# Evidence for a Cascade Model of Lexical Access in Speech Production

Ezequiel Morsella and Michele Miozzo
Columbia University

How word production unfolds remains controversial. Serial models posit that phonological encoding begins only after lexical node selection, whereas cascade models hold that it can occur before selection. Both models were evaluated by testing whether unselected lexical nodes influence phonological encoding in the picture–picture interference paradigm. English speakers were shown pairs of superimposed pictures and were instructed to name one picture and ignore another. Naming was faster when target pictures were paired with phonologically related (bed–bell) than with unrelated (bed–pin) distractors. This suggests that the unspoken distractors exerted a phonological influence on production. This finding is inconsistent with serial models but in line with cascade ones. The facilitation effect was not replicated in Italian with the same pictures, supporting the view that the effect found in English was caused by the phonological properties of the stimuli.

Converging evidence from reaction-time experiments (e.g., Schriefers, Meyer, & Levelt, 1990), error analyses (e.g., Fromkin, 1971; Garrett, 1980), and brain-lesion studies (e.g., Badecker, Miozzo, & Zanuttini, 1995; Goodglass, Kaplan, Weintraub, & Ackerman, 1976; Kay & Ellis, 1987) suggests that there are at least two levels of representation at play during lexical processing in word production. At one level of representation, a node pointing to the word's syntactic features is assumed to exist for each word known by a speaker. The lexical node "cat," for instance, points to the features of number (singular vs. plural) and grammatical class (noun), among other syntactic features. Another level of representation encodes information about a word's phonology—for example, that the word *cat* (a) is composed of the phonemes /k/, /æ/, and /t/; (b) is monosyllabic, (c) has only one vowel, and so forth. Consistent with this form of lexical organization, lexical retrieval in word production appears to involve two distinct stages: One stage is devoted to the selection of a word's lexical node and its syntactic features, and the other stage is aimed at retrieving word phonology (Butterworth, 1989; Caramazza, 1997; Dell, 1986; Garrett, 1980; Levelt, 1989; MacKay, 1987; Stemberger, 1985). We refer to these stages as lexical node[1] selection and phonological encoding. Furthermore, there is little disagreement about two additional assumptions concerning lexical retrieval in word production. First, the semantic system activates a cohort of related lexical nodes. If the speaker wants to say "cat," for instance, the lexical nodes for "tiger," "dog," and "whisker" receive activation along with "cat." In normal circumstances the target lexical node—"cat," in our example—is selected because it reaches the highest level of activation. Second, selection proceeds in a fixed order: The selection of lexical nodes precedes that of word phonology. But the agreement ends here, and relevant issues concerning the architecture of the lexical system and the dynamics of lexical access remain controversial.

An issue that is a matter of debate among theories of word production is how activation flows between the two levels of lexical representation. Is it the case that word phonology is activated only after a lexical node has been selected, or can activation from lexical nodes flow onto the phonological level before lexical node selection has taken place? One major view of speech production contends that speech production is strictly a serial process, with phonological encoding beginning only after a lexical node has been selected (Butterworth, 1992; Garrett, 1980; Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992; Schriefers et al., 1990). From this point of view, only the phonological representations of the selected lexical node will be activated. Another widely held view proposes that although phonological forms can only be activated after lexical nodes, the activation at the lexical level can flow onto the phonological level before lexical selection has taken place (e.g., Caramazza, 1997; Dell, 1986; Harley, 1993; Humphreys, Riddoch, & Quinlan, 1988; MacKay, 1987; Stemberger, 1985). These two views have been referred to as the *serial* and *cascade* hypotheses of speech production.

A crucial distinction between the two hypotheses is whether unselected lexical nodes activate phonological representations: Cascade models posit that unselected lexical nodes do activate phonology, whereas serial models claim that only the selected node can activate phonology. Evidence that the phonology of

[1] In psycholinguistics, this representation has traditionally been referred to as the *lemma* (Kempen & Huijbers, 1983; Levelt, 1989), but we avoid this term because it is associated with specific theoretical views. Instead, we use the more neutral term *lexical node*.

unselected words is activated would falsify serial models. Cascade models, on the other hand, are more difficult to test because when faced with the potentially problematic finding of a lack of such activation, they could hypothesize that the activation exists but is too weak to be detected. In the literature there is mounting evidence consistent with the cascade hypothesis (e.g., Costa, Caramazza, & Sebastian-Galles, 2000; Cutting & Ferreira, 1999; Jescheniak & Schriefers, 1998; Martin, Dell, Saffran, & Schwartz, 1994; Peterson & Savoy, 1998), but there is also some evidence that is apparently consistent with the serial hypothesis (Levelt et al., 1991; Schriefers et al., 1990). For reasons outlined below, the evidence at hand has left the serial–cascade question still a matter of great contention. In this article we address the controversy by first examining some of the evidence cited in support of the cascade hypothesis and then by reporting an experiment we conducted to test a crucial distinction between the two models— whether unselected lexical nodes activate phonology.

Historically, one of the first sets of evidence favoring cascade models came from speech-error analyses. Errors can come in various guises, and they differ as to how the erroneous word is related to the intended word. Some errors are semantic (e.g., saying "tiger" instead of "dog"), and some are phonological (e.g., saying "dof" instead of "dog"). Because some of the semantic competitors may also happen to be phonologically related to the intended word, sometimes a speech error is semantically and phonologically related to the intended word (e.g., saying "rat" instead of "cat"). These are known as mixed errors. According to serial accounts, all other things being equal, a mixed error should be no more likely to occur than a regular semantic error, because its phonological relation to the intended word is purely incidental. Contrary to what is predicted by serial models, analyses of spontaneous and experimental speech, from normal and brain-lesioned speakers, show that this is not the case: Mixed errors occur more frequently than they would if they were selected by chance (Dell & Reich, 1981; Martin et al., 1994). Cascade theories explain mixed errors as a product of the phonological activation of unselected lexical nodes, which is a natural consequence of the cascade architecture (Dell, 1986; Stemberger, 1985). Specifically, these errors result from the phonological activation of lexical nodes that are semantically related to the target word. For instance, when intending to say "cat," the semantically related lexical nodes "rat" and "pig" also send activation to the phonological level. Because *rat* happens to be phonologically related to *cat*, its phonological features will also receive activation from the lexical node "cat." A lower activation level is reached by the phonological features of *pig*, because they are activated only by the lexical node "pig." If the probability of producing an erroneous word is in part proportional to the activation level reached by the word's phonological features, one is more likely to say "rat" instead of "cat" than "pig" instead of "cat." Though mixed errors seem to be incompatible with serial models, by incorporating some additional assumptions such errors can be explained by these models. For example, it is possible to consider mixed errors as reflecting the functioning of a postencoding, speech production editor, whose task it is to monitor the production system (Levelt et al., 1991). The editor is purportedly less accurate in detecting mixed errors because of these errors' semantic and phonological resemblance to the desired target, which makes them less salient and thus less detectable as errors.[2] On these grounds, Levelt et al. (1999) dis-

missed the speech-error data, claiming that it cannot serve as unequivocal support for a cascade model of speech production; instead, they suggested that stronger evidence would come from reaction-time experiments, which more closely elucidate the normal performance of the speech-production process.

Of the reaction-time experiments bearing on the serial–cascade debate (Rapp & Goldrick, 2000, for review), we focus on the experiments that attempted to determine whether unselected lexical nodes activate phonology—the issue directly examined by our experiment. Despite extensive investigations, this remains an issue of contention, in part because serial models have been capable of explaining data that, at first, appeared to be inconsistent with their fundamental tenets. This was the fate, for instance, of the findings obtained by Starreveld and La Heij (1995) with the picture–word interference paradigm. In the picture–word interference paradigm, participants are instructed to name a picture and ignore a written word distractor that is presented along with the picture. Word distractors disrupt picture naming, which takes significantly longer when distractors are shown. However, interference varies as a function of the relation between the picture and the distractor word (for reviews see Deyer, 1973; MacLeod, 1991). If one considers the interference obtained with unrelated picture–word pairs (e.g., *cat–sky*) as baseline, interference is larger with semantically related pairs (e.g., *cat–horse*) and significantly reduced with phonologically similar pairs (e.g., *cat–mat*). Starreveld and La Heij (1995) were interested in the interference produced by picture–word pairs that were both semantically and phonologically related (e.g., *cat–rat*). According to the serial hypothesis, the semantic characteristics of a distractor should affect lexical node selection only, and the phonological features of a distractor should affect phonological encoding only. This is because, from the serial point of view, lexical node selection and phonological encoding are independent processes. Hence, Starreveld and La Heij (1995) reasoned that the effects from a distractor that is both semantically and phonologically related to the target should be additive—that is, no evidence of statistical interaction should be found. Contrary to these predictions, Starreveld and La Heij (1995) observed a statistical interaction for semantically and phonologically related distractors. This interaction was interpreted as problematic for the serial hypothesis and as support for cascade models, but not without controversy.

In response to Starreveld and La Heij's (1995) study, Roelofs, Meyer, and Levelt (1995) pointed out that whether the interaction found with semantically and phonologically related distractors is at odds with serial models depends on the interpretation given to the phonological effects generated by written-word distractors. It is possible that written-word distractors activate phonology through mechanisms that bypass the lexical nodes and reach phonology directly from orthography (for a thorough exposition, see Roelofs et al., 1995). The point is that the data generated by the picture–

---

[2] It is far from clear how the editor would work within the framework of serial models. According to serial models, the phonology of the intended word is not available until it is selected. Hence, it is not obvious how the editor would deem the sound of an erroneous word as incorrect if the phonology of the intended (albeit unselected) word cannot be retrieved. For a detailed discussion of the notion of an editor, see Santiago and MacKay (1999).

word interference paradigm could be interpreted in ways that are consistent with serial models. In short, the arguments raised by Roelofs et al. (1995) reveal that written words are ill-suited as stimuli to adjudicate between serial and cascade models of lexical access—an issue to which we return later.

Another, more recent set of evidence apparently in favor of cascade models comes from research on bilingual individuals. Costa et al. (2000) compared naming performance on two groups of pictures: cognate pictures, the names of which are phonologically similar in Catalan and Spanish (e.g., *gat–gato* [*cat*]), and noncognate pictures, the names of which sound different in Catalan and Spanish (e.g., *taula–mesa* [*table*]). When proficient bilingual individuals named pictures in Spanish, response latencies were faster for pictures with cognate names than for those with noncognate names. Costa et al. offered an explanation of this finding that is in line with the cascade hypothesis: The selected node (from the language being used, Spanish) and the unselected node (from the language not being used, Catalan) both send activation to the phonological level. Because the phonemes corresponding to cognate words receive activation from two sources, they reach a higher activation level than the phonemes of noncognate words. The confluence of activation from the two sources makes cognate words easier to say, all other things being equal.

Yet, it is possible to account for the cognate advantage within the framework of the serial architecture, as admitted by Costa et al. (2000). The advantage for cognate pictures might reflect a frequency effect—for a speaker, the combinations of phonemes constituting cognates tend to be very frequent because they happen to occur in both languages, as in the case of /ga/, which is found in the Catalan–Spanish cognates *gat–gato* [*cat*]. If naming latencies are in part a function of the frequency of phoneme combination, one expects cognate words to be named faster because, on average, they tend to be composed of frequently occurring phoneme combinations. This account, though logically plausible, remains hypothetical, for there is a lack of data showing that the frequency of phoneme combination affects naming latencies (for a discussion on these points, see Costa et al., 2000, and Levelt et al., 1999).

The reaction-time data reviewed thus far are naturally accounted for by cascade models, and serial models can accommodate them only after making ad hoc assumptions about experimental paradigms or about certain aspects of lexical access. If these assumptions hold, the serial hypothesis is saved. It is not so with the data obtained by Peterson and Savoy (1998) and Jescheniak and Schriefers (1998)—data that demand a revision of the serial hypothesis. For reasons of brevity, we illustrate only the finding of Peterson and Savoy here. They used a complex dual-task paradigm. In each trial, participants named a picture after a given cue appeared. On some trials, however, a word appeared instead of the cue, and in those cases participants were instructed to read the word aloud as fast as possible. The critical stimuli in this experiment were pictures with near-synonym names, that is, pictures with more than one acceptable name (e.g., *couch* and *sofa*). Participants were instructed to utter one of the two acceptable names—typically the one used more frequently (e.g., *couch*). The crucial finding was that when presented with the picture couch, for instance, participants were faster at reading the word *soda*, which is phonologically related to the unselected near-synonym of the picture (i.e., *sofa*) than the unrelated word *fork*. Similar findings were reported by Jescheniak and Schriefers (1998).

Are the findings with near-synonyms reconcilable with the serial hypothesis? According to Levelt et al. (1999), they are incompatible with a strong version of the serial hypothesis, a version which restricts phonological activation to words that are effectively produced. In an alternative account, Levelt et al. (1999) proposed what one could call a weak version of the serial hypothesis, according to which there is not only the activation but also the selection of multiple lexical nodes in the special case of near-synonyms. Multiple selection occurs because each of the near-synonyms satisfies the conditions for lexical node selection, and hence, the lexical system inadvertently licenses the selection of more than one lexical node. When faced with the picture couch, for example, both "couch" and "sofa" are selected. In these circumstances, the phonological features of both near-synonyms are activated, leading to the facilitatory effects observed experimentally by Peterson and Savoy (1998) and by Jescheniak and Schriefers (1998).

If the weak serial hypothesis explains the phenomena of phonological facilitation associated with near-synonyms, then evidence for the multiple selection of other types of words raises problems for the hypothesis. This is indeed the case for the data obtained by Cutting and Ferreira (1999) with homophones, words that sound the same but differ in meaning. The words *ball*, as in toy, and *ball*, as in dance, are an example of homophones. In the version of the picture–word interference paradigm adopted by Cutting and Ferreira (1999), pictures appeared along with auditory distractors. Of interest here are picture–word pairs like (toy) ball–*dance*, in which the distractor is semantically related to (dance) *ball*, a homophone of the picture (toy) ball. Picture-naming latencies were faster for picture–word pairs like (toy) ball–*dance*. Phonological activation from unselected lexical nodes explains this facilitatory effect. At the semantic level, the word *dance* activates the related concept "ball" (as in a dance). Activation then flows from the concept (dance) "ball" to its corresponding lexical nodes and then to the phonemes that constitute the word *ball*. If the target picture is a (toy) ball, the phonological activation from (dance) *ball* makes it easier to say "ball." Cascade models seem to provide a natural explanation of the facilitatory effect reported by Cutting and Ferreira (1999). In contrast, it is not obvious how serial models can account for this effect, which may turn out to create serious, if not insurmountable, problems for them.[3]

In summary, there is no shortage of findings claiming to show that unselected lexical nodes can activate their phonological features. These findings have been cited in support of cascade models. The problem, as we have seen, is that these very findings, perhaps with few exceptions, are not completely incompatible with an alternative serial hypothesis that restricts phonological activation to selected lexical nodes. Obviously, attempting to account for this evidence comes with a cost for serial models, which are forced to make additional, severely constraining assumptions. Unfortu-

---

[3] Levelt et al. (1999) attempted to provide an explanation for the findings of Cutting and Ferreira (1999) within the framework of their serial model. Levelt et al. (1999) proposed that "the distractor 'dance' will semantically and phonologically coactivate its associate 'ball' in the perceptual network" (p. 17). It was further assumed that the perceptual network will directly activate the corresponding phonological features in the production lexicon. However, it remains mysterious to us how the word *dance* might activate the semantically related word *ball* at the perceptual level, in which semantic information is not specified.

nately, without evidence clearly demonstrating that unselected lexical nodes can activate phonology, deciding between a weak serial hypothesis and a cascade architecture in speech production remains a matter of taste. Our experiment represents a further effort to show that unselected lexical nodes indeed activate phonology and, as such, attempts to contribute to a resolution of this ongoing debate between serial and cascade models of lexical processing in speech production. Below, we describe the paradigm used in our experiment and how it addresses the issues under investigation.

### Picture–Picture Interference Paradigm

In this paradigm, participants were instructed to name only one of two colored pictures that were presented simultaneously, with one superimposed upon the other (see the examples in Figure 1). The paradigm can be considered a picture–picture variant of the Stroop task (Stroop, 1935), and participants are required to name pictures of a given color (green) and ignore pictures of another color (red). Slightly different versions of the picture–picture paradigm have been used in past studies. Glaser and Glaser (1989) showed picture pairs presented sequentially, and participants were instructed to name the picture that appeared first or second. With this paradigm, slower responses were found with semantically related pairs (e.g., *cat–dog*), whereas facilitation occurred with identical pairs (e.g., *cat–cat*). Tipper (1985) used overlapping composites like ours in a task in which a distractor picture became the target picture of the following trial. Tipper (1985) used semantically related composites to examine the extent to which the processing of the distractor picture was inhibited. In short, picture–
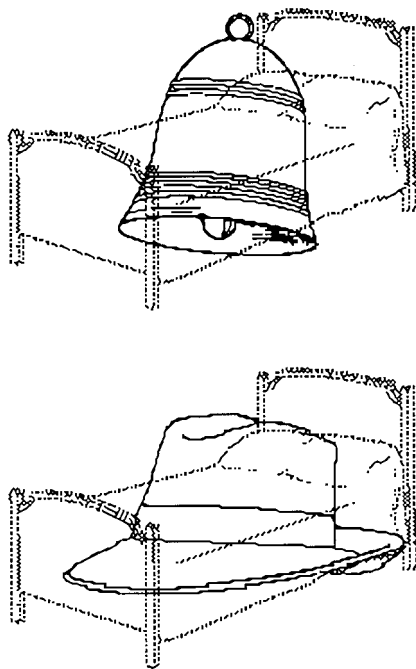


*Figure 1.* Sample stimulus of a phonologically related composite, *BED–bell* (with the lighter drawing [BED] being the green target and the darker drawing [BELL] being the red distractor), and an unrelated composite (*BED–hat*).

picture stimuli had never been previously used in the manner described below for the purpose at hand.

Our experiment examined the effect of pairs formed by pictures whose names are phonologically related as in *BED–bell* and *PIG–pin*. Henceforth, we refer to these pairs as *phonologically related composites*, and the target object is presented in capital letters. The reaction times for such cases were compared with those of phonologically unrelated composites (e.g., *BED–pin* and *PIG–bell*). We were operating under the assumption that because speakers did not name the distractor picture, the lexical node corresponding to the name of the distractor picture was not selected. This is the same assumption underlying the picture–word version of the interference paradigm (see Levelt et al., 1991; Schriefers et al., 1990). Because serial models of word production posit that unselected lexical nodes should not influence phonological processing, phonologically related distractors should not produce any effect whatsoever—neither facilitation nor inhibition. Conversely, any effect found would be incompatible with serial models. Such an effect, however, would be in line with cascade models of word production.

Although an effect in any direction would bear on the serial–cascade issue, we predicted that phonologically related distractors would facilitate naming, because this is what is found in the picture–word interference analogue of this experiment (Klein, 1964; Lupker, 1982; Rayner & Posnansky, 1978; Underwood & Briggs, 1984). It should be noted that in using the picture–picture paradigm we were avoiding the problems inherent in the interpretation of the phonological effects observed in the picture–word paradigm, problems that we briefly discussed above. Indeed, the only possibility for picture distractors to activate their phonology seems to be through the prior activation of their concepts and lexical nodes.

To make sure that, in truth, any effect found would be related to phonology and not to some other property of the composites (e.g., the identifiability of the drawings), a control study was carried out in Italian, a language in which the composites selected for English do not possess a phonological relationship. If the effect found in English is truly due to phonology, then it should not be found in Italian. Presumably, Italian speakers would be as sensitive as English speakers to any visual or semantic properties of the composites. If in Italian, too, responses are faster for related composites, then any effect found in English is probably not due to the phonological relationship between the targets and their distractors. On the other hand, if in Italian we do not replicate the difference observed in English, then phonology is the only feasible variable explaining an effect in English. In sum, the replication in Italian provides an adequate control for possible artifacts stemming from material selection and preparation.

### Experiment

#### Picture Naming in English and Italian

The experiment had two parts: In the first part, English native speakers named a set of picture–picture composites, and in the second part, Italian native speakers named the same set of composites. In some trials, the two pictures had phonologically related names in English (as in *BED–bell*). In another condition, the pictures were also paired with distractors that were phonologically

unrelated in English. Phonologically unrelated composites consisted of the same pictures forming the related condition but re-paired in such a way that there was no phonological relation between the target and distractor in English. By using the same items for both lists, we controlled for any effects that could arise from having lists composed of different materials. In Italian, the pictures forming the composites have phonologically unrelated names. Two procedures have been adopted to maximize the probability of detecting a phonological effect in English. First, the pictures of the phonologically related composites have English names sharing the largest number of phonemes without being homophones (as in *PIG–pin* or *BED–bell*).[4] Second, we selected only pictures that English speakers named consistently. In a preparatory study, 10 native English speakers named a set of pictures at a leisurely rate (none of these speakers took part in the experiment proper). Only pictures that received the same name by all 10 speakers were retained for the experiment.

We had a large number of unrelated filler composites (74%), so that only in a small percentage of the trials (13%) were the names of the target and distractor pictures phonologically related. This measure was aimed at discouraging participants from building strong expectations about the nature of the test materials. Finally, it is important to note that distractor pictures were never named during the experiment. This rules out that the phonological effect comes from having recently named the distractor picture or from having its name as one of the possible responses.

### Method

*Participants.* Thirty-nine native English-speaking students from Columbia University were paid for their participation. Another group of 32 native Italian speakers and students at the University of Padua, Padua, Italy, volunteered their participation in the control study. None of the Italian-speaking participants reported being a native Italian–English bilingual or having participated in any English courses besides the one-semester course that is mandatory at the university.

*Materials.* A total of 152 composites of superimposed red and green line drawings served as the stimuli. Composites were created with the program Aldus Superpaint (1993). The target was green, and the distractor was red. Composites occupied the center of the screen, covering roughly 3.5 in. (8.89 cm) in diameter. Of the 152 composites, 19 (13%) depicted objects that were phonologically related in English (as in *BED–bell* and *PIG–pin*). The pictures of these 19 composites were then randomly paired again with one another to form the unrelated condition, the items of which were not phonologically related in English (e.g., *BED–pin* and *PIG–bell*; in four cases, different pictures served as phonologically related and control distractors). The phonologically related composites and their controls formed the experimental set (see the Appendix for a complete list). Pictures of the phonologically related composites have monosyllabic names sharing the consonant onset and, in most cases, the vowel. Care was taken to make sure that neither the phonologically related nor the unrelated items had a semantic relationship. The 19 pictures were also paired with a new set of 38 distractors, and these composites served as one third of the fillers. We included these 19 target pictures with filler composites so that participants would have a reasonable number of targets to name in the practice phase (see below) and during the experimental session. Additional fillers were obtained by pairing each picture of a new target set ($n = 19$) with 76 distractor pictures. Filler composites were not phonologically related in any apparent way. A filler and an experimental composite were in all respects undistinguishable and named an equal number of times during the experiment (four times). To summarize, of the 152 composites

that formed the experimental session, 19 (13%) were phonologically related, 19 served as their control, and 114 were filler items.

The same materials and procedures used with English speakers were used in the second part of the experiment carried out in Italian. The Italian names of the composites are given in the Appendix. Related and unrelated composites do not, in their translation, bear a phonological relationship in Italian. The only exception is the unrelated pair *CANE–campana* [*DOG–bell*], in which the words share the initial consonant and vowel.

Each participant was run through four experimental blocks, each with 38 composites. Within each block, the presentation of the composites was in a pseudorandom order, according to the following criteria: (a) The first and last three composites of each block consisted of filler items; (b) each experimental target was never named more than once per block, and experimental targets never appeared contiguously within the blocks; and (c) a composite was never preceded by a picture–word pair that contained items that were semantically or phonologically related to it. Two lists were prepared that differed in terms of the order in which items were shown within the blocks. Half of the participants were presented with one list; the other half were presented with the other list. The order of presentation of the four blocks was randomly determined for each participant.

*Procedure.* Participants were run individually. The session began with a familiarization–training phase, in which participants named all of the 38 target pictures (of the experimental and filler pairs) twice at a leisurely rate. The pictures were presented as black-on-white drawings. The experimenter corrected participants whenever they referred to a picture by a name other than what we had selected for our list (e.g., saying "lips" instead of "mouth"). Next, participants practiced the picture–picture naming task. Participants were instructed to name the green picture of the composites, and not the red picture, as fast and as accurately as possible. In the practice block, each picture appeared twice with distractor pictures that were not re-presented during the experimental session. We opted for a rather long training session for two reasons. One reason, put forward in other studies (e.g., Caramazza & Costa, 2001; Starreveld & La Heij, 1995), was to ensure that participants would name the pictures without hesitation. The other reason was to allow participants to master the task of perceptually segregating two overlapping pictures.

In the practice phase and in the actual experiment, the picture–picture trial went as follows. A ready prompt (a question mark) appeared, and as soon as participants pressed the space bar, a fixation point (+) was shown at the center of the screen for 700 ms. The fixation point was replaced by the composite, which remained on the screen until participants responded or for up to 3,000 ms. Stimulus presentation was controlled by the PsyScope experiment software (Cohen, MacWhinney, Flatt, & Provost, 1993) on an iMac Macintosh computer. Naming responses were measured with a microphone (Model 33-3014; Radio Shack; Fort Worth, TX) and a PsyScope button box (Model 2.0.2; New Micros; Dallas, TX). Participant responses were manually coded by the experimenter. Responses in which participants hesitated or used names that were different from those expected, as well as responses presented along with microphone malfunction, were classified as errors and therefore excluded from analyses. Exceedingly long responses (more than 2 s) were also excluded. A trimming procedure removed outliers (responses more than three standard deviations above the mean of a participant's experimental items) from the remaining data.

### Results

*English.* Within the experimental conditions (phonologically related and their controls) erroneous responses were relatively infrequent, accounting for less than 2% of the data points (see

---

[4] In a pilot study with items that shared considerably less phonological overlap (e.g., just the onset, as in *SKIRT–skull*), we found negligible effects. This finding led us to use pairs with larger phonological overlap.

Table 1). Responses longer than 2 s and outliers were observed in 2.2% of the responses to the experimental items.

As shown in Table 1, English speakers were faster at naming pictures paired with phonologically related than with unrelated distractors. Such a conclusion was confirmed by paired $t$ tests that analyzed participants' means, $t(38) = -3.89$, $p = .0004$, and items' means, $t(18) = -2.11$, $p < .05$. Error rates did not differ reliably between the related and unrelated conditions in a by-subject analysis, $t(38) = 1.78$, $p = .08$, and in a by-item analysis, $t(18) = 1.83$, $p = .08$, although the statistics approach significance.[5]

*Italian.* Data were analyzed following the same procedure as in the English condition. The responses of 1 participant were excluded because they were excessively slow. Error rates were less than 1% for both conditions, and the difference in error rates between the two conditions was not statistically reliable ($ts < 1$). Outliers and responses longer than 2 s were observed in 3.1% of the data. Regarding response latencies, no difference was found in Italian between the composites that constituted the phonologically related and unrelated condition in English ($ts < 1$).

### Discussion

In English, we found a sizable phonological effect of 22 ms that was characterized, as we expected, by faster responses for phonologically related than unrelated composites. We have independent evidence that the distractor pictures are spontaneously named with the nouns that we selected. This renders more plausible the claim that the facilitatory effect reflects the phonological overlapping between the composite names. Moreover, by including the same pictures in both the related and unrelated conditions we ruled out that the phonological effect arose because we selected distinct lists for the two conditions.

The effect obtained in English was not replicated in Italian, and these contrasting results suggest two conclusions. On the one hand, it is unlikely that the effect found in English is attributable to variations in the visual characteristics of related and unrelated composites. If this were the case, then identical results should have been obtained in English and Italian. On the other hand, it can be concluded convincingly that the phonological relation between the target and distractors was responsible for the facilitatory effect manifest in English. In fact, word phonology is the only factor varying between the English and Italian conditions, and hence, it is the only possible cause for the discrepant findings. The implications of this experiment for theories of lexical access in speech production are examined in the General Discussion.

Table 1
*Picture Naming Latencies (M and SEM) and Percentage Errors Observed in English and Italian*

| Language | Related pairs | | | Unrelated pairs | | |
| | Response lat. | | | Response lat. | | |
| | M | SEM | % errors | M | SEM | % errors |
|---|---|---|---|---|---|---|
| English | 672 | 11 | 1.50 | 694 | 12 | 0.70 |
| Italian | 700 | 12 | 0.18 | 707 | 12 | 0.54 |

*Note.* lat. = latencies.

### General Discussion

In the picture–picture version of the Stroop task, we found that participants were faster at naming pictures with distractors that are phonologically related. We also showed that there is good reason to believe that this effect is phonological and not an artifact by (a) using the same pictures in the phonologically related and unrelated condition, (b) making sure that distractor pictures were named accordingly in English, and (c) running the experiment with non-English speakers. No effect was found in Italian, in which the picture names of the experimental composites do not bear any phonological similarity. This result is best accounted for by cascade models of speech production, which posit that unselected lexical nodes can influence phonology. In contrast, the effect seems to be problematic for serial models. We chose the picture–picture paradigm because it is not a candidate for the type of criticisms that were aimed at the speech error data and the picture–word interference data. However, it should be noted that our data are limited in at least one sense: The effect was found with a relatively small number of items. This small number was due to the difficulty of finding picture pairs with names that share a high degree of phonological overlap and the pictures of which have high naming consistency. It is very likely that we may have exhausted the pool of such picture pairs in English, making a replication in English with different items untenable. A test of the reliability of the phonological effect should then come from replications in other languages.

Our finding is predicted by the cascade hypothesis of the lexical system, and within its framework, the phonological activation from unselected lexical items could be instantiated in many ways. It could be that activation cascades from the activated lexical nodes onto the phonological level—a type of model assuming a strictly feed-forward propagation of activation (Humphreys et al., 1988). In our experiment, participants may have been faster with related items because the phonological features of these items have been activated by both the target and the distractor. The following example illustrates this point. Suppose that the target picture BED appears along with the distractor bell. By assumption, the distractor, bell, activates the phonological representations /b/, /e/, and /l/. At the moment of selecting the phonology of the target, *BED*, the onset and vowel of the target word have already received activation from the distractor lexical node. The activation sent by the distractor lexical node primes the phonemes of the target, facilitating their selection and ultimately the production of the target word. In another scenario, activation not only proceeds from the lexical nodes to the phonological features but it also bounces back from the phonological features to the lexical nodes (Dell, 1986; Stemberger, 1985). Feedback models can account for the facilitatory effect observed in our experiment by positing that phonologically related distractors not only facilitate the selection of the word phonemes but also the selection of lexical nodes. This second effect emerges as follows. Phonologically related distractors acti-

---

[5] This trend may lead one to suspect that there is the possibility of a speed–accuracy tradeoff. It should be noted, however, that in the vast majority of the participants ($n = 30$) there is no difference in the error rates between the related and unrelated conditions (27 made no errors at all and 3 made one error in each condition). In any event, the difference in the total number of errors in both conditions is minuscule (6 errors).

vate some of the target features, which in turn send activation back to their corresponding lexical nodes, including the lexical node of the target word. The extra activation reaching the target lexical node from the lower phonological level facilitates the selection of the lexical node, and in part explains why phonologically related distractors speed up picture naming. In sum, although both classes of cascading models can account for the phonological effect obtained in the picture–picture interference paradigm, their accounts presuppose slightly different mechanisms underlying the phonological effects observed in the task.

A serial explanation for our findings may appeal to a variant of the multiple-selection hypothesis discussed in the introduction. It could be proposed that the nature of our Stroop interference task is so demanding and unnatural that both the target and distractor lexical nodes are selected for production. Thus, in the case of the composite *BED–bell*, the lexical nodes of both words would have been selected. It could be added that speakers utter only one word because of the functions of a postencoding editor. This multiple-selection account is similar to that proposed by Levelt et al. (1999) to explain the data with near-synonyms (e.g., saying "couch" versus "sofa"; Jescheniak & Schriefers, 1998; Peterson & Savoy, 1998), wherein both lexical nodes are selected but somehow only one is produced. The first problem with this account of our findings is that considering the decision processes involved, it is not clear how the resolution of the multiple selection would lead to faster instead of slower response times for phonologically related composites such as *BED–bell*. If the postencoding editor cannot resolve mixed errors (Levelt, 1989; Levelt et al., 1999), then it is reasonable to expect that it would have more difficulties detecting and correcting similar items than dissimilar ones. A system based on functional principles of this sort seems at odds with our findings. The second problem is empirical and relates to the contrasting results observed between tasks requiring the production of single versus multiple phonologically related words. Stroop tasks are a prototypical example of the first class of tasks, and facilitation is the finding typically associated to phonologically related distractors (Lupker, 1982; Rayner & Posnansky, 1978; Underwood & Briggs, 1984). An example of the second class of tasks is the task devised by Sevald and Dell (1994), in which speakers are required to utter the largest number of four-words-sequences possible within 8 s. Having to produce strings composed of onset-related words (*pick–pin*) hurts performance, which is less efficient compared with strings formed by unrelated words. If in the picture–picture paradigm there is multiple selection, one would expect that in accord with other multiple-selection tasks, phonologically related distractors would lead to interference rather than facilitation. Finally, it would seem unlikely that proponents of the serial hypothesis suppose multiple selection in the picture–picture interference paradigm. Multiple selection was not proposed for the picture–word variant of the task. On the contrary, data from the picture–word interference paradigm were cited in support of the cascade hypothesis (Levelt et al., 1999).

Our results contribute to the increasing series of findings in favor of a cascade model of lexical selection. As we have seen, proponents of the serial hypothesis have been able to accommodate some of the findings that initially appeared to be irreconcilable with their hypothesis. It is not immediately obvious how the serial account could accommodate our data. A major challenge for proponents of the serial account would be to provide evidence that

not only supports a serial account but that would also be incompatible with cascade models.

In conclusion, we have found that words unselected for production, that is, words that will not be spoken, can nonetheless activate phonology. Although it is difficult to generalize from our experimental setting to everyday contexts, it would be interesting to discover whether phonology is always activated for all the things that happen to fall on the perceptual system, or whether this unintentional activation of phonology occurs only in the context of speech tasks. Beyond its implication for models of speech production, such a finding would bear on theories about the nature in which output programs are activated and selected in all actions, not just linguistic ones.

## References

Aldus Superpaint [Computer software]. (1993). Aldus: San Diego, CA.

Badecker, W., Miozzo, M., & Zanuttini, R. (1995). The two-stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. *Cognition*, 57, 193–216.

Butterworth, B. (1989). Lexical access in speech production. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 108–135). Cambridge, MA: MIT Press.

Butterworth, B. (1992). Disorders of phonological encoding. *Cognition, 42*, 261–286.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*, 177–208.

Caramazza, A., & Costa, A. (2001). Set size and repetition in the picture–word interference paradigm: Implications for models of naming. *Cognition, 80,* 215–222.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments, & Computers, 25,* 257–271.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1283–1296.

Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 318–344.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20,* 611–629.

Deyer, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition, 1*, 106–120.

Fromkin, V. A. (1971). The non-anomalous of anomalous utterances. *Language, 47*, 27–52.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language Production. Vol. 1: Speech and Talk* (pp. 177–220). London: Academic Press.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General, 118,* 13–42.

Goodglass, H., Kaplan, E., Weintraub, S., & Ackerman, N. (1976). The "tip-of-the-tongue" phenomenon in aphasia. *Cortex, 12*, 145–153.

Harley, T. A. (1993). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes, 8,* 291–309.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology, 5,* 67–104.

Jescheniak, J., & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24,* 1256–1274

Kay, J., & Ellis, A. (1987). A cognitive neuropsychological case study of anomia. *Brain, 110,* 613–629.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition, 14,* 185–209.

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology, 77,* 576–588.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechman, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98,* 122–142.

Lupker, S. J. (1982). The role of phonetic and orthographic similarity in picture–word interference. *Canadian Journal of Psychology, 36,* 349–367.

MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills.* New York: Springer-Verlag.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203.

Martin, N., Dell, G. S., Saffran, E. M., & Schwartz, M. F. (1994). Origins of paraphasias in deep dysphasia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language, 47,* 609–660.

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 539–557.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107,* 460–499.

Rayner, K., & Posnansky, C. J. (1978). Stages of processing in word identification. *Journal of Experimental Psychology: General, 107,* 64–80.

Roelofs, A. (1992) A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42,* 107–142.

Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1995). Interaction between semantic and orthographic factors in conceptually driven naming: Comment on Starreveld and La Heij. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 246–251.

Santiago, J., & MacKay, D. G. (1999). Constraining production theories: Principled motivation, consistency, homunculi, underspecification, failed predictions, and contrary data. *Behavioral and Brain Sciences, 22,* 55–56.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time-course of lexical access in production: Picture–word interference studies. *Journal of Memory and Language, 29,* 86–102.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition, 53,* 91–127.

Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 686–698.

Stemberger, J. P. (1985). Bound morpheme errors in normal and agrammatic speech: One mechanism or two? *Brain and Language, 25,* 246–256.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Tipper, S. P. (1985). The negative priming effect: inhibitory priming by ignored objects. *The Quarterly Journal of Experimental Psychology, 37*(A), 571–590.

Underwood, G., & Briggs, P. (1984). The development of word recognition processes. *British Journal of Psychology, 75,* 243–255.

## Appendix

### Picture Stimulus Lists

| | English names | | | Italian names | |
| | | Distractors | | | Distractors |
| Target | Related | Unrelated | Target | Related | Unrelated |
|---|---|---|---|---|---|
| BAG | bat | cage | BORSA | pipistrello | gabbia |
| BED | bell | hat | LETTO | campana | cappello |
| BOAT | bone | mouse | BARCA | osso | topo |
| BOW | bowl | cork | FIOCCO | scodella | tappo |
| CAN | cat | pin | BARATTOLO | gatto | spilla |
| CANE | cage | heart | BASTONE | gabbia | cuore |
| CHAIR | chain | bat | SEDIA | corona | pipistrello |
| CLOUD | clown | bone | NUVOLA | pagliaccio | osso |
| CORN | cork | fire | PANNOCCHIA | tappo | fuoco |
| COW | couch | bowl | MUCCA | divano | scodella |
| DOG | doll | bell | CANE | bambola | campana |
| FILE | fire | rain | LIMA | fuoco | pioggia |
| HAND | hat | clown | MANO | cappello | pagliaccio |
| HARP | heart | crib | ARPA | cuore | culla |
| MOUTH | mouse | nest | BOCCA | topo | nido |
| NET | nest | doll | RETE | nido | bambola |
| PIG | pin | towel | MAIALE | spilla | asciugamano |
| RAKE | rain | cat | RASTRELLO | pioggia | gatto |
| WHIP | witch | harp | FRUSTA | strega | arpa |