteness, uniformity of application, and modularity assumed for the symbolic rule system. It proposes that both regular and exceptional aspects of verb inflection (and of other aspects of language, too; see [13,14]) emerge from a single, integrated mechanism. The connectionist approach makes opposite predictions to those of the rule-based approach (Table 1), so that evidence against one is support for the other. It is therefore crucial to examine the evidence.

In what follows we consider whether inflectional morphology exhibits three key aspects of the symbolic

rule/dual mechanism theory: (1) that acquisition of the symbolic rule is sudden; (2) that the rule is uniform in its applicability and independent of phonological, semantic, or other factors; and (3) that the rule-based mechanism is separate from the mechanism that deals with exceptions.

Is Acquisition of the Regular Past Tense Sudden?

Marcus et al. [15] considered the onset of the regular past tense, using Cazden’s [16] analysis of recorded speech from three normally developing children (Adam, Eve, and Sarah: Brown, [17]). Marcus et al. suggest that the first over-regularization in each child’s corpus signals the moment of acquisition of the past tense rule, and state that this over-regularization error is followed by “rapid increases [in inflecting regulars] to high levels [...] shortly afterward. Adam’s first over-regularization occurred during a 3-month period in which regular marking increased from 0 to 100%” [15, p. 103].

Hoeffner (1996 Ph. D. Dissertation, Carnegie Mellon University) evaluated these data (see Figure 1), both as presented by Marcus et al. and as they emerged in a re-analysis using the transcription in the CHILDES database [18]. Considering first the data presented in Marcus et al., Hoeffner noted that one could just as easily say that “Adam’s first over-regularization occurred during a six-month period in which the probability of using the regular ... rose gradually from 24 to 44%” (p. 98). Either statement seems fairly arbitrary; the data are noisy, and spikes occur when relatively few observations were available (Adam’s 100% marking at 37 months is based on 8 observations). Given the noise, the graphs from all three children suggest a process that proceeds from very little marking in obligatory contexts to fairly reliable marking over the course of about one year. Hoeffner’s own analysis (see caption), suggests an even more gradual acquisition process. A good fit to the data was achieved with a logistic regression in which the use of the regular past increases monotonically with age. Use of first over-regularization as a predictor did not reliably improve the account for regularization rates in any of the children.

In short, the acquisition of the regular past tense is not sudden. According to Brown [17, p. 257], reviewing Cazden’s analysis of other inflections, the situation is the same in all cases:

There is always a considerable period ... in which production-when-required is probabilistic. This is a fact that does not accord well with the notion that the acquisition of grammar is a matter of the acquisition of rules, since the rules ... either apply or do not apply. One would expect rule acquisition to be sudden.

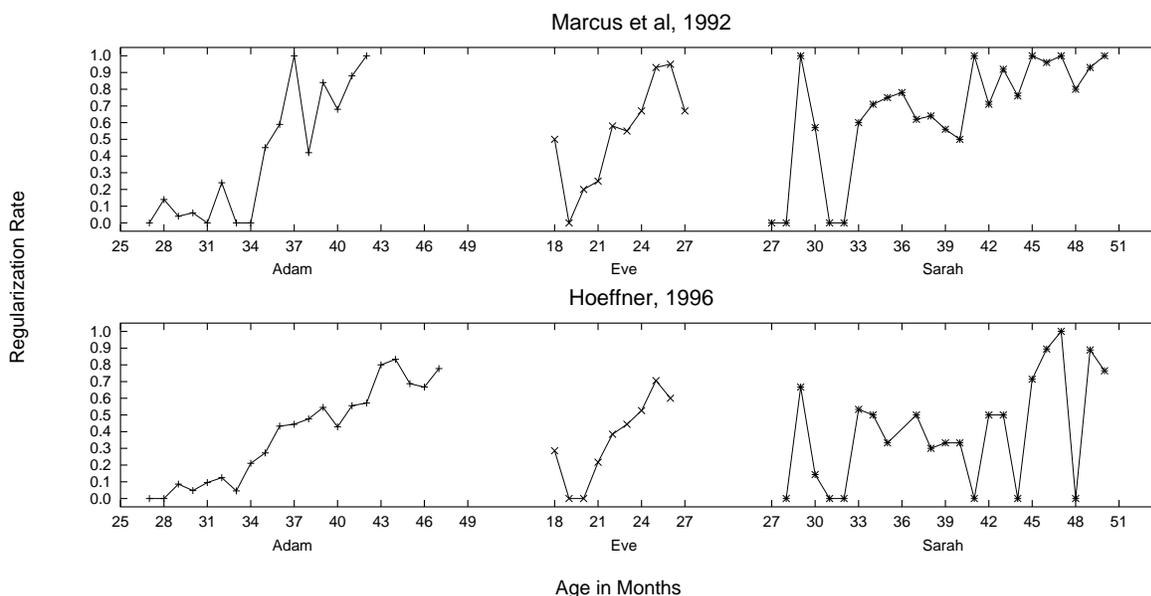


Figure 1. Acquisition of the regular past tense by Adam, Eve, and Sarah, as presented in Marcus et al [15] and in James Hoeffner's Ph. D. Thesis. Percent usage of the regular past tense in obligatory contexts is graphed as a function of the child's age in months. Marcus et al presented data based on scoring by Cazden [16]. Hoeffner repeated the analysis starting from the transcript provided in the CHILDES data base, and included additional time periods. Two independent raters considered each occurrence of a regular verb in the child's speech, first considering the context of occurrence and evaluating whether a past tense was required before seeing the form of the verb actually used, thereby eliminating possible bias in determining whether the context required a past tense and producing an increase in the number of obligatory contexts identified. Data in the upper panel are replotted based on data from Cazden [16] reprinted in Appendix Tables A1, A2 and A3, pages 145-146, of Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F. (1992). Overregularization in Language Acquisition. *Mongraphs of the Society for Research in Child Development, Serial No. 228, Vol 57, No. 4*. Data in the lower panel are replotted with permission from Appendices 1, 2, and 3, pages 311-335 of Hoeffner, J. (1996). Are rules a thing of the past? A single mechanism account of English past tense acquisition and processing. Ph. D. Dissertation, Department of Psychology, Carnegie Mellon University.

Is Application of the Regular Past Tense Uniform?

Pinker stresses that symbolic rules do not vary in their applicability, but depend only on categorical conditions: the past tense applies to any verb stem. Does the evidence support the predicted uniformity? We consider four cases.

Uniformity with respect to phonology. Prasada and Pinker [19] tested judgements on and production of the past tense using nonce forms like 'plip' or 'ploamph', manipulating phonological similarity to existing words. They concluded that there was an effect of similarity to known exceptions on novel irregular inflections, but no effect of similarity to known regulars for the regular inflection. But there was an effect for regulars, which Prasada and Pinker attributed to a confound: their nonce stems, like 'ploamph', that were not similar to other regular items, were also phonologically strange. Even though subjects were asked to judge the inflection and not the stem, Prasada and Pinker claimed that the judgements were affected by the phonological properties of the

stem, and 'corrected' for this by subtracting stem acceptability ratings; but this may be correcting away a real effect. A recent study by Albright and Hayes (manuscript, Department of Linguistics, UCLA) avoided the confound by using nonce stems of high phonological acceptability, and varied whether the item occurred in an 'island of reliability' for the regular or for an exceptional past tense. For example, their corpus contained over 300 verbs ending in an unvoiced fricative (e.g., 'rush' or 'laugh'); this is an island of reliability since every such verb is regular. Both regular and irregular inflections received higher ratings if they came from reliable islands. The effect for regulars survived partialling out any competing influence favoring exceptions. Thus the regular past tense is sensitive to phonological attributes of the stem, violating the prediction of the symbolic rule account.

Uniformity with respect to semantics. A role for word meaning in forming the regular past tense is vigorously rejected in Pinker's theory, since sensitivity to semantic similarity runs counter to the claimed encapsulation of the system that applies phonological transformations to word forms. Yet an influence of mean-

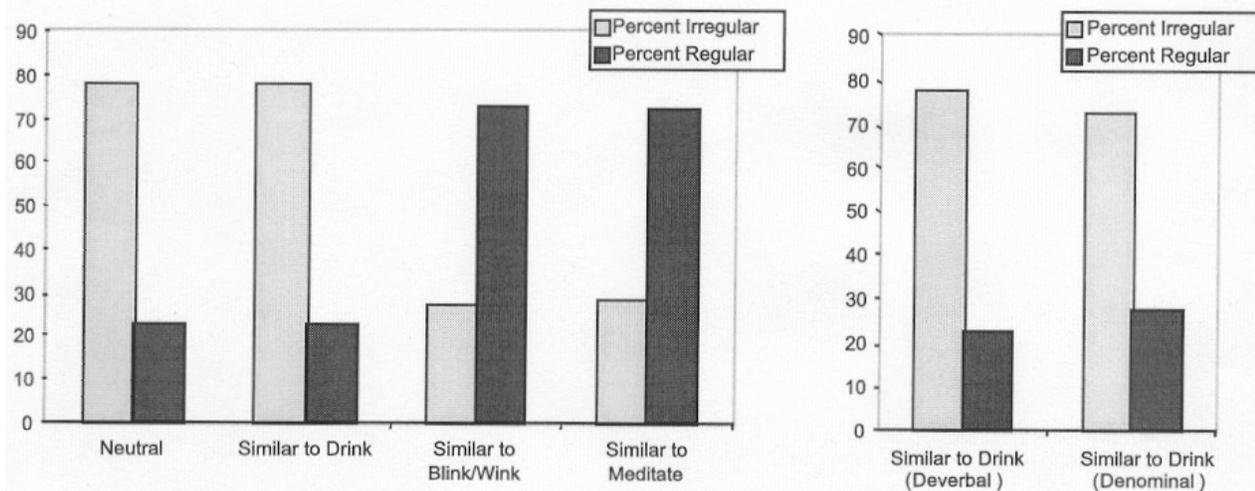


Figure 2. Summary of effects of semantics and grammar on inflections of the nonce verbs 'frink' and 'sprink' from Ramscar [22]. The left panel shows use of irregular ('frink' or 'sprink') or regular ('frinked' or 'sprinked') in four different conditions. Note that in a neutral condition, with no semantic context, participants preferred irregular past tenses, and this trend persisted when context provided a meaning for the nonce verb similar to that of drink. When the context suggested a meaning similar to regular 'wink' or 'blink', or even to the regular word 'mediate', participants shifted to the regular past tense, suggesting that use of the regular past tense can be influenced by semantics. The right panel indicates that subjects' ratings were not affected by their judgement of whether the nonce verb seemed to be denominal. Other experiments in Ramscar (2002) demonstrated strong effects of context specifying a particular meaning of a known polysemous verb like fly, and again there was no effect of denominal status. These findings clearly show that meaning can influence choice of the regular vs irregular inflection, and fail to support the claim [5,23] that denominal status blocks access to lexically-marked exceptions. Reprinted from Figure 1, p. 68, of Ramscar, M. (2002). The role of meaning in inflection: Why the past tense doesn't require a rule. *Cognitive Psychology*, 45, 45-94. Permission Pending.

ing in the selection of regular as well as irregular past-tense forms has often been argued [20-22]. In a recent study, Ramscar [22] placed nonce verbs like 'frink' into semantic contexts that encouraged an interpretation resembling either 'drink' or 'blink'. The former typically elicited 'frink' while the latter increased the likelihood of 'frinked' (see Figure 2). The overall pattern of data contradicts several features of Pinker's account, including the modularity of the regular inflectional system, a purported blocking of lexical/semantic access in denominal verbs [5,23] and the predicted independence of regular inflection from influence by semantic information (caption).

Semantic influences during acquisition. Shirai and Anderson [24] examined the use of the past tense as a function of semantic properties of the situation referred to in children's speech. When it first appears, the use of the past tense (including over-regularization) is largely restricted to descriptions of punctate events that have endpoints and produce results ('I dropped it'); it then gradually spreads to cases in which one of the typical properties (is punctate, has endpoint, produces results) is violated. The children's initial usage corresponds to the typical, but certainly not the only, cases that appear in their mothers' speech, suggesting that initial use of the regular past grows from a semantic prototype.

The exception that proves the rule? In English, the regular past is common, applying to 86% of the 1000 most common verbs [5]. Pinker [5,6] and Marcus et al [25] have suggested, however, that high type frequency is not necessary for the discovery of a regular pattern. Three cases have received the bulk of this discussion: (1) the regular German past participle -t [26]; (2) the Arabic Broken Plural [27]; and (3) the German +s plural [25]. Careful scrutiny of cases (1) and (2) [28,29] indicates that the forms in question are not in fact in the minority. So the case for 'the exception that proves the rule' [25] falls to the German +s plural. Marcus et al claim that the s-plural, despite occurring in only a small fraction of German nouns, is the default used by German speakers whenever there is a "failure of lexical memory". They enumerate 21 separate contexts in which they suppose that lexical memory will fail, and argue that the +s plural should be used in all of these cases because it functions as a symbolic rule independent of the particular characteristics of the item to which it applies.

The +s plural certainly is in the minority in German; but does it apply uniformly as the symbolic rule account predicts? In fact, its usage is not uniform even in [25], which examined assignment of the +s plural to nonce forms treated as (a) unknown but real German words, (b) foreign words, or (c) proper names. For both (b) and (c) only the default rule should be available, yet these two cases do not reveal the same pattern of extension of the +s plural. Hahn and Nakisa [30] (Figure 3) disconfirm the claim that +s acts uniformly across several of the

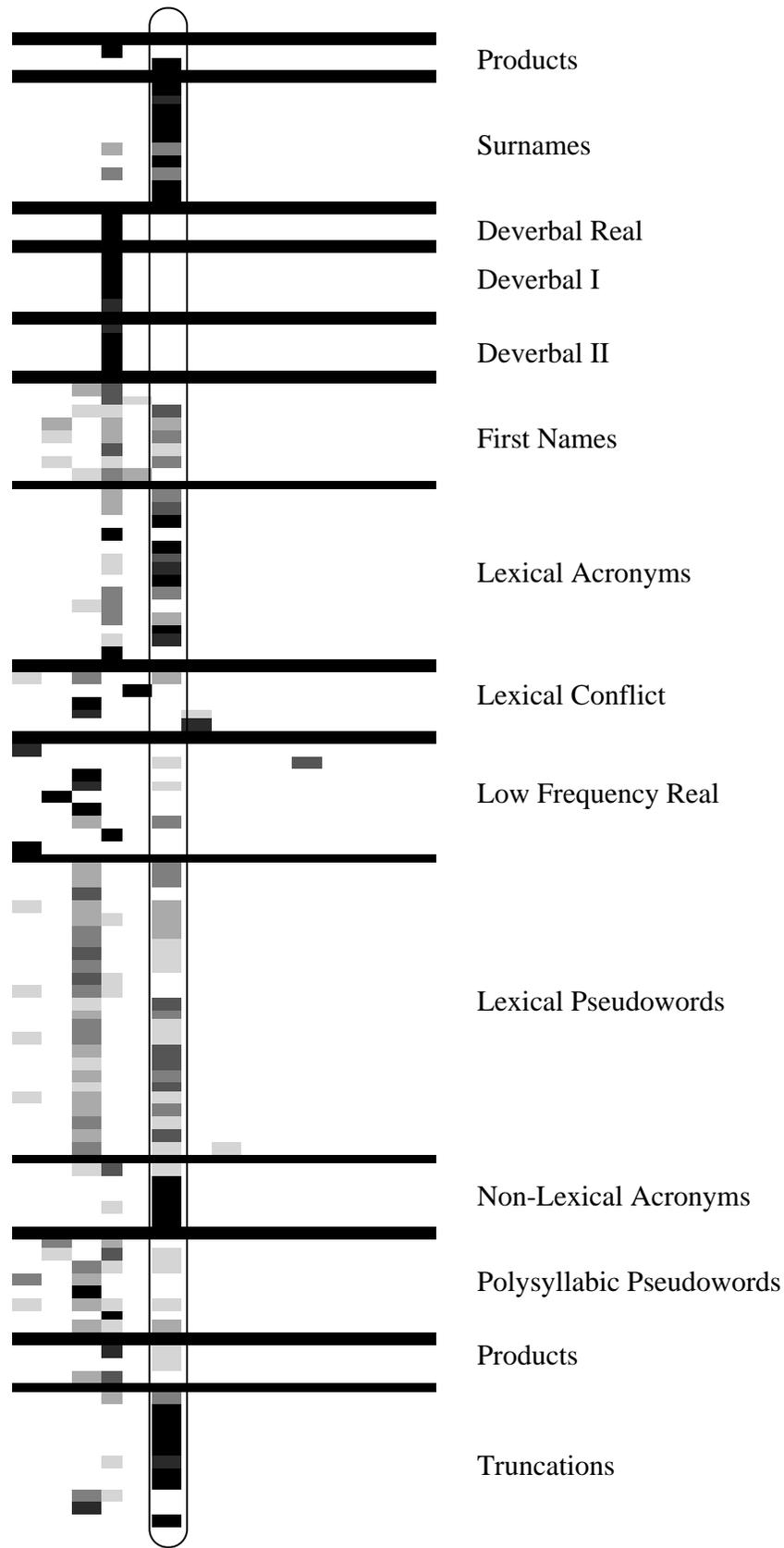


Figure 3. Evidence that the German +s plural is not used uniformly across several situations supposedly calling for the use of a default as proposed by Marcus et al [25]. Each row of the figure represents a different noun form, with the type of the form indicated; the horizontal bars separate the different types. Columns of the figure indicate alternative possible plural inflections, with the +s plural specifically highlighted, and darkness of the entry in each cell indicates the likelihood of using the particular plural for the given item, based on data from native German speaking adults. Reprinted from Figure 9, p. 349, of Hahn, U & Nakisa, R. C. (2000). German inflection: Single-route or dual-route? *Cognitive Psychology*. 41, 313-360. Permission pending.

contexts claimed by Marcus et al. The only case of high and nearly uniform use of +s occurs with surnames and does not extend fully even to first names: two members of the 'Mann' family are called 'Manns' but two girls named 'Ulrike' can be two 'Ulriken'. Bybee [26] also notes relatively high probability for foreign borrowings ending in full vowels. Surnamehood is an arbitrary property that must be associated with a specific use of an item in context, and assigning +s to foreign borrowings ending in full vowels requires sensitivity to phonology and etymology. Such specificity undercuts the notion that the German +s plural is in any sense a default. It is not the exception that proves the rule; instead it is another case with the graded, probabilistic, and context-sensitive characteristics seen in connectionist networks.

Is Regular Inflection Separable from Inflection of Exceptions?

Is there a separate mechanism for regular inflections? In contrast to the connectionist approach, the dual mechanism theory argues yes, and predicts the occurrence of selective deficits in producing/comprehending regular inflections. Pinker [4] considered two putative examples.

Genetic knockouts? A large family (the KE family) consists of some normal individuals and some with an identified single-gene defect [31,32]. Reports based on testing with a small number of stimuli [33,34] suggested that affected individuals had special difficulty with regular compared to irregular inflections. Subsequent investigation by Vargha-Khadem et al [35], however, paints a different picture. Affected family members were found to have a wide range of deficits in linguistic and non-linguistic tasks, and they demonstrated substantial and equal difficulty with regular and irregular forms (Figure 4) when tested with a longer and better-controlled list. There is no sign of selective vulnerability of the regular inflection. We do not rule out the possibility that a developmental phonological deficit could result in difficulty acquiring regular forms [36]. Indeed, if regular inflections are phonetically weak in the input to a network, an impairment in phonological representation can result in a failure to learn the regular past tense [37]. This provides one way of understanding why some children diagnosed with specific language impairment present with an apparent selective deficit in inflectional morphology and other aspects of grammar [38], since many aspects of grammar are signalled by phonetically weak material [39].

Effects of brain damage? Anterior lesions in the left hemisphere often result in dysfluent speech containing few grammatical morphemes or inflections [40]. Ullman et al [41,42] have reported a patient of this type who produced the correct past tense for 69% of exceptions but only 20% of regulars and 5% of nonce forms in a past-tense elicitation task. In collaboration with several others

(Bird et al, [43]) we have considered the possibility that an uncontrolled difference between the regular and exception items in Ullman's study may have influenced the results: the word-final consonant clusters were longer on average in the regular past tenses (2.0 consonants) than in the exceptions (1.2 consonants). This is natural, since regular inflection involves the addition of phonological material to the verb stem, increasing its complexity [44]. In contrast, the formation of exceptions generally involves a vowel and/or consonant change (eat-ate, think-thought) that tends to conserve complexity. Where something is added, there is typically a compensatory reduction in vowel length ('keep'-'kept'), so that exceptional past tenses fall within acceptable phonological bounds.

Bird et al. identified 10 nonfluent aphasic patients who were all significantly better with irregular verbs on a screening list unmatched for phonological factors. The advantage occurred in the elicitation task (37% vs 20% correct), and also in single-word repetition (68% vs 47%) and single-word reading (44% vs 24%). When tested with regular and exception past tenses matched for phonological complexity, none of the patients showed an advantage for irregulars in the elicitation task (means of 26% irregular, 29% regular) or in repetition (65% vs 64%), supporting the view that the initial difference was phonological rather than morphological in origin. A remaining irregular advantage in reading (41% vs 27%) was interpreted as a concreteness effect: past-tense verbs like 'ground' and 'rose' are also concrete nouns.

Ullman et al. [41] also reported a disadvantage in the elicitation task for regular verbs in patients with Parkinson's Disease (PD). Again, however, the effect can probably be interpreted in terms of phonological complexity because, in the specially designed 'PD retest' list, onset consonant clusters were longer in the regular than the irregular verbs. Furthermore, the disadvantage reported for nonwords relative to exceptions cannot be attributed to inflectional processes: the PD patients' responses to nonwords, although often characterized by stem distortions ('pragged' or 'planned' instead of 'plugged'), were correctly inflected 91% of the time (vs. 88% for the exceptions).

Summary of The State of the Evidence

In Table 1 we listed contrasting predictions of the Dual Mechanism and PDP theories. Our review suggests that the onset of the regular past (and all other inflections) is gradual rather than sudden; that both the English regular past tense and the German +s plural are subject to phonological, semantic and other influences rather than being uniform in their application; and that there is no convincing evidence that the inflection of regular verbs can be selectively impaired, except insofar as such impairment is a direct or indirect consequence of a phonological impairment. The evidence seems therefore to be fully compatible with the idea that inflectional processes arise in a single integrated system, in which

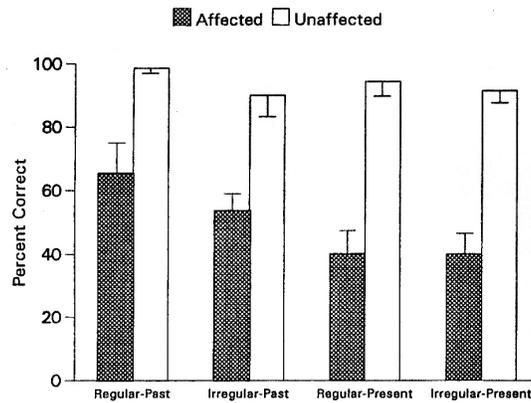


Figure 4. Performance in an elicitation task requiring production of the past tense given the present (“Every day I wash my clothes; yesterday I ___ my clothes.”) or of the present given the past (Yesterday I washed my clothes; every day I ___ my clothes”) for affected and unaffected members of the KE family. Results are based on matched sets of 10 regular and irregular verbs. Reprinted from Figure 2, p. 933 of Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., & Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of the National Academy of Sciences*, 92, 930-933. Permission pending.

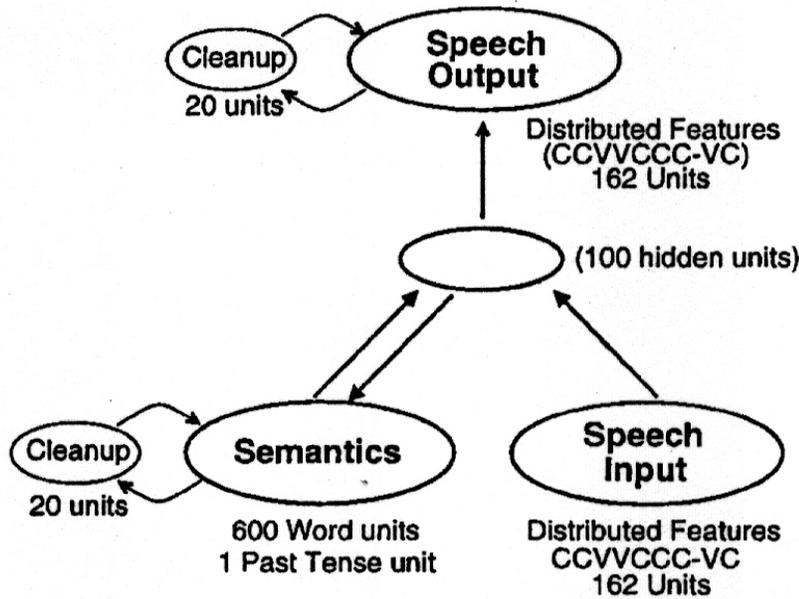


Figure 5. The connectionist model of Joanisse and Seidenberg [46], in which regular and irregular forms are generated by a single system, employing phonological input and output representations and a semantic internal representation. When a verb is presented on the input, the network is trained to generate an appropriate semantic representation (activating the correct word unit and the past tense unit if appropriate) and also to generate the corresponding output representation. The network is also trained to produce the corresponding phonological output when given an input activating an individual semantic unit corresponding to each taught word, and to generate past tenses when the past tense unit is activated and either a verb stem is presented to the phonological input or a word unit is activated in semantics. Reprinted from Figure 2, p. 7594, of Joanisse, M.F., & Seidenberg, M.S. (1999). Impairments in verb morphology following brain injury: a connectionist model. *Proceedings of the National Academy of Sciences*, 96, 7592-7597. Permission pending.

graded and context-sensitive influences of many different types jointly determine whether a regular or an exceptional past tense (or other inflection) will apply. This single system has all of the characteristics of the connectionist framework for inflectional processing.

We do not claim that it would be impossible to construct a rule-based model of inflection formation that has all of the properties supported by the evidence. However, such an account would not be an instantiation of Pinker's symbolic rule account. In fact, rule-based models with some of the right characteristics are currently being pursued [45; Albright & Hayes manuscript]. If such models use graded rule activations and probabilistic outcomes, allow rules to strengthen gradually with experience, incorporate semantic and phonological constraints, and use rules within a mechanism also incorporating word-specific information, they could become empirically indistinguishable from a connectionist account. Such models might be viewed as characterizing an underlyingly connectionist processing system at a higher level of analysis, with rules providing descriptive summaries of the regularities captured in the network's connections.

Toward an Adequate Connectionist Account

Existing connectionist models still have limitations. Given the extent of empirical support for the predictions arising from the connectionist approach, however, we remain convinced of the fruitfulness of pursuing the approach. Our current efforts build on a model by Joanisse and Seidenberg [46] (Figure 5) which incorporates a role for semantic representations [see also 13, 14], something left out of Rumelhart and McClelland [8] as a simplification. This model can explain why a semantic deficit disproportionately disrupts production of exceptional past tenses, as demonstrated by Ullman et al [41, 42] and Patterson et al [47]: word meaning provides information that helps the network to treat a particular item distinctively, counteracting the network's tendency to apply the regular inflection. Some limitations remain, however. Our extensions will use distributed semantic representations that capture similarity in meaning, as well as refinements to phonological processes to address phonological complexity and perceptibility effects. The fact that such a complete model is not yet implemented is scarcely surprising or unique. Encompassing the whole problem is a real challenge for any model, and current rule-based proposals are at best only partially implemented.

In pointing toward a future connectionist account, we note one significant aspect that may be underappreciated. Contrary to some statements [e.g., 4], connectionist networks are not simply analogy mechanisms that base their tendency to generalize on raw item-to-item similarity [48]. Instead, they are sensitive to regularities, so that if an input-output relationship is fully regular, the network can closely approximate a categorical, symbolic rule. Such a property is necessary if these models are to capture the full range of inflectional systems,

since there are cases throughout the world's languages (including the English progressive, -ing form) that are completely regular [49]. These occur interspersed with many other cases with varying degrees of regularity, and networks of the right sort should be able to capture the whole spectrum. This makes the connectionist network fundamentally different from either the symbolic rule or the lexical mechanism considered in the dual-mechanism account.

Box: The Rumelhart-McClelland (1986) Model

The Rumelhart-McClelland model of past-tense inflection [a] consists of a simple pattern associator network [b,c] that learns the relationship between the phonological forms of the stems and past-tenses of English words. This network is flanked by a fixed encoding network on the input side and a fixed decoding network on the output side (see Figure 1). All learning occurs in the pattern associator. The encoding network simply converts a string of phonemes into the 'Wickelfeature' representation used inside the network to represent the stem of each word. Similarly, the decoding network converts the computed Wickelfeature representation of the attempted past tense response back to a sequence of phonemes. The overall theory within which this model arose asserts that processing is meaning- and context-sensitive; for simplicity such influences were not included.

Processing. For a given input, the pattern associator produces an output by a simple neuron-like activation process. Each output unit computes a 'net input' based on the current input pattern and the values of the connection weights. The net input is the sum over all of the connections coming into the unit of the activation of the sending unit multiplied by the weight of the connection. Each unit also has a modifiable threshold. When the net input exceeds the threshold, the unit tends to be turned on, with a probability approaching 1 as net input increases; otherwise, the unit tends to be turned off.

Learning. The network is trained using Rosenblatt's perception convergence procedure [d]. On a learning trial, the model is presented with the stem form of a word and its correct past tense. The stem form is encoded, and the activations of the Wickelfeature output units are computed. This computed representation is compared to the correct representation of the word's past tense. If the computed activation of a given unit matches the correct value, no learning occurs. If a unit that should be active is not, the weights to that unit from each active input unit receive a small fixed increment, and the threshold is reduced. Correspondingly, if a unit that should not be active is on, the weights from each active input unit are decremented and the threshold is increased. As a result

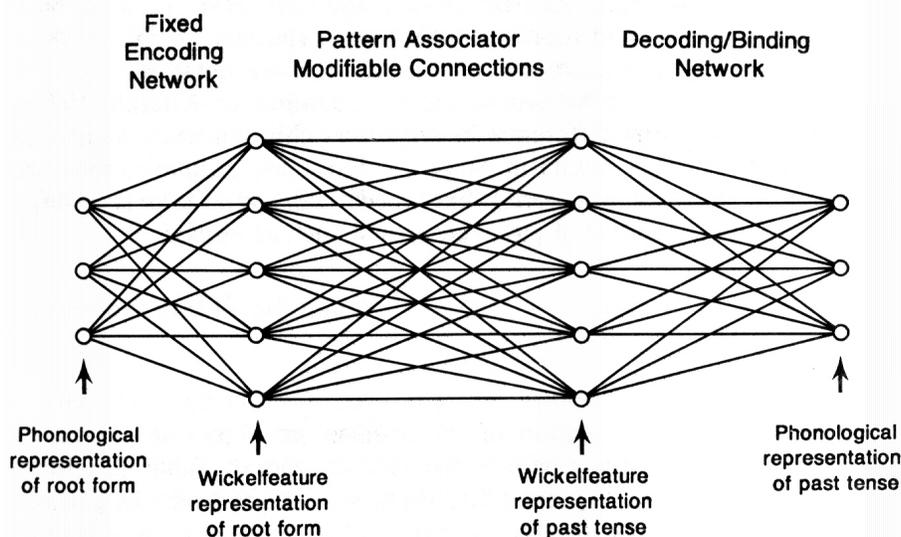


Figure 6. [Box Figure 1]. The Rumelhart-McClelland model of past-tense inflection, Reprinted from Fig. 1, p. 222, of Rumelhart D. E., & McClelland J. L. (1986). On learning past tenses of English verbs. In Rumelhart D.E. and McClelland J.L. (Eds), *Parallel Distributed Processing: Vol 2: Psychological and Biological Models*. Cambridge, MA: MIT press. Permission Pending.

the network gradually improves performance over many learning trials, simulating a gradual developmental process. Later models employ the back-propagation learning algorithm [e], an extension that allows the use of one or more layers of hidden units between inputs and outputs, and/or recurrent connections [f].

Representation. Coding is based on an idea by Wickelgren [g], in which word forms are represented by units designating each phoneme, together with its predecessor and its successor. Thus 'help' would be represented by '.he', 'hel', 'elp', and 'lp.'. The model used units called 'Wickelfeatures' (WFs) each representing a feature from each of the phonemes in such triads. For example, there is a unit representing the feature sequence 'liquid-unvoiced-end', which would be active in representing 'lp.'. In general, words ending in a unvoiced phoneme are represented by several WFs capturing that the final phoneme is unvoiced. For the past tense output 'helped', such WFs should be replaced with others representing the added unvoiced stop /t/ that forms the past-tense inflection.

Capturing regular and exceptional inflections. For regular verbs in English, if the stem ends in a unvoiced sound (like the /p/ in help) the past tense will be formed by adding the unvoiced dental /t/. Through exposure to regular words, the network will repeatedly experience cases where the input contains WFs coding final unvoiced stem phonemes and the output contains WFs coding the added final /t/. The learning process will build up positive connections from the active input units to the ap-

propriate output units, thereby encoding the regular addition of /t/ after unvoiced phonemes. Also, all non-final WFs of the stem are simply maintained in the past tense form, so the network will gradually acquire connections mapping each non-final WF to its counterpart in the output. At the same time, each output unit can be influenced by any input unit. To produce exceptions, connections from units coding specific input features to units coding for exceptional aspects of the inflection will be strengthened, thereby allowing specific properties of the input (such as presence of 'ee' followed by final /p/) to modify specific properties of the output, so that items like creep, keep, and sleep are correctly mapped to the past tenses crept, kept and slept.

Words or Rules Cannot Exploit the Regularity in Exceptions: A Reply to Pinker and Ullman

Pinker and Ullman [10] succinctly restate their position that the English past tense is governed by two competing mechanisms, identified as "words and rules", and taken as examples of distinct procedural and declarative systems. Their mechanisms work separately, so that only one or the other is responsible for yielding a particular past tense form. To produce the past tense of *keep*, words and rules race to generate a response; since the correct past tense of *keep* is not regular, it must be generated by the lexical mechanism. For this reason, we think of their approach as the words or rules theory.

Our approach is different. An integrated connectionist network maps from the stems of all verbs to their past tense forms, using a single network of units and connections. For example, in the original Rumelhart & McClelland (1986) model [8], the same units and connections that produce regular past tenses from regular stems also process the irregulars, so the network has an inherent tendency to do the same thing to the exceptions that it does to regulars—namely copy the features of the stem to the past tense form and add /d/, /t/ or /d/ depending on the final consonant. To produce *kept* instead of *keep*ed (note *keep*ed is pronounced with unvoiced /t/) all that is required is to adjust the activations of the output units representing the vowel, something that the network will have learned to do on the basis of experience with *keep* and neighbors *creep*, *leap*, *sleep*, *sweep*, and *weep*. The network uses the same connection-based knowledge that allows it to perform the regular mapping, and also taps into specific connections activated by the particular properties of 'keep' to produce the vowel adjustment.

A core difference between the approaches is that one exploits the regularity in the exceptions—what we call quasi-regularity—and the other does not. Quasi-regularity is the tendency for an exception to exhibit aspects of the regular pattern [14]. If there were only a few quasi-regular items, one might treat them as accidents, but in fact nearly all exceptional past-tenses in English are quasi-regular to some extent. To demonstrate this, we enumerate the different types (for other taxonomies, see [7,50]):

1. Two very frequent verbs, *have* and *make*, delete a consonant and add the regular /d/ to what remains, forming *had* and *made*.
2. The *-eep* words listed above and others, including *say*, *do*, *tell*, *hear*, *sell*, *flee* and *shoe*, form the past tense by adding regular /d/ or /t/ and making a vowel adjustment, producing *kept*, *did*, *said*, *told*, etc.
3. Twenty-eight verbs like *cut* and *hit* have past tenses identical to their stems; all end in /d/ or /t/, as regular past tenses do.
4. Another set of verbs ending in /d/ or /t/, including

bleed, *breed*, *feed*, *lead*, *read*, *speed*, *hide*, *ride*, *slide*, and *fight*, adjust the vowel to create /d/- or /t/-final *bled*, *slid*, *fought*, etc.

Several sets of verbs (waning in some dialects) use unvoiced /t/ instead of /d/, usually after /l/ or /n/.

5. One such set, including *dwelt*, *smelt*, *spelt*, *burnt*, and *learnt*, would be completely regular except for the de-voicing of the inflection, producing past forms like *spelt* and *burnt*.
6. Another group that takes /t/ instead of /d/, including *meant*, *dreamt*, *dealt*, *felt*, and *kneelt*, adjust the vowel as well, yielding *meant*, *dealt*, etc.
7. A group of verbs already ending in /d/, including *built*, *bent*, *lent*, *rent*, *sent* and *spent*, replace stem-final /d/ with /t/ to make *built*, *sent*, etc.
8. Another, related group, including *brought*, *caught*, *sought*, *taught*, and *thought* adjust the vowel to /aw/ and replace the final consonant cluster with /t/, creating *brought*, etc.

Overall, 59% of the 181 English exceptions listed in Pinker & Prince [7] have past tenses ending in /d/ or /t/, and fall into one classes 1-8.

9. Items like *sang*-*sang*, *rose*-*rose* and *flew*-*flew*, which comprise nearly all of the remaining cases, are also quasi-regular, since the consonants of the stem are preserved, and only a vowel change occurs.

There are only two 'suppletive' verb roots in English (*be* and *go*, with derivatives *forgo* and *undergo* where the past tense form is completely different from the present tense).

As noted above, the Pinker-Ullman theory provides no mechanism for exploiting the aspects of the regular past tense that are so prevalent among exceptions. Pinker [4] did adopt the idea that the lexical system has connectionist-like properties. This provided a way to account for clusters among the exceptions and for creative formation of novel forms consistent with such clusters. This step was in the right direction, but did not go far enough. Since past tenses of exceptions in this account are formed only by the lexical system, the theory still fails to explain why many of the exceptions share properties with regular past tense forms and offers no way to exploit the regular mapping in forming past tenses of these exceptions.

In contrast, connectionist models inherently capture the regularity in the exceptions because the exceptions are processed by the same network that processes the regulars. As already noted for *keep*-*kept*, items that are quasi-regular can make partial use of the same connections that are used in forming exceptions. All nine of the types noted above, encompassing 177/181 forms, exploit to some degree the connection weights that produce regular items. Only the suppletive items fail to make any

use of the connections that produce the regular past tense [51].

The past tense of English is just one domain that exhibits quasi-regularity. In English spelling-sound mapping, virtually every exception has some degree of regularity; exceptional pint, aisle, hymn and champagne all partially adhere to regular correspondences. Quasi-regularity exists in richly inflected languages like Spanish and in derivational as well as inflectional morphology [52, 44]. It is found in language units beyond the word level [53, 54], and beyond language it characterises real-world objects, which have properties shared with other objects of the related types as well as some unique properties [55]. Given these observations, the plausible candidate mechanisms of human linguistic and conceptual processes are those that can exploit quasi-regularity. Single-system connectionist models have this property; the words or rules theory does not.

References

1. Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
2. Chomsky, N. (1959). Review of Skinner (1957). *Language*, 35, 26-58.
3. Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.
4. Pinker, S. (1991) Rules of language. *Science*, 253, 530-535.
5. Pinker, S. (1999) *Words and Rules*. New York: Basic Books.
6. Pinker, S. (2001) Four decades of rules and associations, or whatever happened to the past tense debate? In E. Dupoux (Ed.), *Language, the brain, and cognitive development: Papers in honor of Jacques Mehler*. Cambridge, MA: MIT Press.
7. Pinker, S. & Prince, A., (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28, 73-193.
8. Rumelhart D. E., & McClelland J. L. (1986). On learning past tenses of English verbs. In Rumelhart D.E. and McClelland J.L (Eds), *Parallel Distributed Processing: Vol 2: Psychological and Biological Models*. Cambridge, MA: MIT press.
9. McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The Appeal of Parallel Distributed Processing. In D. E. Rumelhart, J. L. McClelland and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Vol I. Foundations*. Cambridge, MA: MIT Press. pp. 3-44.
10. Pinker and Ullman (2002). The past and Future of the Past Tense. *Trends in Cognitive Sciences*, 6, xxx-xxx.
11. Ervin, S. M. (1964). Imitation and structural change in children's language. In E. Lenneberg (ed.), *New directions in the study of language*. Cambridge, MA: MIT Press.
12. Berko, J. (1958). The child's learning of English morphology. *Word*, 14, 150-177.
13. Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.
14. Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, 103, 56-115.
15. Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F. (1992). Overregularization in Language Acquisition. *Mongraphs of the Society for Research in Child Development, Serial No. 228, Vol 57, No. 4*.
16. Cazden, C. B. (1968). The acquisition of noun and verb inflections. *Child Development*, 39, 433-448.
17. Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.
18. MacWhinney, B. & Snow, C. (1990). The Child Language Data Exchange System: An update. *Journal of Child Language*, 17, 457-472.
19. Prasada, S. & S. Pinker (1993). Generalization of regular and irregular morphological patterns, *Language and Cognitive Processes*, 8, 1-56.
20. Harris, C. L. (1992). Understanding the English past-tense formation: The shared meaning hypothesis. *Proceedings of the Fourteenth Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
21. Daugherty, K., MacDonald, M.C., Petersen, A., & Seidenberg, M.S. (1993). Why no mere mortal has ever flown out to center field but people often say they do. *Proceedings of the fifteenth annual conference of the cognitive science society*. Hillsdale, NJ: Erlbaum, pp. 383-388.
22. Ramscar, M. (2002). The role of meaning in inflection: Why the past tense doesn't require a rule. *Cognitive Psychology*, 45, 45-94.
23. Kim, J. J., Pinker, S., Prince, A. & Prasada, S. (1991) Why no mere mortal has ever flown out to center field. *Cognitive Science*, 15, 173-218.
24. Shirai, Y., & Anderson R. W. (1995). The acquisition of tense-aspect morphology. A prototype account. *Language*, 71, 743-762.
25. Marcus, G., Brinkmann, U., Clahsen, H.,

- Wiese, R., Woest, A., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, *29*, 189-256.
26. Clahsen, H., & Rothweiler, M. (1992). Inflectional rules in children's grammars: Evidence from the development of participles in German. *Yearbook of Morphology*, *1992*, 1-34.
 27. McCarthy, J., & Prince, A. (1990). Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory*, *8*, 209-283.
 28. Bybee, J. L. (1995) Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*, 425-455.
 29. Boudelaa, S. & Gaskell, M. G. (2002) A re-examination of the default system for Arabic plurals. *Language and Cognitive Processes*, *17*, 321-343.
 30. Hahn, U & Nakisa, R. C. (2000). German inflection: Single-route or dual-route? *Cognitive Psychology*, *41*, 313-360.
 31. Vargha-Khadem, F., Watkins, K.E., Price, C.J., Ashburner, J., Alcock, K.J., Connelly, A., Frackowiak, R.S.J., Friston, K.J., Pembrey, M.E., Mishkin, M., Gadian, D.G. & Passingham, R.E. (1998). Neural basis of an inherited speech and language disorder. *Proceedings of the National Academy of Sciences*, *95*, 12695-12700.
 32. Lai, C.S.L., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., & Monaco, A.P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, *413*, 519-523.
 33. Gopnik, M. (1990). Dysphasia in an extended family. *Nature*, *344*, 715.
 34. Gopnik, M. & Crago, M. (1991). Familial aggregation of a developmental language disorder. *Cognition*, *39*, 1-50.
 35. Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., & Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of the National Academy of Sciences*, *92*, 930-933.
 36. Leonard, L. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.
 37. Hoeffner, J., & McClelland, J. L. (1993). Can a perceptual processing deficit explain the impairment of inflectional morphology in developmental dysphasia? A computational investigation. In E. Clark (Ed.), *Proceedings of the Twenty-fifth Annual Child Language Research Forum*. Center for the Study of Language and Information, Stanford University, Palo Alto, CA. 38-49.
 38. van der Lely, H. K., Rosen, S. & McClelland, A. Evidence for a grammar-specific deficit in children. *Current Biology*, *8*, 1253-1258.
 39. Leonard, L.B., McGregor, K.K., & Allen, G.D. (1992). Grammatical morphology and speech perception in children with specific language impairment. *Journal of Speech and Hearing Research*, *35*, 1076-1085.
 40. Marin, O., Saffran, E. M., & Schwartz, M. F. (1976). Dissociations of language in aphasia: Implications for normal function. In S.R. Harnad (ed.), *Origin and evolution of language and speech*. *Annals of the New York Academy of Sciences*, *280*, 864-884.
 41. Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience*, *9*, 266-276.
 42. Ullman, M. T., Izvorski, R., Love, T., Yee, E., Swinney, D. A., & Hickok, G. (in press). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language*.
 43. Bird, H., Lambon Ralph, M. A., Seidenberg, M. S., McClelland, J. L., & Patterson, K. (forthcoming). Deficits in phonology and past tense morphology: What is the connection? *Journal of Language & Memory*.
 44. Burzio, L. (2002). Missing players: Phonology and the past tense debate. *Lingua*, *112*, 157-199.
 45. Taatgen, N. & Anderson, J. (in press). Why do children learn to say "Broke"? A model of learning the past tense without feedback. *Cognition*.
 46. Joanisse, M.F., & Seidenberg, M.S. (1999). Impairments in verb morphology following brain injury: a connectionist model. *Proceedings of the National Academy of Sciences*, *96*, 7592-7597.
 47. Patterson, K., Lambon Ralph, M. A., Hodges, J. R., & McClelland, J. L. (2001). Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. *Neuropsychologia*, *39*, 709-724.
 48. Plaut, D. C. & Gonnerman, L. M. (2000) Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, *15*, 445-485.
 49. Dabrowska, E. (2001). Learning a morphologi-

- cal system without a default: the Polish genitive. *Journal of Child Language*, 28, 545-574.
50. Bybee, J. L. & Slobin, D. L. (1982). Rules and Schemas in the Development and Use of the English Past Tense. *Language*, 58, 265-289.
 51. Plunkett, K, & Marchman, V. A. (1991). U-Shaped Learning and Frequency Effects in a Multi-Layered Perceptron: Implications for Child Language Acquisition. *Cognition*, 38, 43-102.
 52. Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. Philadelphia, PA. John Benjamins.
 53. Harris, C.L. (1994). Coarse coding and the lexicon. In C. Fuchs & B. Victorri, (Eds.), *Continuity in linguistic semantics*. Amsterdam: John Benjamins.
 54. McClelland, J. L. (1992). Can connectionist models discover the structure of natural language? In R. Morelli, W. M. Brown, D. Anselmi, K. Haberlandt & D. Lloyd (Eds.), *Minds, brains & computers*. Norwood, NJ: Ablex Publishing, 168-189.
 55. Rogers, T. T., & McClelland, J. L. (forthcoming). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Box References

- a. Rumelhart D. E., & McClelland J. L. (1986). On learning past tenses of English verbs. In Rumelhart D.E. and McClelland J.L (Eds), *Parallel Distributed Processing: Vol 2: Psychological and Biological Models*. Cambridge, MA: MIT press.
- b. Kohonen, T. (1977). *Associative Memory: A system theoretical approach*. New York: Springer.
- c. Anderson, J. (1977). Neural models with cognitive implications. In D. LaBarge and S. J. Samuels (Eds.), *Basic Processes in Reading: Perception and comprehension* (pp 27-90). Hillsdale, NJ: Erlbaum.
- d. Rosenblatt, F. (1962). *Principles of Neurodynamics*. New York: Spartan.
- e. Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986). Learning Representations by Back-Propagating Errors, *Nature*, 323, 533-536.
- f. Williams, R. J. & Zipser, D., (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, 1, 270-280.
- g. Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, 86, 44-60.