

Chapter 1

LEXICAL ACCESS IN SPEECH PRODUCTION: STAGES VERSUS CASCADING [1]

Willem J.M. Levelt

How does the speaker get from a (lexical) concept to the articulation of the corresponding word? There are at least two major processes involved here. The first one is to select the appropriate word from several thousands of alternatives in the mental lexicon. The second one is to compute an articulatory program for the target word on the basis of its abstract phonological representation. The present paper addresses the following question: Do these two processes overlap in time, thus allowing for interaction between them, or are they strictly successive? The first view is in line with connectionist approaches that predict some degree of phonological activation for all activated semantic alternatives to the target item. The second, two-stage view predicts that only the selected (target) item will undergo phonological activation. The paper reports on a series of experiments in which the time course of semantic and phonological activation during picture naming was measured. The results turned out to support the two-stage view.

INTRODUCTION

Lexical access in speech production proceeds at a rate of, on average, two or three words per second. At this rate words are selected from a production lexicon which contains thousands, probably tens of thousands of words. These words are not only selected, but also phonologically encoded. This happens at a rate of about 15 speech sounds per second. The problem to be addressed in this paper is how these high-rate and fairly accurate processes of lexical selection and phonological encoding proceed over time.

Figure 1 outlines a possible architecture for the organization of these processes of lexical access. There is a so-called "formulator" receiving as input the (lexical) concept-to-be-expressed (usually as part of a larger conceptualization) and producing as output an articulatory plan for item (usually as part of a plan for a larger utterance). The formulator contains two component processors. The first one takes care of selecting the appropriate lexical item from the mental lexicon and of integrating it into the developing syntactic structure (grammatical encoding). The second one generates an articulatory program for the selected lexical item on the basis of its stored phonological code and the developing phonological context of the utterance as a whole.

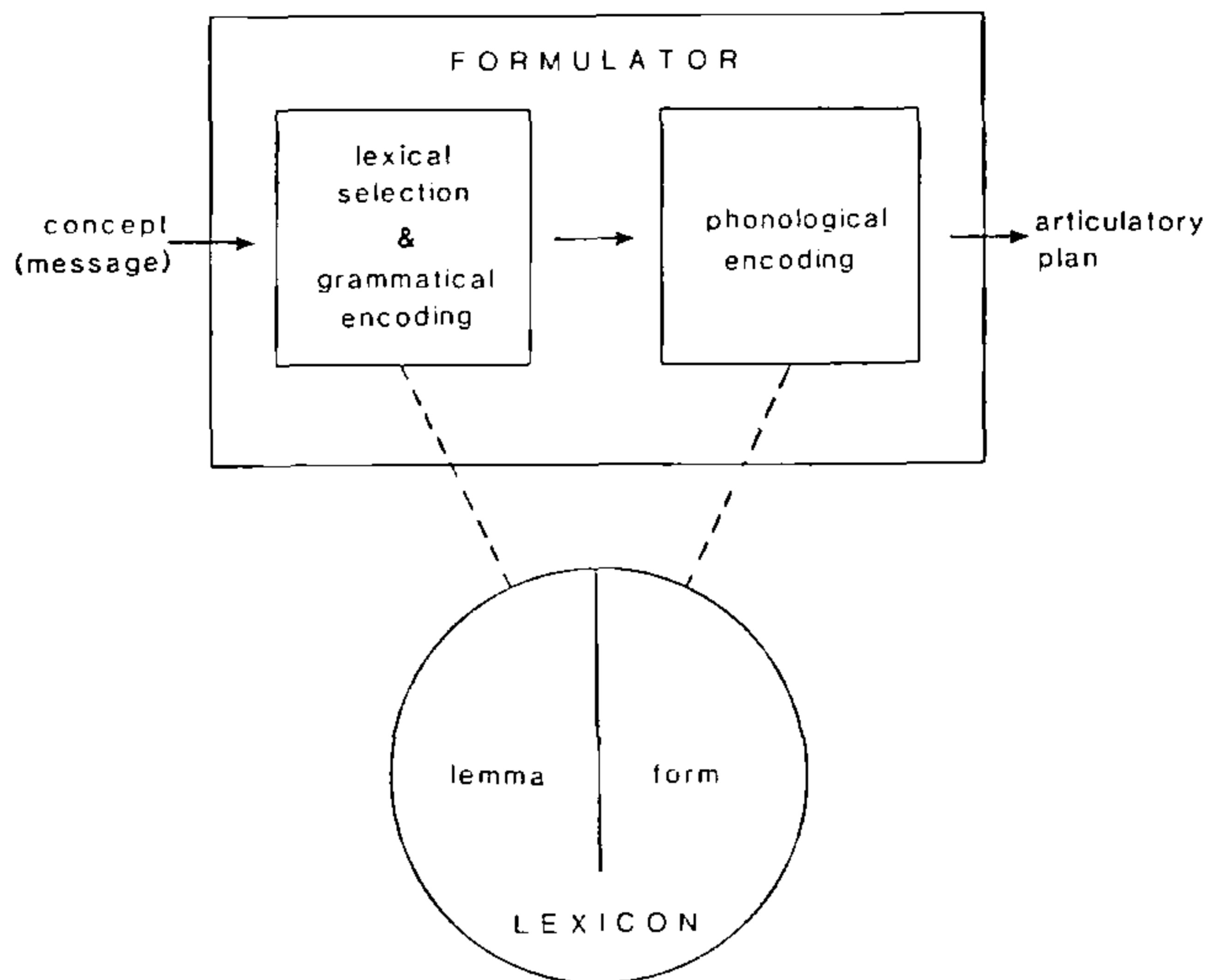


Figure 1. Outline of a possible architecture for lexical access.

Each of these two component processes may occasionally derail. If lexical selection goes awry, the errors such as these may occur: *Errors of lexical selection and grammatical encoding*.

Examples of these errors are:

- Don't burn your toes (intended: fingers)
- Examine the horse of the eyes (intended: the eyes of the horse)

Incorrect phonological encoding leads to a very different kind of error: *Errors of phonological encoding*.

Examples of this type of errors are:

- Fart very hide (intended: fight very hard)
- Face spood (intended: space food)

An extensive review of lexical selection, grammatical encoding, and phonological encoding can be found in Levelt (1989). The issue addressed in the present paper concerns the temporal alignment of lexical selection and phonological encoding.

THE TIME COURSE OF LEXICAL ACCESS

Returning to Figure 1, one can distinguish two views on the time course of lexical access. The first one is the more traditional modular view, which says that there is no phonological encoding before lexical selection and there is, accordingly, no feedback from phonological encoding to lexical selection. On this view, lexical selection and phonological encoding proceed through strictly successive *stages*. The second view is the connectionist picture, which assumes a temporal overlap

of lexical selection and phonological encoding, and a continuing interaction between the two processes. The temporal relation between lexical selection and phonological encoding is one of *cascading*. These two views are depicted in more detail in Figure 2.

In the classical theories (in particular Garrett's, Kempen's, Butterworth's, Levelt's - see Levelt 1989 for a review) there is an early stage of semantic activation, which ends up in lexical selection. It is followed by a stage of phonological encoding where only the selected item becomes phonologically encoded. In the connectionist theories (in particular Dell's, 1986), not only are *selected* items phonologically activated, but any semantically *activated* item.

There are three critical time course predictions proceeding from these theories;

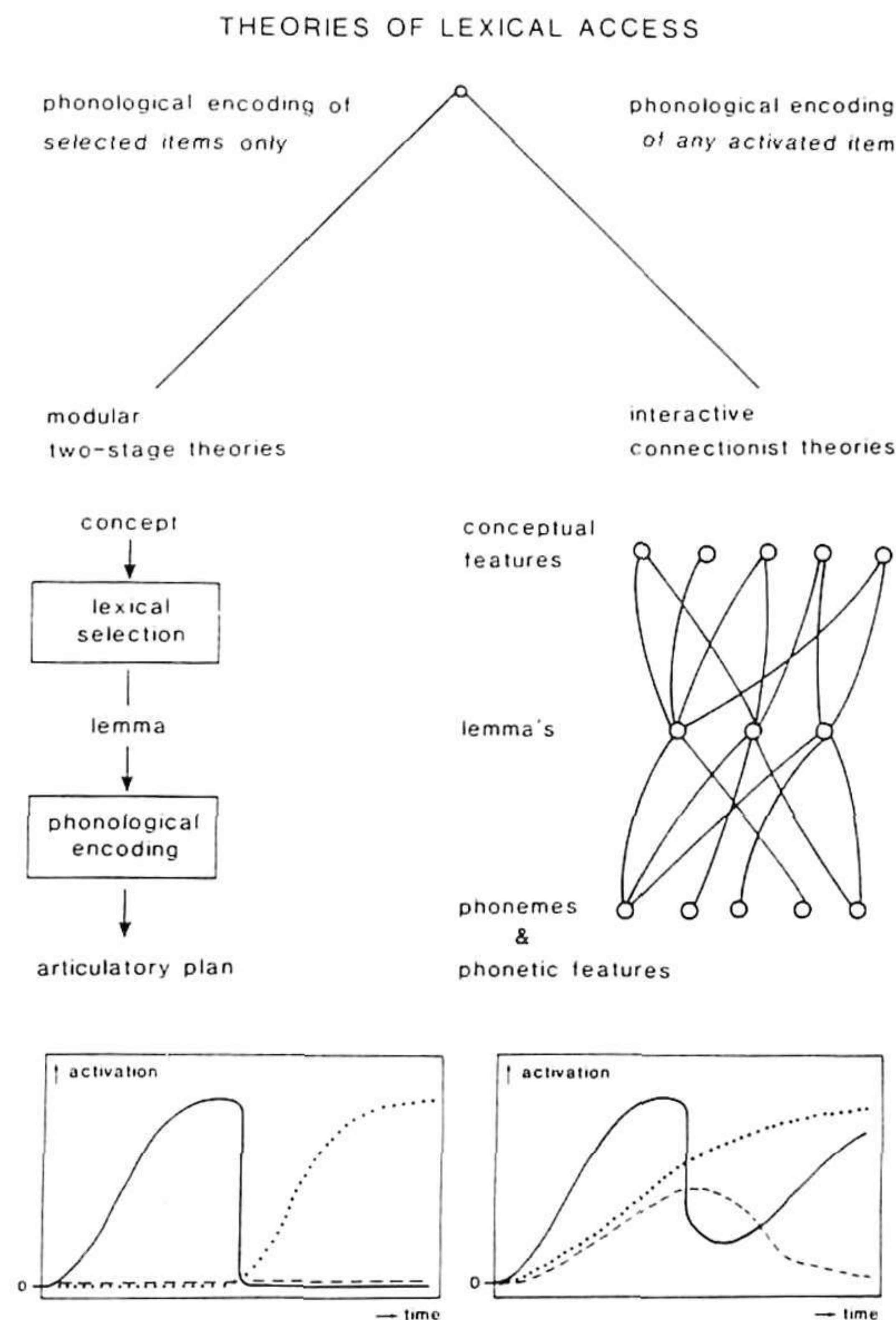


Figure 2. Two theories of lexical access. The modular two-stage theory (top left) allows for the phonological encoding of the selected target item only. The connectionist theory (top right) predicts phonological activation of semantic alternatives to the target item. The corresponding activation curves are presented at the bottom: Semantic activation of target (—), phonological activation of target (.....) and phonological activation of semantic alternative to target (---).

they are given at the bottom of Figure 2. The first one concerns the course of semantic activation. The modular theory predicts early, but no late semantic activation; the connectionist theory predicts early semantic activation and a later rebound of semantic activation, due to feedback from the phonemic to the lemma level. Second, the modular theory predicts late phonological activation, the connectionist theory predicts both early and late phonological activation. Third, the modular theory interdicts phonological activation of semantic alternatives (only the selected item, but no co-activated item becomes phonologically encoded). The connectionist theories, on the other hand, predict phonological activation of semantic alternatives to the target item.

We [2] performed several experiments to sort these predictions out. The experimental paradigm is presented in Figure 3.

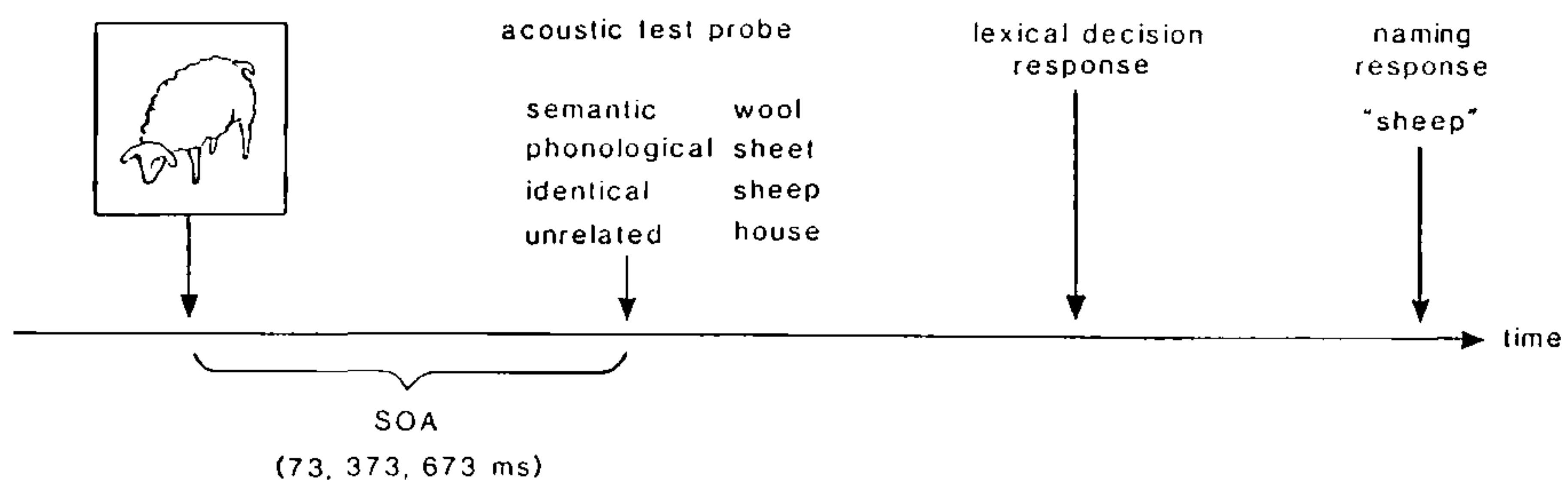


Figure 3. The experimental procedure of the naming-cum-lexical decision task. See text.

Subjects were asked to perform a picture naming task. A long series of pictures was presented, one by one, and the subject would name each picture as soon it appeared. Occasionally a secondary so-called "lexical decision" task was given. Shortly after presentation of the picture an acoustic test probe was presented, which could either be a word or a non-word, like *sip* (word) or *sef* (non-word). When this happened, the subject was supposed to push a "yes" or a "no" button, correspondingly. This task made it possible to probe into the subject's developing representation in his effort to produce the picture's name. For example, if the picture was one of a sheep the subject would internally generate semantic and phonological representations that were appropriate to the target name *sheep*. In order to test semantic activation of *sheep*, we would present as lexical decision probe a word like *wool*. There is reason to expect that semantic activation of *sheep* will delay the lexical decision to the acoustic test probe *wool*. Similarly, we could measure the phonological activation of *sheep* by presenting a test probe like *sheet*. In addition, the experiment contained the target word itself as probe (*sheep* in the example) - which I will not further discuss - and a control condition, namely a test word that is unrelated to the target, for instance *house*.

The critical issue is whether a semantic or a phonological test probe shows longer lexical decision latencies than the unrelated test probe. In order to see how

the semantic and phonological representations develop over time, we presented the acoustic probe at different delays after the picture. There were three moments: 73, 373, and 673 ms (on average) after picture onset. These are called "stimulus onset asynchronies", or SOAs.

The lexical decision latencies that we obtained in the experiment (192 subjects) are presented in Figure 4 (solid lines). In fact, these data are differences between the measured lexical decision latencies and lexical decision latencies for the same items when presented without concurring naming task. (The positive difference values in the figure show that the concurring naming task generally slowed down the lexical decision response).

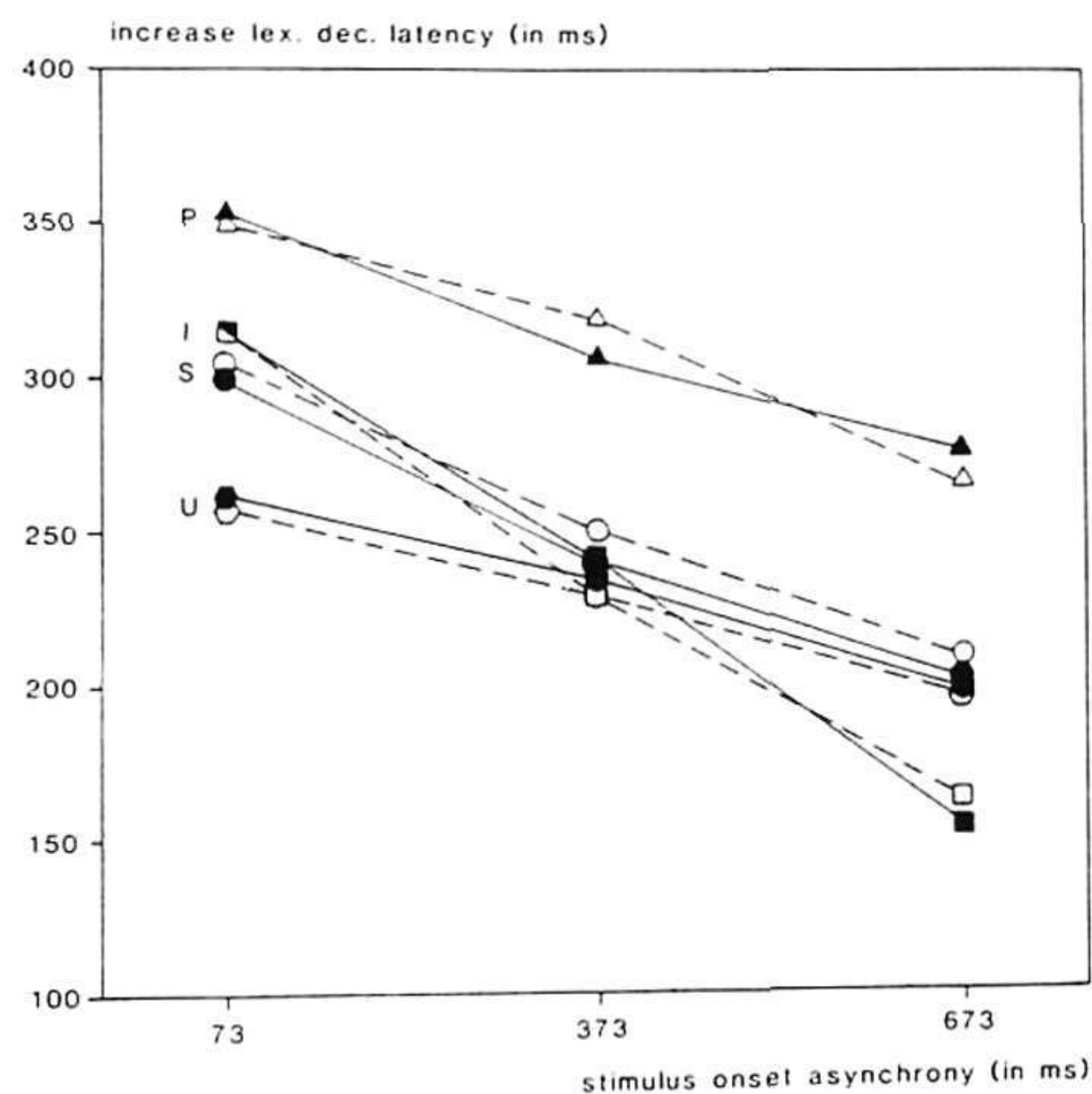


Figure 4. Increase of lexical decision latencies in the dual naming/lexical decision task (solid lines) for three different types of probe and three SOAs. S = semantic probes, P = phonological probes, I = identical probes, U = unrelated probes. The dotted lines show the fit of the two-stage model to the data

As both kinds of model predict, there is good evidence for early semantic activation (at the 73 ms. SOA the latency for the semantic probe is significantly longer than the latency for the unrelated probe). There is, however, no late semantic activation - contrary to the connectionist prediction. As both models predict, there is good evidence for late phonological activation, but seemingly contrary to the modular two-stage model there is evidence for early phonological activation. Hence, these results seem to be equivocal.

I will now argue that the two-stage model should be preferred. There are two arguments. First, I will show that the two-stage model can give a perfect account of these data. Second, I will report experimental results on the phonological activation of semantic alternatives, which are in support of the two-stage model.

The data in Figure 4 are the statistical result of a huge number of

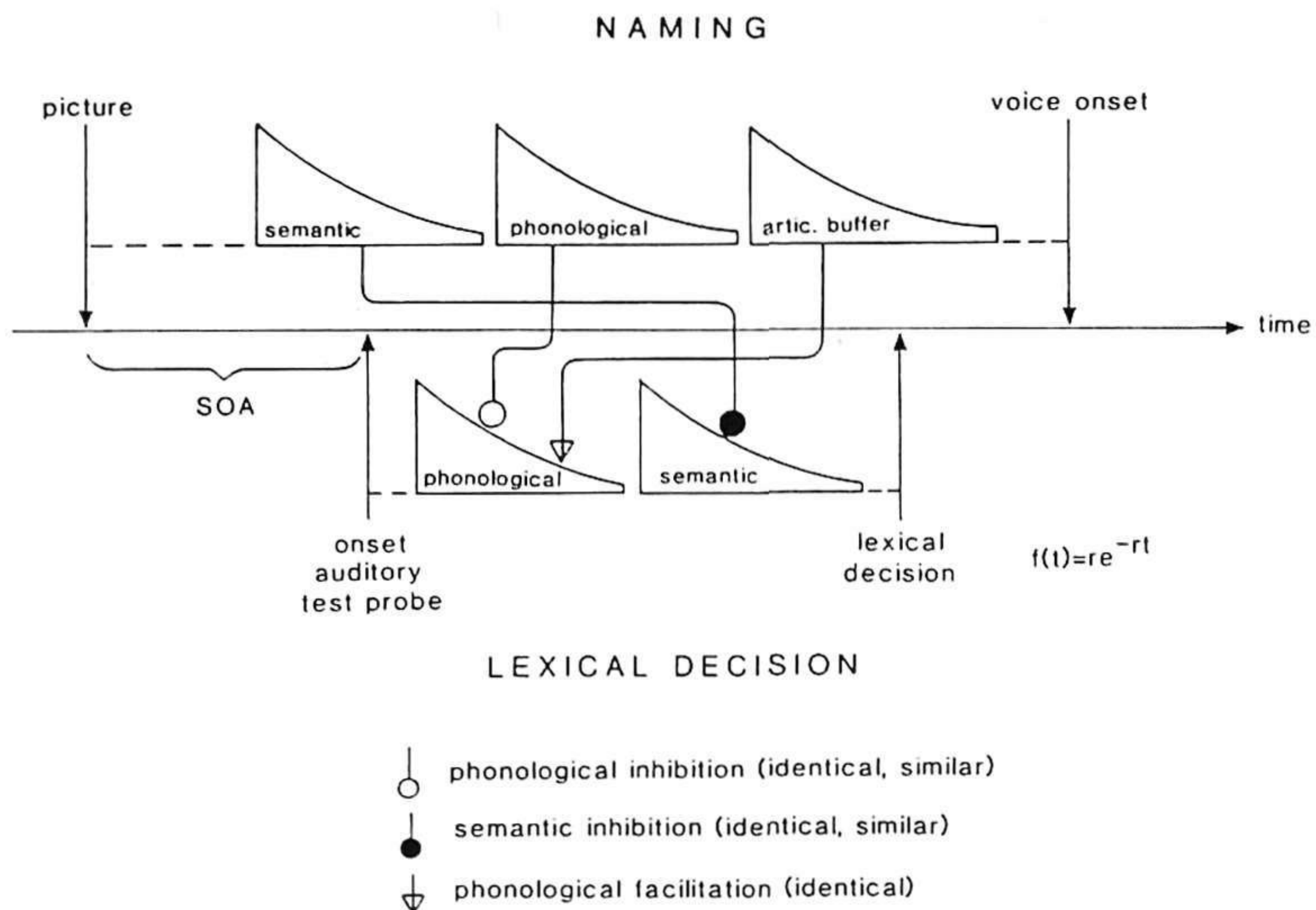


Figure 5. Mathematical rendering of the two-stage model.

measurements. It is therefore necessary to make a statistical model of this naming-cum-lexical decision task. Figure 5 depicts the model we [3] developed.

It incorporates the two-stage modular theory in that there is a strict succession of lexical selection and phonological encoding. The idea is that there will be interference when the semantic stage of naming coincides with the semantic stage of lexical decision, and when the phonological stage of naming coincides with the phonological stage of lexical decision, whenever same or similar items are involved in naming and lexical decision. The statistical time distribution of each of these phases is assumed to be exponential, with a characteristic rate parameter for each of the component processes. These rate parameters and the interference parameters can be estimated in order to find a best fit of the model to the data. This we did, and the result is presented in Figure 4, dotted lines. It showed that the data do not contradict the two-stage model. In fact, the fit is statistically perfect.

Turning now to the issue of phonological activation of semantic alternatives, I will report on an experiment that is quite similar to the previous one. But there are two differences. First, we used the short SOA (73 ms.) only, because it gave us both good semantic and good phonological activation in the previous experiment. Second, we used new acoustic test probes. Using again the example where the picture shows a sheep, we used the acoustic probe *goat* as a semantic probe, and the word *goal* as a phonological probe. This means that we can test whether the semantic alternative *goat* is not only semantically, but also phonologically active. In the latter case we should find an effect on *goal*. And that is what the connectionist theories predict.

Before reporting the results of this experiment, let me first remind you how strong a phonological activation effect we found for targets words like *sheep* in the previous experiment (i.e., the lexical decision latencies for phonological probes like *sheet*). They are given in Figure 6, together with the results for the unrelated test probes (such as *house*) as a comparison.

Now compare this to the phonological activation we found in the present experiment, i.e., for the semantic alternatives (i.e., for probes such as *goat*). These results are presented in Figure 7, together with the results for the unrelated test probes (such as *house*).

There is not the slightest trace of phonological activation (the result for phonological probes is not different from the result for unrelated probes), contrary to the connectionist predictions. One might, of course object that there was no activation of the semantic alternatives (such as *goat*) to start with. But that is not so. Figure 7 also presents the lexical decision latencies for semantic alternatives

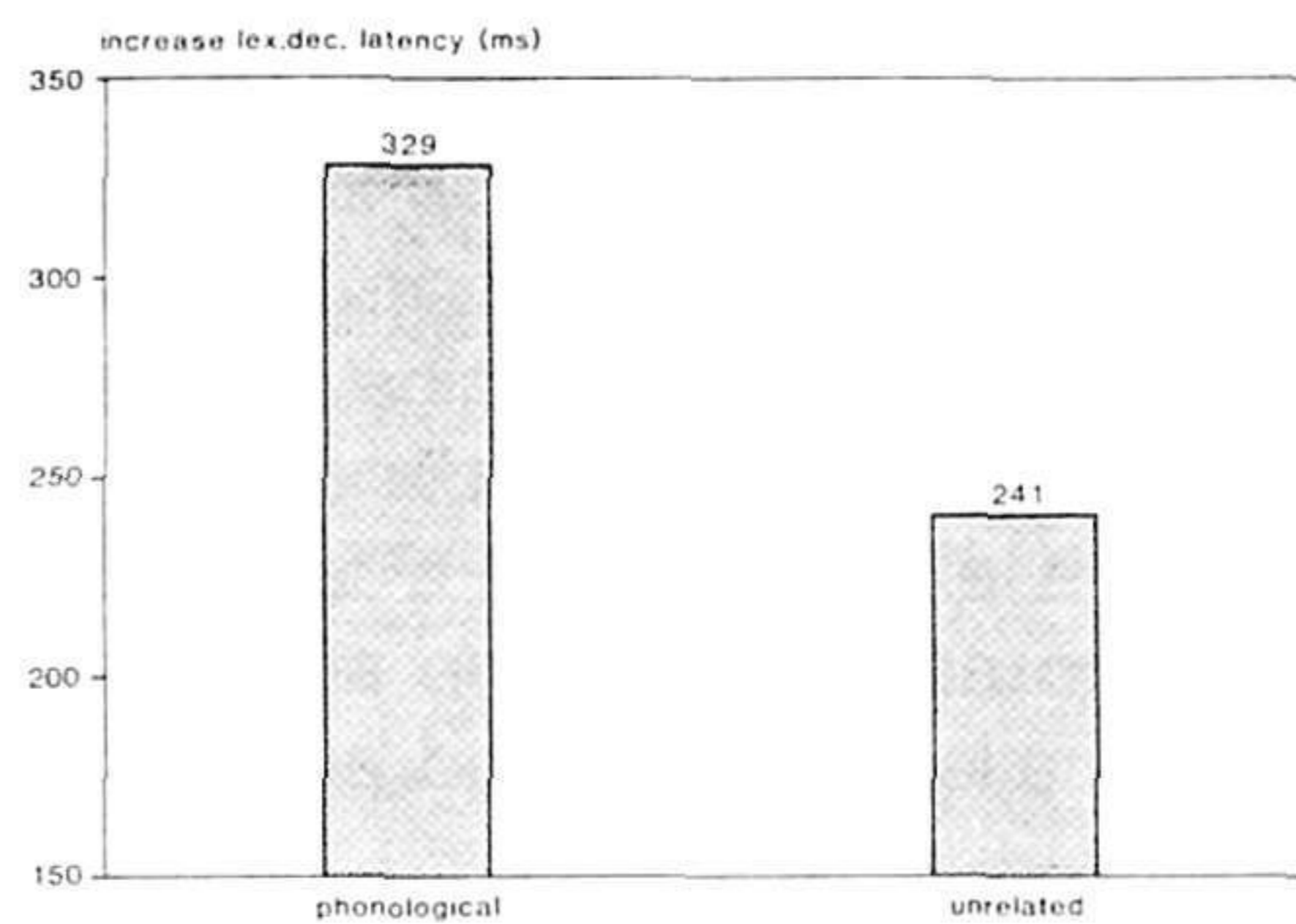


Figure 6. Increase of lexical decision for probes phonologically related to the target and for unrelated probes.

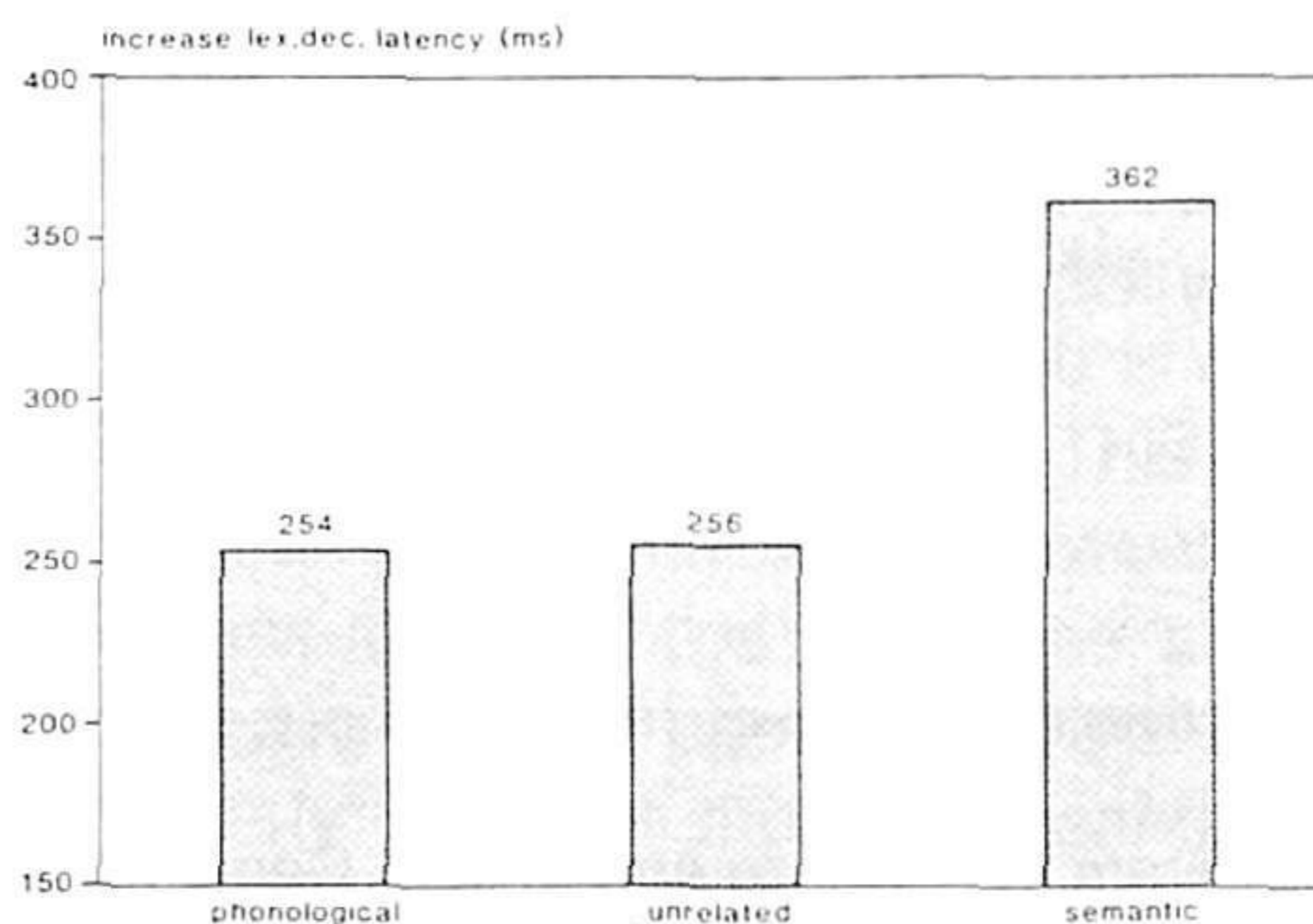


Figure 7. Increase in lexical decision latency for probes that are phonologically related to a semantic alternative, for unrelated probes, and for probes that are semantic alternatives themselves.

such as *goat*. There is a highly significant effect here if one compares the results for these semantic alternatives to those for unrelated lexical decision probes (such as *house*). For a more comprehensive and balanced treatment of the above findings, the reader is kindly referred to Levelt et al. (in press).

CONCLUSION

Taken together, the reported results support the modular two-stage notion of lexical access. (Further experimental support for this notion can be found in Schriefers et al., 1990). An important remaining question is: what could be the biological utility of such a modular architecture for lexical access? The obvious answer is that modularity is nature's protection against error-proneness of a system. The two components of the lexical accessing mechanism have to perform wildly different tasks. Lexical selection involves fast search in a huge lexicon. Phonological encoding involves the creation of a motor program for a single selected lexical item. If these processes were to interact, one would increase mutual interference without obvious functional advantages. Such interference would lead to errors of lexical selection and of phonological encoding. Though errors of these kinds do occur, their rate is astonishingly low for a process so complex and so fast as lexical access. Errors of lexical selection are probably below one per thousand selected items, and errors of phonological encoding are even rarer.

Still, one cannot exclude that Dell's (1986) connectionist theory - which, after all, gives a powerful and unified explanation of a large variety of speech error phenomena - can be adapted to accommodate the above experimental findings. I have been informed that modifications of this kind are in the offing.

NOTES

- [1] This paper is shortened and adapted version of W.J.M. Levelt, "Lexical access in speech production". In: R. Reuland and W. Abraham (Eds.). *Knowledge and Language*, Volume I. Dordrecht: Kluwer Academic Publishers (1991, in press). Reprinted by permission of Kluwer Academic Publishers.
- [2] Apart from myself, the research team involved Herbert Schriefers, Antje Meyer, and Thomas Pechmann.
- [3] The model was largely developed by Dirk Vorberg with the assistance of Jaap Havinga.

REFERENCES

- Dell, G.S. (1986). A spreading activation theory of retrieval in speech production. *Psychological review*, 93, 283-321.
- 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, Th. & Havinga, J. (in press). The time course of lexical access in speech production. A study of picture naming. *Psychological Review*.
- 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.