Positive Feedback in Hierarchical Connectionist Models: Applications to Language Production

GARY S. DELL
University of Rochester

Recent connectionist models of the perception and production of words make use of positive feedback from later to earlier levels of processing. This paper focuses on production and identifies several specific effects of phoneme-to-morpheme feedback. In addition, I argue that there is support for the use of this kind of feedback in production from experimental and naturalistic studies of slips of the tongue.

One feature of highly parallel network models of word and letter perception is the existence of positive feedback from the word to the letter level (Adams, 1979; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). So, not only do activated letter units send activation to all words that contain them, but the reverse happens as well. Words send their activation back to the letters that comprise them. This mutual backscratching between words and letters is an elegant mechanism for allowing lexical knowledge to augment stimulus information at the letter level. More generally, positive feedback from “later” to “earlier” levels of processing is a simple way of letting knowledge influence the identification of perceptual features and objects in models that employ connectionist principles. This paper considers the utility of this kind of positive feedback in another information processing domain, the production of ordered behavior. In particular, I will focus on the production of spoken words and consider, first, what positive feedback can contribute to production models in general, and second, what empirical justification there is for adding it to models of human language production.

1 The author wishes to thank Susan Garnsey, James L. McClelland, and Michael Tanenhaus for helpful comments. This research was supported in part by National Science Foundation grant BNS-8406886.

Correspondence and reprint requests should be sent to Gary Dell, Department of Psychology, University of Rochester, Rochester, NY 14627.
Before turning to language production, I would like to clarify what I mean by positive feedback. Connectionist models dealing with the perception of words and their parts (and many other information processing models for that matter) contain a network with two or more levels of processing. For example the interactive activation model of McClelland and Rumelhart (1981; Rumelhart & McClelland, 1982) includes a feature, a letter, and a word level, each composed of a set of units or nodes as shown in Figure 1.

![Figure 1. Word, letter, and feature nodes and their connections according to the interactive activation model. Excitatory connections are indicated by arrows and inhibitory ones by dots. The subscripts on the letter nodes indicate the letter's word position. The small letters labeling some connections are explained in the text. One should note that only the features of the letter T are shown.](image)

As one can see in the figure each node stands for some particular word or part of a word. At any given time, each node has a potential or activation level which is a real number reflecting the extent to which that node is participating in current processing. The connections between nodes can be of four types, labeled a, b, c, and d in Figure 1. Each connection is either excitatory (a and d) or inhibitory (b and c), and either bottom-up (a and b), lateral (c), or top-down (d). An excitatory connection from node x to node y transmits an excitatory signal to y in proportion to the activation level of x. An inhibitory connection does the same except that it transmits an inhibitory signal to y. The excitatory and inhibitory signals directed to y combine and may raise or lower y's activation level depending on y's current level and the net input. The function of the connections is to transmit a pattern
of activation in the feature nodes to the word level in such a way as to identify the most likely word given the set of activated features.

Connections of type $a$, the excitatory bottom-up ones, are the true workhorses of the network. They get the main job done by activating nodes on a higher level that are consistent with activated lower level nodes. Thus the node for initial $T$ should have type-$a$ connections to word nodes for $TAN$ and $TENT$, but not to $BAN$ and $BENT$. Because any conceivable connectionist word recognition model would require something like these connections, I will call them the primary connections.

The other kind of bottom-up connections in the interactive activation model are labeled as type $b$ and are inhibitory in nature. These serve to communicate information about the presence of a feature (or letter) to those letters (or words) that do not contain it. Thus these connections do much the same thing that the primary connections do, only in an inhibitory fashion.

The two remaining connection types act to control the spread of activation rather than to directly send it from lower to higher levels. The lateral inhibitory connections (type $c$) occur between all pairs of units at the same level that are competitors. Two nodes are competitors if they collect evidence for incompatible hypotheses. For example, in a given word stimulus there can be only one initial letter. So a letter node for initial $T$ would be competitor with one for initial $B$. By having a set of competitors each inhibit the others one tends to create a "winner-take-all" situation, in which one competitor, usually the one that initially had the highest activation, will end up with most of all the activation (Feldman & Ballard, 1982). The general effect is to sharpen and simplify the activation pattern at a given level.

The excitatory top-down connections (type $d$) provide positive feedback from higher to lower levels. Like the lateral inhibitory connections, they clean up the activation pattern, but unlike them, they do it with some degree of sophistication. Whereas the lateral inhibitory connections merely throw most of the activation to the "winner," the positive feedback connections, acting in concert with the primary connections, mold the activation pattern of a lower level until it meshes with information available at higher levels. Consider the following example: Assume that at a certain time the nodes for $G_i$, $C_i$, $A_i$, and $T_i$ are equally activated (the subscripts on the letters indicate the letters' word positions). Later, after activation has spread to the word level and fed back to the letter level, the pattern will change so that $C$, is more activated than $G$, simply because $CAT$ is a word and $CAT$ is not. Thus, through positive feedback, the activation pattern at the letter level is changed into one that makes sense when viewed from the higher word level. In general, a word node and the nodes for that word's letters and features will form a mutually reinforcing group of nodes, what Feldman and Ballard (1982) term a stable coalition, and over time activation will drift toward the nodes in a single such group.
Why is it good to have mutually reinforcing nodes at different levels? There are both functional and empirical reasons. It is a good idea to have letter perception constrained by lexical knowledge simply because input is often noisy. You cannot be sure that the key features of each letter will be detected and so you do not want letter identity decisions made independently of lexical knowledge. A second reason is that a model that employs positive feedback between levels will be able to explain many psychological phenomena. McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982) have shown that the interactive activation model does an excellent job of simulating the perceptual advantage for letters in words over letters in unrelated contexts (the word superiority effect) and a variety of other lexical context phenomena. It is the positive feedback from the word to the letter level that makes this possible.

There is little doubt that positive feedback is valuable in hierarchical perception models, particularly if they are to model psychological processes. What about production models? Is there a role for positive feedback here, too? Highly parallel network models of language production (Dell & Reich, 1977, 1980; MacKay, 1982; Stemberger, 1982, in press) and typing (Rumelhart & Norman, 1982) have been proposed. Of these only Stemberger's and my own model have allotted a role to the kind of positive feedback being discussed here. One thing I hope to do here is clarify the function that positive feedback has in these models.

So, exactly what is positive feedback in production? Consider Figure 1, with its network composed of words, letters, and features. If we change words into morphemes, letters into phonemes, and features into phonemic or phonetic features, we have a rudimentary network for phonological encoding processes in production. There are good reasons to include other intermediate levels corresponding to syllables and syllabic constituents, but these are not important for the discussion here.

To produce a morpheme in this kind of model one would activate its node and allow activation to spread. After a period of time some decision mechanism would select the most highly activated features or phonemes, with their order being determined by either activation levels or a kind of position encoding of the nodes similar to that in the interactive activation model.

Next let us consider whether there is a role for positive feedback in a model like this. Unlike the perception model, in which the a connections are the primary connections, the d connections become the primary ones in a production model and the a connections now provide the positive feedback, since they go from later to earlier levels. However, despite the differences in the functions of the connections in the perception and production models, the spreading activation process is similar. The presence of positive feedback in conjunction with the top-down primary connections enables the nodes
for a single morpheme and its phonemes and features to act as a mutually reinforcing coalition. Earlier we saw that the tendency to make such coalitions is helpful in perception models because it filters out noise. Is there a corresponding role for these coalitions in production? I claim that there is. By allowing feedback from lower levels (e.g., phonemes) to higher levels (e.g., morphemes) one can edit out potential production errors. Specifically, I will show that this kind of feedback enables a system to avoid encoding strings of sounds that are inconsistent with higher level information. Before I elaborate on this claim, I want to present a small “demonstrator” simulation model that will point out some features of phoneme-to-morpheme feedback in a phonological production model.

**A MODEL OF PHONOLOGICAL ENCODING**

This model, a no-frills version of a larger model (Dell, 1980, 1984, 1985), was designed to translate an ordered string of morphemes into an ordered string of phonemes. To keep things simple there was just a morpheme and a phoneme level, and each morpheme was a single CVC syllable. Each morpheme node had excitatory connections to the nodes representing its phonemes and each phoneme node had excitatory (feedback) connections to all the morphemes that contained that phoneme.

Each phoneme node was marked as to whether it represented an onset (initial consonant), vowel, or coda (final consonant). So, although the /k/’s in *cat* and *came* would be represented by a common node (the onset /k/), the /k/ in *tack* would be a different node (the coda /k/). The separation of onsets and codas is a form of position encoding that insures that each phoneme knows its position in the morpheme. I have discussed elsewhere how this kind of scheme can be expanded to handle the general problem of order for any morpheme or word (Dell, 1985). The model’s network contained 12 phoneme nodes (4 onsets, 4 vowels, and 4 codas) and 16 morpheme nodes. Each phoneme was present in exactly 4 morphemes.

The processing in the network occurs in four stages: input, spreading activation, decision, and post-decision clean up. These stages are cycled through for each intended morpheme. In the input stage the intended morpheme is activated, that is, its activation level is incremented by an arbitrary amount, for example, 100 units. Also, upcoming morphemes in the same phrase are primed; their activation levels are increased by a smaller amount. (This was always 50 units.) For our purposes it will be sufficient to assume that this priming occurs only one morpheme in advance of the intended one.

Following input, activation is assumed to spread by the following rules. During a given period of time called a time step, each node sends some fraction of its current activation level to all nodes directly connected to it.
When the activation that is sent out reaches its destination, it adds to that node's current activation level. Because all the connections for this model are excitatory (i.e., the fraction of activation that is sent is positive), activation levels will grow without bound. To combat this tendency it is assumed that activation levels passively decay. During each time step all nodes lose some fraction of their activation. However, I should note that using decay to restrict spreading activation is, in general, not a good idea because it makes it difficult to tune parameters so that activation levels stay in useful ranges.

The spreading activation rule proposed here is among the simplest possible of such rules. There are no thresholds, saturation points, or other non-linearities. Furthermore, I will assume that the fraction of activation sent during each time step, designated by $p$, is constant for all connections regardless of their type, and the fraction of activation lost during each time step, designated by $q$, is constant for all nodes. Thus we can state a very simple rule for the spread of activation for one time step:

$$V_i = [V_i + V_i(pM)][(1-q)I]$$

where $V_0$ is a vector containing the activation levels of all $n$ nodes at a certain time and $V_t$ contains the same after one time step has passed. $M$ is an $n \times n$ matrix of 1's and 0's, a 1 indicating a connection from node $i$ to node $j$, and $I$ is the $n \times n$ identity matrix. After activation has spread for a certain number of time steps (parameter $r$, which reflects the speaking rate), the decision stage is entered. Here the most highly activated onset, vowel, and coda are determined and scheduled for articulation in that order. (For this model I will assume that the sounds are selected simultaneously.) Thus the model can produce $4 \times 4 \times 4 = 64$ different strings, only 16 of which are actual morphemes from the model's point of view. In this way the model captures the generative nature of phonology. Many more strings are phonologically legal than are actual lexical items, and the model is perfectly capable of encoding these legal, but nonmorphemic, strings.

In the final stage, the post-decision cleanup, two things happen. First, the activation levels of the three selected phonemes are set to zero. This is necessary to prevent a large number of perseveratory errors. Second, the next morpheme in the set of planned morphemes becomes the intended morpheme, and the entire process cycles back to the input stage.

If a model like this encodes several morphemes in a row it may have difficulty, particularly if the intended utterance repeats sounds and is encoded at a fast rate (Dell, 1980). This difficulty reveals itself as encoding errors—slips of the tongue, in a sense. It is easy to see how errors might happen by imagining how a phrase such as blue bug's blood would be encoded. When the model is encoding the morpheme bug, the morpheme
nodes for *blue* and *blood* will be somewhat activated — *blue* because it has not yet decayed and *blood* because it is primed. Thus the *bl* onset node has two sources of activation compared with the *b* onset, and depending on the model’s parameters and the speaking rate, *bl* may have a higher activation level than *b*, leading to the selection of *blug* rather than *bug*. Earlier work with similar models (Dell, 1980; Dell & Reich, 1980; MacKay, 1982; Stemberger, 1982) and work in progress have shown that the models can explain the variety of errors that occur, their frequency of occurrence, and the effects of changing the speech rate on error probability. In this paper, I will focus on error effects that arise directly from excitatory feedback from lower (and later) to higher (and earlier) levels of processing—particularly phoneme-to-morpheme feedback. The next section will identify these effects using the model outlined above to demonstrate them.

**EFFECTS OF PHONEME-MORPHEME FEEDBACK**

If we use the demonstrator model to encode the phrase *deal back* with $p = .3$, $q = .4$, and with the speech rate at 4 time steps per morpheme ($r = 4$), it will do so correctly. This assumes that there is no other input to the network and every node starts at zero activation. Next, let us assume that for some reason there is initially some residual activation, 40 units worth, in the node for the onset /b/ and every other node starts at zero. Under these circumstances the model will err by encoding *beal back*, a phoneme anticipation slip, which is the most common kind of phonemic slip of the tongue. Other errors, such as exchanges—for example, *beal dack*—tend to occur if the speech rate is faster and the replaced sound /d/ from *deal* has a chance to bump out the /b/ in *back* (Dell & Reich, 1980). So far, none of this is very interesting. The model can, naturally enough, be made to slip by prodding it in the right way. It turns out that it takes at least 30 units of activation on /b/ at the beginning of the encoding process to create the slip of *deal to beal when the next morpheme is back*. I will call this value the anticipatory threshold for the phrase *deal back* with the above mentioned parameter values.

Next consider the phrase *dean bad*. Its anticipatory threshold using the same parameters is only 25. In the model the slip *dean bad—bean bad* (or *bean dad*) is much easier to generate—that is, it requires less stringent pre-conditions—than the slip *deal back—beal back* (or *beal dack*). The reason is that, from the model’s perspective, *bean* and *dad* are legitimate morphemes and *beal* and *dack* are not. Errors that create morphemes are more likely in the model than those creating phonologically legal nonsense strings. This lexical bias effect comes directly from the feedback loops that develop as activation spreads between phonemes and morphemes. A pattern of activation in which the most highly activated phoneme nodes come from a single
morpheme is continually reinforced. This is true whether the pattern corresponds to the correct morpheme or to some other one. If the pattern does not correspond to a morpheme it is likely to change until it does.

In this way the feedback system acts as what Baars, Motley, and MacKay, (1975) have called a lexical editor. It edits out potential nonmorpheme slips, but it does nothing to prevent slips that make erroneous, but genuine, morphemes. Like an editor, the feedback system only works well if it has enough time. When the speech rate is fast the model’s errors do not show any lexical bias. This can be seen in Figure 2 in which the anticipatory threshold for a morphemic outcome (e.g., dean—bean) is contrasted with that for a nonmorphemic outcome (e.g., deal—beal) for various speech rates ($r = 1$ to $r = 4$ time steps per morpheme). At the fast rates ($r \leq 2$) the thresholds are low and do not differ as a function of lexical status of the outcome. Thus, at those rates errors would be common and they would exhibit no lexical bias. As speech slows ($r > 2$) the thresholds rise and the lexical bias effect emerges. Although only 3 time steps are necessary to create lexical bias (1 for morphemic input to reach the phonemes, 1 for phonemic activation to feed back to the morphemes, and 1 for the effects of feedback to be transmitted back to the phonemes), additional time steps increase the size of the effect. In general, the greater the opportunity for activation to reverberate between morphemes and phonemes the greater the likelihood of editing out nonmorphemes.

![Figure 2. Lexical bias in the model's errors as a function of speaking rate. The anticipatory threshold is lower for error outcomes that create morphemes (dean bad—bean bad) than for those that create nonsense (deal back—beal dock) and this effect increases with the number of time steps per morpheme.](image-url)
Another way to conceptualize the model’s lexical bias is as a production counterpart to the interactive activation perception model’s word superiority effect. The latter model finds it easy to perceive letters that occur in words just as the production model finds it easy to say strings of phonemes that make words, or technically, morphemes. Both effects arise from excitatory connections in both directions between adjacent levels of processing. The influence of word–letter feedback on letter perception, however, extends beyond simple word superiority and encompasses several effects that reflect the simultaneous influence of many lexical items. For example, in the interactive activation model letters are perceived in word-like nonwords (e.g., SIND) nearly as well as in words. This effect is due to feedback from the many words that contain most letters of the stimulus (e.g., SEND, MIND, SING, etc.). Stimuli that are not word-like (e.g., OHSG) would not receive this benefit.

Analogous effects occur in the production model. Slips are biased toward creating morpheme-like strings as well as true morpheme strings. I have shown elsewhere that positive feedback from phonemes to morphemes leads to certain frequency asymmetries in slips (Dell, 1980), that is, a given phoneme or combination of phonemes that is present in many morphemes will tend to substitute for those that are present in fewer morphemes. For example, if the onset /t/ occurs in more morphemes that the onset /f/, then initial /f/ will slip to /t/ more than vice-versa. The same would be true for a phoneme combination. The sequence /æn/, which is present in many morphemes, would tend to replace /æb/ which is not common. A commonly recurring set of sounds will, via feedback, activate the many morphemes that contain them and these morphemes will, in turn, lead to an even higher activation level for those sounds. In this way the model is biased toward creating common phoneme combinations, and more generally, toward creating combinations that reflect the entire morphemic inventory.

So far I have shown that phoneme–morpheme feedback has what seems to be a beneficial quality. Activation patterns at the phoneme level are modified through feedback to form patterns that are likely to be correct, that is, likely to correspond to morphemes. However, not all effects of phoneme–morpheme feedback are editorial in nature. Consider the phrase deal back, which, as I showed earlier, has an anticipatory threshold of 30 under the previously specified conditions. If the phrase is changed to deal beak the threshold for the slip deal—beak is reduced to 28. Clearly this difference has nothing to do with the nature of the error string because it is the same in both cases (beak). Rather, the difference reflects the fact that deal and beak have the same vowel (/i/) while deal and back have different vowels. As deal is being encoded with beak pruned, the node for the vowel /i/ will act as a pathway for activation to spread between deal and beak. The effect will be to equalize the activation levels of the two morphemes which will, in turn,
lead to a greater chance of their phonemes jumping from one morpheme to the other. This effect will be called the repeated phoneme effect.

As with the lexical bias effect the repeated phoneme effect does not occur at fast speech rates. Figure 3 shows the anticipatory thresholds for deal back and deal beak as a function of speech rate. The effect, shown as a difference in anticipatory thresholds for the two phrases, only shows up when \( r > 2 \).

![Figure 3. Repeated phoneme effect in the model's errors as a function of speaking rate. The anticipatory threshold is lower for pairs with a repeated vowel (deal beak) than for those with different vowels (deal back) and this effect increases with the number of time steps per morpheme.](image)

Up to now I have identified two general effects of phoneme-morpheme feedback in connectionist production models, a tendency for errors to create morpheme-like strings and a tendency for phrases with similar morphemes (ones with common phonemes) to lead to slips. With these effects in mind I would now like to consider the question of the usefulness of positive feedback in production models. Does the feedback serve any valuable functions in production and, if so, do these benefits outweigh the costs incurred by adding feedback connections to a model? Following that I will examine the psychological evidence. Does the human language production system—undoubtedly the most efficient yet devised—employ positive feedback from more peripheral, "later" levels of processing to more central, "earlier" levels?

**FUNCTIONS OF POSITIVE FEEDBACK IN PRODUCTION**

Positive feedback seems to be beneficial for at least two reasons. First, it allows for prearticulatory editing of the activation pattern at each process-
ing level so that the pattern reflects constraints from higher levels. Second, some linguistic constraints such as those between particular lexical items and syntactic structures can be nicely handled by mutual excitatory connections.

The first of these functions, prearticulatory editing, has already been discussed with respect to the editing out of potential phonological slips if they do not resemble morphemes. Of course this function is only worthwhile to the extent that there is potential for these and other errors. I would like to claim that errors are always a possibility at all levels of processing in connectionist production models. Production, like comprehension, is inherently a noisy enterprise. One reason is that the basic retrieval mechanism in connectionist models is parallel spread of activation from many units to very many other units. Thus by its basic nature, this kind of model activates many more units than just the ones that are sought, a kind of “many are called, few are chosen” retrieval system. These extra units can act as a source of error because it takes time to sort out the right units from the others. If the speaking rate is rapid or variable, it is likely that this sorting out will often not be finished when a decision is required. The result would be errors if there were no editorial mechanisms in the model.

Another source of noise in production is variability in the input to the language production system. This input would be some kind of semantic-pragmatic representation of the utterance-to-be-spoken, perhaps a speech act plus a set of propositions, for example, ASSERT (COLD (THIS ROOM)). Bock (1982) calls this the interfacing representation. There could easily be variability in the coalition of units forming this representation simply because there are many different reasons for saying a given sentence. For example, one may say “This room is cold” to get someone to close the window, to get sympathy, or to explain why the plants died. Each reason would be associated with different computations and, perhaps, somewhat different representations of the final speech act. However, because the intended sentence is supposed to be the same regardless of its function, the possible variations in the interfacing representation will act as noise from the perspective of the next level down, the syntactic level. Thus, it would be worthwhile to have feedback from the syntactic decisions back up to the interfacing representation to prevent this noise from leading to error.

The second function that positive feedback could have in production is to link word selection with syntactic structure selection. Certain syntactic structures (e.g., those with direct objects) require the selection of certain lexical items (e.g., transitive verbs). Bock (1982) has outlined an elegant model of syntactic processing in which word and syntactic structure selection occur in parallel guided by the interfacing representation. So structures are not necessarily selected before words, nor are words selected before structures. What happens is that many candidate structures and words become activated to the extent that they are consistent with the higher level representation. The words then feed activation to the structures that they
are consistent with and the structures activate words that they are consistent with. For example, transitive verbs would activate syntactic structures with direct objects and vice versa. Thus, in connectionist terms, word units and syntactic structure units are joined by mutual excitatory connections. The feedback loops that would be created would quickly select out the best structure and words to go with it. This function of feedback differs from the editing functions, only in that the two mutually interacting components (word selection and syntactic structure selection processes) are not ordered with respect to each other. In the case of editing through feedback, one level of processing, the earlier one, provides the criteria that govern the editing of the later level.

So far we have seen that positive feedback as described here would be of considerable benefit to connectionist language production models. But there are costs associated with positive feedback. One that has already been identified is that, because of positive feedback, similar items will interfere with one another. We saw that phoneme-morpheme feedback makes it somewhat more difficult to encode a string of morphemes that share phonemes, which was called the repeated phoneme effect.

Other similarity effects would occur for other types of feedback. For example, if feature-phoneme feedback is permitted, phonemes that share features will have a greater chance of substituting for each other than dissimilar phonemes. If phoneme feedback goes all the way to the nodes associated with lexical selection, similar sounding words will interact in errors, either by misordering (e.g., cob on the corn), or substitution (e.g., Lizst's second Hungarian restaurant).

Although each of these feedback effects seems to contribute to errors, I suspect the contribution is small. In the model presented earlier the repeated phoneme effect (Figure 3) is small compared with the lexical bias effect (Figure 2). At least in this case the editorial effect outweighs the similarity interference effect. However, this is a limited example and I do not want to conclusively state that positive feedback prevents more errors than it causes.

Another cost of positive feedback is the increased number of connections. Connectionist modelers must, above all, avoid computational processes that require a tremendous number of nodes and connections when scaled up to real-world size. The use of positive feedback does add connections, but the number added does not accelerate as the complexity of the model increases, at least for the kind of feedback described here. In the worst case there would be one bottom-up excitatory feedback connection (type a in Figure 1) for every top-down primary connection (type d). Thus, there is no combinatorial explosion of feedback connections as the model is made larger.

Perhaps the best way to determine if positive feedback is of value in language production models is to look at the psychological evidence. If the
human language production system employs the kind of feedback described here, there are probably good reasons. The next section will identify some speech error effects that can be attributed to positive feedback. In particular I will examine alternate (nonfeedback) explanations of these effects and consider whether a strong case can be made for feedback models.

**PSYCHOLOGICAL EVIDENCE**

In the previous discussion I identified two general effects of phoneme-morpheme feedback in the model: lexical bias, the tendency for errors to create morphemes, and the repeated phoneme effect, the tendency for slips to occur between morphemes with common sounds. Do these effects occur in people? With regard to lexical bias, it has been suspected since the time of Freud that slips tend to be meaningful. However, lexical bias was first demonstrated experimentally by Baars, Motley, and MacKay (1975) who used visual interference to create initial consonant misorderings in subjects' speech (exchanges and anticipations such as deal back—beal back or beal dack). They compared the probability of errors that created lexical items (dean bad—bean dad) with those that created nonmorphemic strings (deal back—beal dack) and found that errors were more than twice as likely when the outcomes were lexical than when they were not. This and other studies have established that lexical bias is a reliable effect (Berg, 1983; Dell & Reich, 1981). One of the corollary effects of the tendency for errors to create morphemes in the model was a tendency for errors to create strings that look like morphemes, that is, strings with frequently occurring phoneme combinations. This effect has, as well, been shown to be true of people's slips (Motley & Baars, 1975).

The repeated phoneme effect also has substantial empirical support. MacKay (1970) and Nooteboom (1969) have shown that phoneme exchanges are often characterized by a repeated phoneme next to the exchanging ones. For example, in the slip left hemisphere—heft lemsphere the phoneme /e/ is next to both of the exchanging ones, /l/ and /h/. In addition, it has been demonstrated in an experiment using a similar procedure to that of Baars, Motley, and MacKay (1975) that initial consonant misorderings are more likely between two words that share a vowel (deal beak) than those that do not (deal back) (Dell, 1984).

Thus both the lexical bias and repeated phoneme effects occur in human language production. The fact that they do occur, however, does not guarantee that there is phoneme-morpheme feedback, or something like it, in production. There are other possible explanations that do not involve feedback, and these must be considered. In order to identify these alternatives it is useful, first, to distinguish between multi-level interactive and single-level structural explanations of psycholinguistic phenomena. Many
empirical effects involving the processing of linguistic information are of the form: "the perception (or production) of \( X \) is influenced by \( Y \)." Numbers 1-3 below give some examples.

1. The perception of letters is influenced by lexical context (example—word superiority effect).
2. The perception of words is influenced by preceding semantically related words (example—semantic priming in lexical decision).
3. The production of a string of phonemes is influenced by whether or not that string forms a lexical item (example—lexical bias in sound speech errors).

Explanations for these kinds of effects can take two general forms. First, one can say that \( X \)-type units are processed at a different level than the one where the \( Y \) influences come from, but these influences are nonetheless made available during the processing of \( X \). This is what I call a multi-level interactive explanation. The second kind of explanation incorporates the \( Y \) influences directly into the structure of the \( X \)-level, hence the term single-level structural explanation.

Consider the effects of semantic priming. The time to decide whether nurse is a word (lexical decision) or the time to read it aloud (naming) is faster if it is preceded by a semantically related word such as doctor (e.g., Meyer & Schvaneveldt, 1971). I will restrict the discussion to priming in naming rather than lexical decision because naming seems to be the simpler task (Stanovich & West, 1983). There is considerable debate as to why priming in naming occurs, whether it is due to an influence of the semantic level of processing on lexical access (a multi-level interactive explanation) or "intra-lexical" connections between highly related words (a single-level structural explanation). If it turns out that this priming is not semantic, but rather just a structural feature of the mental lexicon, then one can propose a simple modular account of lexical access (Forster, 1976, 1979; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Tanenhaus, Leiman, & Seidenberg, 1979). In fact the issue of the "modularity of language" in general hinges upon distinguishing the two explanations of priming (Fodor, 1983).

Returning to the speech error effects under consideration (lexical bias and repeated phoneme effects) it should be clear that the account of these effects using phoneme-morpheme feedback is an interactive explanation. The alternative explanations, which I will now consider, are all single-level structural explanations. These accounts claim that either the phonemic level or the lexical level is structured so as to produce the effects. The next three sections describe some of these accounts and contrast them with the feedback account.
A Single-level Account of the Repeated Phoneme Effect

Wickelgren (1969) proposed that repeated item effects arise from associative contextual influences between adjacent items. In his theory deal beak might slip to beal deck because the units for both /d/ and /b/ are labeled as being associated with a following /i/ sound. Hence they would be more likely to switch places than if they had come from the phrase deal back, in which case they would have been labeled for different following vowels.

This explanation of the repeated phoneme effect can be empirically distinguished from the feedback one. If the effect is due to associations between adjacent phonemes as Wickelgren (1969) proposed, then repeated phonemes should only be instrumental in causing slips in those phonemes adjacent to the repeated ones. So, in the phrase rolling ball the repeated /l/ should only induce slips in the preceding adjacent vowel sounds, (e.g., rall- ing boll), not in the nonadjacent initial consonants, (e.g., boiling rail). If the effect is the result of feedback from the phonemes to some higher level then the repeated /l/ could induce boiling rall. The scope of influence of the repeated sound should extend to all phonemes within the relevant higher order unit, and if that unit is the morpheme or syllable, the repeated /l/ can affect /r/ and /b/.

The data clearly support the interactive feedback account of the repeated phoneme effect. Repeated sounds can induce slips in nonadjacent sounds from the same syllable (or morpheme) to about the same extent as they can in adjacent sounds (Dell, 1984). Although it is not clear whether the relevant higher order unit is the syllable or the morpheme, the effect is definitely not limited to adjacent sounds. This result effectively eliminates structural explanations of the effect that appeal to a phonemic adjacency mechanism.

A Single-level Account of Lexical Bias

Like the repeated phoneme effect, the lexical bias effect might be explainable through properties of the phonemic level alone. For example, it could be that frequently co-occurring phonemes have excitatory lateral connections among them—a phonemic analogue to the intra-lexical connections mentioned earlier. These connections would bias for activation patterns that resemble actual morphemes. Adams (1979) has, in fact, proposed a similar explanation for the effect of pseudowords on letter perception based on these kinds of connections among letters. Although I cannot conclusively reject this as an account of lexical bias I can provide some data that can be more easily explained through phoneme-morpheme feedback than through lateral connections between phonemes. These data concern the time-depen-
gence of lexical bias, and the sensitivity of the effect to expectations and semantic factors.

It is generally true that feedback-caused effects need time to become established. Earlier, we saw that the lexical bias effect in the feedback model increased as speech slowed (Figure 2). If it can be shown that lexical bias actually is dependent on speech rate in this way it would support the feedback explanation. Although such a finding would not be inconsistent with a single-level structural explanation of the effect based on lateral phonemic connections, the feedback explanation provides a more natural account.

The data showing the time dependence of lexical bias come from an experimental study of sound errors (Dell, 1985). I used a modification of the procedure used by Baars, Motley, & MacKay (1975) to generate initial consonant exchanges and anticipations at three different speech rates. Subjects saw pairs of words, one pair at a time, at a 1-second rate. Certain critical pairs were designed so that anticipation or exchange of their initial phonemes created either words (e.g., *dean bad—bean dad*) or meaningless syllables (*deal back—beat back*). Each critical pair was preceded by three or four interference pairs that biased the subject toward making a slip. For example, the critical pairs *dean bad* or *deal back* were preceded by *big dumb*, *bust dog*, and *bet dart*. Following the presentation of a critical pair, subjects saw a series of question marks. This directed them to say aloud the most recent pair. They had to complete the utterance of the pair before a certain deadline, which was either 500, 700, or 1000 ms for different groups of subjects. The probability of anticipations or exchanges of initial consonants was dependent on both the deadline (shorter deadlines led to more errors) and whether the error outcome was meaningful. However, as can be seen in Figure 4, the lexical bias effect was only present at the two longer deadlines. When subjects had only 500 ms to complete their utterance there was no lexical bias in their errors. Thus the lexical bias effect is dependent on time, as would be expected if the effect were due to phoneme-morpheme feedback. If the effect were due to some structural property of the phonological level, it is not likely that the effect could be nullified simply by speeding up the speech.

In addition to the time-dependence of lexical bias there is other evidence that favors a feedback explanation of the effect. Baars, Motley, and MacKay (1975) found that lexical bias is reduced in a context where most of the subjects' intended utterances consist of nonwords. This kind of flexibility seems more consistent with an explanation in which lexical bias is imposed on phonological encoding from outside, that is, from a separate level, than an explanation that is internal to a level. Finally, there is evidence that lexical bias in slips is sensitive to the kind of information that should only be available at morphemic or higher levels. If subjects have just read a phrase such as *dump rifle* they are more likely to make sound errors such as *get*
Figure 4. Empirical demonstration of the interaction of lexical bias with the speaking rate. Data are from Dell (1985) and come from 132 subjects. Each point is based on 440 error opportunities.

one—wet gun (Motley & Baars, 1976). This result can be interpreted as evidence for phoneme-to-semantics feedback and this is how Motley and Baars interpreted the effect. However, if we assume that the process of comprehending the phrase dump rifle activates morpheme nodes for wet and gun for some period of time, then the effect can be seen as just lexical bias and the feedback need only go from phonemes to morphemes, not up to the semantics. In either case, however, it is difficult to reconcile the effect with a single-level structural explanation based on lateral phonemic connections without making the connections very flexible and complex.

In conclusion, it seems that explanations of the lexical bias and the repeated phoneme effects based on properties of the phonemic level do not fare so well, at least when compared with the feedback model. Neither of the alternatives accounts for as much data in a natural fashion as the feedback model and, in the case of the contextual association view of the repeated phoneme effect, some data actually contradict it.

Thus far, the single level alternatives to the feedback model that were considered were phonemic, that is, the structure necessary to explain the effects was given to the phonemic, not the lexical level. The next section considers a lexical explanation of the effects—one that produces much of the effects of the feedback model, but without using feedback.
The Lexical Neighborhood Model

In the feedback model, when a particular morpheme is activated phonologically similar morphemes become activated through bottom-up feedback. This is essentially how the feedback leads to the lexical bias and repeated phoneme effects. What if the lexicon (the morpheme nodes) happened to be “arranged” so that the activation of one morpheme would directly, without feedback, lead to the activation of a set of phonologically similar neighbors? This will be called the lexical neighborhood model. This model proposes that the lexicon is structured so that phonologically similar items are “close” and that retrieving one item partially activates that item’s neighbors. The only difference between this and the feedback model is that the activation of the neighbors is accomplished through feedback from the phonological level in the one case, and it is simply a structural property of the lexical level in the other case. Thus any particular feedback model could be mimicked by a lexical neighborhood model whose lexicon had the right structure.

When contrasted with feedback models, however, lexical neighborhood models have some shortcomings. First, feedback models can explain the fact that lexical bias does not occur at fast speaking rates (Figure 4). There is just not enough time for feedback to work at the fast rate. Lexical neighborhood models have no natural account of the time dependence of the effect. (One could say that the activation of the neighbors is delayed relative to the activation of intended morpheme, but this certainly does not fall out of the lexical neighborhood concept.) Second, Baars, Motley, and MacKay (1975) found that under some circumstances, lexical bias holds even when the intended utterance is composed of nonwords. How would the lexical level become activated when one intends to say a nonsense string if not through a bottom-up (i.e., feedback) route? Finally, one can fault lexical neighborhood models for not providing a distinct mechanism for the activation of the neighbors. Whatever mechanism does activate the neighbors, it will have to work exactly as does phoneme-morpheme feedback if it is to account for the data. If it must act like feedback why not just admit that it is feedback?

In conclusion, there is good evidence for feedback during production that goes from a phonemic level to a morphemic or lexical level and that this feedback affects the selection and ordering of phonemes. Alternative explanations of the evidence appear to be weaker than the feedback explanations and they are certainly not as completely specified. All of the evidence discussed so far can be explained by phoneme-to-morpheme feedback as outlined in the phonological encoding model presented earlier.

Evidence for Other Kinds of Feedback in Production

Other speech error effects have been attributed to positive feedback, and I would like to briefly consider them. One of the most robust facts about
phonological slips is the sound similarity effect—similar sounds (e.g., /t/ and /k/) slip with each other much more readily than dissimilar sounds (e.g., /z/ and /k/). Because similar sounds can be said to share features, theorists have suggested that the effect may be due to feature-phoneme feedback (e.g., Dell & Reich, 1980; Meyer & Gordon, in press; Stemberger, 1982). Common features would provide pathways for activation spreading between competing similar phonemes in much the same way that occurs in the repeated phoneme effect.

The case for feature-phoneme feedback, however, is not nearly as strong as that for phoneme-morpheme feedback. As one would expect, the sound similarity effect can be given single-level structural explanations, as well as interactive ones. For example, Shattuck-Hufnagel and Klatt (1979) argue that phonemic features are not units of production, but instead are just names of categories of phonemes. Similar phonemes are just those that are in the same category and similar phonemes slip because phoneme selection processes use these categories. In order to rule out this and other structural explanations for the sound similarity effect we need data on the effect’s time dependence and its boundary conditions—the same kind of evidence that was used to argue against structural explanations for the lexical bias and repeated phoneme effects. Unfortunately, this evidence is lacking, so the case cannot be strongly made for feature-to-phoneme feedback during production.

The situation is similar with regard to phoneme-to-lexical selection feedback. Dell and Reich (1980, 1981) and Stemberger (1982, 1983) have argued for feedback from the phonemic level to lexical selection processes based on the fact that similar sounding words often substitute for one another (e.g., saying propose for propel). Again, there is a possible structural explanation that is difficult to rule out. Fay and Cutler (1977) and Fromkin (1971) have argued that lexical neighborhood models of the kind discussed earlier can easily explain phonologically related word errors—the substituted word was simply in the intended word’s neighborhood. Although there are some data favoring the feedback explanation (Dell & Reich, 1981; Harley, 1984; Stemberger, 1983) there is, at present, no resolution (see Garrett, 1976).

**CONCLUSIONS**

Despite the difficulty in establishing that human language production makes use of positive feedback, I think the case can be made that such feedback exists, and moreover, that its purpose is to establish coalitions made up of units from separate levels. Establishing these coalitions can prevent errors that would have gone uncorrected if production were purely top-down in nature.
REFERENCES


