22 Morphology in Language Production with Special Reference to Connectionism

JOSEPH PAUL STEMBERGER

Modern work on morphology in language production was begun by Meringer and Mayer (1895) and Meringer (1908). They observed that adult (and child) speakers of German occasionally regularize irregular verbs:

(1)	(a)	*erzieht*	for	*erzogen*	(infinitive *erziehen* 'educate')
	(b)	*gedenkt*	for	*gedacht*	(infinitive *denken* 'think')
	(c)	*heisste*	for	*hiess*	(infinitive *heissen* 'name')

After a long hiatus, MacKay (1970) brought attention back to these data, using them to argue that speakers actively use morphological rules to construct inflected forms, and that misapplication of those rules underlies the errors in (1). MacKay's work set the pattern for more recent work in several ways. First, most of it has focused on errors, rather than speed (reaction time). Second, most of it has focused on the question of whether speakers use morphological rules.

Early work was heavily based on linguistic theory, using the phonological and morphological concepts of linguistic theory and trying to demonstrate that they were psychologically real. Some researchers had no overt psychological theory of performance (e.g. Fromkin 1971), while others used a symbol-based serial theory of psychological processing in which it would be possible to instantiate a performance version of linguistic theory (e.g. M. F. Garrett 1975). G. S. Dell (1986) and Stemberger (1985b, c) introduced local connectionist models into research on morphology, but still focused on evidence for linguistic concepts such as rules. Rumelhart and McClelland (1986) brought distributed connectionist models into the picture, along with an attempt to eliminate rules from consideration. Researchers using symbolic models (e.g. Pinker and Prince 1988, Pinker 1991, Marcus et al. 1992) have argued that rules are necessary.

In this chapter, I explore the data and issues that have been brought up. First, I review the studies of the characteristics of morphology in language production. Given the focus of the field, most of this discussion addresses inflectional morphology. Next I review psychological models of morphology in language production, with an emphasis on connectionist models. A common theme is the question of whether there is any evidence that speakers use morphological rules during language production. I provide a summary of work thus far and comments about where the future will lead.

1 Empirical work on language production

Work on morphology in language production has focused heavily on errors. Much of it has to do with the expression of particular morphosyntactic features in a particular position in the syntactic structure. The phenomena can be typologized as follows:

(1) Right features, right place
  (1) wrong morphological pattern (regularization)
  (2) wrong base lexical item (and attendant agreement)
(2) Wrong features, right place
  (1) no affix when one should be there
  (2) an affix when none should be there
  (3) wrong affix (expressing the wrong features)
(3) Right features, wrong place
  (1) affix shifts, exchanges, anticipations, and perseverations
  (2) overtensing

Most of these error types occur with both inflectional and derivational affixes. A second area of interest is more overtly phonological (sections 1.4 and 1.5): when there is allomorphy, the allomorphy can be incorrect (right affix, wrong allomorph), or the allomorphy can be altered to accommodate a conditioning factor that was in error.

These errors provide information about several diverse areas. Errors where the wrong lexical item is present, or where the affix is expressed in the wrong syntactic position, reveal how syntax constrains the expression of morphosyntactic features. Getting an incorrect alternative pattern that expresses the same features has more to say about lexical processing. And allomorphy errors bear on the phonological processing of lexical items. As I discuss the phenomena, I briefly note what they reveal about processing. More details are provided in the later discussion of models of language production.

1.1 Right features, right place

1.1.1 Wrong morphological pattern: regularization and irregularization

Regularizations (as in (1) above in German) are commonly observed in English and other languages. In English, they have been observed with all inflectional affixes where irregulars are found: past tense, perfect aspect, present tense, and plurals.

(2) (a) I carefully looked at 'em & *choosed* – *chose* that one.
    (b) She's always *goed* – *gone* into these weird things.
    (c) She goes and *do-s* – *does* that.
    (d) . . . but two *childs* – two *children* usually isn't.

Regularizations are easily obtained from adults in experimental tasks such as *morphonaming*, in which speakers are asked to produce a particular morphological variant of a word given a different morphological variant presented visually or auditorily (MacKay 1976; Bybee and Slobin 1982; Stemberger and MacWhinney 1986a, b). To derive such errors, there must be some mechanism by which morphological patterns can generalize. One such mechanism, rules, adds an affix to a base form, and is blocked by irregular forms (e.g. Aronoff 1976, Kiparsky 1982a). In regularizations, a speaker fails to access the irregular form and instead accesses the base; the addition of the regular suffix is then automatic.

In some cases, the irregular form is accessed, but there is a failure of blockage: the regular rule applies anyway, to yield a *partial* regularization.

(3) (a) It *tooked* a while. '*took*'
    (b) . . . he doesn't have any . . . *lices*. '*lice*'

Full regularizations as in (1)–(2) are more common than partial regularizations (MacKay 1976, Bybee and Slobin 1982, Stemberger 1985c). From this, we can infer that failure to access an irregular is more of a problem than failure of blockage. This does not derive from any independent principle, but must be stipulated as an empirical finding.

Less common, and receiving less attention in the literature, are *irregularizations*: when an irregular pattern generalizes to another word, whether that word is properly regular or irregular (but following a different pattern; Bybee and Slobin 1982):

(4) (a) *cloamb* '*climbed*'
    (b) *crull* '*crawled*'

(5) (a) *flang* '*flung*'
    (b) *brung* '*brought*'

There must be some means to generalize irregular patterns as well. One issue is whether regular and irregular patterns can be treated (and generalized) in a parallel fashion, differing primarily in terms of likelihood of generalization, or whether they are produced via different mechanisms.

There are strong phonological influences on both regularization and irregularization. Bybee and Moder (1983) reported that nonce forms are more likely to be irregular if they bear a family resemblance to known irregular verbs (often referred to as the *hypersimilarity effect*; the errors in (5) above rhyme with known irregulars like *rang* and *slung*, and so are more common than the errors in (4) above, which are far less common. Stemberger and MacWhinney (1986b) showed that regular verbs with base forms that rhyme with families of irregulars (e.g. *thank*, which rhymes with *drank*, *sank*, etc.) are more likely to be produced without the *-ed*, so that they appear to be irregular: *thank* rather than *thanked*. Daugherty and Seidenberg (1994) show that regulars whose base forms rhyme with the base forms of irregulars take longer to produce than other regular verbs, suggesting interference from these groups of irregulars. Overall, data suggest that regulars and irregulars are both considered as candidates for the pathway used for irregulars.

Stemberger (1993), Stemberger and Setchell (1994), and Marchman (1995) show that the relationship between the vowel of the base form and the vowel of the irregular past-tense form is important. Given any two phonemes involved in a phonological error, phoneme A (the recessive phoneme) tends to be mispronounced more as phoneme B (the dominant phoneme) than the reverse (see Stemberger 1992b for discussion of English vowels and the underlying mechanisms). If the dominant vowel is in the base form, there is a phonological bias for the vowel of the past-tense form to be mispronounced as the vowel of the base form. Given that mispronunciation, regularization often follows; regularizations like *falled* (/ɑː/ dominant over the /ɛ/ of *fell*), *sinked* (/ɪ/ dominant over the /æ/ of *sank* or the /ʌ/ of *sunk*), and *throwed* (/ow/ dominant over the /uː/ of *threw*) are common. By contrast, if the dominant vowel is in the past tense, regularizations are rarer: *getted* (/ɛ/ recessive to the /ɑː/ of *got*) and *see-ed* (/iː/ recessive to the /ɑː/ of *saw*). Stemberger (1994a) further found that phonological priming from the subject NP also affects regularizations: when subjects must take an NP like *THE BALL* and a verb like *FALL* (where the noun and base form of the verb rhyme) and construct a sentence in the past tense, regularizations like *falled* are common (between 6 and 11 percent of tokens, depending on the nature of the priming); but regularizations are much less common if the noun is phonologically unrelated (as in *THE CONE* and *FALL*; *c.* 4 percent of tokens in error) or if the noun rhymes with the past-tense form (as in *THE BELL* and *FALL*; *c.* 3 percent of tokens). Such effects reveal that the processing system considers both the base and past-tense forms when producing the past-tense form, and that factors that favor the vowel of the base form (dominance or priming) lead to the system failing to retrieve the correct irregular past-tense form. As discussed below, not all models predict such phonological effects.

1.1.2 Stranding and accommodation There are errors that occur that could be analyzed as either lexical or syntactic, in which a word is inserted into a sentence at the wrong place. For example, given a sentence with an embedded VP under the main VP (as in *wanted to watch*), the two verbs involved could be reversed. When this occurs, we commonly observe two phenomena: *stranding*, in which the morphological affixes remain at the place in the sentence where they belong, and *accommodation*, in which the misordered words take on the inflected form that is appropriate to their new locations in the sentence (M. F. Garrett 1976, 1980). In (6a), for example, when *sound* is anticipated, it leaves -*ing* behind and acquires -*s*.

(6) (a) It *sounds* up – *ends* up sounding like 'split'.
(b) You just count *wheels* on a *light*. ('. . . *lights* on a *wheel*')

Stemberger (1985c) points out that these two phenomena are equally true of purely lexical errors, in which the wrong lexical item is accessed, since the form produced almost invariably has the inflected form appropriate to the syntactic context.

(7) That *understands* why he was that way. ('. . . *explains* why . . .')

Stemberger (1985c) also notes that there is one common exception to both stranding and accommodation: the plural -*s* affix often is exceptional, a fact that Stemberger attributes to the low level of syntactic constraints on plurality (8) and (9b) below).

(8) Your *teeth are* all red. ('Your *tongue is* all red.')

Stemberger argues that stranding and accommodation invariably occur when words of different syntactic categories are involved, but there can be exceptions when the two words are of the same category.

These facts are expected, given that we know from linguistic studies of syntax that inflections are often limited to particular positions in the sentence, such as tensed clauses, heads of NPs (but not nouns embedded in compounds), etc. They have no bearing on the issue of morphological rules, but reveal only that syntax constrains where inflections appear. If a word appears at an unexpected place in the sentence, it is subject to the constraints on that location in the sentence, so it loses any inflections that it would have had if it had appeared in its correct location, and takes on any inflections demanded by its erroneous syntactic position, with any irregularity appropriate to the word involved.

In syntactic environments in which one word agrees with another along some dimension, these lexical errors can lead to changes in other words. Stemberger (1985c) shows that changes in the subject (pro)noun can lead to changes of the verb from singular to plural or vice versa.

(9) (a) *You're* too good for *that*! (for '*That's* too good for *you*!')
(b) *Most cities are* true of *that*. ('*That's* true of *most cities*.)

Berg (1987), however, has shown that articles in German rarely accommodate to the gender and number of the displaced noun.

(10) Die wollen auch das *Welt* – das *Licht* der Welt erblicken.
('They want to see the (n.) world (f.) – light (n.) of day, too.')

Accommodation seems to be strong within a lexical item, but is less common in agreement with other words, which often agree with the word that they would have agreed with had the error not occurred.

1.2 Wrong features, right place

1.2.1 Base form errors In some cases, a speaker drops out inflections and produces the uninflected base (as found in English in singular nouns and in non-third-person present verb forms, infinitives, and imperatives).

(11) (a) They had cute little *mouse* on – *mice* on it.
(b) Boy, that *draw* him out. – *Drew* him out.

1.2.2 Affix addition Stemberger (1985a) reports that speakers are more likely to drop affixes than to randomly add them. He argues that this is a frequency effect, since the base form in English is usually more frequent than any particular inflected form. One contrast is particularly interesting: in the present-tense forms of most verbs, speakers tend to drop the *-s* affix, substituting the (more frequent) plural for the singular; but with the irregular verb *to be*, most errors involve replacing the plural form *are* with the (more frequent) singular form *is*. While such errors are compatible with the notion of rules, they occur with irregularities as well (to an even greater extent than with regular affixes; Stemberger and MacWhinney 1986a), and reflect only lexical accessing errors. Congruent with this, Stemberger and MacWhinney (1986a) show that base form errors are more likely for low-frequency verbs, even for regulars (though there is a much larger frequency effect for irregulars than for regulars).

Prasada et al. (1990), using reaction-time data rather than error data, have argued that irregular forms show frequency effects, but regular forms do not. They argue that this difference suggests that regular and irregular forms are processed in fundamentally different ways, with the irregular forms showing more word-like properties than the regular forms. Daugherty and Seidenberg (1994) also report this difference, but show (via a connectionist simulation) that a single mechanism can handle the differences. (See below for further discussion.)

1.2.3 Wrong affix Speakers sometimes produce forms with the wrong inflectional affix, given the meaning and/or the syntactic position of the word.

(12) She has *paying* – *paid* $15 for a blouse.

This seems to be an accessing error, where either the wrong inflected form is accessed in the lexicon, or the wrong inflectional rule is applied.

1.3 Right features, wrong place

1.3.1 Affix shifts, exchanges, anticipations, and perseverations Affixes can themselves be misordered: anticipated or perseverated, either being added to a word or replacing a different affix, being exchanged with another affix, or simply shifting so as to appear on the wrong word; these are rare but do exist (Stemberger 1985b, *contra* M. F. Garrett 1975).

(13) (a) I can't keep their *name straights*. ('. . . *names straight*')
(b) . . . where the safe *part* of the *cities* are. (. . . *parts* of the *city*)[1]
(c) . . . may take several years to be *masters*. ('to be *mastered*')
(d) I *wind* up *rewroting* 12 pages. ('I *wound* up *rewriting*')
(e) Rosa always *date shranks*. ('*dated shrinks*')
(Fromkin 1971)

Stemberger argued that such errors suggest that the affixes are separate morphemes, but the errors involving irregularities suggest rather that a syntactic error has been made, whereby the morphosyntactic features of the inflection are expressed in the wrong place in the sentence, in a way that violates the normal mapping rules of English syntax, rather than movement of affixes per se. Fromkin's example in (13e) suggests that shifts of features can even take place between words of different syntactic classes.

1.3.2 Overtensing In some instances, the morphosyntactic features are erroneously expressed twice. In particular, verbs are normally inflected for tense, but are uninflected when embedded under an auxiliary or modal. In overtensing errors, the verb is erroneously inflected for tense.

(14) (a) Did you *found* her? (for 'did you *find* her')
(b) Who does he *thinks* he is? (for 'does he *think* he is')

Irregular verbs are more likely to be overtensed than regular verbs. Stemberger (1992a) has shown experimentally that this may be due to the vowel-changing nature of most irregular past-tense patterns versus the suffixation nature of regular patterns; perfect *-en* is much less often involved in overtensing than other irregulars. Stemberger further shows that overtensing is more common with low-frequency verbs than high-frequency verbs. Stemberger and Setchell (1994) have demonstrated that vowel dominance has an effect: verbs with the dominant vowel in the past tense are more likely to be overtensed. Apparently, the phonological bias to produce the vowel of the past-tense form combines with the semantic appropriateness of the error to make the error more likely.

1.4 Allomorphy

1.4.1 Accommodation to other errors In many cases, an affix has multiple phonological realizations, depending on the phonological environment. The English *-ed* affixes have three allomorphs: /əd/ after /t/ and /d/, /t/ after other voiceless phonemes, and /d/ after other voiced phonemes. The *-s* affixes have /əz/ after strident phonemes, /s/ after other voiceless phonemes, and /z/ after other voiced phonemes. When phonological errors lead to a crucial change in the final segment of the base morpheme, the affix usually (but not always) takes on the appropriate shape for the new phonological environment.

(15) (a) The infant *tucks* (/tʌks/) – *touches* (/tʌčəz/) the nipple.
(b) The Swedish got *gooed* (/uːd/) up – *goofed* (/uːft/) up.

This accommodation also is observed with the suppletive variants of the articles: *a* and /ðə/ before consonants versus *an* and /ði/ before vowels.

(16) (a) . . . gets 20 miles *an allon* – *a gallon*.
(b) Put *the* (/ðə/) *hoven* – put *the* (/ði/) *oven* on 'hot'.

The most extreme form of phonologically conditioned allomorphy in human languages is reduplication, in which the affix takes on consonants or vowels from the base form. Stemberger and Lewis (1986) investigated reduplication in Ewe, using an experimental task that combined morphonaming with phonological priming designed to lead to phonological errors, and found that the affix usually took on the form of the initial consonant and vowel of the base when that was in error.

(17) (a) *haha fo* (for '*fafa ho*')
(b) *xaxa si* (for '*sasa xi*')

Phonological accommodation demonstrates that the *form* of the affix has not (usually) been finalized at the point in processing where phonological errors occur; allomorphy is determined either at the same time as or subsequent to phonological errors.

1.4.2 Wrong allomorphy In some instances, the wrong allomorph appears.

(18) Queen *Elizabeth'es* (/θəz/) – Queen *Elizabeth's* (/θs/) mother

Such errors are unusual, because incorrect allomorphs usually lead to a phonologically illegal sequence in the language (such as */kd/ or */čs/); speakers rarely produce illegal sequences (see Fromkin 1971) even in phonological errors. The allomorphs /əz/ and /əd/ are the only ones that are always legal, but even these rarely replace the other allomorphs, because /əz/ and /əd/ are low in frequency, and because these allomorphs require the addition of a syllable, something the phonological system is biased against (Stemberger and MacWhinney 1986a).

1.4.3 Affix checking Base form errors are especially prevalent when there is a close phonological similarity between the base and the affix. Bases that end in /s/ and /z/ (like the nonsyllabic allomorphs of the *-s* affixes) tend to drop the *-s* affix. Bases that end in /t/ and /d/ (like the nonsyllabic allomorphs of the *-ed* affixes) tend to drop the *-ed* affix.

(19) (a) So we *test* 'em on it. (for '. . . we *tested* 'em')
(b) It just *lose* something. (for '. . . just *loses* something')

This is true of both adult speech (Bybee and Slobin 1982, Stemberger and MacWhinney 1986b), and child speech (Berko 1958, Bybee and Slobin 1982). MacWhinney (1978), Stemberger (1981), Bybee and Slobin (1982), and Menn and MacWhinney (1984) suggest that such errors arise because the base already appears to be inflected, and so the speaker does not add an "additional" inflection. Pinker and Prince (1988) suggest that it might be phonological dissimilation between the base and affix consonants (but see Stemberger 1981, Menn and MacWhinney 1984). Rumelhart and McClelland (1986) and Pinker and Prince (1988) suggest that it might be the overgeneralization of the no-change pattern of verbs like *hit* and *hurt*; but this is not the whole story, since such errors are also common with the *-s* affixes, where no-change irregulars do not exist.[2]

1.5 Derivational morphology and compounding

The bulk of studies of language production in normals has focused on inflectional morphology, but something is known about derivational morphology (MacKay 1978, 1979; Stemberger 1985c). The small number of studies is related to the fact that inflectional morphology is more frequent in spontaneous speech, so errors there are more noticeable.

1.5.1 Derivational morphology: base form/ wrong affix errors As with inflectional morphology, derivational affixes can be left out, added, or replaced by an inappropriate affix.

(20) (a) It's not *mass* enough – *massive* enough for a sun.
(b) If you're *hunger* – *hungry*, you should've . . .

(21) (a) I'm Tony's *brother-in-law*. (for 'Tony's *brother*')
(b) It was when they were first *marriaged* – *married*.

(22) (a) He was a *philosophist*, wasn't he? (for 'a *philosopher*')
(b) It's an arbitrary *decidal*. (for 'an arbitrary *decision*')

Inappropriate affixes are usually only inappropriate lexically; they represent a competing affix with the same meaning that simply cannot be used with that particular word; these might be called "regularizations," though it is often hard to determine which derivational affix is the "regular" one. With semantically transparent derivational affixes, loss errors are common, but addition errors are not (Stemberger 1985c). Apparently, the transparency of the semantics leads to the base form being a strong competitor which, because of its higher frequency, is more likely to win out when inappropriate (similar to the "hypernym problem" of Levelt 1989); but the derived form is unlikely to win out when it is inappropriate. With semantically opaque affixes, however, loss and addition errors are more balanced. Apparently, the differences in meaning between base and derived form lead to less semantic interference between them, so that the base is not activated as much when the derived form is appropriate, and vice versa; thus, substitution of one for the other is likely in either direction.[3]

Affixes can be anticipated or perseverated from a nearby word.

(23) (a) This longish *woodish* – longish wooden object is . . .
(b) . . . brood reduction and hatching *asynchrontion* – asynchrony.

1.5.2 Stranding and accommodation As with inflections, derivational affixes can be left behind when a word is inserted into the wrong syntactic position in the sentence, whether regular or irregular (MacKay 1979, Stemberger 1985c).

(24) . . . makes no pretense of *pretending* – of *preparing* you for . . .

This may be purely lexical, with the syntactic demands of a particular position in the sentence requiring a particular derived form. It should be noted that the *addition* of a derivational affix to render the correct part of speech seems to be uncommon, while the *loss* of a derivational affix to yield the right part of speech seems to be common. It appears that derivational affixes *require* that the words containing them be of a particular part of speech, and they can readily be eliminated when the word appears in the wrong syntactic position. However, errors rarely occur in which a derivational affix is added *solely* to change the part of speech of a word. The part of speech of a monomorphemic word in English is quite labile, and can change (e.g. from noun to verb, or vice versa) without the addition of any overt affix. If a monomorphemic word erroneously finds itself in a syntactic position requiring a different part of speech, it follows this no-affix pattern. This suggests that derivational affixes may be added partly for semantic or pragmatic reasons, and that syntax alone does not force the addition of derivational affixes.

1.5.3 Stress and vowel patterns Fromkin (1971) first noted errors where the wrong stress and vowel patterns were present.

(25) (a) . . . inherent linguistic *suPERiority* in women. (for '*superiORity*')
(b) *ecoNOMists* (for '*eCONomists*')

(26) (a) Oakland gets all the *inDUStry*. (for '*INdustry*')
(b) . . . in the paper at the *secreTARy* jobs . . . (for '*SECretary*')

Fromkin interpreted these as demonstrating that the stress and vowel reduction rules posited for English by Chomsky and Halle (1968) were psychologically real. However, the exact details of how the errors occur were not given. Those in (25) seem reasonable, since an affix is simply ignored (thus shrinking the phonological domain of the rule, which entails a different resulting stress pattern). But it is unclear how the errors in (26) occur, where the rules act as if additional syllables were present.

Cutler (1980) (replicated by Stemberger 1985c) showed that such errors rarely occur unless the target has a derivationally related word with a different stress pattern. It seems as if speakers are "borrowing" the stress and vowel patterns from another form of the word. Cutler argues that suppletive stems are stored and retrieved without application of stress or vowel reduction rules (e.g. *econom-* as both /ˌiːkəˈnɑm/ and /ˌiːˈkɑnəm/), and that different suffixes choose different allomorphs. Stress errors arise when a suppletive stem is accessed that does not normally go with the suffix (or lack of a suffix) that is present in the target word. No one has yet proposed a stress-rule-based solution that accounts for the facts.

1.5.4 Compounding Errors of compounding are rare. They are of two types. First, one member of a compound may be dropped, sometimes leading to a blend between the two parts of the compound.

(27) (a) This is *scotch*. – *Hopscotch*.
(b) How many *blerries* did you get? (for '*blueberries*')
(c) . . . a lot of Welch's *jape* commercials . . . (for '*grape jam*')

Second, the two words in the compound can be reversed (Meringer and Mayer 1895, Stemberger 1985c).

(28) (a) You were just closing the *lidboxes*. – The *boxlids*.
(b) That's a *busbike*. – I mean, a *bikebus*.

These errors implicate a processing of compounds that is vaguely syntax-like, with two positions available. Loss of a position (yielding a more frequent non-compound noun) leads to loss of the occupying noun, or to a word blend. Errors can occur in which the two nouns exchange their positions, just as errors occur in which any two nouns in a sentence are exchanged. In compound-noun reversals the stress is stranded (so that both the target and the error compounds have the same stress pattern; stress does not stay with the particular nouns, so that the stress pattern would change when the nouns exchange; the stranding of stress is also observed with true word exchanges (Stemberger 1985b). Lexical processing in compounds seems to engage some of the same mechanisms as lexical processing in general.

2 Morphology in models of language production

2.1 Rule-based models

There has been relatively little work on morphology in language production in nonconnectionist models that involves the building of explicit models. Fromkin (1971) assumed something roughly corresponding to morphology in linguistic theory, but gave few details about organization, how regularizations occur, etc. Cutler (1980) also assumes such a system, but focuses primarily on derivational affixes; she assumes that, for example, *-ity* and *-ness* are produced via rules, but that the bases to which they attach may be suppletive from the independent base word (i.e. /owpæs/ is suppletive from /owpeyk/ in *opacity*). Butterworth (1983) addresses regularizations: irregular forms and known regular forms are stored in the lexicon; if a speaker is unable to access a stored form (in error, or because the target word is a novel word), then the base form is accessed, and the regular inflectional rules of the language are applied. M. F. Garrett (1975, 1976) discusses morphology and stresses the difference between regular and irregular forms, but the details of the morphological component of the system are left unspecified. MacKay (1970, 1976, 1978, 1979) argues that rules are used for both regular and irregular forms, but provides little detail about the workings of the rule component.

Currently, the main psycholinguistic work in this area is being done by Pinker and his colleagues (Pinker 1991; Pinker and Prince 1988; Kim et al. 1991; Marcus et al. 1992, 1995), with a focus on language acquisition. Although the details of the model are still being worked out, the basic outline is clear.

There are two different ways in which inflected forms are processed, one path for irregulars and one for regulars.

Regular inflected forms are not stored in the lexicon at all (except at *very* early stages in language acquisition). The base form of the word is accessed. Then, the regular default rule is applied. The exact details of the rules have never been stated, but in English they are simple and concatenative (such as "add /d/ at the right edge of the base verb"). Rules are not sensitive to accidental phonological properties of the base word (though they *can* apply to words with only specific phonological characteristics, such as those that end in consonants).

There is a separate system to handle irregular words, set up in such a way that it has priority over the regular rule. The exact nature of this priority has not been spelled out. It could be serial priority: try the irregular system first to see if the word being processed is irregular; if it is irregular, bypass the regular system; if it isn't irregular, proceed to the regular system. It could be speed-of-processing priority: try both the regular and irregular systems, so that both a correct irregular like *chose* and the regularization *choosed are computed simultaneously, and select for production whichever pathway finishes first (a standard horse-race model); the irregular pathway would be faster for known irregular forms, so *chose* would usually win.

The irregular system is *not* one in which the irregular form is simply stored in the lexicon, either as an independent lexical entry or as a sub-entry under the base form. Instead, it is an associative network (much like a distributed connectionist model, on which see below). The reason for this is that some account must be given for hypersimilarity effects (section 1.1.1 above), and the symbolic processing model that is assumed apparently cannot compute such similarity naturally (though no details of the system are given, so this is uncertain). Because all verbs (regular or irregular) are checked for irregularity, regulars are subject to the same effects of similarity to families of irregular verbs (error rates and reaction times) as irregular forms (Daugherty and Seidenberg 1994). On occasion, an irregularization results, when an irregular pattern generalizes incorrectly. Note that a concession is made that connectionist models are needed to account for language production, but only for the irregular forms of a language. Bybee (1985, 1995a) disagrees with this concession, and provides a mechanism for accounting for these effects within a symbolic model (based on schema theory).

The standard linguistic account of irregular forms (e.g. Kiparksy 1982a) assumes that irregular forms are also produced via rules, but the rules are lexically bound (i.e. they apply only to words that are marked to undergo them). There is one argument against this: the lexical frequency effect, whereby error rates are higher on low-frequency words. A lexical frequency effect seems easy to derive as a lexical-accessing effect if irregulars are stored in the lexicon, but seems difficult to derive if all irregulars of the same pattern are produced via the same rule. The facts seem to demand lexical storage for irregulars rather than minor rules. However, this is not necessarily the case. If a verb has a feature such as [+æ-rule), that feature will be strong in frequent words and weak in infrequent words. If the mechanism responsible for applying irregular rules has a more difficult time reading the weak feature of the low-frequency verb, then low-frequency verbs would be more likely to be regularized. A lexical frequency effect is thus compatible with a system that uses rules for irregulars, and this is still a possibility (however unpopular).

None of the (nonconnectionist) rule-based accounts have really attempted to work out in detail how morphology works in a psychological processing model, or how most of the phenomena surveyed above are derived by the model. There has been a general feeling that this would not be difficult, but no one has actually attempted to do it.

2.2 *Connectionist models*

Within connectionism, one can distinguish several major types of models. Although there are many dimensions involved, the major division between classes of models lies in the way that information is represented. If a concept (semantic, lexical, etc.) is represented as a *single node*, then the representations are *local*; thus the word *dog* is represented with a discrete node in the system (McClelland and Rumelhart 1981, Dell 1986, Stemberger 1985b). If a concept is represented as a *pattern of activity* across *a set of nodes*, however, then the representations are *distributed*; there is no single node that corresponds to the word *dog*, but only a particular pattern of activation over a set of units that are used for all words (and each word has its characteristic pattern of activation). The local/distributed distinction is not that clear; as noted below, no local connectionist model has ever been truly local. However, the distinction is correlated with another very important factor: learning. There is no explicit learning algorithm for local connectionist models, while distributed models are linked to learning algorithms (such as back propagation). The field underwent a shift in the latter half of the 1980s from local to distributed models, primarily because it was judged important to have a model of learning.

There are other major dimensions that distinguish models, however, which cut across the local versus distributed distinction. The one that I consider the most important deals with the direction of flow of information. In local models, information flows in both directions between any two connected nodes. This means that there is no strict modularity between adjacent levels (though Dell and O'Seaghdha 1991 demonstrate that functionally there is modularity between levels that are not directly connected). Lexical and phonological processing are done simultaneously and interactively, with each influencing the other; such interactions dominated the local connectionism literature. Early distributed models, by contrast, were entirely unidirectional (or *nonrecurrent*). In language production models, information flowed from semantics, through the lexicon, to the phonology, and never in the reverse direction. More recently, *recurrent* models have been developed, in which lexical and phonological processing are intermixed.

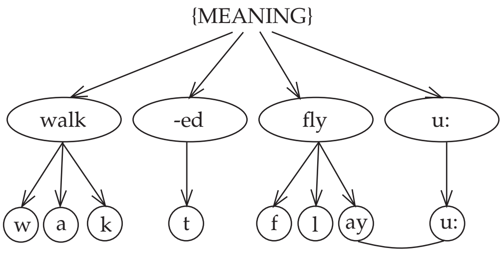

Figure 22.1 General rules within a local connectionist model.

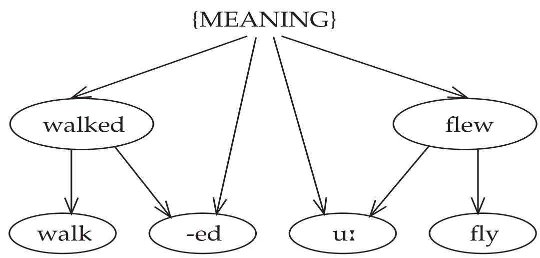

Figure 22.2 Lexically bound rules within a local connectionist model.

All varieties of connectionist models have a lot in common, and show many of the same behaviors. The current theoretical preference is for distributed models. However, in models of language production, local models are at this point more complete; their usefulness is limited primarily by the fact that there is no learning algorithm associated with them. Recurrent distributed models show many of the strengths of local models, and future morphology models will probably be of this type.

2.2.1 Local connectionist models All lexical and phonological elements are represented as discrete nodes and are activated simultaneously. For example, the node *walk* connects to the nodes /w/, /ɑː/, and /k/, and /k/ connects to the nodes [Dorsal], [-voice], [-continuant], etc. There are many schemes for encoding serial order (so that the words *cat*, *tack*, and *act*, which have identical sets of segments in different orders, can be distinguished; Rumelhart and McClelland 1981, Dell 1986); this issue goes far beyond morphology, and I will not address it here. Within a local model, it is possible to have something corresponding to a morphological rule: there may be a node for a suffix such as *-ed* (Stemberger 1985b, Dell 1986); if this node is activated along with the node for a verb, the verb is inflected for the past tense. Stemberger (1985b) distinguishes three positions: general rules (with nodes for affixes activated only by semantic/syntactic information, see figure 22.1), lexically bound rules (with the affix activated both by semantic/syntactic information and by lexical items, see figure 22.2), and non-rule-based representations (where inflected

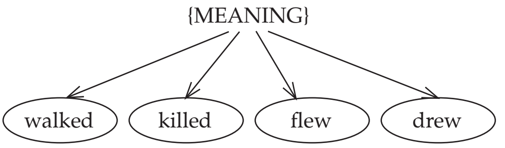

Figure 22.3 A rule-less local connectionist model.

forms are just stored as words; *walk* and *walked* are separate nodes, as are *tree* and *sky*, see figure 22.3). In theory, the three positions can be distinguished, but in practice, this is difficult, because they predict many of the same phenomena. Irregular patterns are lexically bound and cannot appear freely with just any word. This rules out only the general-rule approach. Regular patterns could in theory be instantiated in any of the three ways.

In local connectionist models, co-activated items reinforce each other. A co-activated base verb reinforces the syntactic slot with which it is associated, and thus indirectly reinforces any rule that is based on that slot. In the case of a lexically bound rule, the base verb also directly reinforces the rule. The amount of activation that a node passes is correlated with the frequency of the node: low-frequency nodes pass less activation than high-frequency nodes. Consequently, low-frequency elements provide less activation for all other connected nodes in the system, putting all other nodes at greater risk for error. Stemberger and MacWhinney (1986a) and Dell (1990) report that low-frequency words are more likely to undergo phonological speech errors than high-frequency words. This explains why low-frequency verbs (whether regular or irregular) are more prone to morphological errors: regularizations, base form errors, and overtensing errors (sections 1.1.1, 1.2.1, 1.3.2). Lexical frequency effects are compatible with any of the ways to store inflected forms.

If the system uses rules, generalization of patterns is easy to derive. The most frequent rule will tend to be accessed with most verbs, especially if it is not lexically bound. The less frequent (irregular) rules tend to be suppressed by the regular rule except under those circumstances where they get the most reinforcement. The most favorable circumstances include phonological reinforcement by groups of irregular verbs that take that rule: hence hypersimilarity effects (section 1.1.1). (Any hypersimilarity effects on regulars are swamped out by frequency effects, so it is not surprising that none have been detected.) Also favorable would be those situations where the past-tense vowel is dominant over the base vowel, so that it does not require as much activation to suppress the base vowel: the vowel-dominance effect. And phonological priming (as from the subject noun) should have an effect, positive or negative.

Local connectionist models do not require morphological rules. Inflected forms can be stored as words, parallel to any other word (fig. 22.3). The processing of known words is straightforward. Stemberger (1994b) discusses how generalization occurs via *gang effects*. Within local models, similarity between different lexical items leads to a group of partially activated nodes called a *gang*;

gangs form on the basis of both semantic and phonological similarities. Via inhibition between competing lexical items, nontarget words are kept at a low level of activation, but they still contribute small amounts of activation to the phonological level. In general, nontarget words contain sounds that cover the full spectrum of the phonological space of English, and activation from nontarget words has little effect on the output; different words cancel each other out, and just raise the level of noise in the system. However, if the words in a semantically based gang are correlated phonologically, then the effects are quite different. For example, if we consider 1,000 nontarget verbs that are past-tense forms, we would find that about 850 of them end in *-ed*, with the frequency of allomorphy being /d/ > /t/ > /əd/. Hundreds of words contribute a small amount of activation to /d/. The /d/ unit sums this activation, and the result is that /d/ gets more activation than any phoneme within the target word. However, when /d/ is unlikely because it creates a consonant cluster that is impossible in English (as in **walk-d*), it attains a lesser degree of activation, and the second most frequent past-tense pattern, /t/, wins. If neither /t/ nor /d/ is phonologically possible (as in **need-d* or **need-t*), then the third most frequent pattern, /əd/, wins. No irregular pattern has more than thirty exemplars, and so none gains enough activation to win – *unless* the irregular lexical item suppresses the regular gangs (which is usually the case in adults) or *unless* a phonologically based gang can also form, reinforcing a particular sequence of vowel + consonants that does not end in *-ed*; thus families that end in *ank* and *unk* can generalize to new forms, but only by supplementing semantic information with phonological information. The regular patterns are present in so many lexical items that phonological information is not needed for generalization; phonological effects are swamped out, and hypersimilarity effects are not observed with regular patterns. If the target irregular verb fails to suppress the regular gang, a regularization results; failure should be greatest for low-frequency verbs, whose phonological information is least well encoded. Failures decrease if phonological effects like vowel dominance favor the past-tense form.

Gangs give the lie to the characterization of local models as "local," as opposed to "distributed." All past-tense forms are processed whenever any past-tense form is processed, and together all these forms influence the output. One could say that the representation of any particular past-tense form is distributed across all the past-tense-form nodes in the system. Stemberger (1994b) raises an interesting possibility. How does a learner know when to add a new word to the system? Suppose that a new word is added only when the speaker would otherwise produce the wrong output. When regular gangs reach a certain size, they automatically cause generalization, so that the correct (regular) past-tense form is produced for less-frequent and novel verbs. The speaker would reach a point where new regular past-tense forms would not be added to the system, because the right output would arise without them. The gang comes to function as a distributed rule, and like a rule, the *-ed* pattern is relatively independent of any individual lexical item.

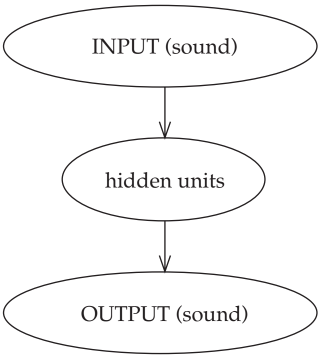

Figure 22.4 A nonrecurrent distributed connectionist model.

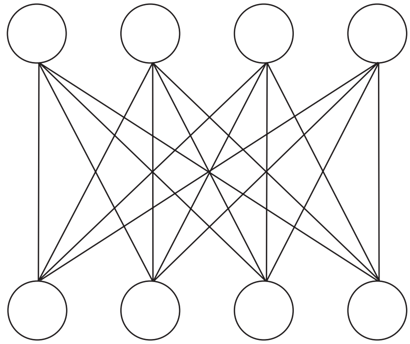

Figure 22.5 Connections between layers in a distributed model.

An important aspect of this model is the interaction between the lexical and phonological levels. Any phonological bias is predicted to affect morphology, including vowel dominance and priming. Additionally, the morphological system should tend to avoid outputs that contain less-frequent phonological patterns. Thus, there would be a tendency to avoid consonant clusters, to output words with fewer syllables, to favor high-frequency phonemes, etc. In this context, it is interesting that Clahsen et al. (1992) and Marcus et al. (1995) argue that for German the regular suffixes are perfect -*t* and plural -*s*, and that this would not be predicted on the basis of frequency. Perfect -*t* and -*en* are equally frequent as perfect suffixes for the 1,000 most frequent verbs of German (tied in both type and token frequency), so (morphological) frequency cannot explain why -*t* is the regular suffix for the perfect form, and not -*en*.[4] Plural -*s* is the least frequent plural suffix in German, but, they argue, is the regular suffix. However, phonological frequency may be relevant here. It happens that -*t* and -*s* are the only overt suffixes that (in general) do not add a syllable to the word; I note below that shorter words are more frequent than longer words in German. The predictions of connectionist models are unclear when two types of frequency conflict: high morphological frequency combined with low phonological frequency, versus low morphological frequency combined with high phonological frequency. German morphology may be an instance where phonological frequency outweighs morphological frequency in terms of which pattern is preferentially generalized; see below.

2.2.2 Distributed connectionist models I: nonrecurrent networks A typical nonrecurrent net is shown in figures 22.4 and 22.5. McClelland and Rumelhart (1986) developed the first nonrecurrent distributed model of morphology. The input was a distributed representation of the base word's pronunciation (e.g. /waːk/ *walk*, and /gow/ *go*), with all phonemes represented simultaneously. These were mapped directly onto an output representation of the pronunciation of the past-tense form (/waːkt/, /wɛnt/); all input nodes connected to all output nodes, as in figure 22.5. Both regular and irregular forms were stored in the same set of units and connections. The advantage of this model over local models was that there was an algorithm for learning (back propagation). The model produced regularizations, and (given family resemblance) irregularizations. Regularizations predominated, because the learning algorithms of distributed models extracted the most frequent pattern as the default one.

This was a primitive model (lacking even a layer of hidden units between the input and output layers), and has been heavily criticized, for example, by Pinker and Prince (1988). Most of the criticisms are of non-essential details, such as (a) the nature of the phonological units (wickel features) used to encode serial order, (b) the inability to differentiate homophones, (c) the temporal structuring of the training (done in an artificial way so that a U-shaped learning curve was derived), (d) the presence of a "teacher," disagreements over whether certain types of errors occur during acquisition, and (e) "cheating" by including standard aspects of linguistic representations in the input and output. (a) More recent models have used more conventional phonological representations (MacWhinney and Leinbach 1990, Plunkett and Marchman 1991, Hare and Elman 1992, Daugherty and Seidenberg 1994), and have derived similar results. (b) MacWhinney and Leinbach include semantic information, and have found that this allows the system to differentiate homophones (such as *ring* – *rang* vs. *wring* – *wrung* vs. *ring* – *ringed*) without compromising the system's ability to generalize patterns on a phonological basis. Marcus et al. (1995) have criticized the fact that MacWhinney and Leinbach used only a handful of semantic features, but this is a non-essential aspect of the model, deriving from limitations on the size of simulations; Marcus et al. admitted that the homophone problem can be solved in this way. (c) The artificial and incorrect way that Rumelhart and McClelland (1986) derived U-shaped learning is not a concern, because it now appears that U-shaped learning of the sort that had been assumed is unattested in children's acquisition of morphology (Marcus et al. 1992). (d) The "teacher" that told the system whether the output was correct or incorrect has been criticized, because children rarely get overt correction from adults. This is a misunderstanding of the nature of the "teacher," which is simply another cognitive subsystem, possibly the comprehension system. *All* models require some subsystem that can recognize that an error has occurred. (e) Although Pinker and Prince protested against outputs like *membled* as the past-tense form of *mail*, it should be noted that young children often pronounce words in ways that are quite different from adults (e.g. Priestly 1977, in which a child pronounced words such as *panda* /pandə/ as [pajan], and *dragon* /drægən/ as [dajak]), and all theories of acquisition probably predict that child forms like [mɛmbəld] for adult /meyld/ are possible.[5] This criticism is not based on any theory of language or of acquisition, and seems to lead to a position that would be unable to account for normal language acquisition. As a result, it cannot be taken as a failing of connectionist theories.

(f) Connectionist models often make the same assumptions about representations as linguistically based models. This is not a problem if we accept that all approaches to cognition have something of value to offer. It is no more "cheating" to include such analyses in connectionist models than it is to include them in symbolic models.

Current arguments about the feasibility of connectionist models concentrate on three issues: the lack of hypersimilarity effects for regular patterns, the role of semantics in determining whether a verb form is regular or irregular, and whether it must always be the most frequent morphological pattern that is the regular one. (a) As with non-rule-based local models, the most frequent pattern tends to generalize so readily that it is difficult to detect effects of factors like hypersimilarity or even lexical frequency (Daugherty and Seidenberg 1994). (b) The role of semantics is unclear. Kim et al. (1991) and Marcus et al. (1995) have maintained that semantic effects are irrelevant, and that symbolic rule-based deletion of features for irregularity (as when a verb is nominalized and then subsequently changed back into a verb, or when an irregular noun is made into a proper noun) are unaffected by how similar the resulting form is semantically to the regular verb or noun. They also maintain that metaphorical and semantic extensions of verbs and nouns always preserve the irregularity of the word. However, extensions of words always involve close semantic similarity, whereas denominal and deverbal verbs are usually more distant semantically. Kim et al. and Marcus et al. actually show that their predictions are incorrect: the probability that a denominal verb will be irregular is linearly correlated with the semantic similarity of the denominal verb to the prototypical usage of the homophonous verb. Thus, *sink* 'to put in a sink' (which has no semantic similarity to the usual verb *sink* 'to go down') has regular past *sinked*, but *fly* 'to hit a fly ball' (which is more similar to the usual verb *fly*, almost 'to cause a ball to fly high') has two equally acceptable past-tense forms, regular *flied* and irregular *flew*. The role of semantics is far from clear. And while it is possible to make the effects of semantics indirect (as in Kim et al.'s and Marcus et al.'s models, where they arise only during the process of learning lexical items), it remains possible that semantics has a direct effect.[6] (c) As noted above in the discussion of local models, it is an oversimplification to focus just on the frequency of the morphological pattern, ignoring the frequency of the phonological patterns that result. In some cases, a less frequent morphological pattern may be preferred because it creates the most frequent phonological pattern. When different measures of frequency conflict, predictions are not clear, for any connectionist model.

In relation to this last point, it should be noted that many nonrecurrent models (e.g. Rumelhart and McClelland 1986) do not predict that general phonological frequency in the language will affect morphological patterns. The reason is that such models have a special dedicated network, the sole purpose of which is to create past-tense forms. Such a network will pick up on statistical properties of past-tense forms only. If, however, the network produces all words (singular nouns, plural nouns, possessive nouns, infinitives,

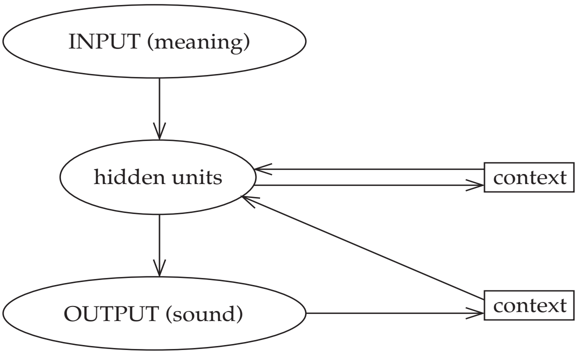

Figure 22.6 A recurrent distributed connectionist model.

present-tense forms, past-tense forms, progressive forms, adjectives, adverbs, etc.), then the statistics for all words will matter, not just the statistics of past-tense forms. In all probability, models cannot have subnetworks dedicated to particular morphological patterns. However, nonrecurrent networks do not require dedicated subnetworks (MacWhinney and Leinbach 1991), and so are still viable candidates for describing morphology in human languages.

2.2.3 Distributed connectionist models II: recurrent networks Recent connectionist models are most often recurrent in nature: they contain loops that allow multiple passes through the system, with each pass corresponding to a phoneme (in models that generate the phonological output of words). Only a few models of morphology have been recurrent (e.g. Corina 1991), but this is likely to change. The basic structure of such a system is illustrated in figure 22.6 with the model of Dell et al. (1993), which was designed to address phonological processing; the model takes meaning as input and gives phonological features as output.[7] Each word is a pattern of activity across the meaning feature nodes and a pattern of activity across the phonological feature nodes, and these two layers are mediated by a layer of hidden units that maps the input onto the output. This model has two context layers to keep track of previous outputs.

Early distributed models, like local models, output all the segments of a word simultaneously. Recurrent models output a single segment at a time. The first pass through the system outputs the first segment of the word. The pattern of activity in the hidden units is saved in an *internal context* layer. The pattern of activity in the output layer is stored in an *external context* layer. Both context layers are then input into the hidden units on the second pass through the system, along with the *same* semantic pattern that was used during the first pass. That semantic pattern, in combination with context information about what was just produced, leads to the output of the second segment of the word on the second pass; without the context layers, the first phoneme would be output again. The pattern of activation in the hidden units and output units in the second pass are stored in the context units and input through the hidden units on the third pass, resulting in the third segment of the word. This continues until the system returns a null element that corresponds to the word boundary.

This recurrent model has interesting ramifications for morphology, especially for the regular/irregular distinction. First, the system can learn statistical probabilities between meaning units and phonological segments. Thus, it learns that there is a strong statistical correlation between the meaning {past tense} and a word-final /t/, /d/, or /əd/. Second, the system extracts statistical properties of a phonological nature as well, including *phonological constraints*. This phonological information can interact with morphology in predictable ways. The difference between the three allomorphs depends on phonological information contained in the context units; it is predictable from general phonological distribution patterns on voicing and sequences ending in /t/ and /d/ that are present even before any morphology is learned by the system. The correlation between any particular semantic unit and these final segments is small, and is equal for most input patterns; thus, the system will not tend to learn that a particular verb (like *walk*) takes *-ed*, but will learn that {past tense} maps onto *-ed*. The two local models in figures 22.1 and 22.2 above resemble this model; it is most like figure 22.2, in that there is input from both the meaning unit {past tense} and the /d/ output unit, and there is a lexical-item-specific component; but it approximates figure 22.1 in that the lexical-item-specific component for *-ed* is small. Unlike the distributed models with dedicated subnetworks for each inflectional category, this network outputs all morphological variants of the base word, including uninflected forms. As with local models, there is no mapping from the phonology of the base onto the phonology of the past tense.

Irregulars behave quite differently. Statistically, irregulars like *sang* and *fell* contain vowels that in general do not correlate with past tense any more than any other vowel (cf. base verbs such as *crack* and *rest*), and they fail to end in the *-ed* which correlates very strongly with past tense. The system must learn to output a vowel different from that of the base, as well as to suppress the final *-ed*. It learns both of these by being sensitive to the co-occurrence of elements in the semantic input; thus, given the co-occurrence of {past tense} and {fall}, the output will contain the vowel /ɛ/ but not the final consonant /d/. The final /d/ will also be suppressed using phonological context; given that the vowel /ɛ/ occurred with this semantic input, the system suppresses the final /d/, and outputs just /fɛl/.

Consider what happens when the system attempts to output the past tense of the word *fall*. The meaning maps onto /f/ in the first pass. In the second pass, in combination with the context units, the meaning of {fall} could map over onto either /ɑː/ (*fall*, *falls*, *falling*, *fallen*) or /ɛ/ (*fell*). Given that {past tense} is present, /ɛ/ will usually be accessed, but consider what happens if /ɑː/ is erroneously accessed. This will alter the information in the context units. This will not affect the phoneme accessed on the third pass, since /l/ appears in all forms of the word. However, the altered feedback makes the system less able to suppress the /d/ that is activated in the word-general meaning–form mapping. If this word-general mapping is weak, as in early speech, a base form error (*fall*) will usually result. If this mapping is strong, as later in learning, then the /d/ will not be suppressed, and a full regularization (*falled*) will usually result. Which output results depends on two things: the strength of the word-specific suppression of /d/ and the strength of the word-general mapping to /d/. In adults, the word-specific suppression is better learned, and base form errors are relatively more likely than they are in later child speech.

This system is capable of generalizing morphological patterns, whether inflectional or derivational, whether regular or irregular, using a single pathway. In fact, the system does not need to know that there are rules.[8] All that is necessary is the learning of particular lexical items, something that is present in all models of language.

This model is sensitive to a variety of factors, including the frequency of a pattern, but also including phonological factors. Marcus et al. (1995) have recently argued that connectionist models are inherently wrong, because frequency is irrelevant to the generalization of inflectional patterns in German. In fact, their own data require some frequency sensitivity (since Clahsen et al. 1992 are forced to normalize the competing noun-plural affixes for frequency in order to account for how affixes generalize in the speech of children with language disorders), but show that there are additional factors. Recurrent models, like local nets, intermix lexical and phonological processing, and do not predict that morphological frequency is the *only* factor that affects processing, just that frequency is *one* factor. For German, it appears that there is a preference for patterns that keep words short. Regular perfect *-t* and irregular perfect *-en* are about equally frequent as morphological markers of the perfect form, but *-t* generalizes preferentially because it leads to words with fewer syllables. Similarly, noun-plural *-s* generalizes more than expected when compared to the more frequent suffixes *-en*, *-er*, and *-e*, because those other suffixes add a syllable to the word and *-s* does not. The preference for shorter words may be based on frequency; Zipf (1935) reports that 49.8 percent of word tokens in German are monosyllabic, 22.9 percent are disyllabic, 12.9 percent are trisyllabic, and only 8.4 percent are longer. Frequency may be important here, but not the frequency of the morphological pattern.

This recurrent system is basically driven by meaning–sound mappings, just like the local model discussed above. It avoids the problems with the Rumelhart and McClelland model raised by Pinker and Prince (1988), Pinker (1991), Marcus et al. (1992, 1995), and Clahsen et al. (1992), in the same way that local models do. There is no basis for the claim that data from language production (whether of adults or of children) require the use of discrete inflectional rules and show that connectionist models (which lack such rules) are wrong in principle. Further, by highlighting the interactions of different sources of knowledge, connectionist models are inherently more likely to provide a non-arbitrary, non-stipulated *explanation* of why the facts are the way they are. Pinker's model requires us to accept that all the basic properties of morphology are random, due to the capricious nature of the genes that by chance have become a part of the human genetic endowment that controls the acquisition of language (e.g. Pinker 1991). One would hope that there is a more interesting reason for the basic properties of morphology than that.

3 Summary and conclusions

I have reviewed the types of morphological errors observed in language production in normal adults. Many of these phenomena derive from the way that syntax constrains morphology, and I did not dwell on them in the section of the chapter on models, since none of the models have addressed the syntax–morphology interface. But many phenomena concern the way that particular morphological patterns generalize, and all models have focused on that. Recent debates have focused on lexical frequency effects, semantic effects, and phonological effects of various types. We looked at three types of connectionist model: local, distributed nonrecurrent, and distributed recurrent. All can, in principle, handle the data that are currently known. However, extant nonrecurrent models have included dedicated subnetworks for inflected forms, and seem unable to handle phonological priming effects from nouns to verbs. Local models have the drawback of lacking an explicit algorithm for learning, putting them at a disadvantage relative to distributed models (though symbolic models suffer from the same problem).

Connectionism is still a young field. Models, especially ones that have actually been implemented, are still quite primitive, and have addressed relatively small sub-areas of any cognitive domain. This is a reflection of the short timespan in which connectionist models have been addressing these questions, and of the great complexity of interactions in these models, rather than a limitation of the models per se. Basic issues concerning the organization of these models are still being worked out. Models will be in a state of flux for many years to come. Morphology is being addressed more and more frequently, both in modeling and in following up empirical predictions of models. Future models will focus even more on lexicon–phonology interactions. Whether they will ultimately be able to account for all the known facts without morphological rules remains to be seen. But any facts that ultimately derail this endeavor will be subtle in nature, and have yet to be found.

ACKNOWLEDGEMENT

This study was supported in part by NSF Research Grant #DBS-9209642 to the author.

NOTES

1 Note that the verb is plural *are*, and does not accommodate to *is*.

2 The four irregular present-tense singular forms of English (*is*, *has*, *does*, and *says*) all arguably end in the *-s* suffix.

3 Productivity is not the culprit here, since causative/inchoative *-en* (as in *redden*), which occurs in only a small number of words but is semantically transparent, is lost but never added.

4 The *-t* suffix is overwhelmingly more common than *-en* for less frequent verbs, but this is probably irrelevant. Young children generalize the *-t* suffix even when their lexicon is limited to the most frequent verbs.

5 All words containing a long vowel or diphthong followed by /l/ are often pronounced as two syllables by young children, and an onset is provided for the second syllable, e.g. mailed [meyːowd]. A harmonic onset is sometimes present, e.g. *piano* [pɪnænow]; a similar harmonic onset in *mailed* would yield [meymowd]. Other attested pronunciations in child phonology are intervocalic /m/ as [mb] and the simplification of diphthongs in closed syllables (such as before intervocalic [mb]); if we add in those two processes, *mailed* becomes [mɛmbod]. This is quite close to the "desired" output. Since no theory of phonology rules out such a pronunciation, it does not constitute a failing on Rumelhart and McClelland's model.

6 The mechanism is termed "short-circuiting." Semantic similarity leads a learner to conclude that a verb such as *fly* 'to hit a fly ball' is an extended use of the verb *fly* rather than a denominal verb based on the noun *fly* 'fly ball'. Any learner who (erroneously?) draws that conclusion will prefer the irregular past-tense form *flew*. Any learner who concludes that it is a denominal verb will prefer the regular past-tense form *flied*. Semantics has an effect only on learning in that model. It never has an effect on on-line processing of past-tense forms. However, the same predictions seem to be made whether semantics affects processing or learning. The only difference (so far untested) is the following: if semantics has a direct effect on processing, responses should be monomodal, with all speakers showing the same effect; if semantics has only an indirect effect, via learning, responses should be bimodal, reflecting the two different learning outcomes outlined above.

7 Dell's model uses two context layers. Corina (1991) used only internal context, and other models use only external context.

8 Rumelhart and McClelland (1986) make a similar claim, but their system actually needs to know that a dedicated subnetwork for each inflectional category is necessary.