

Normal and Deviant Lexical Processing: Reply to Dell and O'Seaghdha (1991)

Willem J. M. Levelt
Max-Planck-Institut für Psycholinguistik
Nijmegen, The Netherlands

Herbert Schriefers
Freie Universität Berlin
Berlin, Federal Republic of Germany

Dirk Vorberg
Technische Universität Braunschweig
Braunschweig, Federal Republic of Germany

Antje S. Meyer
Nijmegen University and
Max-Planck-Institut für Psycholinguistik
Nijmegen, The Netherlands

Thomas Pechmann
Universität des Saarlandes
Saarbrücken, Federal Republic of Germany

Jaap Havinga
Nijmegen University
Nijmegen, The Netherlands

In their comment, Dell and O'Seaghdha (1991) adduced any effect on phonological probes for semantic alternatives to the activation of these probes in the lexical network. We argue that that interpretation is false and, in addition, that the model still cannot account for our data. Furthermore, and different from Dell and O'Seaghdha, we adduce semantic rebound to the lemma level, where it is so substantial that it should have shown up in our data. Finally, we question the function of feedback in a lexical network (other than eliciting speech errors) and discuss Dell's (1988) notion of a unified production-comprehension system.

Until recently, models of lexical access in speech production were almost exclusively based on speech error data. This is true both for the modular two-stage models and for the interactive connectionist models of lexical access. Both kinds of models were initially designed to account for the distributions of naturally observed or experimentally elicited speech errors. From the start, however, they were conceived as *process* models of normal speech production. Therefore, the ultimate test of such models cannot lie in their account of infrequent derailments of the process. Rather, the proof of their efficacy should be sought in the account of the normal process itself. An exclusively error-based approach to lexical access in speech production is as ill-conceived as an exclusively illusion-based approach in vision research. One should, of course, hope that an ultimate theory of the normal process also has the potential of explaining observed error distributions (or visual illusions, for that matter), but it should not be one's main concern.

In our earlier article (Levelt et al., 1991) we tried to take the *process* aspect of existing models seriously, and we developed a reaction time paradigm that allowed us to trace the time course of semantic and phonological activation in normal lexical access.¹ We used this paradigm to compare two kinds of theories: discrete two-stage theories and connectionist network theories. The latter kind was further partitioned into forward-only

spreading networks and forward-and-backward spreading networks. Dell's (1986, 1988) theory is an activation-spreading theory of the latter kind; it is, moreover, the most explicit and the only quantified connectionist model of lexical access in existence. It should, therefore, have been no surprise that Dell's theory was a target of much attention in our article. The process aspects of other connectionist models are not as well spelled out and hence are less accessible to experimental testing.

What we found was, first, that all our data were in full agreement with the predictions of the two-stage theory. It holds that lexical selection strictly precedes phonological encoding and, in particular, that only the selected item becomes phonologically encoded. The theory also excludes feedback from phonological encoding to lexical selection.

Second, we found that our data were not in obvious agree-

¹ Although our picture-naming task is, on all accounts, certainly a "normal" lexical accessing task, one can disagree about how "normal" our dual task (lexical decision during naming) is. The task is obviously not daily practice for most people, but we do believe that until the moment of probe presentation, our subjects were involved in normal lexical access. There is as yet no reason to doubt that our method measures the *normal* state of activation at the moment of probing. The problem with dual tasks is not so much their normality but the necessity to work with an explicit model of the dual task—that is, a model of the interaction between the normal process and the probing task. As Gary Dell mentioned to us, the data-theory link is more direct in single-task situations involving normal error-free speech (such as naming latency measurement).

Correspondence concerning this article should be addressed to Willem J. M. Levelt, Max-Planck-Institut für Psycholinguistik, Wundtlaan 1, NL-6525 XD Nijmegen, The Netherlands.

ment with the connectionist models. They all predict phonological activation of items that are semantically coactivated with the target, and we found no trace of that. Models (such as Dell's) that allow for backward spreading of activation moreover predict a semantic rebound after phonological activation; no trace of that was found either. In their comment, Dell and O'Seaghdha (1991) take up these two points one by one, and our reply follows the same order.²

The Phonological Activation of Semantic Alternatives

In our General Discussion, we commented on our negative finding on the phonological activation of semantic alternatives as follows:

Although our negative evidence is clearly supportive of the latter notion [i.e., the two-stage theory], one cannot a priori exclude the possibility that a connectionist account can be reconciled with this finding. One should choose the model's parameters in such a way that the phonological activation of the target becomes substantially stronger than the phonological activation of its semantic associates or competitors (p. 140). But, we added, there are limits. "A nonnegligible phonological activation of semantic alternatives is, for instance, necessary to handle the occurrence of mixed errors (such as *oyster* for *lobster*)" (p. 140).

In their reply, Dell and O'Seaghdha (1991) precisely follow this line. They argue that if the target word (e.g., *sheep*) is activated by the picture to value 1.0 and a semantic alternative (e.g., *goat*) to value $s \leq 1$, then an item that is phonologically related to the semantic alternative (e.g., *goal*) will be activated to the extent $s \cdot f$ if feedback through the phonological level has value $f \leq 1$. This they call *mediated priming*. However, a word that is both semantically and phonologically related to the target (e.g., *steed*) will be activated to the amount $s + f$. This they call *convergent priming*. When s and f are both substantially smaller than 1 (e.g., when both are 0.2), the difference in activation is substantial. This would account for both the occurrence of a relatively high rate of mixed errors and the absence of measurable phonological activation of semantic alternatives.

This argument, however, assigns our findings entirely to the activation produced by our picture stimuli. It presupposes that the explanations for both our negative findings with respect to phonological activation of semantic alternatives (such as *goat*) and the occurrence of mixed errors (such as *steed*) should be based on the activation the network produces at the word level *in response to the picture*.

The mixed error (such as *steed*) arises through convergent priming of the corresponding word node; any effect for the phonologically related test probe (such as *goal*) in our task is caused by the mediated priming that this word receives from the picture. But whereas the occurrence of a mixed error should indeed be so explained, our experimental finding for the test probes such as *goal* must have a completely different origin. In our dual-task situation *these items were activated by the acoustic stimulus* (while seeing the picture of a sheep, the subject *heard* the word *goal*). Whether these probe words were also activated by the picture through mediated priming is irrelevant. Even if they were not activated at all, they still were valid as test items for the measurement of phonological activation of semantic alternatives (such as *goat*). Remember how we explained the effect of phonological similarity on lexical decision latency. If *goat* is phonologically active and *goal* is presented as lexical decision item, the activation of *goat* delays reduction of the cohort to the single element *goal* (in comparison with the control condition). This has nothing to do with whatever mediated activation that *goal* might receive from the picture. The latter may add to the effect, but it cannot be the sole or main cause. Hence the comparison between activation levels $s \cdot f$ and $s + f$ is not the correct approach for comparing our experimental null effect to the above-chance occurrence of mixed errors. Therefore, we still

do not know whether we can find model parameters that can account both for the absence of phonological activation of semantic alternatives in our experiments and for the above-chance occurrence of mixed errors.

But even on their own interpretation of our dual task, we must deny Dell and O'Seaghdha's (1991) contention that the spreading-activation model can produce the relevant priming effects. Their account is based on the so-called indexing problem. The subject must distinguish the probe from the name of the picture. For instance, the picture of a sheep produces some activation in the word node *goat*, and this makes it difficult for the system to recognize that *goat* is also the lexical decision probe. According to this interpretation of the dual task, "lexical-decision probe processing is inhibited to the extent that the probe's word node is already activated by the picture naming task" (Dell & O'Seaghdha, 1991, p. 607). The more activated the word node that corresponds to the lexical probe is, the larger the inhibitory effect will be. But now observe our data in Figure 3:

1. For all stimulus onset asynchrony (SOA) conditions, we find that phonological probes (*sheet* when the picture is one of a sheep) are more inhibited than identical probes (*sheep*). According to Dell and O'Seaghdha's (1991) one-shot model, this means that $f + sf > 1 + sf$ for all SOAs. Hence $f > 1$, which is contrary to the model's assumptions ($0 \leq s, f \leq 1$). (The same follows if we ignore the mixed node.)

2. For short and medium SOAs, we found about the same inhibition for identical (*sheep*) and semantic (*wool*) probes. This means that $1 + sf \approx s$. But given the model's assumptions, this holds only if $s \approx 1$ and $f \approx 0$, which is contrary to the requirements of Observation 1 and contrary to the additionally imposed requirement that s and f are both less than 0.5. For long SOAs, identical probes are less inhibited than semantic probes; that is, $1 + sf < s$. Hence it cannot be the case that both $0 \leq s \leq 1$ and $0 \leq f \leq 1$; this violates the model's assumptions. (If we ignore the mixed nodes, we get $s \approx 1$ for the short and medium SOAs and $1 \leq s$ for the long SOA, violating the same requirements.)

It follows, given their own account of the dual task, that the model proposed by Dell and O'Seaghdha (1991) cannot handle our data, whatever the choice of parameters. We are aware of the fact that both Observations 1 and 2 hinge on the findings in our identical (I) condition (i.e., in which the probe word is the name of the object pictured). Dell and O'Seaghdha may want to exempt that condition from the indexing interpretation of the dual task. That is arbitrary (especially because the I condition is so similar to the mixed case), but no more so than our own account of this condition (which we labeled as "by stipulation" in the original article). The point here is only that it is still a long shot for the connectionist model to generate both our SOA curves and the zero phonological activation of semantic alternatives. We are not saying that it is impossible, but it is still to be done.

Whatever the solution is going to be, a central issue will be what ratio of word node activation one should expect between semantic and alternative. In Dell and O'Seaghdha's (1991) formalization (and without regard to the mixed items—i.e., those that are both phonologically and semantically related to the target), this ratio is s . On the network account, one should expect the same ratio for the phonological node activation of those items. In Experiment 3, the short-SOA condition, we found an 88-ms effect for the phonological activation of the target. (In terms of the current example, if the picture was one of a sheep, the test word *sheet* produced an 88-ms longer lexical decision latency than did an unrelated control word.) In Experiment 6, in which we measured the phonological activation of semantic alternatives (such as *goat* when the target was *sheep*) by presenting a phonologically related test probe (such as *goal*), the average effect was -2 ms. So what value should s

² We express our appreciation for the extensive and constructive e-mail exchange that we could have with Gary Dell before the writing of the present comment and reply. We may not have reached complete agreement, but certainly we have moved considerably closer in the evaluation of our theoretical positions and of their empirical support.

have in order to produce this substantial difference? In our article we argued that s cannot be tiny, and it certainly cannot be on Dell and O'Seaghdha's indexing account: We found a 106-ms effect when we used the semantic alternative itself as a test probe (in the example, *goat*). On the indexing account, this indicates a very substantial activation of the word node of the semantic alternative. Still, there was no trace of that alternative's phonological activation. Although we agree that more empirical research is needed to estimate s , it is a priori unlikely, given Dell's (1986) model, that one would obtain a zero effect. The two-stage explanation of that finding, however, is straightforward: only the selected item will be phonologically encoded.

The Absence of a Late Semantic Rebound

Dell and O'Seaghdha's (1991) Figure 5 shows the semantic rebound that Dell's (1986) model predicts for one choice of parameters. The critical case is the bump under "phonological" in the second panel. The word node is reactivated shortly after time instant 18 because of the external jolt that the phonological nodes receive at that instant (see panel 3 under "phonological").

In our experiments, we found no evidence for a late reactivation of the target lemma. In the General Discussion, we wrote,

The question is whether a parameter estimation can be found that simultaneously satisfies two borderline conditions. The first one is that the feedback from the phonological to the lemma level is weak enough to prevent a measurable semantic reactivation of the target lemma. The second one is that the same feedback is still strong enough to explain the speech error phenomena for which it was proposed to start with. (Levelt et al., 1991, p. 139)

In their comment, Dell and O'Seaghdha (1991) interpret the semantic rebound in a different way than we did. For them, semantic rebound means reactivation of the semantic level nodes—that is, of the concepts. The relevant bump, then, is the one under "phonological" in the first panel. Clearly, that rebound is about zero. But is this the correct assignment of our experimental findings to the model? We do not think it is, but this needs some explanation.

In the experiment under concern (Experiment 3), we measured activation of the target lemma by presenting an associate as a lexical decision probe (e.g., *wool* when the target was *sheep*). At short (but not at medium or long) SOAs, we found a response delay for these probes (relative to unrelated control probes). The obvious question here is whether we measured conceptual-level (semantic) activation or lemma-level activation. This was tested in our Experiment 4, in which the subjects had a conceptual-level task on the same items (a recognition task) that did not involve naming. If the effect for associated targets (such as *wool*) were to disappear, we would have evidence that we were measuring at the lemma level, not at the conceptual level. That is what we obtained. In other words, the associates probed into the lemma level, and we found no reactivation at that level at late SOAs (i.e., when the target was phonologically active).

Dell and O'Seaghdha (1991) argue that their model is locally but not globally interactive. There is one-level-up feedback but little or no two-level-up feedback. We return shortly to this important conclusion, but we first notice that if our semantic probes indeed measure activation at the lemma level and not at the conceptual level, we have not been able to find evidence for

local interactiveness either. In their simulation, Dell and O'Seaghdha used parameters that are in the general range that Dell (1986, 1990) used for modeling of his evidence of speech error. Thus if the same choice of parameters had resulted in a sheer absence of semantic rebound, the conditions that we requested in our General Discussion would have been fulfilled. But the rebound of lemma activation (Figure 5, second panel under "phonological") was certainly not insubstantial, and this contrasts with the absolute absence of any rebound effect in our data.³ We admit, however, that our experimental procedure may not have been sensitive enough to pick up this degree of activation. Further experiments are needed to decide on this issue.

Where the interaction between Dell's (1986) modeling and our experimentation has been most profitable is in the emerging recognition that the connectionist models of lexical access addressed in our article must incorporate a substantial degree of modularity. One could say that the lexical networks themselves should be relatively "cold" networks. Their pattern of performance is to a substantial degree produced by staged structural input from outside the network proper.

But at this point one should ask whether a "cold" network can indeed handle the speech error evidence. If interactiveness is largely restricted to occur between adjacent levels (i.e., between levels N , $N + 1$, and $N - 1$), as Dell and O'Seaghdha (1991) proposed in their comment, then Dell's (1986) model may not be able to account for the statistical biases in speech errors that it set out to explain (we are grateful to Paul Meyer for bringing this point to our attention). In contrast to the simplified model in Dell and O'Seaghdha's Figure 4 (which generated the curves in their Figure 5), the original model had two intervening levels between the lemma level and the phonemic level: a level of syllable nodes and a level of syllable constituent nodes (clusters and rimes; cf. Dell 1986, Figure 2). If that network were a "cold" one with only local interactiveness between levels, then lexical bias and the repeated phoneme effect would disappear because they involve nonlocal interaction between the phonemic and the morphological (lemma) level.⁴

³ Gary Dell let us know that our long-SOA condition may be comparable to Time Steps 28–35 in Dell and O'Seaghdha's (1991) Figure 5. In that case the semantic rebound is to be found under "word" in the fourth panel, and that is clearly negligible. This solution may or may not work. The only way to find out is to try to generate our SOA curves on the basis of both this assumption and the indexing interpretation of our dual task. That is still to be done.

⁴ In our continuing exchanges, Gary Dell conceded this point and turned it into two nontrivial predictions. The first one is that the repeated phoneme effect should be a within-syllable effect. The second one is that the lexical bias effect is in fact a bias toward frequent syllables. The latter would be consistent with the fact that nobody has found any word-frequency effect in lexical bias (see Dell, 1990). (Both predictions hold only if the level of syllable *constituents* is removed from Dell's 1986 model). Dell suggested that if these predictions are not substantiated, both levels of syllable representation may have to be removed. Syllable structure effects could then be adduced to the construction of frames, or—in parallel distributed processing fashion—emerge from existing correlations.

Concluding Remarks

Although Dell and O'Seaghdha's (1991) comment has substantially clarified the issues that our original article addressed, it has not yet produced the full "balancing act" required to reconcile our findings with the speech error data that Dell's (1986) model was designed to explain. Still, we do not exclude the possibility that the act may eventually be feasible. Here, however, we want to return to the issue raised in our introduction. The ultimate test for a process model, we argued, is how it accounts for the normal process.

In that normal process, lexical selection and phonological encoding fulfill wildly different functions. The former involves fast search in a huge lexicon for an appropriate item; the latter creates an articulatory program for that item. Any feedback from the latter to the former level only makes the system more error prone than necessary. Modularity, one could say, is nature's protection against error.

So what could be the functional sense of the interactive feedback in Dell's (1986) model? Clearly, Dell introduced it in order to account for statistical properties of speech error distributions, particularly the relative rate of mixed errors, the lexical bias effect, the repeated phoneme effect, and so forth (see Dell, 1988). But it is hardly satisfactory to say that the interactive feedback is there in order to produce special kinds of speech error.

Dell's (1988) discussion of this issue alludes to one other possibility: that the feedback is there for the network to serve as a word recognition device as well. That makes good sense. There is nothing in speech production that specifically requires interactive feedback. But if the same lexical network subserves speech comprehension, then two-way traffic is unavoidable. A certain kind of error proneness is then a natural consequence of the unitary lexicon. Allport (1984), reviewing the case for a unitary lexicon, concluded that the study of language pathology has not (yet) produced convincing evidence against that notion (such as cases of double dissociation between phonological processing in comprehension and production of speech.⁵ MacKay's (1987) node structure theory involves a proposal for such a two-way lexicon. Unfortunately, unlike Dell's model, MacKay's model is only a qualitative one and is therefore not easily evaluated against process data in speech production and comprehension. Levelt (1989), however, argued against the node structure theory on the grounds that it cannot account for the often long delays in a speaker's detection of a self-produced

error. Still, it is only fair to say that the case of a unitary lexicon is as yet completely undecided. In fact, our own model of the dual task implies that there is substantial lexical interaction between production and comprehension.

Our conclusion is this: The two-stage theory is in full agreement with our reaction time data. We recognize that error distributions should eventually also be accounted for by a process model, but the ultimate testing ground is in the model's handling of normal processing. As far as speech error biases are concerned, the two-stage theory treats them as postlexical effects, originating in a postlexical editing device (the speech comprehension system). As far as the feed-forward and two-way connectionist theories are concerned, it is now time to hunt for normal processing evidence in their support. We have not found it in our experiments, but perhaps, as the saying goes, he that seeketh findeth.

⁵ Howard and Franklin's (1987) conduction aphasic patient MK may represent such a case of lexical dissociation.

References

- Allport, D. A. (1984). Speech production and comprehension: One lexicon or two? In W. Prinz & A. F. Sanders (Eds.), *Cognition and motor processes* (pp. 209-228). Berlin-Heidelberg: Springer.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283-321.
- Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, 27, 124-142.
- Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, 5, 313-349.
- Dell, G. S., & O'Seaghdha P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review*, 98, 604-614.
- Howard, D. & Franklin, S. (1987). *Missing the meaning*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, 98, 122-142.
- MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. New York: Springer.

Received March 14, 1991

Accepted March 14, 1991 ■