Models of word production

Willem J.M. Levelt

Research on spoken word production has been approached from two angles. In one research tradition, the analysis of spontaneous or induced speech errors led to models that can account for speech error distributions. In another tradition, the measurement of picture naming latencies led to chronometric models accounting for distributions of reaction times in word production. Both kinds of models are, however, dealing with the same underlying processes: (1) the speaker's selection of a word that is semantically and syntactically appropriate; (2) the retrieval of the word's phonological properties; (3) the rapid syllabification of the word in context; and (4) the preparation of the corresponding articulatory gestures. Models of both traditions explain these processes in terms of activation spreading through a localist, symbolic network. By and large, they share the main levels of representation: conceptual/semantic, syntactic, phonological and phonetic. They differ in various details, such as the amount of cascading and feedback in the network. These research traditions have begun to merge in recent years, leading to highly constructive experimentation. Currently, they are like two similar knives honing each other. A single pair of scissors is in the making.

How do we generate spoken words? This issue is a fascinating one. In normal fluent conversation we produce two to three words per second, which amounts to about four syllables and ten or twelve phonemes per second. These words are continuously selected from a huge repository, the mental lexicon, which contains at least 50–100 thousand words in a normal, literate adult person[1]. Even so, the high speed and complexity of word production does not seem to make it particularly error-prone. We err, on average, no more than once or twice in 1000 words[2]. This robustness no doubt has a biological basis; we are born talkers. But in addition, there is virtually no other skill we exercise as much as word production. In no more than 40 minutes of talking a day, we will have produced some 50 million word tokens by the time we reach adulthood.

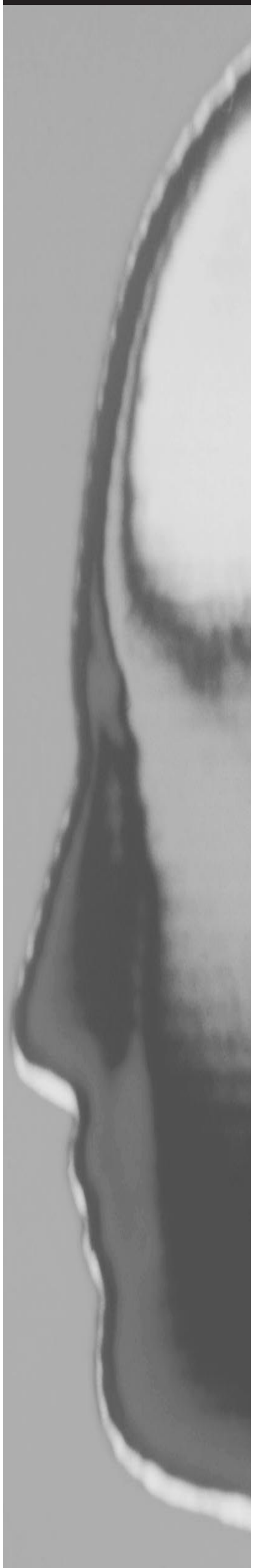
The systematic study of word production began in the late 1960s, when psycholinguists started collecting and analyzing corpora of spontaneous speech errors (see Box 1). The first theoretical models were designed to account for the patterns of verbal slips observed in these corpora. In a parallel but initially independent development, psycholinguists adopted an already existing chronometric approach to word production (Box 1). Their first models were designed to account for the distribution of picture naming latencies obtained under various experimental conditions.

Although these two approaches are happily merging in current theorizing, all existing models have a dominant kinship: their ancestry is either in speech error analysis or it is in chronometry. In spite of this dual perspective, there is a general agreement on the processes to be modeled. Producing words is a core part of producing utterances; explaining word production is part of explaining utterance production[3,4]. In producing an utterance, we go from some communicative intention to a decision about what information to express – the 'message'. The message contains one or more concepts for which we have words in our lexicon, and these words have to be retrieved. They have syntactic properties, such as being a noun or a transitive verb, which we use in planning the sentence, that is in 'grammatical encoding'. These syntactic properties taken together, we call the word's 'lemma'. Words also have morphological and phonological properties that we use in preparing their syllabification and prosody, that is in 'phonological encoding'. Ultimately, we must prepare the articulatory gestures for each of these syllables, words and phrases in the utterance. The execution of these gestures is the only overt part of the entire process.

This review will first introduce the two kinds of word production model. It will then turn to the computational steps in producing a word: conceptual preparation, lexical selection, phonological encoding, phonetic encoding and articulation. This review does not cover models of word reading.

Two kinds of model

All current models of word production are network models of some kind. In addition, they are, with one exception[5], all 'localist', non-distributed models. That means that their



W.J.M. Levelt is at the Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands.

fax: +31 24 352 1213
e-mail: pim@mpi.nl

1364-6613/99/$ – see front matter © 1999 Elsevier Science. All rights reserved. PII: S1364-6613(99)01319-4

Box 1. Historical roots of word production research

The study of word production has two historical roots, one in speech error analysis and one in chronometric studies of naming.

The speech error tradition

In 1895, Meringer and Mayer published a substantial corpus of German speech errors that they had diligently collected (Ref. a). The corpus, along with the theoretical analyses they provided, established the speech error research tradition. One important distinction they made was between meaning-based substitutions [such as *Ihre* ('your') for *meine* ('my')] and form-based substitutions [such as *Studien* ('studies') for *Stunden* ('hours')], acknowledging that there is often a phonological connection in meaning-based errors (i.e. the over-representation of mixed errors was observed over a century ago). Freud was quick to confuse the now generally accepted distinction between meaning- and form-based errors by claiming that innocent form errors are practically all meaning-driven [why does a patient say of her parents that they have *Geiz* ('greed') instead of *Geist* ('cleverness')? Because she had suppressed her real opinion about her parents – oh, all the errors we would make!]. A second, now classical distinction that Meringer and Mayer introduced was between exchanges (***m**ell **w**ade* for *well made*), anticipations (***t**addle tennis* for *paddle tennis*), perseverations (*been a**b**ay* for *been away*) and blends or contaminations (***evoid***, blending *avoid* and *evade*).

Many linguists and psychologists have continued this tradition (Ref. b), but an ebullient renaissance (probably triggered by the work of Cohen; Ref. c) began in the late 1960s. In 1973, Fromkin edited an influential volume of speech error studies, with part of her own collection of errors as an appendix (Ref. d). Another substantial corpus was built up during the 1970s, the MIT–CU corpus. It led to two of the most influential models of speech production: (1) Garrett discovered that word exchanges (such as *he **left** it and **forgot** it behind*) can span some distance and mostly preserve grammatical category as well as grammatical function within their clauses (Ref. e). Sound/form exchanges (such as ***r**ack **p**at* for *pack rat*), on the other hand, ignore grammatical category and preferably happen between close-by words. This indicates the existence of two modular levels of processing in sentence production, a level where syntactic functions are assigned and a level where the ordering of forms (morphemes, phonemes) is organized; (2) Shattuck-Hufnagel's scan-copier model concerns phonological encoding (Ref. f). A core notion here is the existence of phonological frames, in particular syllable frames. Sound errors tend to preserve syllable position (as is the case in ***r**ack **p**at*, or in ***p**ope **sm**iker* for *pipe smoker*). The model claims that a word's phonemes are retrieved from the lexicon with their syllable position specified. They can only land in the corresponding slot of a syllable frame.

In 1976, Baars, Motley and MacKay (Ref. g) developed a method for eliciting speech errors under experimentally controlled conditions, ten years after Brown and McNeill had created one for eliciting tip-of-the-tongue states (Ref. h). Several more English-language corpora, in particular Stemberger's (Ref. i), were subsequently built up and analyzed, but sooner or later substantial collections of speech errors in other languages became available, such as Cohen and Nooteboom's for Dutch (Ref. c), Berg's (Ref. j) for German, Garcia-Albea's for Spanish (Ref. k) and Rossi and Peter-Defare's for French (Ref. l).

A final major theoretical tool in this research tradition was supplied by Dell (Ref. m), who published the first computational model of word production, designed to account for the observed statistical distributions of speech error types.

The chronometric tradition

In 1885, Cattell (Ref. n) discovered that naming a list of 100 line drawings of objects took about twice as long as naming a list of the corresponding printed object names. This started a research tradition of measuring naming latencies, naming objects and naming words. Initially, most attention went to explaining the difference between object and word naming latencies. It could not be attributed to practice. It could also not be attributed to a visual differences between line drawings and words. Fraisse showed that when a small circle was named as 'circle' it took, on average, 619 ms, but when named as 'oh' it took 453 ms (Ref. o). Clearly, the task induced different codes to be accessed. They are not graphemic codes, because Potter *et al.* obtained the same picture-word difference in Chinese (Ref. p). The dominant current view is that there is a direct access route from the word to its phonological code, whereas the line drawing first activates the object concept, which in turn causes the activation of the phonological code – an extra step. Another classical discovery in the picture-naming tradition (by Oldfield and Wingfield; Ref. q) is the word frequency effect (see main article).

In 1935, Stroop introduced a new research paradigm, now called the 'Stroop task' (Ref. r). The stimuli are differently colored words. The subject's task is either to name the color or to say the word. Stroop studied what happened if the word was a color name itself. The main finding was this: color naming is substantially slowed down when the colored word is a different color name. It is, for instance, difficult to name the word *green* when it is written in red. But naming the word was not affected by the word's color.

Rosinski *et al.*, interested in the automatic word reading skills of children, transformed the Stroop task into a picture/word interference task (Ref. s). The children named a list of object drawings. The drawings contained a printed word that was to be ignored. Alternatively, the children had to name the printed words, ignoring the objects. Object naming suffered much more from a semantically related interfering word than word naming suffered from a meaning-related interfering object, confirming the pattern typically obtained in the Stroop task. Lupker set out to study the nature of the semantic interference effect in picture/word interference (Ref. t). He replaced the traditional 'list' procedure by a single trial voice-key latency measurement procedure – which is the standard now. Among many other things, Lupker and his coworkers discovered that it is semantic, not associative relations between distracter word and picture name that do the work. The interference is strongest when the distracter word is a possible response to the picture, in particular when it is in the experiment's response set. Also, Lupker was the first to use printed distracter words that are orthographically (not semantically) related to the picture's name (Ref. u). When the distracter had a rhyming relation to the target name, picture/word interference was substantially reduced. This also holds for an alliterative relation between distracter and target. In other words, there is phonological facilitation as opposed to semantic inhibition. Glaser and Düngelhoff were the first to study the time course of the semantic interaction effects obtained in picture/word tasks (Ref. v). They varied the stimulus-onset asynchronies (SOAs) between distracter and picture. They obtained characteristic SOA curves that were different for picture naming, picture categorization and word naming. These results were taken up by Roelofs in his WEAVER modeling of lemma access (see main text). A final noteworthy experimental innovation was the paradigm developed by Schriefers *et al.* (Ref. w). Here, the distracter was a *spoken* word, aurally presented nodes represent whole linguistic units, such as semantic features, syllables or phonological segments. Hence, they are all 'symbolic' models. Of the many models with ancestry in the speech error tradition[6–8] only a few have been computer-implemented[9–11]. Among them, Dell's two-step interactive activation model[9] has become by far the most influential. Figure 1 represents a fragment of the proposed lexical network.

The network is called 'two-step', because there are two steps from the semantic to the phonological level. Semantic feature nodes spread their activation to the corresponding word or lemma nodes, which in turn spread their activation to phoneme nodes. Activation 'cascades' from level to level over all available connections in the network. The type of model is called 'interactive', because all connections are bi-directional; activation spreads both ways. Interactiveness is a property shared by all models in this class. One of the original motivations for implementing this feature is the statistical over-representation of so-called mixed errors in speech error corpora. They are errors that are both semantic and phonological in character. If, for example, your target word is *cat* but you accidentally produce *rat*, you have made a mixed error. The network in Fig. 1 can produce that error in the following way. The lemma node *cat* is strongly activated by its characteristic feature set. In turn, it spreads its activation to its phoneme nodes /k/, /æ/ and /t/. A few of the semantic features of *cat* (such as 'animate' and 'mammalian') co-activate the lemma node of *rat.* But the same lemma node *rat* is further activated by feedback from the now active phonemes /æ/ and /t/. This confluence of activation gives *rat* a better chance to emerge as an error than either the just semantically related *dog* or the just phonologically related *mat*. Interactiveness also gives a natural account of the tendency for speech errors to be real words (for example *mat* rather than *gat*). Still, bi-directionality needs independent motivation (its functionality can hardly be to induce speech errors). One recurring suggestion in this class of models is that the network serves in both word production and word perception[6]. That would, of course, require bi-directionality of the connectivity. However, Dell *et al.*[12] argue against this solution because many aphasic patients show both good auditory word recognition and disturbed phonological encoding. The functionality of bi-directional connections (and hence interactivity) would rather be to support fluency in lemma selection. Some word forms, in particular the ones that are infrequently used, are less accessible than others. It will be advantageous to select a lemma whose phonological form will be easy to find. Feedback from the word form level will provide that functionality (and might explain a recent chronometric result[13]). Still, one should consider the possibility that interactiveness is merely a property of the error mechanism: an error might occur precisely then when undue interactivity arises in an otherwise discrete system.

Most implemented computational models in the chronometric tradition extend no further than accessing the word's whole name from a semantic or conceptual base[14–16]. There is no activation of phonological segments, no phonological encoding. Only Roelofs's WEAVER model[17,18] has a fully developed phonological component. A fragment of the WEAVER lexical network is shown in Fig. 2.

to the subject at different SOAs with respect to picture onset. The distracter words were either semantically or phonologically related to the target word, or unrelated. This paradigm and its many later variants made it possible to study the relative time course of the target name's semantic and phonological encoding in much detail.

References

a Meringer, R. and Mayer, K. (1895) *Versprechen und Verlesen*, Goschenscher-Verlag (Reprinted 1978, with introductory essay by A. Cutler and D.A. Fay, Benjamins)

b Cutler, A. (1982) *Speech Errors: A Classified Bibliography*, Indiana Linguistics Club

c Cohen, A. (1966) Errors of speech and their implications for understanding the strategy of language users *Zeitschrift für Phonetik* 21, 177–181

d Fromkin V.A. (1973) *Speech Errors as Linguistic Evidence,* Mouton

e Garrett, M. (1975) The analysis of sentence production, in *Psychology of Learning and Motivation* (Bower, G., ed.), pp. 133–177, Academic Press

f Shattuck-Hufnagel, S. (1979) Speech errors as evidence for a serial ordering mechanism in sentence production, in *Sentence Processing: Psycholinguistic Studies Dedicated to Merrill Garrett* (Cooper, W.E. and Walker, E.C.T., eds), pp. 295–342, Erlbaum

g Baars, B.J., Motley, M.T. and MacKay, D. (1975) Output editing for lexical status from artificially elicited slips of the tongue *J. Verb. Learn. Verb. Behav.* 14, 382–391

h Brown, R. and McNeill, D. (1966) The 'tip of the tongue' phenomenon. *J. Verb. Learn. Verb. Behav.* 5, 325–337

i Stemberger, J.P. (1985) An interactive activation model of language production, in *Progress in the Psychology of Language* (Vol. 1) (Ellis, A.W., ed.), pp. 143–186, Erlbaum

j Berg, T. (1998) *Linguistic Structure and Change*, Clarendon Press

k García-Albea, J.E., del Viso, S. and Igoa, J.M. (1989) Movement errors and levels of processing in sentence production *J. Psycholinguist. Res.* 18, 145–161

l Rossi, M. and Peter-Defare, É. (1998) *Les Lapsus: Ou Comment Notre Fourche a Langué,* Presse Universitaire France

m Dell, G.S. (1986) A spreading-activation theory of retrieval in sentence production *Psychol. Rev.* 93, 283–321

n Cattell, J.M. (1885) Über die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben *Philosophische Studien* 2, 635–650

o Fraisse, P. (1967) Latency of different verbal responses to the same stimulus *Q. J. Exp. Psychol.* 19, 353–355

p Potter, M.C. *et al.* (1984) Lexical and conceptual representation in beginning and proficient bilinguals *J. Verb. Learn. Verb. Behav.* 23, 23–38

q Oldfield R.C. and Wingfield, A. (1965) Response latencies in naming objects *Q. J. Exp. Psychol.* 17, 273–281

r Stroop, J.R. (1935) Studies of interference in serial verbal interactions *J. Exp. Psychol.* 18, 643–662

s Rosinski, R.R., Michnick-Golinkoff, R. and Kukish, K.S. (1975) Automatic semantic processing in a picture–word interference task *Child Dev.* 46, 247–253

t Lupker, S.J. (1979) The semantic nature of response competition in the picture–word interference task *Mem. Cognit.* 7, 485–495

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

v Glaser, M.O. and Düngelhoff, F-J. (1984) The time course of picture–word interference *J. Exp. Psychol. Hum. Percept. Perform.* 7, 1247–1257

w Schriefers, H., Meyer, A.S. and Levelt, W.J.M. (1990) Exploring the time course of lexical access in production: picture–word interference studies *J. Mem. Lang.* 29, 86–102

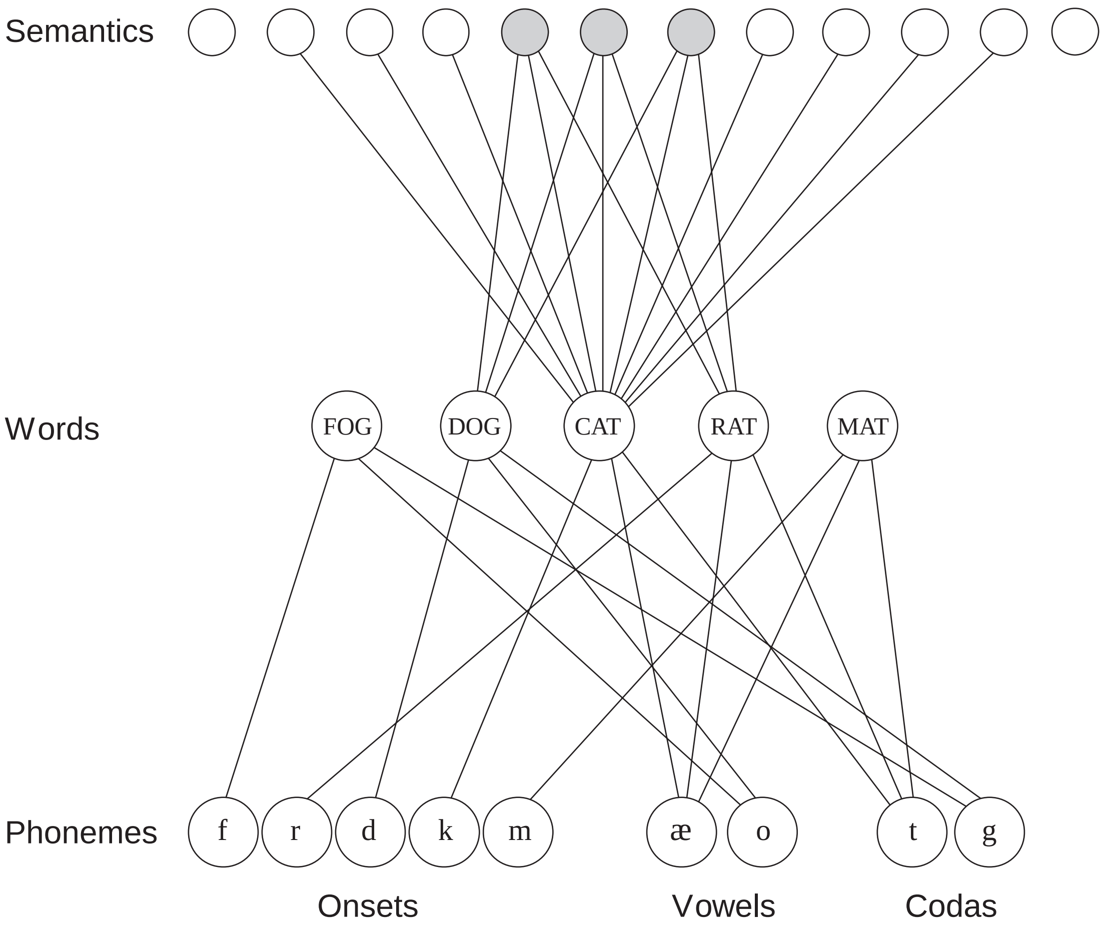


Fig. 1. Fragment of Dell's interactive lexical network. The nodes in the upper layer represent semantic features. The nodes in the middle layer represent words or lemmas. The nodes in the bottom layer represent onset, nucleus and coda phonemes (in particular consonants and vowels). All connections are bi-directional and there are only facilitatory, no inhibitory, connections. Activation spreads throughout the network without constraints; there is full cascading. It is always the most highly activated word or lemma node that gets selected. The moment of selection is determined externally, by the developing syntactic frame of the utterance. Upon selection the node receives an extra jolt of activation, which triggers its phonological encoding. The computational model has many more features than represented in the present figure. There is a further layer representing phonological features (such as 'voiced' or 'nasal') and there are versions of the model with a layer of syllable nodes. (Adapted from Dell *et al.*[12])

The main strata in this network are the same as those in the interactive model. There is a conceptual/semantic level of nodes, a lemma stratum and a phonological or form stratum. But the model is only partially interactive. There are good reasons for assuming that conceptual and lemma strata are shared between production and perception[18], hence their interconnections are modelled as bi-directional. But the form stratum is unique to word production; it does not feed back to the lemma stratum. Therefore it is often called the *discrete* (as opposed to 'interactive') two-step model. Although the model was designed to account for response latencies, not for speech errors, the issue of 'mixed' speech errors cannot be ignored and it has not been. The explanation is largely[18] post-lexical. We can strategically monitor our internal phonological output and intercept potential errors. A phonological error that happens to create a word of the right semantic domain (such as *rat* for *cat*) will have a better chance of 'slipping through' the monitor than one that is semantically totally out of place (such as *mat* for *rat*). Similarly, an error that produces a real word will get through easier than one that produces a non-word. There is experimental evidence that the monitor is indeed under strategic control[19]. Still, the causation of mixed errors continues to be a controversial issue among models of word production.

Conceptual preparation

The first step in accessing content words such as *cat* or *select* is the activation of a lexical concept, a concept for which you have a word or morpheme in your lexicon. Usually, such a concept is part of a larger message, but even in the simple case of naming a single object it is not trivial which lexical concept you should activate to refer to that object. It will depend on the discourse context whether it will be more effective for you to refer to a cat as *cat*, *animal, siamese* or anything else. Rosch[20] has shown that we prefer 'basic level' terms to refer to objects (*cat* rather than *animal*; *dog* rather than *collie*, etc.), but the choice is ultimately dependent on the perspective you decide to take on the referent for your interlocutor[21]. Will it be more effective for me to refer to my sister as *my sister* or as *that lady* or as *the physicist*? It will all depend on shared knowledge and discourse context. This freedom of perspective-taking appears quite early in life[22] and is ubiquitous in conversation.

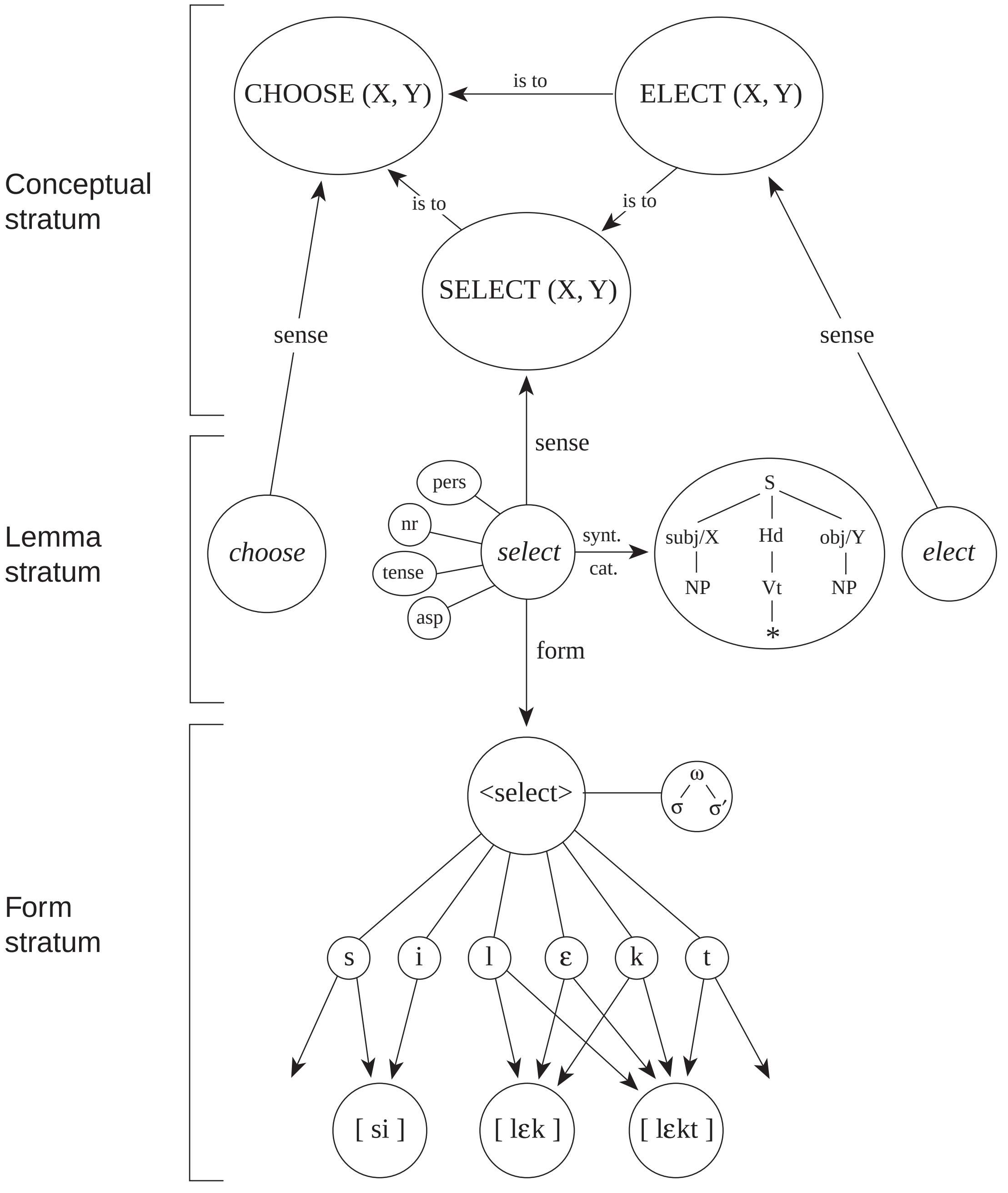


Fig. 2. Fragment of Roelofs's WEAVER network model of the lexicon. The nodes in the upper layer represent whole lexical concepts. The arrow connections represent the semantic relations holding among them. There is bi-directional activation spreading at this level. The mid stratum is syntactic. The nodes represent lemmas, that is syntactic words and their features. Among these properties are the word's syntactic frame, its variable inflectional features (such as number), its gender (in gender marking languages), etc. The activation spreading is uni-directional from lemma node to feature nodes. The lemma's connection to its concept node represents the sense of the word. The connection allows for bi-directional activation spreading. The arrow down from the lemma points to its form. Activation spreading is uni-directional here, which is maintained all the way down through the network. Only a selected lemma can spread its activation to the form level. The bottom stratum represents morpheme nodes with their connections to metrical and phoneme nodes. In their turn, the phoneme nodes point to all (stored) phonetic syllables in which they participate; they are not specified for their syllable position. There are no inhibitory connections in the network. (Adapted from Levelt[4].)

Working models of word production begin where perspective-taking ends: at the activation of a target concept to be expressed. The representation of a target concept, however, varies among models. The two preferred variants are just the ones exemplified in Figs 1 and 2. Concepts are either represented as decomposed, or as non-decomposed or 'whole'. The issue is controversial[23], but arguments have been accumulating for using whole-concept representations in models of word production[24]. One argument is the so-called 'hyperonym problem'[3]. If you activate some set of semantic features as a representation of the notion 'cat', the notion 'animal' will involve a proper subset of these features. Hence, it is indeterminate which of the two will ultimately be expressed. This is not an advantage: hyperonym speech errors are rare in any case[25] and you need extra machinery to prevent the hyperonym problem from arising[26].

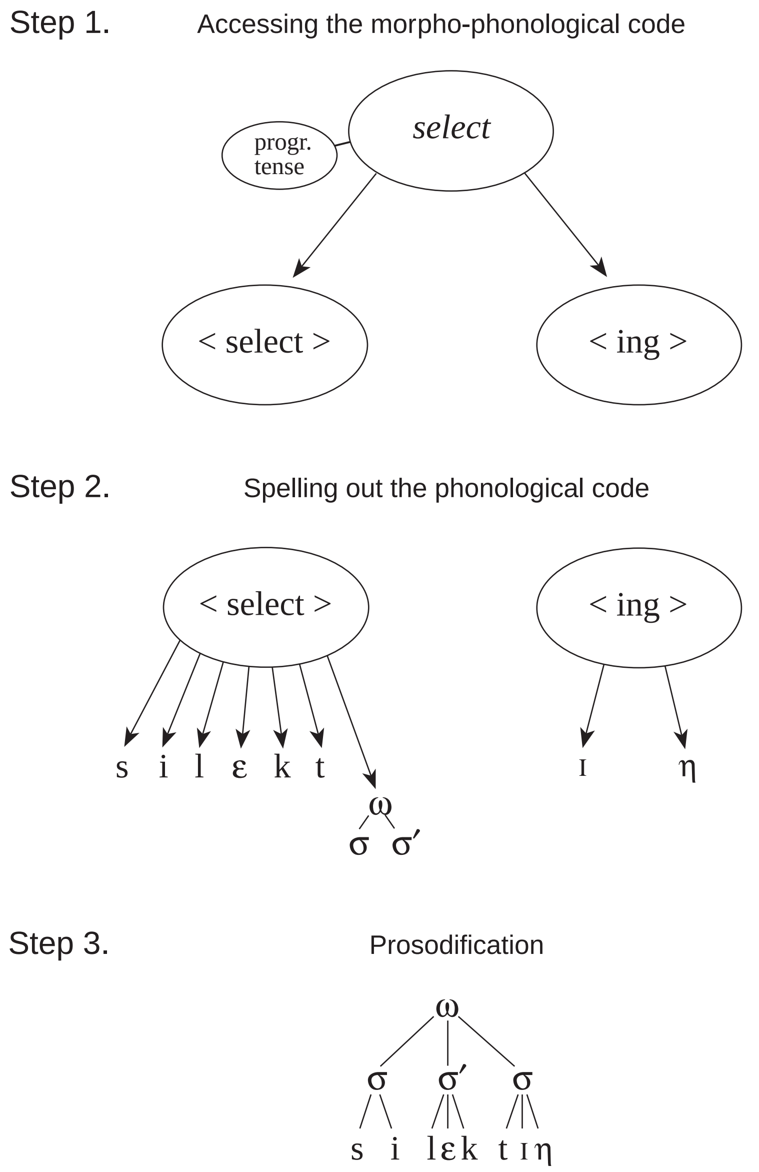


Fig. 3. Three steps in the morpho-phonological encoding of the word *selecting*. Step 1. The target lemma is *select*, marked for progressive tense. Two codes are successively accessed, first the code for the head morpheme <select>, then the code for the suffix morpheme <ing>. For each code the speed of access is dependent on its frequency of usage.
Step 2. Each morphemic code is 'spelled out'. Each morpheme's segments (s, i, l, ...) are simultaneously selected. Also spelled out is the metrical code of <select>. It specifies that word stress must go to the second syllable. The symbol ω stands for 'phonological word' and the symbol σ stands for 'syllable'; σ′ is the stressed syllable. The affix does not have a metrical code.
Step 3. The spelled out segments are one by one attached to the metrical code. Following the rules of the language, s becomes the onset of the first syllable and i its nucleus or vowel; that completes the first syllable. Then l becomes the onset of the second syllable and ϵ its nucleus. The next segment k cannot become the onset of a new syllable for two reasons. First, ϵ is a short vowel; an English syllable cannot close on a short vowel. Second, k is a more sonorous consonant than t; English doesn't allow a more sonorous consonant to precede a less sonorous one at syllable onset; kt cannot become the onset of a syllable. Hence, k becomes the offset of the second syllable. Notice that the syllabification process must have access to segmental features such as vowel length or sonority. The next segment, t, can become the onset of a new syllable, which must be created anyway because there is a further vowel, i, in the offing. The new syllable node, σ, is attached to ω, which means that it becomes part of the phonological word under construction. Then i is attached as nucleus and the remaining η as a coda of the third and last syllable.

Both whole-concept and featural representations allow for precise semantic inferencing (of the type 'a dog is an animal'), but this inferential potential plays no role in the factual word production process.

Lexical selection

In the chronometric tradition lexical selection has been studied with interference paradigms, in particular picture-word interference (see Box 1). The recurring finding has been that naming an object is slowed down when a distracter word is presented with the picture; the effect is stronger when the distracter word is semantically related to the target than when it is semantically unrelated and it is at maximum when picture and distracter word are presented simultaneously[27]. The WEAVER model provides an accurate quantitative account of a wide range of picture-word interference data[28], with only a few free parameters. How does it work? When you are naming a picture of a sheep and you decide to go for the basic level term, you will activate the lexical concept *sheep* as your target and activation spreads to the corresponding lemma. In the semantic network activation spreads to related concepts, such as *goat* and *llama*. They, in turn, spread activation to their lemmas. During any unit time interval the probability of selecting the target lemma *sheep* from the mental lexicon is the ratio of that lemma's degree of activation and the total activation of all lemmas (including *goat*, *llama* and *sheep*). This is called Luce's ratio[29], and it allows for the computation of an expected selection latency. In other words, there is competition between semantically related lemmas. Active alternatives slow down the selection process (even though a special checking mechanism in WEAVER normally prevents them from replacing the target). If you present the semantically related word 'goat' as a distracter, the already co-activated lemma *goat* will receive an additional boost, thereby becoming a strong competitor to *sheep*. By contrast, if you present a semantically unrelated word, such as 'chair', as distracter, there will be no convergence of activation and, correspondingly, competition will be relatively weak. That explains the semantic-inhibition effect.

Activation spreading through a semantic network (of whatever type) is also the obvious explanation for semantic naming errors, the dominant speech error type (about two-thirds of errors in a normal picture naming task are semantic in character[25]). But what is a semantic error? A particular choice of words may have its cause in perspective-taking[30]. If a speaker decides to name a depicted dog an 'animal' or a 'collie', that may well be an intentional act rather than an error.

There is a substantial literature on the types of semantic (and other) errors produced by aphasic patients[25,30], which will not be covered in the present review. It is a major challenge to predict these error distributions by 'damaging' the normal network[31]. Dell *et al.*[12] have set an impressive example. They successfully modeled the naming errors (semantic and other) of a diverse set of aphasic patients by manipulating no more than two parameters in their interactive two-step model: the weight on the network connections and the decay rate of the nodes' activation.

The timing of lexical selection is not explicitly modeled in the speech-error based models. In the interactive two-step model the selection moment is determined from outside. When you produce a sentence, the moment of selecting the most activated lemma is dictated by when it is to be inserted in the grammatical frame. The selection moment is usually given a constant default value in modeling error distributions.

Box 2. Implicit priming

The method of implicit priming was introduced by Meyer to study the time course of phonological encoding, that is the speaker's construction of a spoken word's form (Refs a,b). The initial and major discovery, which has been repeatedly reconfirmed, was that a word's form is built up incrementally, starting with the first segment. Apparently, phonological word shapes do not come as whole templates; rather they are generated afresh, time and again, from beginning to end.

The method is exemplified in Table 1. Subjects learn a set of three semantic word-associations (A–B), for instance set 1 in the leftmost column. Then, an A-word from the set appears on the screen and the subject produces the corresponding B-word as fast as possible. The word onset latency is measured by voice key. The A-words from the set are repeatedly presented in random order and at each trial the naming latency of the B-word is registered. Then the subject is presented with set 2, the triple in the second column of the table below, and the same procedure is run for that set. Finally, set 3 is run in the same way.

The response words in a set share a phonological property. The B-words in set 1 are *loner, local* and *lotus*; they share the initial syllable *lo*. Similarly, the B-words in set 2 share the initial syllable *bea*, and those in set 3 share the initial syllable *ma*. Such sets sharing a phonological property are called 'homogeneous' and the shared property is called the 'implicit prime'.

Can the subject use this implicit prime when running through the set? Whether the subject can prepare for the first syllable of the response word can be tested by comparing the homogeneous condition with a heterogeneous condition; that is, one in which there is no implicit prime. The heterogeneous condition is created by reordering the A–B pairs in such a way that they no longer share their first syllable. For instance, the first set of the heterogeneous condition (fourth column in the table) has *loner*, *beacon* and *major* as response words. Each word-pair is its own control in the experiment: it appears in both the homogeneous and the heterogeneous condition.

In the heterogeneous condition there is no implicit prime, hence the subject cannot prepare anything. When Meyer did the experiment exemplified in Table 1 (in Dutch), she found that response latencies were significantly shorter in the homogeneous condition than in the heterogeneous condition. Apparently, subjects can prepare for the response word's first syllable.

If prosodification is really incremental (i.e. starting at the beginning of the word), subjects should not be able to prepare for the second syllable of a bisyllabic word. And indeed, a test using response words ending in the same syllable, such as *murder, ponder, boulder*, showed no implicit priming whatsoever. Generally, there was always implicit priming for words that shared any beginning part of the word, but never for words sharing any final part, not even for monosyllabic rhymes, such as *deed*, *feed, seed*. In addition, the longer the shared word-beginning stretch, the stronger the priming.

These robust findings have led to the suggestion that, normally, the speaker does not initiate articulation before the whole word has been encoded. If encoding is incremental, which is now well-established, this should predict a robust word-length effect. But this is rarely obtained (Ref. c). It is unknown under which conditions a speaker does complete a word's phonological encoding before initiating speech.

Implicit priming can also be used to test whether a speaker must know which syllable to stress in prosodification. For instance, in Table 1 all response words have the same metrical shape: they are all first-syllable-stressed. In this case, there is strong implicit priming. But will there still be implicit priming if the response words do not share stress position? Roelofs and Meyer tested this in Dutch, using sets of response words such as: *ma-ri'-ne* – 'navy', *ma-nus-cript'*–'manuscript', *ma-te'-rie* – 'matter', *ma-de-lief'* – 'daisy' (Ref. d). Here two words in the set have second syllable stress and two have third syllable stress. There was not the slightest hint of implicit priming in this condition of variable metrics. This means that you cannot prepare for the first syllable if you don't know where the word's stress is to go – in order to prepare you must know the word's 'metrical frame'.

References

a Meyer, A.S. (1990) The time course of phonological encoding in language production: the encoding of successive syllables of a word *J. Mem. Lang.* 29, 524–545

b Meyer, A.S. (1991) The time course of phonological encoding in language production: phonological encoding inside a syllable *J. Mem. Lang.* 30, 69–89

c Bachoud-Lévi, A-C. *et al.* (1998) Where is the length effect?: a cross-linguistic study of speech production *J. Mem. Lang*. 39, 331–346

d Roelofs, A. and Meyer, A.S. (1997) Metrical structure in planning the production of spoken words *J. Exp. Psychol. Learn. Mem. Cognit.* 24, 922–939

Table 1. The implicit priming method: priming the first syllable of bisyllabic words

Homogeneous condition			Heterogeneous condition		
Set 1	Set 2	Set 3	Set 1	Set 2	Set 3
single–loner	signal–beacon	captain–major	single–loner	place–local	fruit–lotus
place–local	priest–beadle	cards–maker	signal–beacon	priest–beadle	glass–beaker
fruit–lotus	glass–beaker	tree–maple	captain–major	cards–maker	tree–maple

Morpho-phonological encoding

When you are planning the sentence '*they are selecting me*', you must retrieve from your lexicon the morpho-phonological codes for each of the selected words, among them the two morpheme-size codes *select* and *ing* (see Fig. 3), and compute their syllabification and accent structure in context (*se-léc-ting*). This naturally divides the process into 'code retrieval' and 'prosodification'.

Code retrieval

An item's morpho-phonological code consists of its morphological make-up, its metrical shape and its segmental make-up (see Fig. 3, Step 1 and Step 2). Retrieving that information *follows* activation/selection of the lemma[32]. Much ink and many subjects have been spilled over this issue. In the WEAVER model, the activation and retrieval of a phonological code is strictly conditional on selecting the corresponding lemma. For instance, when your target word is *cat*, you first select its lemma and only then spread activation to its phonological code (kæt). This predicts that alternative active, but non-selected lemmas (such as the lemma for *dog*) do not spread any activation to their phonological codes. Initial experimental evidence[33] showed that, in picture naming, there is semantic but indeed no phonological activation of same-category alternatives (if *cat* is the target, *dog* is semantically but not phonologically active). All speech-error based models of word production, however, assume that there is free cascading of activation throughout the network. Hence, active alternatives should also become active phonologically, at least to some extent[34]. However, the original finding was reconfirmed in a quite critical replication[35]. Still, evidence for phonological co-activation of semantic alternatives was obtained for one restricted case: if the alternative is a synonym of the target[35,36]. When you name the picture of a couch, the phonological code of *sofa* is measurably co-activated. The cause of this robust finding is unclear. It shows that cascading exists, but not that it is a general property of the lexical network – after all, it doesn't show up for same-category items such as *cat* and *dog*. I suggest that the phenomenon is related to perspective taking. When you have two equivalent ways of making reference to an object, you may occasionally select both lemmas and hence spread activation to both phonological codes. This means that WEAVER's special checking mechanism (see above) can occasionally fail if two highly competitive lemma's are about equally activated. There is suggestive speech error evidence that this indeed occurs: phonological word blends tend to be blends of near-synonyms (such as *close* and *near* blending into *clear*), hardly ever of same-category items.

There is a strong and robust word-frequency effect in word production (which is in part an age-of-acquisition effect)[37,38].Controlling for conceptual biases, you are typically faster in producing a high-frequency response such as *mouth* than a low-frequency response such as *moth*. It is now known that the effect arises in accessing the phonological code (Fig. 3, Step 1), not in selecting the lemma[39]. This fact has a suggestive relation to the so-called tip-of-the-tongue (TOT) phenomenon. It happens occasionally that, while normally speaking, you get stuck on the name of a person, flower, instrument or whatever. The phenomenon can be experimentally induced by presenting a subject with the definition of an object to be named. If the name is low-frequency, you often induce the TOT state in the subject. When the target language is gender-marking (such as Italian), there is a good chance that the subject knows the gender of the problem word[40] and this also holds for the much amplified case of word finding trouble in many anomic patients[41]. This has been used as one of many arguments for the distinction between an 'earlier' syntactic lemma-level and a 'later' phonological code level in the lexical network[18]. But that argument has provoked some controversy[42–44] which is, so far, unresolved. Probably more relevant speech error evidence for the precedence-of-syntax claim is the repeated finding of almost absolute gender preservation in phonological word substitution errors (such as *lasagna* for *lavagna* in Italian)[44–46]. Most of these errors are real, on-line productions of the lexical network. So far, however, they have not been modeled.

Outstanding questions

- How should error-based and chronometric models be further reconciled computationally and empirically?
- What causes a speech error? Is it caused by *occasional* cascading or *occasional* feedback in a normally non-cascading, feed-forward system? Is it the product of noise in a normally cascading interactive system? Or is the origin of speech error something else entirely?
- How does the word-production network relate to the word-perception network? How is self-monitoring realized in this combined system?
- How are syllabic and larger gestures computed from a syllabified phonological code? Is there anything like a repository of syllabic gestural scores?
- If phonological word encoding is an incremental process, why is it that naming a short word is hardly faster than naming a long word?
- Which brain regions subserve the core components of conceptual/semantic preparation, lexical selection, phonological code retrieval, prosodification, phonetic encoding, articulation and self-monitoring?

Prosodification

The core process here is incremental syllabification. Let us return to the target sentence *they are selecting me* (Fig. 3, Step 3). The morpho-phonological code of the progressive lemma *select* consists of two morpho-phonological packages, (s, i, l, ϵ, k, t) and (ɪ, η). Syllabification proceeds 'from left to right'. You first chunk the first two phonemes to create the syllable /si/. You then take the next three to compose the syllable /lϵk/ and, finally, you chunk the remaining segments to compose /tɪη /. The best evidence for the strict incrementality of this process comes from experiments using the 'implicit priming paradigm' (see Box 2). Notice that the last syllable, /tɪη/, straddles two morphemes, *select* and *ing.* This can also happen across words. When you utter *They will select us*, the syllabification will be /si-lϵk-tʌs/, where /tʌs/ straddles the words *select* and *us*. But when you produce *they select me*, the syllabification is /si-lϵkt- mi/, without straddling. Apparently, the syllables are not given in the phonological code of the morpheme, but depend on the context in which the word and its morphemes appear. The word's phonemes are not marked for a fixed position in their syllables; the /t/ in *select* will appear as syllable onset or as syllable offset, dependent on the context. The domain of syllabification (such as *selecting*, *selectus*, *select*) is called the 'phonological word'. It can be larger or smaller than the lexical word. The incremental 'chunking' of segments in the on-line composition of syllables follows a strict set of rules, which varies among languages[47]. These rules are rapidly applied, time and again, in the fluent generation of speech. When you are a speaker of Papuan Hua, all your syllables consist of a consonant (C) followed by a vowel (V), CV. Other languages have one or more other syllable frames in addition, such as V, CVC, CCV and so on. Traditionally, syllabification was conceived of as filling such syllabic frames (see Box 1), but arguments for this view have become less convincing[5,18,48]. In particular, the idea that phonemes in the phonological code (such as /t/ in *select*) are marked for a particular syllable position creates more problems than it solves. The preference of sound exchanges (such as *moggy barsh* for *boggy marsh*) to preserve syllabic position can be explained differently, as a combination of word onset vulnerability, phoneme similarity and phonotactic restrictions.

There is good chronometric evidence, however, for the existence of *metrical* frames (see Fig. 3). For Dutch, and probably for other stress-assigning languages such as English and German, there is a dominant metrical pattern:

word stress goes to the first full-voweled syllable (*mórning, yéllow, forgét* – the 'o' in the latter word is not full-voweled, but rather a neutral 'schwa'-sound). This can be automatically produced in incremental syllabification. But when a word has a deviant stress pattern, the automaticity breaks down[18,49] (see Box 2 for an example). A word's deviant metrical frame is probably stored as part of its phonological code; it guides the deviant prosodification. Languages differ, however, in their default metrics.

The distinction between accessing a word's phonological code and its subsequent rapid syllabification is crucial for understanding the neural achitecture of word production. A meta-analysis of imaging studies in word production[50] suggests that accessing the code involves Wernicke's area, whereas prosodification involves the posterior inferior frontal cortex.

Phonetic encoding and articulation

As incremental prosodification proceeds, the resulting syllabic and larger prosodic structures should acquire phonetic shape. As a speaker you will incrementally prepare articulatory gestures for the syllables in their prosodic context. A core feature of the WEAVER model is the notion of a *syllabary*[51]. Statistics show that native speakers of English or Dutch do 80 percent of their talking with no more than about 500 different syllables[18] (although these languages have many more than 10 000 different syllables). The syllabary is postulated as a repository of such overused, high-frequency syllabic gestures, one 'syllabic score' for each. Each time a new phonological syllable, such as /si/, /lɛk/, or /tʌs/, is composed, the corresponding gestural score is triggered. The score specifies which motor tasks (such as closing the glottis or releasing lip closure) are to be performed[52] in order to generate the syllable. In WEAVER there is always competition among gestural scores. The activation spreads from individual segments to all syllabic scores in which they participate (see Fig. 2). Hence, similar syllabic scores tend to be co-activated. The occasional mis-selection will resemble the target gesture. Selection latency is determined by Luce's rule (as it was the case for lemma selection).

There are further restrictions in selecting a syllabic score for execution. Repeated use of a particular *type* of syllable, for instance in producing the nonsense phrase *kem-til.fler* (where *kem* and the following *til* are both CVC syllables), may facilitate articulation[53]. Gestural scores of similar types (such as CV or CVC) can apparently co-activate one another. Finally, WEAVER and the two-step interactive model have a featural representation of each segment. In both models the units of phonological encoding are whole phonemes (for which there is good experimental evidence[54]), but their features, such as 'voiced', 'nasal', 'sonorous', are already 'visible' to the process of syllabification (see legend to Fig. 3). During the next stage, phonetic encoding, these features function in the construction of articulatory gestures. The study of speech movement planning has become a discipline of its own[55,56] and is not covered in the present review.

Conclusion

There is still a long way to go before the two research traditions emerging from speech error analysis and from naming chronometry are fully reconciled. But there has been lively and highly constructive interaction, leading to a much improved understanding of the processes involved in lexical selection and phonological encoding. One unifying force has been computational modeling. Current implemented models share their major strata, they are localist and symbolic; they compute quite similar linguistic representations. Another unifying force will hopefully proceed from brain imaging (see Ref. 57 for a recent review of imaging studies of word processing). It is the processing models that should guide the design of brain imaging experiments in word production, not naive intuition as is still too often the case[50]. The return will be convergence of evidence for or against particular processing components and their interactions.

Acknowledgement

I gratefully acknowledge helpful commentary by Antje Meyer and by Gary Dell.

References

1 Miller, G.A. (1991) *The Science of Words*, Scientific American Library

2 Garnham, A. *et al.* (1981) Slips of the tongue in the London–Lund corpus of spontaneous conversation *Linguistics* 19, 805–817

3 Levelt, W.J.M. (1989) *Speaking: From Intention to Articulation*, MIT Press

4 Levelt, W.J.M. (1999) Language production: a blueprint of the speaker, in *Neurocognition of Language* (Brown, C. and Hagoort, P., eds), pp. 83–122, Oxford University Press

5 Dell, G.S., Juliano, C. and Govindjee, A. (1993) Structure and content in language production: a theory of frame constraints in phonological speech errors *Cognit. Sci.* 17, 149–195

6 Mackay, D.G. (1987) *The Organization of Perception and Action: A Theory for Language and other Cognitive Skills,* Springer-Verlag

7 Stemberger, J.P. (1985) An interactive activation model of language production, in *Progress in the Psychology of Language* (Vol. 1) (Ellis, A.W., ed.), pp. 143–186, Erlbaum

8 Berg, T. (1998) *Linguistic Structure and Change*, Clarendon Press

9 Dell, G.S. (1986) A spreading-activation theory of retrieval in sentence production *Psychol. Rev.* 93, 283–321

10 Harley, T.A. (1993) Phonological activation of semantic competitors during lexical access in speech production *Lang. Cognit. Process.* 8, 291–309

11 Schade, U. and Berg, T. (1992) The role of inhibition in a spreading-activation model of language production: II. Simulational perspective *J. Psycholinguist. Res.* 21, 435–462

12 Dell, G.S. *et al.* (1997) Lexical access in aphasic and non-aphasic speech *Psychol. Rev.* 104, 801–837

13 Damian, M.F. and Martin, R.C. (1999) Semantic and phonological codes interact in single word production *J. Exp. Psychol. Learn. Mem. Cognit.* 25, 1–18

14 Humphreys, G.W., Lamote, C. and Lloyd-Jones, T.J. (1995) An interactive activation approach to object processing: effects of structural similarity, name frequency and task in normality and pathology *Memory* 3, 535–586

15 Glaser, M.O. and Glaser, W.R. (1982) Time course analysis of the Stroop phenomenon *J. Exp. Psychol. Hum. Percept. Perform.* 8, 875–894

16 La Heij, W., Dirkx, J. and Kramer, P. (1990) Categorical interference and associative priming in picture naming *Br. J. Psychol.* 81, 511–525

17 Roelofs, A. (1997) The WEAVER model of word–form encoding in speech production *Cognition* 64, 249–284

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

19 Motley, M.T., Camden, C.T. and Baars, B.J. (1982) Covert formulation and editing of anomalies in speech production: evidence from experimentally elicited slips of the tongue *J. Verb. Learn. Verb. Behav.* 21, 578–594

20 Rosch, E. *et al.* (1976) Basic objects in natural categories *Cognit. Psychol.* 8, 382–439

21 Levelt, W.J.M. (1996) Perspective taking and ellipsis in spatial descriptions, in *Language and Space* (Bloom, P. *et al.*, eds), pp. 77–108, MIT Press

22 Clark, E.V. (1997) Conceptual perspective and lexical choice in

language acquisition *Cognition* 64, 1–37

23 Bierwisch, M. and Schreuder, R. (1992) From lexical concepts to lexical items *Cognition* 42, 23–60

24 Roelofs, A. (1997) A case for nondecomposition in conceptually driven word retrieval *J. Psycholinguist. Res.* 26, 33–67

25 Martin, N. *et al.* (1996) Phonological faciliation of semantic errors in normal and aphasic speakers *Lang. Cognit. Process.* 11, 257–282

26 Bowers, J.S. (1999) Grossberg and colleagues solved the hyperonym problem over a decade ago *Behav. Brain Sci.* 22, 38–39

27 Glaser, M.O. and Düngelhoff, F-J. (1984) The time course of picture–word interference *J. Exp. Psychol. Hum. Percept. Perform.* 7, 1247–1257

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

29 Luce, R.D. (1959) *Individual Choice Behavior,* John Wiley & Sons

30 Garrett, M.F. (1992) Disorders of lexical selection *Cognition* 42, 143–180

31 Plaut, D.C. and Shallice, T. (1993) Perseverative and semantic influences on visual object naming errors in optic aphasia: a connectionist account *J. Cogn. Neurosci.* 5, 89–117

32 Van Turennout, M., Hagoort, P. and Brown, C. (1998) Brain activity during speaking: from syntax to phonology in 40 milliseconds *Science* 280, 572–574

33 Levelt, W.J.M. *et al.* (1991) The time course of lexical access in speech production: a study of picture naming *Psychol. Rev.* 98, 122–142

34 Dell, G.S. and O'Seaghdha, P.G. (1991) Mediated and convergent lexical priming in language production: a comment on Levelt *et al. Psychol. Rev.* 98, 604–614

35 Peterson, R.P. and Savoy, P. (1998) Lexical selection and phonological encoding during language production: evidence for cascaded processing *J. Exp. Psychol. Learn. Mem. Cognit.* 24*,* 539–557

36 Jescheniak, J.D. and Schriefers, H. (1998) Discrete versus cascaded processing in lexical access in speech production: further evidence from the coactivation of near-synonyms *J. Exp. Psychol. Learn. Mem. Cognit.* 5, 1256–1274

37 Oldfield, R.C. and Wingfield, A. (1965) Response latencies in naming objects *Q. J. Exp. Psychol.* 17, 273–281

38 Levelt, W.J.M. *et al.* (1998) An MEG study of picture naming *J. Cogn. Neurosci.* 10, 553–567

39 Jescheniak, J.D. and Levelt, W.J.M. (1994) Word frequency effects in speech production: retrieval of syntactic information and of phonological form *J. Exp. Psychol. Learn. Mem. Cognit.* 20, 824–843

40 Vigliocco, G., Antonini, T. and Garrett, M.F. (1997) Grammatical gender is on the tip of Italian tongues *Psychol. Sci.* 8, 314–317

41 Badecker, W., Miozzo, M. and 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

42 Caramazza, A. and Miozzo, M. (1997) The relation between syntactic and phonological knowledge in lexical access: evidence from the 'tip-of-the-tongue' phenomenon *Cognition* 64, 309–343

43 Roelofs, A., Meyer, A.S. and Levelt, W.J.M. (1998) A case for the lemma/lexeme distinction in models of speaking: comment on Caramazza and Miozzo (1997) *Cognition* 69, 219–230

44 Vigliocco, G. *et al.* Is 'count' and 'mass' information available when the noun is not?: an investigation of tip-of-the-tongue states and anomia *J. Mem. Lang.* (in press)

45 Berg, T. (1992) Prelexical and postlexical features in language production *Appl. Psycholinguist.* 13, 199–235

46 Marx, E. Gender processing in speech production: evidence from German *J. Psycholinguist. Res.* (in press)

47 Goldsmith, J.A. (1990) *Autosegmental and Metrical Phonology*, Blackwell

48 Meyer, A.S. (1992) Investigation of phonological encoding through speech error analysis: achievements, limitations, and alternatives *Cognition* 42, 181–211

49 Roelofs, A. and Meyer, A.S. (1997) Metrical structure in planning the production of spoken words *J. Exp. Psychol. Learn. Mem. Cognit.* 24, 922–939

50 Indefrey, P. and Levelt, W.J.M. in *The Cognitive Neurosciences* (2nd edn) (Gazzaniga, M., ed.), MIT Press (in press)

51 Levelt, W.J.M. and Wheeldon, L. (1994) Do speakers have access to a mental syllabary? *Cognition* 50, 239–269

52 Browman, C.P. and Goldstein, L. (1988) Some notes on syllable structure in articulatory phonology *Phonetica* 45, 140–155

53 Sevald, C.A., Dell, G.S. and Cole, J.S. (1995) Syllable structure in speech production: are syllables chunks or schemas? *J. Mem. Lang.* 34, 807–820

54 Roelofs, A. (1999) Phonological segments and features as planning units in speech perception *Lang. Cognit. Process.* 14, 173–200

55 Kent, R.D., Adams, S.G. and Turner, G.S. (1996) Models of speech production, in *Principles of Experimental Phonetics* (Lass, N.J., ed.), pp. 3–45, Mosby

56 Guenther, F.H., Hampson, M. and Johnson, D. (1998) A theoretical investigation of reference frames for the planning of speech movements *Psychol. Rev.* 105, 611–633

57 Price, C.J. (1998) The functional anatomy of word comprehension and production *Trends Cognit. Sci.* 2, 281–287

Corrigendum

In the September 1998 issue of *Trends in Cognitive Science*, the Review article entitled 'Dysmetria of thought: clinical consequences of cerebellar dysfunction on cognition and affect' by J.D. Schmahmann (Vol. 2, pp. 362–371) contained a textual error.

On p. 363, it was incorrrectly stated that 'Children with attention deficit hyperactivity disorder have also been shown to have statistically smaller vermal lobules VI and VII on MRI (Ref. 18)…' (right col., l. 22). The sentence should instead read 'Children with attention deficit hyperactivity disorder have also been shown to have statistically smaller vermal lobules VIII to X on MRI (Ref. 18), and a similar observation has been made in fragile-X syndrome[19].'

We apologize to readers for this error.

PII: S1364-6613(99)01326-1