Cognition, 42 (1992) 1–22

Accessing words in speech production: Stages, processes and representations*

Willem J.M. Levelt

Max-Planck-Institut für Psycholinguistik, Wundtlaan 1, NL 6525 XD Nijmegen, Netherlands

> Ich glaube, daß mancher großer Redner, in dem Augenblick, da er den Mund aufmachte, noch nicht wußte, was er sagen würde (Heinrich von Kleist, 1809).

Abstract

Levelt, W.J.M., 1992. Accessing words in speech production: Stages, processes and representations. Cognition, 42: 1–22.

This paper introduces a special issue of Cognition *on lexical access in speech production. Over the last quarter century, the psycholinguistic study of speaking, and in particular of accessing words in speech, received a major new impetus from the analysis of speech errors, dysfluencies and hesitations, from aphasiology, and from new paradigms in reaction time research. The emerging theoretical picture partitions the accessing process into two subprocesses, the selection of an appropriate lexical item (a "lemma") from the mental lexicon, and the phonological encoding of that item, that is, the computation of a phonetic program for the item in the context of utterance. These two theoretical domains are successively introduced by outlining some core issues that have been or still have to be addressed. The final section discusses the controversial question whether phonological encoding can affect lexical selection. This partitioning is also followed in this special issue as a whole. There are, first, four papers on lexical selection, then three papers on phonological encoding, and finally one on the interaction between selection and phonological encoding.*

Issues of lexical access

How do we access words when we speak? This question has not received serious scrutiny until relatively recently. But as soon as it was asked, a whole range of issues emerged.

*I am grateful to Aditi Lahiri for her important remarks on syllable representation and the association process, and to Antje Meyer for her helpful comments on an earlier version of this paper.

What is the rate of lexical access in normal conversation? Some 120–150 words per minute on the average (Maclay & Osgood, 1959), but there are spurts of up to double this rate (Deese, 1984).

How many words do we have to select from? We don't know. There are reliable ways of estimating the size of our word recognition lexicon (Oldfield, 1963, estimated the vocabulary size of Oxford undergraduates at about 75 000 words), but no such tests exist for measuring the active production lexicon. Levelt (1989) estimated the production lexicon of normal educated adults at about 30 000 words, but this can easily be out by a factor two. Still, there is no doubt that we can access a huge lexical database at high rates, over long stretches of time, and without signs of fatigue worth mentioning. This alone characterizes lexical access as a cognitive skill par excellence. The skill is further marked by an astonishingly low error rate. Garnham, Shillcock, Brown, Mill, and Cutler (1982) found 86 errors of lexical selection in a spoken text corpus of 200 000 words, and 105 other slips of the tongue. That is an error rate of about one per thousand. Butterworth (this issue) gives similar data. It is important to stress this low error rate, because much of what we know about lexical access is based on careful analyses of naturally occurring speech errors. Reading this literature may create the misleading impression that felicitous lexical access is a matter of good luck rather than of exquisite design.

Are we aware of how we do it? As for most other high-speed skilled behaviour, the answer is "no". We can muse about the meanings of lexical items. We can even reject a word that jumps to mind and go for a more appropriate one. But we cannot trace the process by which we retrieve a word to start with. Introspection is largely useless in the study of lexical access.

This being so, another important issue became how to study the process. Since the 1960s and 1970s (and in fact since Meringer & Mayer, 1895), the dominant answer has been to study failures of access, slips of the tongue, speech errors (Cohen, 1965; Fromkin, 1971, 1973; Garrett, 1975; MacKay, 1970; Nooteboom, 1967; Shattuck-Hufnagel, 1979; see Cutler, 1982, for a bibliography of the early work). And indeed, this work has provided us with the main outlines of the processing architecture subserving speech in general, and lexical access in particular. Another approach has been the analysis of pre-lexical hesitations in spontaneous speech (Beattie and Butterworth, 1979; Goldman-Eisler, 1968).

It took longer until issues of lexical access were put to experimental test at any scale, or at least so its seems. The initial steps were to *elicit* speech errors in the laboratory, with Baars, Motley, and MacKay (1975) as the pioneers, or to elicit tip-of-the-tongue effects (Brown & McNeill, 1966). But in addition, reaction time paradigms intruded the study of lexical access, with Oldfield and Wingfield (1965) as pioneers and discoverers of the word frequency effect (see also for early reaction time studies Glaser & Düngelhoff, 1984; Klapp, 1974; Levelt & Maassen, 1981; Lupker, 1979). In fact, the reaction time study of lexical access was much

older, going under headings such as picture naming, colour naming, or even more disguised under the name of "Stroop effect" (cf. La Heij, 1988).[1] Glaser (this issue) reviews this history.

No less important, finally, is the study of the neuropathology of lexical access. Over the last decade or two there has been an increasing integration of the study of normal lexical access and its pathology. Garrett (this issue) and Butterworth (this issue) review the state of the art, but see also Caramazza and Hillis (1990) and Bub and Caplan (in press, chapter 4).

At present, research in lexical access has a pluralistic methodology, ranging from the analysis of naturally observed slips of the tongue, via error elicitation, to picture naming and picture–word interference studies. In addition, the pathology of lexical access in aphasic patients is increasingly contributing to our understanding of the underlying mechanisms.

This brings us to the main issue that emerged. What kind of processing mechanism governs the skill of accessing words? If we cannot introspect the mechanism, we are at the mercy of our theoretical inventiveness. The first serious proposal was Morton's (1969) logogen theory, which is still a significant competitor on the theoretical battleground. The mental lexicon was conceived of as a pandemonium, a collection of so-called *logogens*, each sensitive to its own specific information. For speech production (exclusive of reading) a logogen's relevant information stems from the "cognitive system", which is semantically active. The logogen becomes activated by semantic information relevant to "its" word. When the activation exceeds some threshold value, the logogen fires, and sends the phonological code of its word to a so-called "response buffer", from which an overt articulatory response can be initiated.

The logogen theory has (at least) two attractive features. One is that all logogens are simultaneously active in "watching" the cognitive system. There is parallel processing, which makes the speed of access largely independent of the size of the lexicon. The other is that lexical access is a two-step process. The first step, the logogen's activation to threshold, is semantic in nature. The second step, the logogen's firing and the preparation of response execution, is phonological in nature.

This two-step approach to lexical access is, in one guise or another, common to all modern views of lexical access (cf. Butterworth, 1989). There are two component processes to lexical access. The first one is *lexical selection*, retrieving the one appropriate word from among thousands of alternatives. The second one is *phonological encoding*, computing the phonetic shape from the selected item's phonological code or form specification as it is stored in the mental lexicon (Kempen & Huijbers, 1983, called this stored phonological code the *lexeme* as

[1]The Stroop effect: it is relatively hard to name the colour of a printed word (for instance red) if that word is itself the name of a different colour (for instance green), a case of lexical interference or competition. See Glaser (this issue) for details.

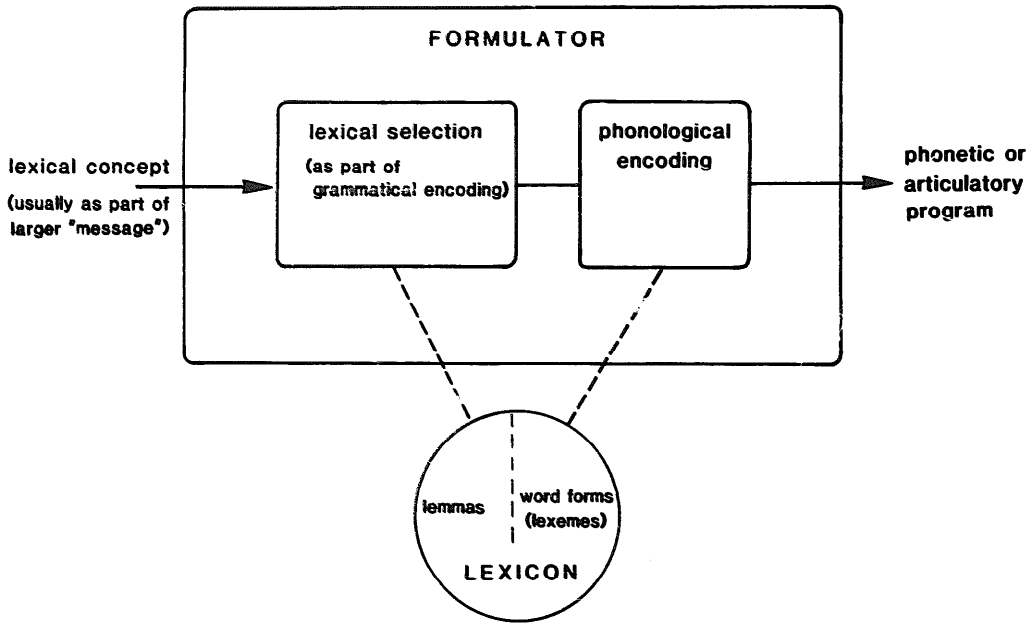


Figure 1. *An outline of lexical access in speech production.*

opposed to the *lemma*). This scheme is presented in Figure 1; it will be used as a guideline for the present introductory chapter of this special issue. In the following section I will first present some thoughts on lexical selection. This will be followed by a section on phonological encoding. The final section will consider some aspects of interaction between these two component processes.

This ordering corresponds to the layout of the present special issue. There are, first, four contributions on lexical selection. This is followed by three contributions on phonological encoding. And, finally, there is one paper on the staging and potential interaction between lexical selection and phonological encoding.

Lexical selection

A speaker's mustering of words usually serves the performance of some speech act. And a speech act is a way of revealing some communicative (and hence interactive) intention by means of spoken language. It is important not to ignore this larger perspective when discussing matters of lexical selection.

Recent years have seen substantial convergence on the following general picture of the initiation of a speech act. In order to reveal some communicative intention (e.g., to commit oneself or the interlocutor to some action, to share certain experiences with the interlocutor), the speaker will encode a so-called "message" whose expression can be effective in revealing that intention. So, for instance, if the speaker intends the addressee to recognize that his intention is to

let her know that her sister has arrived, an effective message might be the declaration that her sister has arrived. But it need not be, dependent on the context. It may, for instance, be more effective to declare that an angel or a witch (as the matter may be) has arrived. Or the context may make it even more effective for the speaker to express the question whether he might just have seen a woman entering the door.

The choice of message is a subtle function of the relation between the interlocutors, their common ground, the existence of secondary intentions, such as to understate or to express irony, and other factors (cf. Clark & Wilkes-Gibbs, 1986; Levelt, 1989). And these choices have an immediate impact on lexical selection (*sister*, *angel*, *witch*, *woman* in the above example, all intended to make reference to the same person).

It is widely held that a message is a conceptual structure, cast in a propositional language of thought. It forms the input to the so-called *formulator*, whose task it is to map the message onto linguistic form. Its final output is a phonetic plan that can be executed by the articulatory motor system. The formulator involves two component processes: *grammatical* and *phonological encoding* (see Figure 1).

Grammatical encoding takes a message as input, retrieves lexical items from the mental lexicon, and delivers a surface structure as output. A surface structure is a hierarchical organization of syntactic phrases. Its terminal elements are *lemmas*. These are lexical items unspecified for phonological form. They are, however, semantically and syntactically specified. Their semantic specification is a set of conceptual conditions whose fulfilment in the message is a necessary requirement for their retrieval. Their syntactic specification involves category and subcategorization information, as well as the way in which grammatical subcategory functions of the lemma are mapped onto the conceptual arguments in its semantic description (the thematic role assignments, see especially Jackendoff, 1990). So, for instance, the lemma *swallow* has as part of its semantics that some theme Y is ingested by some agent X. That is, these conceptual conditions must be fulfilled for the appropriate use of that lemma. Syntactically *swallow* is a transitive verb, subcategorizing for two grammatical functions: subject and direct object. The canonical thematic mapping for this verb is agent (X) to subject and theme (Y) to direct object.

Lexical selection drives grammatical encoding. Lemmas are retrieved (activated, selected) when their semantic conditions are met in the message. In their turn, they call (activate, trigger) syntactic procedures that correspond to their syntactic specifications. A verb will instigate the construction of a verb phrase, a noun the construction of a noun phrase, etc. Grammatical encoding is somewhat like solving a set of simultaneous equations, simultaneously realizing the appropriate thematic role assignments for all lemmas retrieved. Not quite simultaneous, however, because lemmas can become available at different moments in time, dependent on the speaker's unfolding of the message. Different orders of lemma

selection can lead to vastly different syntactic constructions (see Levelt, 1989, for an extensive review of grammatical encoding).

But there is a special set of lemmas whose retrieval is not conceptually driven. They all belong to the closed class vocabulary. In *the woman that arrived* the relative pronoun *that* is itself called by the syntactic procedure that constructs relative clauses. The retrieval of *that* is not semantically driven, such as the retrieval of *woman*. Here, in other words, grammatical encoding drives lexical selection. Notice, however, that many other closed class items do have some semantic specification, such as sex or definiteness; these specifications must be met at the message level for the item to be selected (except when they are syntactically derived by agreement).

I will now formulate some major problems for a theory of lexical selection. They are my own selection; others in this special issue will add several more of their own.

The hyperonym problem

It was noticed by Levelt (1989) that no existing theory of lexical selection, including Morton's logogen theory and Miller and Johnson-Laird's (1976) decision table theory, can account for unique convergence of the conceptual information onto the one appropriate lexical item. This is due to the hyperonym[2] problem, which was formulated as follows:

> When lemma A's meaning entails lemma B's meaning, B is a hyperonym of A. If A's conceptual conditions are met, then B's are necessarily also satisfied. Hence, if A is the correct lemma, B will (also) be retrieved.

So, when a speaker intends to express the concept CAT, then all conceptual conditions for the retrieval of the lemma *animal* are also satisfied, because the meaning of *cat* entails the meaning of *animal*. Why then does the speaker not say *animal* instead of *cat*?

There are two types of approach to this problem. The first one is to maintain some degree of compositionality in semantic representations. Bierwisch and Schreuder (this issue) argue this to be imperative anyhow. But then one should implement what I called a *principle of specificity* (Levelt, 1989). It should be clear that the hyperonym problem derives from an assumption of compositionality: if A is a lemma, and B one of its hyperonyms, then the meaning of B is assumed to be a constituent part of the meaning of A. As long as one maintains this much compositional semantics in the mental lexicon, then any theory of lexical selection

[2]In Levelt (1989) I introduced the term *hypernym* (as opposed to *hyponym*). The correct Latin–Greek compound should, however, be *hyperonym*, as Michael Studdert-Kennedy explained to me.

must involve a principle of specificity, which says that of all lemmas whose conditions are satisfied by the concept-to-be-expressed the most specific one (the most entailing one) should be selected. So, if the intention is to express the concept CAT, which satisfies the semantic conditions of the lemmas *cat*, *animal*, *thing*, etc., the most specific lemma (*cat*) should be retrieved. Notice that distributed connectionist representations of word meaning are also subject to the hyperonym problem (cf. Roelofs, this issue). I am not aware of any explicit effort to implement a specificity principle in such network representations.

The second approach is to give up compositionality altogether, as has been proposed by Fodor, Garrett, Walker, and Parkes (1980) and Dik (1987). The classical non-compositional implementation of lexical knowledge is by Collins and Quillian (1969) and Collins and Loftus (1975). Here concept nodes (such as for CAT) are directly linked to word form nodes (such as for /kat/). Intending to express the concept CAT means activating the corresponding node, which then spreads its activation to the appropriate word form node. Here, the hyperonym problem doesn't arise. The node for CAT may spread some of its activation via the *is-a* link to the node for ANIMAL, but the latter node's activation will normally be less than the former node's, so that the connected word form node for *animal* will receive relatively minor activation. Roelofs (this issue) has extended a non-compositional Collins/Quillian/Loftus-type network to incorporate lemma-level nodes. Then it becomes not only possible to correctly predict many of the naming results reviewed by Glaser (this issue), but also some new and unexpected ones in the realm of picture–word interference. Garrett (this issue) also argues for a lemma level as interface between conceptual and form representations, but his arguments stem from the pathology of lexical selection in aphasic patients.

It is hard to anticipate which of these two approaches will ultimately be more successful. Bierwisch and Schreuder (this issue) at least make us cautious not to give up compositionality altogether, and so does Jackendoff (1990). But to the extent that compositionality is maintained, the hyperonym problem will keep haunting us. Garrett (this issue) analyses hyperonym substitutions in Alzheimer patients and featurally defined semantic field effects in aphasics. These deficits underline the importance of the compositionality issue.

The dissection problem

How do we dissect an idea to be expressed into *lexical concepts*, that is, into conceptual "chunks" that map neatly onto lemmas? What makes a speaker say *John is Mary's father* rather than *John is Mary's male parent*? There are certainly thinkable contexts in which the latter is the more appropriate expression. Is it the same idea here that is dissected into one or into two lexical concepts? Or is there

no full synonymy here between *father* and *male parent*, following Eve Clark's (1987, 1990) "principle of contrast"? Clearly, the dissection problem is related to the issue of compositionality as Bierwisch and Schreuder (this issue) explicate. If each lemma has a unique relation to a single concept, the dissection problem will not arise (cf. Roelofs, this issue). The speaker will either have an active FATHER node, or two active nodes MALE and PARENT. This "solution", however, shifts the weight of the problem to the speaker's conceptual intentions. Is this the right level of analysis? Our preference for using so-called "basic level" terms (Glaser, this issue; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Seymour, 1979) indicates that we have a preference for chopping up our message in such a way that it can be mapped onto this basic level vocabulary. But it may be intractable what is the chicken and the egg here. We may, after all, prefer to think in "basic level concepts", whether or not we intend to express them in language.

The imitation problem

There is both conversational (Schenkein, 1980; Harley, 1984) and experimental (Levelt & Kelter, 1982) evidence that speakers tend to have some preference for words that have recently been used by the interlocutor or by themselves. This tendency cannot be fully explained by the interlocutors sharing the same topic of discourse (if the topic is for instance the American president, it is likely – anno 1991 – that the partners in speech will tend to use the word *Bush* a lot). The experimental evidence shows that the preference for re-using words even extends to words that are semantically non-discriminative. There is, in addition, a strong recency effect (Levelt & Kelter, 1982), which makes it likely that the effect is caused by a temporary extra activation of the relevant lemma, due to the speaker's hearing or using the word. Bock and Loebell (1990) found that, similarly, syntactic constructions can be induced without there being any conceptual-level grounds for it. This syntactic induction may, in turn, induce the selection of particular closed class items (such as *by* when a passive is induced). This might explain some of the Levelt and Kelter results, which concerned closed-class items. More generally, the fluency of formulating seems to be served by re-using recently activated words. In short, as long as a theory of lexical selection only acknowledges semantic or syntactic reasons for selecting words, these imitation phenomena cannot be explained.

The problem of collocations

This problem was recently formulated by Ward (1988, 1991). The selection of one word can depend on the selection of another word, without there being conceptu-

al reasons for this. Ward gives the example of *strong air currents* versus *high winds*. Conceptually, the modification is identical in the two cases, but different words (*strong*, *high*) have to be selected in dependence on the head that is modified. Another example is *to fall into disuse* versus *to sink into oblivion*. In this context *fall* and *sink* are virtually synonymous, but cannot be exchanged. Probably, the lemmas for *disuse* and *oblivion* carry phrasal information involving *fall* and *sink* respectively.

The collocation problem is not well separable from what one might call the *idiom problem*. Certain concepts are well expressed by idiomatic phrases, such as *red tape*. That this is a phrase, not a compound, appears from its iambic stress pattern. It differs from the trochaic stress pattern of a compound like *hot dog*. But an idiomatic phrase is peculiar as a phrase for two reasons. First, its meaning is opaque, not compositional (*red tape* has very little to do with either *tape* or *red*), and second it allows for only restricted syntactic variations (**my tape is even redder than yours*, **two red tapes*, etc.). It is not entirely clear how the speaker's production of idioms should be modelled. Probably, an idiom is a special kind of lexical entry, specified for the (opaque) meaning. If that semantic condition is met by the message, the idiom is retrieved, just as for any other content word. It calls the constituent lemmas (like *red* and *tape*) and imposes its degraded or limited syntax on the further process of grammatical encoding, in ways that are still to be explored.

These are some problems of lexical selection to be solved, and more will be added in the final section of this chapter and in the paper by Dell and O'Seaghdha (this issue), where the potential influence of phonological encoding on lexical selection is discussed. The main problem, however, is to develop theoretical frameworks in which these problems can be addressed. Such frameworks will be at issue in the following papers by Bierwisch and Schreuder, Glaser, Roelofs, Garrett, and Dell and O'Seaghdha.

Phonological encoding

The second phase of lexical access in speech production is phonological encoding. Eventually, the selected lexical item must be given phonetic shape. A word's phonetic form is not a ready-made template that can be retrieved as a whole. Speech error research has made it abundantly clear that a word's ultimate shape is to be *constructed* time and again. An error such as *peel like flaying* (instead of the intended *feel like playing*) reveals that a word's "skeleton" can (at least to some extent) be specified independently from the segments that have to fill it. In the process of constructing *feel* the speaker apparently missed the segmental information /f/. But the fact that the speaker didn't say *eel* then, but *peel* makes it likely

that there was already an active word skeleton requiring an onset consonant. The onset slot was then erroneously filled by the already available segment /p/. Though the segment /p/ had been used now, the speaker did not proceed to say *laying* instead of *playing*. Rather, the word *flaying* was constructed. Presumably also here the word's skeleton was already available; its first consonantal slot was then filled by the now available /f/, thus creating a second error. Probably the most fundamental insight from modern speech error research is that a word's skeleton or frame and its segmental content are independently generated. Shattuck-Hufnagel's (1979) slot-and-filler theory was the first formal rendering of this insight (see Meyer, this issue, for a review).

Far less attention has been paid in the literature to the question why this should be so. In fact, the frame-filling notion seems quite paradoxical. Why would a speaker go to the trouble of first generating an empty skeleton for the word, and then filling it with segments? In some way or another both must proceed from a stored phonological representation, the word's phonological code in the lexicon. Isn't it wasteful of processing resources to pull these apart first, and then to combine them again (on the risk of creating a slip)? As Meyer (this issue) shows, this question has been essentially ignored in the standard accounts of phonological encoding.

The answer must probably be sought in the generation of connected speech (cf. Levelt, 1989). Talking is mapping discrete linguistic representations onto pronounceable and continuous phonetic programs. The construction of frames serves the purpose of creating a pronounceable metrical pattern for the utterance as a whole. And that metrical pattern is *not* just a concatenation of individual word frames. It is rather more the exception than the rule that a word's stored skeleton will eventually turn up as a frame to be filled. The speaker produces frames for *phonological words* (ω). These are metrical units, not lexical units. A phonological word is the domain of syllabification (see below) and of word stress assignment. It is never smaller than a morpheme, but it can be larger. In English (but not in all other languages) a phonological word is composed of a head word with its affixes and clitics; there may even be two or more head words involved (as in certain compounds).[3] In *Black Bear gave it him*, there are two phonological words: *Black Bear* and *gavitim*. The former one is a compound with its characteristic trochaic word stress, and corresponds to a single (compound) item in the mental lexicon. The latter one derives from a head word (*gave*) and two dependent words (*it* and *him*) that are cliticized to the head word.

The domain of syllabification in speech production is precisely the phonological

[3]Nespor and Vogel (1986) distinguish between a phonological word and a clitic group. The phonological word is the domain of syllabification. The clitic group allows for more limited syllabic interactions only. However, Lahiri, Jongman, and Sereno (1990) argue that the clitic group notion might be superfluous, at least for the phonology of Dutch. I will follow the latter authors in assuming that cliticization results in the creation of a phonological word.

word. So, for instance, *gavitim* is syllabified as *ga-vi-tim* (here the /v/ may be ambisyllabic), which violates all lexical boundaries. This shows that at the level of frames-to-be-filled lexical boundaries have lost their significance. It is therefore not generally the case that phonological encoding consists of filling pre-existing *lexical* skeletons. Rather, new phonological word frames are constructed, dependent on the context of utterance. It is these newly constructed frames that have to be filled with segmental materials. Hence, there is no paradox.

There are four major questions to be answered by a theory of phonological encoding. They are:

Question 1: how are a word's segments made available and to what detail are they specified in the lexicon?
Question 2: how are phonological frames constructed?
Question 3: how are segments associated with slots in the frame?
Question 4: how is a filled frame translated into a phonetic or articulatory program?

Of these, Questions 1 and 3 have received abundant attention. Questions 2 and 4, however, have been largely disregarded.

Let us first consider Question 1, the spelling out of a word's phonological segments. Here, the most important addition on last decade's theoretical scene has been the connectionist modelling of segmental activation and selection (Dell, 1985, 1986; MacKay, 1987; Stemberger, 1985; and others). What these accounts have in common is the notion of the lexicon as a multilayer network of nodes, connected by arcs. The nodes can be in different states of activation, and they can spread their activation over (weighted) arcs to connected nodes. As far as phonological encoding is concerned, the relevant part of the network consists of a layer of lexical nodes at the "top" level, a layer of phonological feature nodes at the bottom level, and a number of layers mediating between these two. In these models, the phonological segments, or rather their features, are made available through activation spreading from the lexical nodes. More details can be found in Dell and O'Seaghdha (this issue).

But the theories differ substantially in detail. They differ in the kinds and numbers of mediating layers (morpheme, syllable, cluster, segment and other layers), in the directionality of activation spreading (one-way or two-way), in the presence or absence of inhibitory connections between same-level nodes, in the amount of extraneous (structurally determined) activation impinging on the network, in the amount of over- or underspecification of segmental information in the network, and in their degree of explicit computer modelling and quantification. In short, theorizing is very much in flux here, and it seems to me that an exclusive reliance on speech error data will not suffice to sort out the theoretical differences (see Meyer, this issue, for more details).

More in particular, there is an increasing need for reaction time studies of

phonological encoding. Connectionist models, if sufficiently explicit, may lend themselves well to experimental test by reaction time paradigms, as the work by Dell (1986, 1988), Meyer (1990, 1991), Schriefers, Meyer, and Levelt (1990), and Levelt et al. (1991a) has shown (see also below). In addition, as Butterworth (this issue) argues, the pathology of phonological encoding may tell us something about the underlying, stored representations and their spell-out mechanisms in lexical access. Segments may well be underspecified in our word form lexicon, as Stemberger (1983) already suggested. And this is in accordance with recent phonological theory (cf. Archangeli, 1988). How a complete phonetic specification arises from such underspecified segments is part of our Question 4, to which we will return below.

Turning now to Question 2, the connected speech perspective introduced above requires one to ask: how does a speaker generate the frame of a phonological word? Levelt (1989) assumes the existence of a prosody generator that takes as input the phrasal syntactic information and the metrical spell-out of words, and produces as output an organization of metrical units (in particular phonological words and phonological phrases).

For each incoming metrical pattern, the prosody generator will decide whether the pattern is to stay alone as a phonological word, or whether it is to be attached to the previous, or maybe a following head element. Here the syntactic information accessible to the prosody generator is crucial. For instance, pronouns can be attached to the main verb of the same clause, but nouns cannot. In *leave me alone*, *leave* and *me* can compose a phonological word, but in *leave Maureen alone*, *leave* and *Maureen* cannot (cf. Nespor & Vogel, 1986). It is also impossible to form phonological words across phonological phrase boundaries. For instance, *I wanna go* is possible, but the sentence *What I want, to be honest, is to go* cannot be uttered as **What I wanna be honest, is to go*. There is a phonological phrase boundary between *want* and *to* here.

For the construction of phonological words, the prosody generator must further have access to the metrical information that is stored with the words in the mental lexicon. A lexeme's metrical information is "spelled out" at an early stage in phonological encoding, according to Levelt (1989). So, for instance, a speaker in the "tip-of-the-tongue" state often knows a word's number of syllables and stress pattern without having access to most or all of the segmental information. Apart from being non-segmental (or "non-melodic" as the unhappy phonological term goes), the precise character of the metrical spell-out is as yet undecided.

Following Hayes (1989), the word's metrical spell-out could, among other things, contain its σ,μ pattern, that is, its syllable/mora structure. This amounts to saying that the metrical spell-out contains at least the following two pieces of information:

(1) the word's *number* of syllables (not the syllables themselves);

(2) the weight of the subsequent syllables (strong/weak – a two-morae syllable is strong, a one-mora syllable is weak).

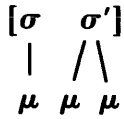
In addition, the metrical spell-out may contain:

(3) the word's foot structure or stress pattern (this is only necessary in languages where a word's stress pattern is not fully determined by the weights of its syllables).

The latter is in phenomenological agreement with the just mentioned tip-of-the-tongue state of metrical information.

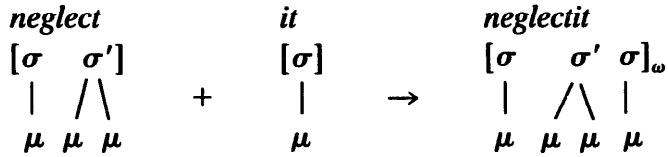
To represent the above three pieces of information, one minimally needs a string of σ's (syllable nodes) each specified for weight (i.e., number of morae), and one for the word's main stress.

Hence, something like



for the word *neglect*. The word has two syllables, the last of which is heavy and stressed. It is notationally convenient to have word boundary symbols ("[" and "]"). This makes it easier to discuss the formation of phonological words, to which we turn now.

If the word *neglect* in the utterance is part of the verb phrase *neglect it* (as in *I neglect it*), the prosody generator can construct the phonological word *neglectit*. The first step here is to concatenate the metrical frames of the two constituting lexical items, which can be diagrammed as follows:



Here the subscript ω indicates that *neglectit* is a phonological word frame.

The next step is to fill this frame with the spelled-out segmental information stemming from the two matrix lexical items (*neglect* and *it*) and to syllabify the resulting phonological word.

This brings us to Question 3: how are the segments (or "planning units") associated with positions in the phonological word frames? The reader is again referred to Meyer's review in the present issue. There is, in addition, Shattuck-Hufnagel's paper, which argues for the special status of the word-initial slot in this process of association. That slot is the most vulnerable position in the process of filler-to-slot association, testifying to the reality of word frames in phonological

encoding. Nothing in the latter paper is in disagreement with the notion introduced here that these word frames are in fact *phonological word* frames. As we will shortly see, the vulnerability of the word-initial segment may in part be due to the restrictions these phonological word frames impose on their filler segments. In his paper for the present issue, Butterworth discusses various disorders of assembling and syllabifying words in phonemic paraphasias.

Here I will only exemplify the filling process and the concurrent syllabification by returning to the phonological word *neglectit*, whose frame was presented above:

neglectit

$[\sigma \quad \sigma' \quad \sigma]_\omega$

$\mu \; \mu \; \mu \; \mu$

Experimental reaction time evidence (Meyer & Schriefers, 1991) shows that the process of filler-to-slot association is a rather strictly "left-to-right" one. What I propose now is that syllabification takes place "on the fly" as this left-to-right association of segments to metrical positions is taking place. For this to occur, the spelled-out segments from the matrix items are assumed to be ordered – or more precisely, to become available in the right order (speech errors show that this ordering can be occasionally disrupted). So, for the present phonological word, the planning units /n/, /i/, /g/, /l/, /ɛ/, /k/, /t/, /I/, /t/ (or their less specified equivalents) are spelled out in this order. They are then one by one associated to the frame, from "left to right" and following a set of association rules that are – in part – language specific. A general convention here is that attachment to σ, the syllable node, can only occur *on the left-hand side of a syllable*, that is, to the left of any unfilled morae of that syllable.[4]

Among the rules for English are:

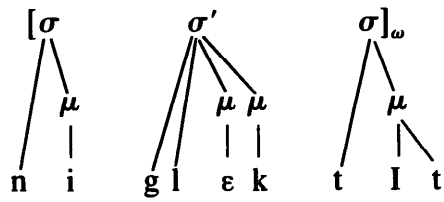
(1) A vowel only associates to μ.
(2) The default association of a consonant is to σ. A consonant associates to μ if and only if any of the following conditions holds:
    (a) the next element is lower in sonority;
    (b) there is no σ to associate to;
    (c) associating to σ would leave a μ without associated element.

Rule (1) says that a vowel is always involved in a syllable's weight. Or in

[4]On a traditional account, this means that a consonant attaches by default to the onset of a syllable. In Hayes' (1989) theory, which I follow here at least notationally, there are no onsets and rhymes. Still, essentially the same process of association can be notated in an onset/rhyme representation of the syllable. In that case syllable weight is not represented by the number of morae, but by the branchingness of the rhyme.

traditional rhyme terminology: a vowel is always part of a syllable's rhyme. The default association of rule (2) is what is traditionally known as "maximization of onset" (cf., for instance, Selkirk, 1984). Consonants between two vowels in a word (i.e., between two syllabic nuclei) are as much as possible associated with the second syllable (i.e., maximizing the second syllable's onset). Condition (a) of rule (2) is traditionally known as the sonority gradient. In a syllable's onset the sonority of segments increases (or at least does not decrease) towards the nucleus. After the nucleus, sonority decreases again (or at least doesn't increase). This makes, for example, /slorp/ a possible syllable but /lsopr/ an impossible one. Checking for condition (a) requires a one-element look-ahead in the association process. Condition (b) takes care of "left-over" consonants at the end of a word; they have no new syllable to go to, and are added to the last μ (traditionally: to the rhyme of the final syllable). Condition (c) takes care that, where necessary, a consonant will carry the weight of a syllable. A mora should not stay unfilled because that would change a syllable's weight.

These rules suffice to exemplify the association process and syllabification of the phonological word *neglectit*. The first spelled-out element is /n/. Since neither of the conditions (a) through (c) of (2) hold, /n/ is associated to the first σ on the left in the frame. The next element is /i/. It is a vowel, and must be associated to μ according to rule (1). The next element /g/ is a consonant. It will, by default, be attached to σ (condition (2a), for instance, doesn't apply, because /l/ is higher in sonority than /g/). Since right attachment to the first σ is excluded by general convention, /g/ attaches to the next σ. The next consonant, /l/, also attaches to (the same) σ, following the default of rule (2). The vowel /ɛ/ then attaches to μ according to rule (1). The next element, /k/, must associate to μ for two reasons. First, attaching it to the next σ would leave the second μ of the current syllable without associate (condition (2c)). Second, the following element (/t/) is lower in sonority (condition (2a)). The /t/, however, will by default be attached to the following σ, thus creating a syllable boundary between /k/ and /t/. (The fact that /t/ is syllable initial and preceding a vowel has as an important phonetic consequence that it will be aspirated.) The vowel /I/ attaches to the final μ, and the last element /t/ attaches to the same μ, following condition (b) of rule (2). The end result (with syllables indicated) is:



Using the same set of rules, the syllabification of the phonological word *regret it* will be re-gret-it. The first /t/ is not syllable initial here, because that would leave

one μ of the second syllable without dependent, violating condition (2c). As a consequence, this /t/, not being syllable initial, will not be aspirated.

This picture of the association process differs from the standard accounts (reviewed in Meyer's paper) in that the one-slot-one-segment idea is given up. The frame doesn't completely specify the number of slots to be filled. For instance, the number of segments to be associated with σ is not specified beforehand. It depends on the ordered string of spelled-out segments and on the fulfilment of conditions in rule (2) whether a syllable's onset will be a single segment or a cluster.

Another difference is that we are dealing now with *phonological* words, not *lexical* words. On the traditional account, the error *peel like flaying* (for *feel like playing*) involves exchanging consonants across an intermediate word. But the speaker who produced this error was probably planning just two phonological words here, $[feelike]_\omega$ and $[playing]_\omega$. On the phonological word account, Shattuck-Hufnagel's finding (this issue) that word onsets are more vulnerable to error than other syllable onsets now predicts that the mentioned error is a more likely one than, for instance, *leel fike playing*. The /l/ in *like* is internal to the phonological word; it is *not* word initial and hence less vulnerable.

The sketched account, moreover, provides us with one possible reason for Shattuck-Hufnagel's finding (1987, this issue) that word-final consonants are less vulnerable to error in phrases than in lists. For instance, the word-final consonants in a tongue-twister like *parade fad foot parole* are relatively more vulnerable than in a tongue-twister like *the parade is a fad and the foot has parole*. This may be due to the formation of phonological words in the phrasal context. In the example, $[paradis]_\omega$ and $[footas]_\omega$ are potential phonological words. When they are, the critical consonants are no longer word final, as they are in the list. However, under the phrasal condition nothing changes for the word-internal consonants, such as /r/ in *parade*. They keep being word internal. And indeed, as Shattuck-Hufnagel (this issue) shows, the relative vulnerability of these word-internal consonants is not affected by phrasal context.

The present account of the association process in connected speech is a highly preliminary one. It needs further elaboration, but is presented here to highlight the notion that syllabification is a late process in phonological encoding, that it is a consequence of the left-to-right association of segments to a metrical frame, and that the domain of syllabification is the phonological word.

Let us now turn to the fourth issue formulated above: how is a filled frame translated into a phonetic or articulatory program? The slot-and-filler theory has nothing to say about this problem. It is a theory about how phonemic segments find their ultimate positions in a metrical frame. It does not specify the allophonic phonetic shapes of segments within the word or syllable. Similarly, none of the connectionist theories address this issue. On first view, they do seem to take a step in the right direction by adding a level of feature nodes at the bottom of the

network. But these feature nodes represent discrete *phonological* features, not scaled phonetic ones. In particular, there is no mechanism for making feature values dependent on a segment's context in the syllable or word.

Here, I only want to signal this hiatus in theory construction. In Levelt (1989) I adopted an important suggestion by Crompton (1982), which in my view indicates one way in which this hiatus can be filled. Crompton proposed that the syllables composed in phonological encoding function as addresses for stored phonetic syllable templates. One can conceive of these syllable templates as motor instructions for complex articulatory gestures. Following Browman and Goldstein (1990) one would call these "gestural scores", that is, specifications of articulatory tasks to be performed in pronouncing the syllables. I added that these syllable scores still have a few free parameters to be fixed, such as stress, rate and pitch parameters. Still, they are genuine phonetic, not phonological representations. The idea that we have a "phonetic syllabary" is certainly not obvious. It is quite an attractive idea for a language such as Chinese, which has no more than about 400 different syllables. But what about English or Dutch, which have somewhere between 6500 and 7000 different syllables? Would they all be stored in the speaker's head as phonetic templates or scores? One straightforward prediction from the theory is a frequency effect. It should take more time to retrieve a low-frequency syllable template than a high-frequency template. In our laboratory, Linda Wheeldon and I (in preparation) could confirm that prediction for Dutch. Naming latencies (not reading latencies) were slower for words consisting of low-frequent syllables than for words consisting of high-frequent syllables. This effect was completely independent of the word frequency effect, but is (as could be expected) related to the phonetic complexity of the syllables.

But even if the notion of an independent phonetic syllabary receives further experimental support, it cannot be the whole story. The syllable is not the only context of phonetic variation; there are cross-syllable and cross-word phonetic effects that are still to be explained.

The interaction between lexical selection and phonological encoding

Lexical selection and phonological encoding are wildly different processes. Lexical selection is semantically (or syntactically) driven search for an appropriate item in a huge lexical store. Phonological encoding is the creation of an executable phonetic program for a single item in context. On first view, it would not seem like a great feat of psychological engineering if these two kinds of process were to interact with one another. It would add unnecessary error proneness to both aspects of the accessing system. Still, in an important paper, Dell and Reich (1981) presented statistical evidence from speech errors showing that errors of selection and errors of phonological encoding were not entirely

independent. Moreover, there is a lexical bias effect (already reported by Baars, Motley, & MacKay, 1975), which means that trouble in phonological encoding tends to create real words more often than should be expected by chance. These findings have been replicated by Stemberger (1983), Harley (1984), Dell (1986), Martin, Weisberg and Saffran (1989) and form a challenge to the above modular view of the accessing process.

The observed interactions between lexical selection and phonological encoding found a natural theoretical explanation in connectionist models of lexical access. In particular, models that allowed for both forward and backward spreading of activation between levels of representation (Stemberger, 1985; Dell, 1986) could account for the statistical speech error evidence.

But Levelt et al. (1991a) argued that models of that kind also have specific implications for the time course of lexical access – implications that can be tested by appropriate reaction time experiments. Specifically, all connectionist models would predict that coactivated semantic alternatives to the target item (e.g., *goat* when the target item is *sheep*) would, at some moment during lexical access, undergo some degree of phonological activation. In addition, the models that allow for backward spreading of activation should predict that there is a late rebound of semantic activation following phonological activation. However, neither of these predictions could be experimentally substantiated.

The authors then argued for two directions in further research. The first one would be a careful exercise in the parametrization of connectionist models. Or, would it be possible to have *just* enough interactiveness between levels in the network to account for the above-mentioned statistical effects, but still so little that no measurable phonological coactivation of semantic alternatives and no measurable semantic rebound would occur? In his contribution to the present issue, Dell argues that appropriate parameters can indeed be found. For a full appreciation of this claim, the reader is referred to an ongoing discussion: Dell and O'Seaghdha (1991) and Levelt et al. (1991b). One issue raised there is what functional sense feedback could have in a lexical production network (it surely cannot be merely to cause a specific type of speech error). Dell (1988) suggested that a deep reason for feedback could be that the same lexical network is also used for lexical access in comprehension, which obviously would involve activation spreading in the reverse direction. For this double use of the same network we pay by occasionally making specific kinds of speech errors. This is a challenging suggestion: are the accessing mechanisms of production and comprehension making use of the same unified lexical network, or are there independent input and output networks? The latter view would find support if a double dissociation could be found in the pathology of lexical access. Allport (1984) explicitly raised this issue, but could not find convincing evidence of this kind. On the other hand, Howard and Franklin's (1988) patient MK seems to provide one half of this double dissociation.

A second direction, proposed by Garrett (personal communication), is to look more carefully into the situations in which mixed errors arise. They may be "environmentals", that is, intrusions of words that happen to be in the speaker's span of attention. In the Martin et al. (1989) study, for instance, the response set in the experiments contained a highly apparent subset of items that were not only semantically related to one another, but also phonologically. No surprise that the subjects tended to make confusions among these items. This probably has nothing to do with the fact that these items were *phonologically* related; any other marked relation among the items would have produced the same result. If Garrett is right, the occurrence of mixed errors may, after all, not exceed chance level if "environmentals" are excluded.

The third direction would be to give renewed attention to the possibility that the observed interactions are due to *post-lexical* processes. The speaker monitors his lexical productions to some extent (see also Butterworth, this issue), and can intercept imminent errors before they are overtly produced. This might account for some of the above statistical findings on speech errors. So, for instance, if a phonological error creates a non-word, it is more likely to be intercepted than when it creates a word; a word is well formed, a non-word is not. This "editor" theory was originally proposed by Baars et al. (1975), and elaborated in several subsequent publications. See Levelt (1989, chapter 12) for a review, and for a comparison between editor theories and connectionist accounts of the phenomena under discussion. My conclusion there was that it will not be easy to distinguish these two approaches experimentally. And indeed, the game is still as open as it was at the time. Only new, sophisticated research on the speaker's self-monitoring can end this deadlock.

Let me, finally, add one more issue to the already disquieting list of problems reviewed in this introductory paper. What we have just discussed with respect to phonological encoding and lexical selection, namely whether there is a real feedback there, can also be considered at the next higher level. Is our selection of concepts-to-be-expressed to any extent dependent on lexical selection? There is not only the general Whorfian problem looming on the horizon here (cf. Schlesinger, 1990, for an excellent review); but there is also a more specific processing issue involved. The aphorism from Kleist (1809) heading this paper suggests that there can be spontaneous activity in a speaker's formulator, generating words or phrases that present themselves to the speaker as potential issues to talk about.[5] One important question then is, What is the "routing" of that feedback? Is an active lemma directly feeding back to the conceptual level? The present evidence for such a direct link is minimal (cf., Levelt, 1989, pp. 275 ff.). Or is it feeding back via internal speech, i.e., does it involve (internal)

[5]Daniel Dennett (personal communication) alerted me to the idea of spontaneous activity in the formulator feeding back to the conceptual level (see also Dennett, 1991).

phonological encoding of the activated word? The latter view has some phenomenological face value. But as I said earlier, phenomenology is not of much help in dissecting the process of lexical access.

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: Springer.

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology, 5*, 183–207.

Baars, B.J., Motley, M.T., & MacKay, D. (1975). Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14*, 382–391.

Beattie, G., & Butterworth, B. (1979). Contextual probability and word frequency as determinants of pauses in spontaneous speech. *Language and Speech, 22*, 201–211.

Bock, J.K., & Loebell, H. (1990). Framing sentences. *Cognition, 35*, 1–40.

Browman, C.P., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. Haskins Laboratories Status Report on Speech Research. SR-103/104, 95–110.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325–337.

Bub, D., & Caplan, D. (in press). *Aphasiology for speech–language pathologists.*

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

Caramazza, A., & Hillis (1990). Where do semantic errors come from? *Cortex, 26*, 95–122.

Clark, E.V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–33). Hillsdale, NJ: Erlbaum.

Clark, E. (1990). On the pragmatics of contrast. *Journal of Child Language, 17*, 417–431.

Clark, H., & Wilkes-Gibbs, D.L. (1986). Referring as a collaborative process. *Cognition, 22*, 1–39.

Cohen, A. (1965). Versprekingen als verklappers bij het proces van spreken en verstaan. *Forum der Letteren, 6*, 175–186.

Collins, A.M., & Loftus, E.F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407–428.

Collins, A.M., & Quillian, M.R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 240–247.

Crompton, A. (1982). Syllables and segments in speech production. In A. Cutler (Ed.), *Slips of the tongue and language production.* Berlin: Mouton.

Cutler, A. (1982). *Speech errors: A classified bibliography.* Bloomington: Indiana Linguistics Club.

Deese, J. (1984). *Thought into speech: The psychology of a language.* Englewood Cliffs, NJ: Prentice-Hall.

Dell, G.S. (1985). Positive feedback in hierarchical connectionist models: Applications to language production. *Cognitive Science, 9*, 3–23.

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., & O'Seaghdha (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review, 98*, 604–614.

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.

Dennett, D.C. (1991). *Consciousness explained.* Boston: Little, Brown & Co.

Dik, S.C. (1987). Linguistically motivated knowledge representation. In M. Nagao (Ed.), *Language and artificial intelligence* (pp. 145–170). Amsterdam: Elsevier Science Press.

Fodor, J.A., Garrett, M.F., Walker, E.C.T., & Parkes, C.H. (1980). Against definitions. *Cognition, 8*, 263–367.

Fromkin, V.A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 27–52. (Reprinted in: Fromkin, V.A. (Ed.) (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.)

Fromkin, V.A. (Ed.) (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.

Garnham, A., Shillcock, R.S., Brown, G.D.A., Mill, A.I.D., & Cutler, A. (1982). Slips of the tongue in the London–Lund corpus of spontaneous conversations. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 251–263). Berlin: Mouton.

Garrett, M.F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of learning and motivation* (Vol. 9, pp. 133–175). New York: Academic Press.

Glaser, W.R., & Düngelhoff, F.J. (1984). The time course of picture–word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 640–654.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.

Harley, T.A. (1984). A critique of top-down independent levels of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science, 8*, 191–219.

Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry, 20*, 253–306.

Howard, D., & Franklin, S. (1988). *Missing the meaning? A cognitive neuropsychological study of the processing of words by an aphasic patient*. Cambridge, MA: MIT Press.

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.

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

Klapp, S.T. (1974). Syllable-dependent pronunciation latencies in number naming, a replication. *Journal of Experimental Psychology, 102*, 1138–1140.

Kleist, H. von (1809). Über die allmähliche Verfertigung der Gedanken beim Reden. In H. von Kleist, *Sämtliche Werke und Briefe* (pp. 320–327). München: Carl Hanser Verlag.

Lahiri, A., Jongman, A., & Sereno, J. (1990). The pronominal clitic [dər] in Dutch: A theoretical and experimental approach. *Yearbook of Phonology, 3*, 1–13.

La Heij, W. (1988). Components of Stroop-like interference in picture naming. *Memory and Cognition, 16*, 400–410.

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

Levelt, W.J.M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology, 14*, 78–106.

Levelt, W.J.M., & Maassen, B. (1981). Lexical search and order of mention in sentence production. In W. Klein & W.J.M. Levelt (Eds.), *Crossing the boundaries in linguistics: Studies presented to Manfred Bierwisch* (pp. 221–252). Dordrecht: Reidel.

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

Levelt, W.J.M., Schriefers, H., Vorberg, D., Meyer, A.S., Pechmann, T., & Havinga, J. (1991b) Normal and deviant lexical processing: A reply to Dell and O'Seaghdha. *Psychological Review, 98*, 615–618.

Lupker, S.J. (1979). The semantic nature of competition in the picture–word interference task. *Canadian Journal of Psychology, 36*, 485–495.

MacKay, D. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8*, 323–350. (Reprinted in Fromkin, V.A. (Ed.) (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.)

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

Maclay, H., & Osgood, C.E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*, 19–44.

Martin, N., Weisberg, R.W., & Saffran, E.M. (1989). Variables influencing the occurrence of naming errors: Implications for a model of lexical retrieval. *Journal of Memory and Language, 28*, 462–485.

Meringer, R., & Mayer, K. (1895). *Versprechen und Verlesen*. Stuttgart: Goschensche Verlag. (Re-issued, with introductory essay by A. Cutler and D.A. Fay (1978). Amsterdam: John Benjamins.)

Meyer, A.S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language, 29*, 524–545.

Meyer, A.S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30*, 69–89.

Meyer, A.S., & Schriefers, H. (1991). Phonological facilitation in picture–word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*.

Miller, G.A., & Johnson-Laird, P.N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review, 76*, 165–178.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Nooteboom, S. (1967). Some regularities in phonemic speech errors. *Annual Progress Report. Institute for Perception Research IPO, 2*, 65–70.

Oldfield, R.C. (1963). Individual vocabulary and semantic currency: A preliminary study. *British Journal of Social and Clinical Psychology, 2*, 122–130.

Oldfield, R.C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 17*, 273–281.

Rosch, E., Mervis, C.B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Schenkein, J. (1980). A taxonomy for repeating action sequences in natural conversation. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk* (pp. 21–47). London: Academic Press.

Schlesinger, I.M. (1990). *The wax and wane of Whorfian views*. Working paper #32. The Goldie Rotman Center for Cognitive Science in Education. Jerusalem: Hebrew University.

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.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Seymour, P.H.K. (1979). *Human visual cognition*. New York: St Martin's Press.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial order mechanism in sentence production. In W.E. Cooper & E.C.T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Hillsdale, NJ: Erlbaum.

Shattuck-Hufnagel, S. (1987). The role of word onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processing in language* (pp. 17–51). Hillsdale, NJ: Erlbaum.

Stemberger, J.P. (1983). *Speech errors and theoretical phonology: A review*. Bloomington: Indiana Linguistics Club.

Stemberger, J.P. (1985). An interactive action model of language production. In A.W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143–186). Hillsdale, NJ: Erlbaum.

Ward, N. (1988). *Issues in word choice*. Budapest: COLING-88.

Ward, N. (1991). *A flexible, parallel model of natural language generation*. Report No. UCB/CSD 91/629. Computer Science Division (EECS), UC Berkeley.