

I

Theories of Language Processing

Wugs and Goed: Evidence for the Child Acquirer's Use and Overuse of Rules

The basic distinction between morphological regularity and irregularity has figured in child language-acquisition studies since at least the 1950s. For instance, researchers such as Anisfeld and Tucker (1967), Berko (1958), Bryant and Anisfeld (1969), and Ervin (1964) all noted that at a certain point in their linguistic development (roughly between the ages of four and seven), child acquirers of English are able to productively inflect novel verbs with regular past-tense endings and novel nouns with regular plural endings. To take one example, Berko (1958) found that when child acquirers of English were shown a picture of a fictitious animal and told that it was a *wug*, they were able to produce its correct plural form: *wugs*. Since the children in these studies had never heard the word *wug* before (and presumably had never heard its plural form either), the results were taken

as evidence that children are able to use a suffixation rule in order to create regular English plural and past-tense forms.

There is further naturalistic evidence that child acquirers of English productively use rules. This evidence comes from documented instances of overgeneralizations of the inflectional endings of the English regular past-tense and plural forms. For example, at about the same time that they become able to productively inflect novel verbs with regular past-tense endings and novel nouns with regular plural endings, children acquiring English will overgeneralize the past-tense suffix *-ed* both to the root forms of irregular verbs (e.g., *sing-singed*) and to irregular verbs already in the past tense (e.g., *broked*; see, e.g., Anisfeld, 1984; Slobin, 1985). Curiously, these children will have previously produced many of these same past-tense forms correctly. Again, these observations have been interpreted as supporting the notion that children are overapplying rules of inflectional morphology to create ungrammatical yet (apparently) rule-generated English regular past tenses and plurals.

The Status of Regular Morphological Rules in the Adult Mind

Of course, pathologically normal children do eventually retreat from the above-described overgeneralizations (see Marcus et al., 1992, for a discussion of how this might happen). Nevertheless, regular-irregular differences in English past-tense morphology have remained interesting to researchers in cognitive science, linguistics, and psycholinguistics who seek to establish the extent to which differences in the morphological structure of regularly and irregularly inflected items reflect differences in how these forms are represented in the minds of both children and adults.

Generally speaking, there are currently two types of explanation for the acquisitional, distributional, and (as we shall see) processing differences between English regular and irregular past-tense forms. *Single-mechanism* theories hold that both regulars and irregulars are processed—that is, produced and understood by speakers of a language—by a single associative memory system. *Dual-mechanism* theories instead propose that regulars are the product of symbolic, algebraic rules, while irregulars are stored in and retrieved from a partly associative mental lexicon. Below, I present an overview of and supporting evidence for both theoretical perspectives. Mirroring recent cognitive scientific discourse on this topic (see, e.g., Pinker and Ullman, 2002; McClelland and Patterson, 2002), this book will use the theory of *connectionism* to represent the single-mechanism perspective and *words and rules theory* to represent the dual-mechanism perspective.

Example of a Single-Mechanism Theory: Connectionism

As stated above, in this book connectionism (Rumelhart and McClelland, 1986; Elman and McClelland, 1986; Rumelhart, McClelland, and PDP Research Group, 1986; Elman et al., 1996) will represent the single-mechanism perspective. The origins of connectionism lie in research on artificial intelligence (AI), which has been defined as “the branch of computer science that investigates the extent to which the mental powers of human beings can be reproduced by means of machines” (Dunlop and Fetzer, 1993, p. 6). This includes designing machines that will engage in intelligent behavior when made to perform such actions as solving problems, playing chess games, or doing other similar activities.¹ Whatever its purpose, AI re-

search can be classified according to the extent to which it is either strong or weak AI, or to the extent to which it is symbolic (top-down) or connectionist (bottom-up) AI. A strong AI system not only exhibits intelligent behavior but also is sentient or self-aware. Strong AI systems exist only on the silver screen, with one of the more recent (and malevolent) examples being the snappily dressed, pistol-packing Agents of the *Matrix* movie trilogy. If any strong AI systems have in fact been created, either their designers have not gone public with the news, or the strong AI systems themselves have not volunteered evidence of their own existence. By contrast, weak AI systems exhibit intelligent behavior but are not self-aware. Designers of weak AI systems view their creations not as complete models of human consciousness, but as models of ways in which information processing might proceed in the mind. Weak AI systems have been designed that play games, solve problems, or, as we will see below, learn the English past tense. An example of a product of weak AI research would be Deep Blue, the chess-playing computer that defeated Garry Kasparov. While Deep Blue was clearly able to execute chess moves, it was neither aware that it was playing chess nor capable of other recognizably human behaviors.

The other dimension by which AI research may vary is the extent to which it is symbolic or connectionist. Symbolic AI research views human intelligence as emerging from the brain's manipulation of symbols. In this account of human intelligence, these symbols are variables, or placeholders; as such, they do not represent individual instantiations of objects but instead abstract categories according to which the objects or entities in the world have been classified. The symbol manipulations performed by the brain are conceived as being algebraic-like rules. As a result, symbolic AI researchers at-

tempt to model human intelligence by designing AI programs that mirror this view of human cognition. Symbolic AI programs have built-in, higher-order representations of entities and objects as well as representations of classes of entities or objects, and they include rules specifying the possible operations that can be performed upon an entity or an object.

The reason that the symbols and rules of a symbolic AI program must be supplied by the programmer is because such programs have no built-in capacity to learn from their environment. Thus in most cases, these programs cannot ever know anything more about their world than what was built into them; as a result, they typically have neither the capacity to learn new information nor the capacity to learn through making mistakes. Possibly the most difficult obstacle for symbolic AI programs to overcome, however, is representing and implementing the background or commonsense knowledge of the world that all humans acquire and are able to reason with (Dreyfus, 1992). Attempts have been made to overcome this shortcoming of symbolic AI, usually either by carefully limiting the world within which the AI system must function, or by attempting to capture within sufficiently abstract semantic frames or scripts a common core of the seemingly diverse range of situations in daily life. Examples of such attempts include building programs with a tightly constrained, rule-based micro-world such as that found in a chess game (see also, e.g., the block world of SHRDLU; Winograd, 1970) or with real-world knowledge but only of select, highly specific situations (such information is contained within semantic frames, or collections of information about a stereotyped situation such as a restaurant setting, a birthday party, etc.; e.g., Minsky, 1975; see also Schank and Abelson, 1977, for a related script-based approach).

In these and other examples of symbolic AI programs, little concern is shown for modeling actual brain processes. Symbolic AI researchers are usually not concerned with how their programs might acquire the knowledge they have of their microworlds, frames, or scripts; in many cases the symbols and rules used to manipulate them are simply taken as a given and are built into the system from the start.

Connectionist AI, by contrast, seeks to create AI systems that to some degree model the structure and functioning of the human brain and human learning processes themselves. In this view of human intelligence, learning occurs through both supervised and unsupervised interaction with the environment. At its most basic level, this learning is thought to occur in a bottom-up fashion: a flow of simple or low-level information enters the system from the outside world, and as it is processed by the brain, the low-level information is transformed into higher-order representations.

The human brain possesses a staggeringly complex circuitry. A recent study estimates that there are approximately twenty-one billion neurons—the number varies by sex and age—in the adult neocortex (Pakkenberg and Gundersen, 1997). Far more important than the sheer number of neurons, however, is the degree of interconnectivity between them: some estimates have suggested an average of two thousand neuronal connections, or synapses, per neuron (Pakkenberg et al., 2003). This means that on average, the twenty-one billion neurons in an adult's neocortex are networked by forty-two trillion synapses. These neuronal connections can either be excitatory (that is, they cause a neuron to fire, or release chemical neurotransmitters across synaptic gaps to the neurons surrounding it, thereby exciting those neurons to fire as well) or inhibitory (that is, they discourage a neuron from

firing). It is thanks to this massive interconnectivity at the cellular level that the brain is able to process and integrate information the way it does.

Thus, connectionist AI researchers attempt to model human intelligence by designing AI programs that mirror our current understanding of how the brain processes information at the cellular level. In contrast to symbolic AI programs, connectionist AI programs usually consist of computational matrices that mimic the networked structure and information flow patterns of interconnected neurons in the brain. One part of the computational matrix represents an array or layer of input neurons, or nodes, while another part represents a layer of output nodes. In recent neural network designs, there may also be a “hidden” layer of nodes between the input and output layers. (They are referred to as hidden because they never have direct contact with the outside world.) As in a real brain, the input and output nodes in an artificial neural network are connected; however, they are not connected physically but mathematically, such that as an input node is turned on (i.e., is told to fire by the program), it sends an excitatory message to the output node(s) it is connected to. Through one of several possible learning algorithms, the network receives feedback during its course of training on the accuracy of its output activations. This feedback is then used to mathematically adjust the strengths of the connections between nodes. In this way, for a given input, connections for correct answers are strengthened while connections for incorrect answers are inhibited. The result is that the network is able to gradually tune itself such that only the correct output is likely to be activated for any particular input.

With respect to efforts to model the acquisition and processing of language, possibly the best known connectionist

networks are the parallel distributed processing (PDP) models of Rumelhart and McClelland (Rumelhart, McClelland, and PDP Research Group, 1986). Two terms may require explanation: *parallel* means that processing of input occurs simultaneously throughout the network, and *distributed* means that there is no command center, or central location in the network where executive decisions are made. These models have also been called *pattern associators* in that they associate one kind of pattern (for instance, a pattern of activated input neurons) with another kind of pattern (for instance, a desired pattern of activated output neurons).

Two well-known PDP models have been designed to cope with different aspects of language processing. The first model, McClelland and Ellman's (1986) TRACE model of phoneme perception and lexical access, was a PDP network of detectors and connections arranged in three layers: one of distinctive-feature detectors, one of phoneme detectors, and one of word detectors. The connections between layers were bidirectional and excitatory, meaning that while lower-level information could feed forward to higher-level layers, higher-order information could also percolate back down and thereby influence detection at the lower-level layers. By contrast, the connections within a given layer were inhibitory. In detecting a word or a phoneme, the feature detectors first extracted relevant information from a spectral representation of the speech signal. That information then spread according to the relative strength of activation of certain distinctive-feature detectors over others to the layer of phoneme detectors. In the case of word detections, the phoneme detectors then activated words in the third layer of the network. The bidirectional, excitatory nature of the connections between layers, coupled with the inhibitory nature of the connections within a given layer, con-

spired to increasingly activate one lexical candidate over all others while simultaneously suppressing competing candidates. Information related to representing the temporal unfolding of a word was modeled by repeating each feature detector multiple times, thereby allowing the possibility of both past and possible upcoming information to be activated along with current information. Using this interactive network architecture, McClelland and Ellman (1986) showed that the TRACE model could successfully simulate a number of phoneme and word detection phenomena that had been observed in experiments with human participants. For example, in phoneme detection experiments, it was observed that factors pertaining to phonetic context (i.e., the phonemes that precede and follow a target phoneme) appeared to aid people in detecting phonemes or in recovering phonemes that had been partially masked with noise.² The TRACE model also produced this finding.

In a second simulation, Rumelhart and McClelland (1986) used a pattern associator network to model not speech perception, but the acquisition of the English past-tense system. For this study, the researchers began by noting that although the findings surrounding the acquisition of the English past tense have been interpreted as bearing the hallmarks of rule-based development (with clearly marked and nonoverlapping stages of development), the documented facts (some of which were reviewed above) could also be argued to suggest instead that acquisition proceeds on a less categorical, more gradual basis. To a certain extent, therefore, the observed facts would also appear to be in line with the predictions of a more probabilistically based acquisition account. Thus, one of the goals of Rumelhart and McClelland (1986) was to successfully model

this kind of gradual convergence upon the correct forms of the English past tense.

The task facing their pattern associator network was the successful acquisition of 506 (408 regular and 98 irregular) English verbs, which were sorted according to frequency (low, medium, and high). This meant that the same network had to associate both irregular infinitives with their largely idiosyncratic past tenses, and regular infinitives with the regular *-ed* past tense. In order to have a relatively constrained number of naturalistic representations that would capture differences between the root and past-tense forms of both regular and irregular English verbs while allowing for the emergence of any possible generalizations between present- and past-tense forms, both the input and output banks of the network were designed to “comprehend” and “produce” individual letters of infinitives and past-tense forms as clusters of phonological features and word-boundary information.

The network was trained by presenting it with infinitival forms (in the form of sequences of letters, which were themselves represented by phonological features) together with a random activation of the output nodes and the infinitives’ correct past-tense forms (representing a “teacher” function, the idea being that children receive only correct input from the environment).³ Using a learning algorithm, the network gradually adjusted its input-output connection weights such that there was a high probability that its output would eventually match the desired output pattern.

The training period was designed to emulate Rumelhart and McClelland’s (1986) characterization of the input that a child would be exposed to during the time he or she was acquiring English. According to them, a child will first learn about

the present and past tenses of the most frequently occurring verbs. This was represented by first training their network for ten epochs, or cycles, on the ten most frequently occurring English verbs of their training set. Following this, their network was trained for 190 cycles on an additional 410 medium-frequency verbs. During this time, the network progressed in its learning such that by the end of its training, it had all but mastered the infinitive-to-past-tense mappings for the 420 verbs. Close examination of the progression of learning revealed some surprising results, among which were the following:

- Similar to what has been noted to occur during child acquisition of English, the network exhibited “u-shaped” learning behavior with respect to irregular verbs. While the network coped equally well with regulars and irregulars for the first ten verbs, performance with irregulars initially dropped upon the introduction of the 410 medium-frequency verbs such that the network incorrectly overgeneralized the regular *-ed* ending to the infinitival forms of irregulars, including both new irregulars and those that had been among the first ten verbs (i.e., irregulars that the network had previously gotten right). The network needed an additional thirty cycles of training before it began to gradually retreat from its overgeneralizations. By the time the two-hundred-cycle training phase had been completed, the network performed almost flawlessly with both regulars and irregulars, though it continued to perform somewhat better with regulars than with irregulars.

- Across nine different morphological classes of irregular verbs and three classes of regular verbs, the network showed overall error patterns similar to those made by preschool children in a child language-acquisition study (Bybee and Slobin, 1982). The resemblance between preschool-child and network performance was particularly striking with verbs that are the same for the present and past tenses (e.g., *beat*, *fit*, etc.), but it was also observable with irregulars having stem-vowel changes (e.g., *deal*, *hear*, etc.).
- In terms of the pattern of occurrence for the two types of regularizations of irregular verbs that have been observed in the child language-acquisition literature (infinitive + *-ed*, for instance, *eated*, and past tense form + *-ed*, for instance, *ated*; Kuczaj, 1977), the network performed similarly to child language-learners, although this was true of only a subset of the verbs that the network was trained upon. Much as older children produce increasing proportions of forms such as *ated* relative to forms such as *eated*, over time the network also produced an increasing proportion of incorrect forms such as *ated*.
- With respect to the eighty-six low-frequency verbs (seventy-two regular, fourteen irregular) that were not a part of the network's original training set, Rumelhart and McClelland (1986) found that the network coped quite well with these novel items: 84 percent of the correct past-tense features were chosen with the novel irregulars, and 92 percent of the correct past-tense features were

chosen with the novel regulars. What is more, when the network was allowed to make unconstrained responses (that is, when the network was allowed to generate its own past-tense forms rather than select one possible past-tense form from among several provided to it), it very often freely generated correct irregular and regular past-tense forms.

Rumelhart and McClelland's (1986) network had accomplished all of its learning strictly on the basis of allowing its connection weights to be tuned by the statistical regularities present in its input (here, the 420 regular and irregular English verbs it had been trained upon). Crucially, the network did not rely upon rules to learn regular past-tense forms for novel verbs or to learn semiregularities among the irregulars (e.g., *sing-sang*, *ring-rang*, etc.). This is because in connectionist accounts of language processing (and in connectionist accounts of human cognition in general), instances of rulelike behavior may be observed, but these instances cannot be attributed to the application of rules because there are no rules anywhere in a connectionist system. Thus, in a system where all knowledge, linguistic or otherwise, is reducible to the sum total of inhibitory and excitatory connections between and within layers of nodes in a network, the notion of "rule" is viewed as nothing more than a convenient fiction devoid of ecological validity; while it may be possible to state a rule that adequately captures a widespread generalization such as "to form the English regular past tense, add *-ed* to the end of a regular verb's infinitival form," rules play no role in how the connectionist network actually learns that generalization. In other words, a rule may describe what the network knows, but it does not describe

how the network learns what it knows. Instead, learning occurs strictly through a probabilistic, data-driven (i.e., bottom-up) process.

An Answer to the Connectionist Challenge: Pinker and Prince, 1988

The results of Rumelhart and McClelland's (1986) study were highly provocative in that they called into question the classical (i.e., symbolic) approaches to the learning and mental representation of language. Symbolically oriented researchers were quick to respond to the connectionists' claims. In fact, an entire special issue of the journal *Cognition* was devoted to a critical examination of connectionism (volume 28, 1988). In that special issue, Steven Pinker and Alan Prince (Pinker and Prince, 1988) published a critique of Rumelhart and McClelland's study in which nearly a dozen objections were raised, pertaining to virtually all of the findings obtained with the neural network. For the sake of brevity, and to focus on the main shortcomings of the model, I will discuss here only Pinker and Prince's objections to the four main findings of Rumelhart and McClelland outlined above. With respect to the first of these findings, Pinker and Prince argued that in contrast to the input characteristics of the data that the Rumelhart and McClelland network was trained upon, parental speech to children does not change in its proportion of regular to irregular verbs over time, as longitudinal studies such as Brown (1973) indicate. Thus, there is a mismatch between observed input patterns to human children and the patterns in the input with which Rumelhart and McClelland trained their network. Pinker and Prince further reasoned that if children do not receive input pertaining to regular and irregular past

tenses in the proportions that the neural network did yet still go through a period of overgeneralization before gradually sorting out the regular and irregular past tenses, then children's overgeneralizations and u-shaped development must be due to factors other than environmental ones. Put differently, the network was able to model a period of overgeneralizations and u-shaped development, but it did so on the basis of input patterns that no human child is likely to be exposed to.

Concerning the second point above, that the network error patterns of nonchanging verbs such as *hit* were similar to error patterns observed with human children, Pinker and Prince argued that there are several equally plausible accounts (two of which are actually rule-based) for the network's early acquisition and overgeneralization of these verbs. They further argued that the results may be attributable not to the network's design *per se*, but to an unintended consequence of the phonological features used to represent infinitives and past-tense forms within the network. In the absence of evidence from studies of children's acquisition of groups of words that do not embody a confound between a common phonological property and the phonological shape of a suffix (for example, the group of nonchanging English verbs ending in *t* and *d*, when *t* and *d* are also regular past-tense allomorphs), Pinker and Prince argued that it is impossible to tease apart an input-based account from a rule-based account of how children learn to cope with this class of verbs. For these reasons, Pinker and Prince held that the network's being able to model the acquisitional data is itself not sufficient to demonstrate the superiority of the network over other possible accounts.

Concerning the third point made above, that the network performed similarly to child language learners by producing an increasing number of incorrect irregular past-tense

forms such as *ated* relative to forms such as *eated*, Pinker and Prince noted that this output pattern was observed only during the trials when the network was given a choice of responses (i.e., not during testing with the eighty-six novel verbs). They further observed that if the same strength-of-activation criterion used for determining a likely response from the network during testing with the eighty-six novel low-frequency regular and irregular verbs had been applied to these irregulars, the network would not actually have activated many forms such as *ated*. Pinker and Prince also cite a more tightly controlled follow-up child study (Kuczaj, 1978) in which forms such as *ated* were, over time, observed to increase but then decrease relative to forms such as *eated*. By this measure, the error patterns of children and of the network do not match so closely. The one plausible hypothesis for the source of the network's behavior with these items was incorrect blending. That is, with infinitival forms such as *eat* the network applies both the irregular past-tense change, producing *ate*, and the regular past-tense change, producing *eated*; the double marking results from a blending of the two different past-tense forms. Following a close examination of the relevant child data, Pinker and Prince observed that incorrect blending does not appear to be an active process in children. Instead, Pinker and Prince argued that in children, such errors most likely result from correctly applying an inflectional rule to an incorrect base form (i.e., applying *-ed* to *ate* rather than to *eat*) and not from incorrectly blending *ate* and *eated*.

With respect to claims of how well the Rumelhart and McClelland network coped with novel verbs, Pinker and Prince observed that approximately a third of the seventy-two novel regular infinitives prompted some form of an incorrect response from the network. Oddly, in a very few cases the model

failed to provide a response at all. Among actual responses that were incorrect, some involved stem vowel changes that had not been strongly represented in the training set (e.g., *shipt* from *shape*), some included double markings of the past tense (e.g., *typeded* from *type*), and some were simply difficult to classify (e.g., *squakt* from *squawk*). From these findings, Pinker and Prince concluded that the network had failed to learn many of the productive patterns represented in the 336 verbs (2 high frequency, 334 medium frequency) that it had been trained on. Moreover, Pinker and Prince suggested that the errors the network made were not reminiscent of errors that humans are likely to make.

Ultimately, Pinker and Prince concluded that the Rumelhart and McClelland PDP network fell short of offering a viable alternative to classical, symbol-based approaches to the learning and mental representation of the English past tense. Aside from the above-noted differences between human and network performance arguing against the network's viability as a model of human performance, Pinker and Prince also made a further, crucial observation: while the network was able to represent the phonological features of both regular and irregular infinitival and past-tense forms (through its use of distributed representations), by design it was unable to represent higher-order constructs (i.e., symbols) such as roots, infinitives, and past-tense forms. The inability of the network to represent these and other classes of linguistic objects meant that unlike humans, it was unable to form new words through well-attested processes of word formation such as reduplication (i.e., creating a new word by repeating the sound sequence of another word, for example, *bam-bam*, *boo-boo*, *can-can*, etc.). Consequently, since the model was limited to representing an infinitive not as a particular kind of linguistic object

(i.e., as an untensed verb) but as an undifferentiated set of activated phonological features, this means that it was unable to provide the past tense of any novel verb for which the infinitival form does not share enough features with the infinitives the network was trained on. Children are in fact able to provide past tenses of novel verbs, however (see the discussion at the beginning of this chapter). Pinker and Prince also offered evidence that a prediction the model makes for human language—that if in fact all verbs (irregular and regular) are simply mappings of phonological features to past-tense meanings, there should be no homophony (i.e., same-sounding words that have different meanings) among irregular verbs or between regular and irregular verbs—is not borne out by the facts of English. To take two examples, there are in fact irregular verbs, such as *ring* and *wring*, that have homophonous infinitival forms but different meanings and different past-tense forms. Additionally, there are regular-irregular inflectional contrasts such as with the verb *hang*, which in the past tense can be either regular *hanged* (i.e., executed) or irregular *hung* (i.e., suspended). Such contrasts should not exist according to a straightforward mapping of features to past-tense meaning.

Other language items, such as words, roots, and affixes, that argue in favor of representations, and by extension in favor of symbolic representations, include verbs that have been derived from nouns or adjectives. For example, one of the senses of the verb *fly* takes an irregular past tense (i.e., *flew*); another, baseball-related sense of the verb *fly* takes a regular past tense with *-ed* (i.e., *flied*, as in *He flied out to center field*). How can it be that one past-tense form of this verb is irregular, while the other is regular? The explanation for these and other similar regular and irregular past-tense pairings is that the root form of one of these senses—the derived sense—is not an irregular

verb but is instead a noun or an adjective. In the case of the derived sense of the verb *fly*, the root of the verb is most likely a noun, *fly ball*. Since its root is not an irregular verb, its past tense is not irregular, and thus its past tense is formed through adding the regular past-tense ending *-ed*. Through these and other examples, Pinker and Prince argued for the necessity of symbolic representations for words, roots, and affixes; comparisons of human performance and network capability suggest that distributed phonological features alone are not sufficient to give rise to the observable facts of the English past tense. The fact that the network would fail where people succeed points to a need to be able to represent words, roots, and affixes as abstract symbols.

In a subsequent set of observations in the same 1988 critique, Pinker and Prince systematically highlighted contrasts between regular and irregular past-tense verbs, with the goal of supporting their hypothesis that as a class, irregular verbs constitute a “partly structured list of exceptions” (p. 114). In contrast to regular verbs, classes of irregular verbs often bear phonological resemblance to one another (e.g., *blow*, *know*, *grow*, *throw*; *take*, *mistake*, *foresake*, *shake*; p. 115); often have a prototypical structure (e.g., all the *blow*-group verbs appear to have a prototypical phonological shape in that they tend to be consonant-liquid-diphthong sequences; p. 116); may have past-tense forms that are markedly less used, and therefore markedly more odd-sounding, than their present-tense forms (e.g., *bear-bore*), suggesting that for these verbs, the present- and past-tense forms have split in the minds of English speakers (p. 117); often have unpredictable membership, meaning that while a verb may have a strong phonological resemblance to a class of strong verbs, there is no way to predict from the verb’s phonological shape whether it is in fact an irregular verb

(e.g., *flow*, which resembles the verbs of the *blow* group but is regular; p. 118); and often exhibit morphological changes that appear to have no phonological motivation (e.g., the [ow] to [uw] present-past vowel change in the *blow* group is not conditioned by the vowel's surrounding phonemes; p. 119). Taken together these facts suggest that, again in contrast to regular verbs, the irregular verbs constitute a closed system, that is, aside from a very few exceptions, it remains a group of verbs to which no new members have been admitted in the recent history of the language. To explain the status of irregular and regular verbs in the minds of English speakers, Pinker and Prince (1988, p. 122) advanced the following proposal: in contrast to regular verbs, which are generated by a regular rule of past-tense formation and are thus not subject to factors related to human memory, the past-tense forms of irregular verbs are memorized, and as such are subject to memory-related factors—in particular, to variations in lexical frequency and to family resemblance.

Example of a Dual-Mechanism Theory: Words and Rules Theory

While not tested by Pinker and Prince (1988), words and rules theory would later be revisited and gradually elaborated upon by Pinker and his collaborators (see, e.g., Prasada and Pinker, 1993). In its most recent form (Pinker, 2000; Pinker and Ullman, 2002), words and rules theory holds that regular forms in English are the product of abstract, algebraic rules, while irregular forms in English are items that are stored in and retrieved from a partly associative lexical memory. Thus, in its division of language into words and rules, words and rules theory appeals to a traditional linguistic distinction between a

repository of words (a mental lexicon) and a productive, rule-based combinatorial system (a grammar).

Words and rules theory contrasts with connectionism and other single-mechanism accounts of language processing in that it does not hold that all processing proceeds from pattern association. As we saw earlier, a network relying upon a simple association of phonological features and meaning would simply be unable to produce (and therefore be unable to account for) a number of existing irregular forms and regular-irregular past-tense contrasts in English. Again, as we saw earlier, what the pattern associator lacked was a way of representing higher-order constructs such as roots, verbs, nouns, and words—that is, symbolic representations. As Pinker and Prince (1988) argued, it is only through appealing to these and other symbolic representations that the language data they discussed can be accounted for.

However, words and rules theory also contrasts with traditional, rule-based accounts of the morphological structure of English (Chomsky and Halle, 1968). Such accounts attempt to characterize both regular and irregular inflection in terms of underlying, or base, forms and phonetic, or surface, forms with a series of ordered rules applying to underlying forms in order to produce the surface forms we actually utter. For example, according to Chomsky and Halle (1968), the regular past-tense suffix *-ed* has an underlying form of /d/; depending upon the final consonant of the verb it is attached to, either a devoicing rule (that is, a rule that changes a segment from voiced to unvoiced) or a vowel epenthesis rule (that is, a rule that inserts a schwa between the final consonant of the root and the past-tense suffix when the two share common phonological features, specifically, place and manner of articulation) may apply to produce the appropriate surface form from the underlying regular past-tense form /d/. With a verb that ends

in a voiced consonant, such as *beg*, neither rule applies, and the regular past tense is *begged*, pronounced [bəgd]. With a verb ending in a voiceless consonant, such as *bake*, the devoicing rule will apply, producing the surface form *baked*, pronounced [beikt]. With verbs ending in a consonant that matches /d/ in terms of its place and manner of articulation, such as *fade*, the rule of vowel epenthesis applies, producing the surface form *faded*, pronounced [feidɪd]. Phonological rules such as these can accurately predict the form of any regular English verb. However, as mentioned earlier in the discussion of Pinker and Prince (1988), many of the present-past sound changes among the irregular verbs do not appear to have phonological motivation, i.e., they are not phonologically conditioned the way they are for the regular past-tense forms. Additionally, as observed by Pinker (2000), Chomsky and Halle's determination to capture all of the semiregularities with a very few rules ultimately led to their having to posit some underlying forms that, while plausible in the sense that they often reflected documented sound changes in the historical development of English, were almost unbelievable as viable representations in the minds of living speakers of English. To take one example, according to Chomsky and Halle the underlying form of the English verb *fight* [faɪt] is /fiçt/, in which the /ç/ represents a sound similar to the Germanic-sounding *ch* in *Bach*. Furthermore, adhering to rules at all cost sometimes meant having to make rather surprising claims concerning irregulars—for instance, that an irregular verb such as *keep* is in fact regular, in the sense that its past-tense form is fully derivable through vowel shifting, vowel laxing, and devoicing rules.

Words and rules theory holds that irregulars are neither the product of traditional phonological rules, nor the product of an unstructured associative memory. Instead, irregulars are said to be organized in memory in a fashion somewhat reminiscent

of semipredictive “lexical redundancy rules” (see, e.g., Aronoff, 1976; Jackendoff, 1975). Such rules do not build structure but instead impose constraints upon the range of possible structures specified by the structure-building rules. Recall that families of irregular verbs appear to have a prototypical phonological shape, yet they often have unpredictable membership and tend moreover to exhibit morphological changes that appear to have no phonological motivation. Pinker and his colleagues hold that these facts reflect the organization of irregulars in a partially associative memory that captures certain similarities among the irregulars. This organization serves to enhance a speaker’s recall of members of the irregular verb family; it also allows limited generalizations, by analogy, to novel forms from existing irregular forms. In contrast to irregulars, regular items tend to be the product of symbolic rules. In the case of the regular past tense, a rule-based (i.e., grammatical) process combines the past-tense suffix *-ed* with the root form of any regular verb to produce a regular past-tense verb. During actual language processing, these two systems are said to work in parallel. If a search of the partly associative memory turns up a stored form, then a signal from the associative-memory system inhibits the grammatical system from computing a regular past tense. If the memory-system search fails to turn up a lexical entry, however, then the grammatical system computes the regular past tense.

There are a number of empirical implications for the words-and-rules-theory-based claims made above. In the context of behavioral studies, which is the main focus of this book, these include the following:

1. To the degree that it is a symbolic, combinatorial process and not the product of a partly associative

- memory, regular inflection should pattern with acknowledged indices of symbolic processing.
2. To the degree that it is the product of a partly associative memory and not the product of symbolic processing, irregular inflection should pattern with acknowledged indices of associative memory.

Differently stated, regular and irregular inflection should be psychologically distinguishable.

Pinker (2000), Ullman (2001a), and Pinker and Ullman (2002) all offer evidence not only of the psychological separability of regular and irregular English inflectional processes, but also of their distributional and neurological separability. In the present work, I will be focusing on the applicability of words and rules theory to the online processing of French. As a backdrop to the experiments that I report in chapters 3–6, I will first introduce the notion of *priming* in the context of online studies. Following this, I present several English language behavioral studies that in addition to the work of Pinker and colleagues discussed above, have often been cited in support of words and rules theory.