EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY, 2000, *12* (4), 433–452

Word for word: Multiple lexical access in speech production

Willem J.M. Levelt and Antje S. Meyer
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

It is quite normal for us to produce one or two million word tokens every year. Speaking is a dear occupation and producing words is at the core of it. Still, producing even a single word is a highly complex affair. Recently, Levelt, Roelofs, and Meyer (1999) reviewed their theory of lexical access in speech production, which dissects the word-producing mechanism as a staged application of various dedicated operations. The present paper begins by presenting a bird eye's view of this mechanism. We then square the complexity by asking how speakers control multiple access in generating simple utterances such as *a table and a chair*. In particular, we address two issues. The first one concerns dependency: Do temporally contiguous access procedures interact in any way, or do they run in modular fashion? The second issue concerns temporal alignment: How much temporal overlap of processing does the system tolerate in accessing multiple content words, such as *table* and *chair*? Results from picture–word interference and eye tracking experiments provide evidence for restricted cases of dependency as well as for constraints on the temporal alignment of access procedures.

We are experts at producing words. Our estimate is that by the age of 21, a normal person in our culture has produced close to 50 million word tokens (assuming an average of 45 talking minutes a day and 2.5 words per second). That is because most of us are talking addicts, forever hooked since the word spurt set in during our second year of life. There is no other cognitive-motor skill exercised so much as the production of words.

Requests for reprints should be addressed to W.J.M. Levelt, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. Email:pim@mpi.nl

This paper was presented as Broadbent Lecture, ESCOP, Gent, 1 September 1999.

We gratefully acknowledge the contributions of our PhD students Femke van der Meulen, Astrid Sleiderink, and Simone Sprenger, and of our student assistants Frouke Hermens and Gijs van Elswijk.

Antje S. Meyer is now at School of Psychology, University of Birmingham.

Linguists, neurologists, and the occasional psychologist began studying this skill over a century ago, but the systematic analysis of word production is only three decades old. Two initially quite unrelated approaches put the production of words again on the research agenda in psycholinguistics: word production chronometry and speech error analysis.

Two classical insights arose from the chronometric tradition. The first one is that semantically related words compete for production, which is saliently the case in the famous Stroop task, and convincingly demonstrated by way of picture/word interference experiments (Lupker, 1979; Rosinski, Michnick Golinkoff, & Kukish, 1975). The paradigm here is to present a picture to be named and to measure the speech onset latency. Simultaneously with the picture, or somewhat earlier or later (at different stimulus onset asynchronies, SOAs), a distractor stimulus is presented, which the subject should try to ignore. It can be a word that is in some way related to the picture, or an unrelated word. It can be a visual stimulus appearing in the picture, or an auditory stimulus. When you present a competing, semantically related distractor, such as *cow* when the target picture presents a dog, you get inhibition, that is slower naming (as compared to presenting an unrelated distractor word, such as *stone*). The second classical insight is that the speed of accessing words for production is frequency dependent. A much-used word, usually one that has been acquired early in life, is produced faster than a low-frequency word (Oldfield & Wingfield, 1965). John Morton (1969) combined these two insights in the very first chronometric model of lexical access: his Logogen theory.

The existence of semantic competition was also apparent from the speech error tradition. Word substitution errors, such as "don't burn your toes" (intended: fingers—from Garrett, 1975), are mostly semantic in nature (Meringer & Mayer, 1895/1978; Fromkin, 1973). Another contribution, where speech error analysis was far ahead of the chronometric tradition, was the demonstration that words are not produced as indivisible wholes, but composed segment by segment, as appears from phonological speech errors such as *Yew Nork* for *New York*, in which single segments exchange position. Stephany Shattuck-Hufnagel (1979) accounted for these data in the first theory of phonological word encoding: the scan-copier model. The confluence of these two approaches (see Levelt, 1999 for more detail) has led to rather sophisticated theories of word production, of which we will presently review one in some more detail.

However, words usually don't come alone. We skilfully combine them in the act of speaking, at rates of some two to three per second. What happens if we produce two, or a few words in close temporal contiguity? Do we simply concatenate the access procedures for each of the words, finishing one before starting the next? Or do these procedures overlap in

ways to be discovered? To make these issues more precise, we need some of the details of our theory of lexical access.

SINGLE WORD ACCESS

A review of our theory, its computational implementation, and empirical support are available as a BBS target paper (Levelt et al., 1999). The theory has two features that are particularly relevant for the present purpose. The first one is the staged character of lexical access (Figure 1), the second one is its realisation through spreading activation (Figure 2).

Conceptual preparation

Producing words is a staged process. Take the verb *defend* as an example. Imagine a city under siege and its pugnacious general declaring: *We'll defend it*. The very first stage of preparing this utterance is to decide what notion to express. Would it be clearer, or more appropriate for the general, to present his state of mind in terms of the notion DEFEND, or

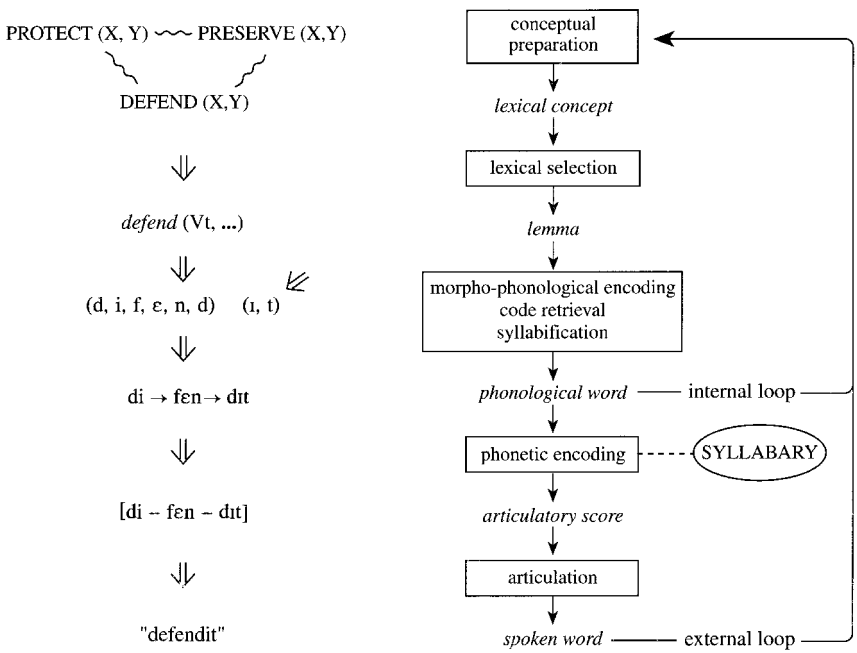

Figure 1. Stages of lexical access in spoken word production.

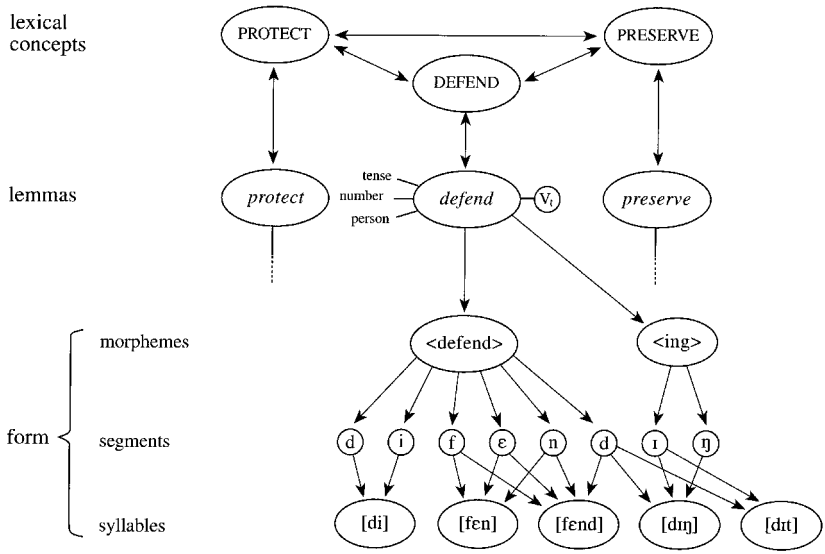

Figure 2. A fragment of the lexical network in the WEAVER model of lexical access in speech production.

rather PROTECT or PRESERVE? That depends on the strategic circumstances, in particular the condition of the addressee. Such ruminations are called "perspective taking". Let us assume the general opts for the notion of defending. It is, then, the target concept. It is also a lexical concept, because the general has a word for it in his language, which happens to be English. Figure 2 shows the general's active concept DEFEND in the network model (called WEAVER, Roelofs, 1992, 1997). From lexical decision experiments (Levelt et al., 1991) we know that, at least during picture naming, semantically related concepts are co-activated. For DEFEND they might be PROTECT and PRESERVE. This completes the first stage in Figure 1, conceptual preparation.

Lexical selection

During the next stage, lexical selection, the general selects from his mental lexicon a word that corresponds to his target concept. There is precisely one word in the lexicon that fits it, the word *defend*. Actually, the general doesn't immediately retrieve the whole word; initially only the word's syntax becomes available plus a pointer to the word form. That bit of information is called the "lemma". The corresponding node, the lemma for *defend*, is depicted in the lemma stratum of Figure 2. It specifies that it is a

verb, a transitive one. The lemma also has variable options for tense, number, and person features, which are set during grammatical encoding. Now remember that the general had a few alternative concepts in mind, among them PROTECT and PRESERVE. They have been sending activation to their lemmas, hence these lemmas are now in competition with *defend*. In WEAVER, classical word competition is competition among lemmas. Competition is a main determinant of selection latency. Roelofs (1992) defines the probability of selecting the target lemma during any unit time interval by Luce's (1959) ratio: the activation of the target lemma divided by the summed activation of all lemmas. This rule made it possible to predict response latencies for a large variety of picture/word interference conditions, predictions that were well met by the experimental data (Levelt et al., 1999; Roelofs, 1992, 1993). The selection of the target lemma, the transitive infinitive verb *defend*, completes this second stage of the general's lexical access, lexical selection.

Morpho-phonological encoding

During the subsequent stage, morpho-phonological encoding, the general's job becomes a very different one. Up to this point, he had to deal with a whole army of active, competing words. From now on he has only one target left: to prepare the articulation of a single word, "defend", in its context. This begins by retrieving the word form, that is the phonological code to which the selected lemma points. There is good evidence now that a word's lemma is accessed before its phonological code is retrieved. One of several relevant studies is by van Turennout, Hagoort, and Brown (1998), which shows that accessing a word's syntactic gender in picture naming precedes accessing its phonological code by about 40 ms.

Our general will retrieve the phonological code of *defend*, depicted at the word form stratum of Figure 2. It largely consists of the word's phonological segments: /d/, /i/, /f/, /ɛ/, /n/, and /d/, an ordered set. Picture/word interference experiments (among others Meyer & Schriefers, 1991) have made it most likely that these segments are simultaneously activated. Still, phonological codes come in successive packages. You retrieve them per morpheme. For instance, we have as yet unpublished experimental evidence showing that when you access a multimorphemic word, such as *doorstep*, you first retrieve the code for *door* and then the code for *step*. Sequentiality may also hold for stems and inflections. If the general decides to say *we will be defending the city*, the progressive tense feature of the lemma *defend* causes the code for *defend* and the code for the inflection *ing* (see Figure 2) to be successively selected.

Accessing the word's or morpheme's phonological code is frequency dependent. When you have two different, but homophonic words, such as

skate, the ice skate, and *skate*, the fish, you get frequency summation, as Jescheniak and Levelt (1994) have shown. In other words, the low-frequency word *skate*-the-fish is accessed just as fast as the high-frequency word *skate*-the-ice-skate. Accessing the phonological code can be a major problem for anomic patients, and, occasionally, for all of us, namely when we slip into a tip-of-the-tongue state.

Let us return to our general. Upon retrieving the phonological code of his target word *defend*, the general can start the real work of phonological encoding. The core business of encoding a phonological word is *syllabi**fi**-cation*. Syllables are the units of articulation and the general will have to prepare the syllable structure of his utterance, *we'll defend it*. Let us consider the part *defend it*. According to our theory of phonological encoding, the general will encode the syllables incrementally, one by one. The general takes the first segment, /d/, and keeps adding segments till he has a first syllable. That is quickly done, /d/ and /i/ form a syllable, /di/, according to the phonology of English. Then the general starts building the next syllable by successively chaining the segments /f/, /ɛ/, and /n/ː /fɛn/. Finally, he encodes the last syllable, /d/, /I/, /t/ː /dIt/. Notice that this syllable straddles the lexical boundary between *defend* and *it*; the general doesn't say *defend-it*, but *defen-dit*. A word's syllabification is therefore not fixed. Whether /fɛnd/ or rather /fɛn/ will be a syllable of *defend* depends on the word that follows. It is, therefore, unlikely that a word's parsing into syllables is stored in our mental lexicon. It is, rather, created on the fly, dependent on the context in which the word appears. There is convincing experimental evidence that syllabification is indeed an incremental process, running from the beginning to the end of words (Meyer, 1990, 1991).

Phonetic encoding

Let us now turn to the next stage, phonetic encoding. As successive syllables are created, the general quickly turns them into the specification of successive articulatory targets. The targets for a syllable are called the syllable's articulatory score. How is this score computed? We don't know, but one notion that we have been pursuing is a claim, first put forward by Crompton (1982), that such articulatory syllable scores are stored. The repository of syllabic scores has been called the speaker's *mental syllabary* (Levelt & Wheeldon, 1994). There are various reasons why such a notion is attractive. Crompton argued from speech errors. You can also argue from statistical evidence. We do some 80 per cent of our talking with no more than 500 different syllables (Levelt et al., 1999). Applying this to the previous word count, we can estimate that on reaching adulthood, we have on average produced each of these syllables some 200,000 times, or

30 times every single day. Syllables are among the most exercised motor patterns we produce. It is precisely such high-frequent, overlearned motor patterns that, according to Rizzolatti and Gentilluci (1988), get stored in the premotor cortex.

The syllabary notion has been incorporated in the WEAVER model (Roelofs, 1997). The idea is that during phonological encoding, each phonological segment spreads activation to all syllable scores in which it partakes (see Figure 2). The speed of retrieving a syllable's articulatory score is, again, determined by Luce's ratio. At this moment, the chronometric evidence in support of this model is still incomplete.

Articulation

The final move of our general is to loudly articulate what he has been composing. The general will, at some moment, begin to execute the speech movements corresponding to the articulatory scores. We will refrain from discussing the physiology and mechanics of articulation. The one relevant issue for our present purpose is this: At what stage in the process is the general to *initiate* his articulation? There are various possibilities here. The general's utterance consists of four syllables: *we'll-de-fen-dit*. The most diligent way would be for the general to begin articulation right after the phonetic encoding of the first syllable [we'll]. The other extreme is that the general will wait till the whole utterance has been planned. Recent research by Bachoud-Lévi, Dupoux, Cohen, and Mehler (1998) and by Schriefers and Teruel (1999) suggests that articulation can be initiated as soon as the first syllable has been phonetically encoded.

Self-monitoring

One final remark on the theory: As the general is preparing and articulating his utterance, he is also monitoring his own production. He can obviously listen to his own overt speech, but also monitor some aspect of "inner speech", probably the syllabified phonological code that he composed incrementally (Wheeldon & Levelt, 1994). This monitoring system has been called "the internal monitoring loop" (Levelt, 1983). So far for our bird's eye view of the lexical access theory. It suffices for approaching some issues of multiple access.

MULTIPLE WORD ACCESS

The issue of multiple access is now easily posed. When you produce an utterance such as *the baby and the dog*, you access two different content words, *baby* and *dog*. For each of them you must run through the stages

we just considered. There are now two essential questions to be asked (see Figure 3). First, will the two processes interact or will they both run in modular fashion? In particular, will the initiation of articulation of the first word be dependent on any aspect of accessing the second word? We will call this the issue of *dependency*. Second, dependency or no dependency, how will the stages of the two access procedures be temporally aligned? The two extreme possibilities are: they run fully in parallel (except, of course, for the words' articulation), or the first accessing process is completed before the second one begins. This we will call the issue of *temporal alignment*.

These two issues have been studied in various contexts. We just mentioned the case of NP-coordination (*the baby and the dog*), but other cases that have received attention are adjective–noun constructions,

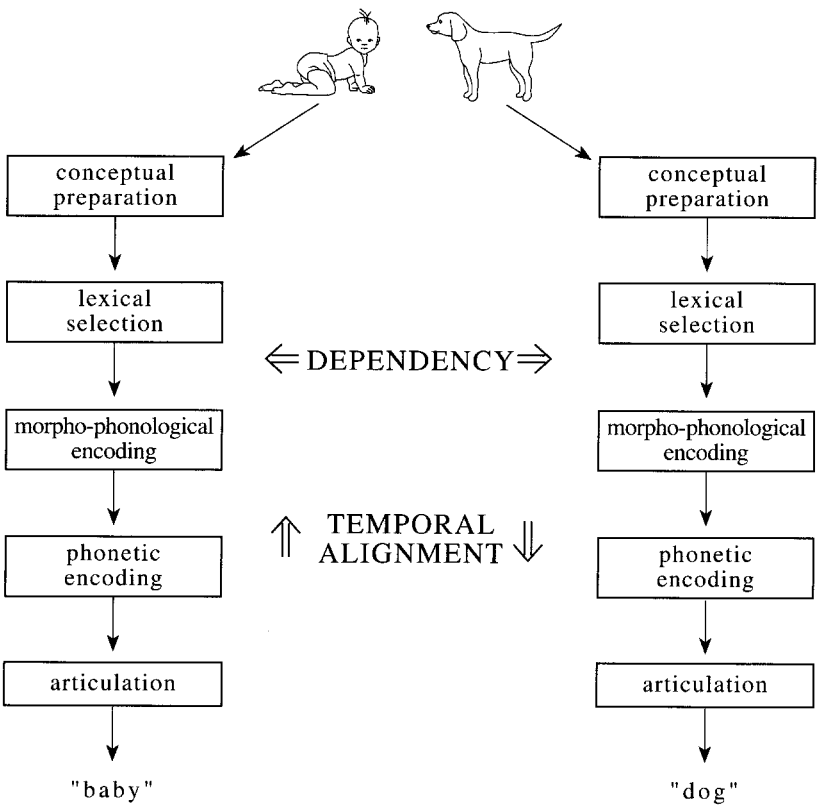

Figure 3. Two issues in multiple lexical access: Mutual dependency and temporal alignments of processing stages.

subject–verb constructions, and, recently, idiomatic constructions. We will first discuss the issue of dependency and then turn to the temporal alignment of stages.

DEPENDENCY

Adjective–noun phrases

Schriefers (1993) was the first to approach the issue of dependency. The crucial initial experiments were on noun phrase production. One central finding of these experiments concerned gender marking in the generation of adjective–noun phrases. There are two different genders in Dutch (neuter and non-neuter) that go with two different definite articles. If you produce a noun phrase like *red chair* without an article, then, in Dutch, you must mark the noun's gender on the adjective (*rodə stoel*, where the schwa-ending on the adjective marks the non-neutral gender of the noun). It is therefore a matter of linguistic necessity to access the noun lemma, which includes gender information, in order to gender mark the adjective. In other words, the dependency goes from the noun's to the article's lemma parameters—gender parameters in this case. Herbert Schriefers tested this dependency in a picture/word interference experiment. Subjects named pictures, such as a picture of a red chair, while trying to ignore an auditory distractor word. That word could be gender congruent or gender incongruent to the target word. Gender mismatch of target and distractor noun resulted in significantly longer naming latencies when the stimulus onset asynchrony was zero. This is the first well-established case of dependency. Setting the features for lemma 1 (the adjective) depends on retrieving features of lemma 2 (the noun) which was apparently delayed by the presentation of a mismatching distractor. The reason for this dependency is syntactic. It is a requirement of Dutch grammatical encoding.

Could it be the case that the Dutch speakers went as far as phonological encoding of the noun before turning to accessing the adjective? That would predict a latency effect of a distractor that is phonologically related to the noun. However, in spite of much testing, Schriefers and colleagues never found an effect of phonological noun primes on the production of adjective–noun phrases. Hence, in this case dependency seems to be limited to the lemma level.

Idiomatic expressions

A rather different case of dependency arises in the production of idioms. Recent experiments by Sprenger, Levelt, and Kempen (1999) addressed

the issue of how speakers access fixed expressions, such as *to skate on thin ice*. Fixed expressions, among them idioms, abound in our vernacular language. Rough estimates indicate that we know as many fixed expressions as single words. A theory of lexical access if not complete till we understand the mental storage and generation of fixed expressions. It is, therefore, a theoretical challenge to extend the WEAVER network theory of lexical access to include idioms. Take the example of *to skate on thin ice*. How is this idiom represented? The theoretical proposal is depicted in Figure 4. A first assumption is that the idiom relates to a specific concept, roughly meaning "to take great risks". It is a concept like any lexical concept. The only difference is that we have an expression for it in our language, not a single word. The second assumption is that you access the expression through a "superlemma". It represents the idiom's restricted syntax and points to a set of simple lemmas, among them *skate*, *thin*, and *ice* in the example. Selecting the superlemma is like selecting any other lemma. It is in competition with semantically related lemmas, such as *risk* or *gamble*; the chronometry of its selection is determined by Luce's ratio. Upon selection of the superlemma, the simple lemmas it points to, such as *skate*, *thin*, and *ice*, are selected. This selection process is special in that the superlemma can impose restrictions on the syntactic potential of the composing lemmas. For instance, *ice* is not allowed to take any other adjective than just *thin* (this can be formally realised by means of a co-indexing device).

The third assumption is that, from here on, grammatical and phonological encoding proceed standardly: The selected lemmas combine into phrases, their phonological codes are retrieved, followed by syllabification, and so on.

Fixed expressions present a curious instance of multiple access. It is no longer the case that each simple lemma is triggered by its own lexical concept. Rather, multiple lemmas are selected in concert, triggered by the idiomatic superlemma. Is this a plausible story? The theory leads to a host of predictions, of which we will mention two. First, idioms should be produced at longer latencies than corresponding non-idiomatic expressions. Compare *skating on thin ice* and *skating on smooth ice*. In the idiomatic case, lemma selection is a two-step affair. You first select the superlemma and then the simple lemmas. In the non-idiomatic or literal case, lemma selection is a one-step procedure, not involving the extra level of superlemmas.

A second prediction is that identity priming will be more effective for idioms than for non-idiomatic constructions. Imagine the following experimental situation. The subject learns a set of prompt-phrase pairs, among them *RISK—to skate on thin ice* or rather *WINTER—to skate on smooth ice*. Then, in the production task, you present the subject with the

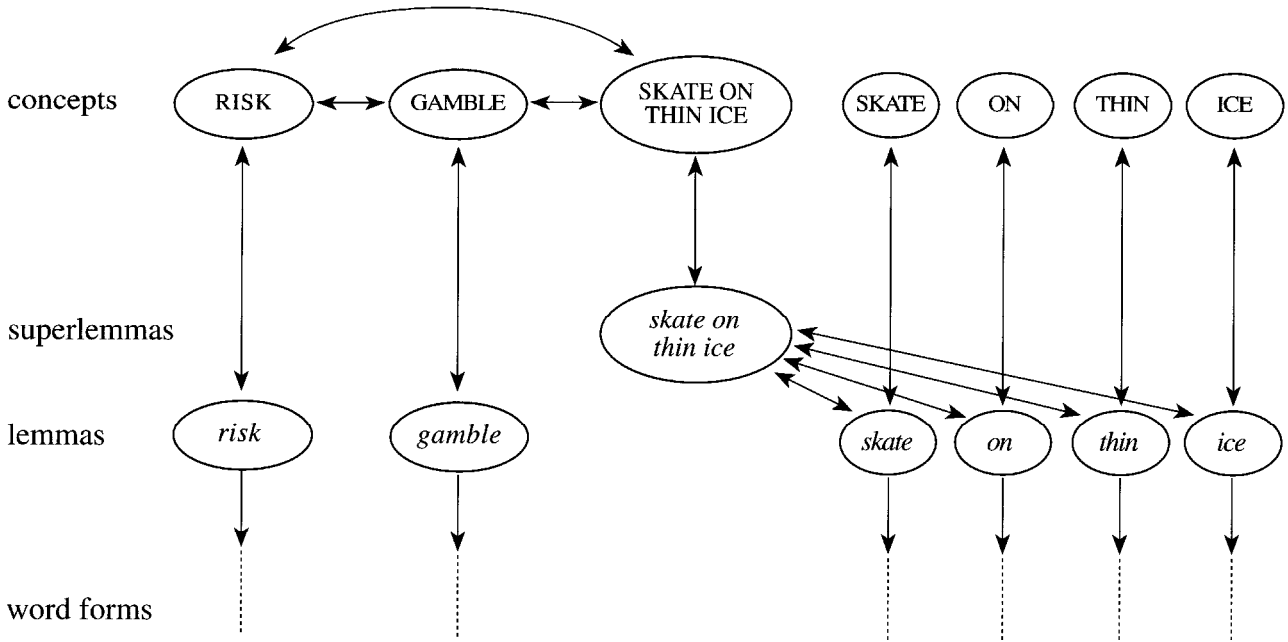

Figure 4. Representation of idiom *to skate on thin ice* as a fragment of the WEAVER model's lexical network.

cue on the screen, for instance the word *RISK*, and the subject responds with the utterance *to skate on thin ice*; the speech onset latency is measured. What will be the effect of presenting the subject with an auditory identity prime, for instance *skate* during the execution of this task? The subject sees *RISK* on the screen and simultaneously hears *skate*. In the model this auditory identity prime will trigger activation of the lemma *skate*, which in turn will activate the superlemma of the idiom, which in turn will activate all other lemmas in the idiom, speeding up their selection, given Luce's rule. If you trigger the lemma *skate* in a non-idiomatic expression, such as *to skate on smooth ice*, no such chain effect will result. You activate just one of the lemmas. Hence one should expect a stronger priming effect for idioms.

Sprenger et al. (1999) tested these two predictions in an experiment of the type just described. The Dutch idioms were roughly of the type *to skate on thin ice* and they were matched with non-idiomatic phrases, such as *to skate on smooth ice*. The identity prime was the same in the two cases, *skate* in the example. As a control, there was also an unrelated prime (such as *cat* or *house*). In a further experiment, Sprenger et al. used phonological primes, such as *scale* for *skate*. Given the theory, a phonological prime should not affect idioms and non-idioms in different ways; one should find only that idioms are slower than non-idioms and that phonological priming facilitates the production of both utterance types. The experiments confirmed all predictions. First, idioms are significantly slower to be initiated than non-idioms, for both kinds of prime. Second, identity primes are more effective for idioms than for non-idioms—we have a significant interaction here. Third, no such interaction arises for phonological primes; they are neither facilitative for idioms nor for non-idioms. These results support the theoretical account, which is that producing idioms involves a particular type of dependency in selecting the composing lemmas; they are selected in joint dependence on the selection of a superlemma.

Expressions containing multiple noun phrases

A third and final case of dependency has been reported by Meyer (1996). Subjects described pictures of two objects, such as a baby and a dog, by a coordination of two definite noun phrases: *the baby and the dog*, or—in another experiment—by a locative sentence: *the baby is next to the dog*. The results were the same for these two cases and we will combine them here. The experiments were again of the picture–word interference type. Presentation of the picture was combined with auditory presentation of a distractor word, whose onset could be simultaneous with picture onset or 150 ms earlier. The distractor could be semantically related to either the

first or the second content word (for instance *child*—related to baby, or *cow*—related to dog). Alternatively, there could be an unrelated control distractor, such as *pipe*. As discussed earlier, semantic distractors can have an inhibitory effect in single object naming. In the WEAVER model this effect is caused at the level of lemma selection. Figure 5(a) presents a summary of the experimental data, ignoring the SOAs. Both semantic distractors produced highly significant interference, that is, there was interference with both the first and the second lemma. In other words, in these experiments speakers make the initiation of their utterance dependent on having accessed both the first and the second lemma.

Or is it rather the case that they make it dependent on the encoding of both word *forms*? Of course, the first word form must be encoded before the initiation of speech, but what about the second one? To test this, the semantic distractors were replaced by phonological ones, for instance *babel* for baby and *doll* for dog. These primes were presented at four different SOAs. It has been repeatedly shown that such primes speed up access to the phonological codes, relative to phonologically unrelated primes (see Levelt et al., 1999 for a review). Figure 5(b) presents the results, averaged over utterance types and SOAs. The initiation of articulation was faster when the first content word (*baby*) was phonologically primed, but priming the second content word (*dog*) was without effect. Hence the dependency is one of accessing the second lemma only, not its word form. Different from the gender case discussed previously, there is no reason to suppose that accessing the first lemma is made dependent on accessing the second one; there is no syntactic reason for doing so. It is

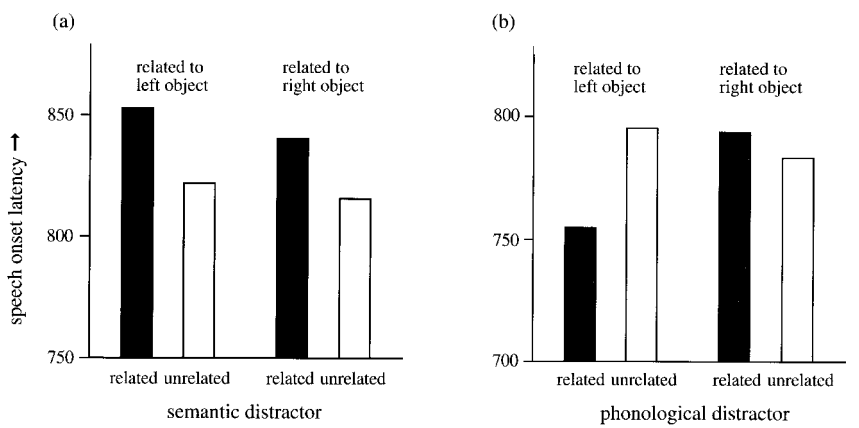

Figure 5. Naming latency effects of semantic (a) and phonological (b) distractors in double object naming.

either the phonological or phonetic encoding of the first content word, or even just the initiation of articulation that waits for the second lemma to be available.

Why would a speaker do so? In all of these experiments, but also in many real-life situations, a speaker is under two pressures. One is to respond fast, the other is to respond fluently, that is without hesitations or interruptions. It serves speed to begin articulation as soon as the first content word has been phonetically encoded. It serves fluency to check whether the crucial second lemma has been selected; that guarantees diligent phonological encoding of the second noun phrase while you are articulating the first one. This type of fluency-motivated dependency seems to be strategic in nature. In fact, in some subsequent experiments no semantic priming effect was obtained for the second noun, even when the description was as simple as *the baby and the dog* (Meyer, 1997). Subjects apparently often value speed more than fluency. We will return to that issue.

Conclusion

Cases of dependency are relatively rare. First, there are no known cases of phonological dependency, that is the initiation of articulation waiting for the phonological encoding of a later content word. Second, there are cases of lemma dependency: The initiation of articulation depends on the retrieval of two or multiple lemmas. This is necessarily the case in Dutch adjective–noun constructions, where gender marking of the adjective depends on the gender of the noun lemma. It is, apparently, also the case in the encoding of idioms; a superlemma theory of idiom access can explain this. Third, there are strategic cases where speakers hold the initiation of articulation till they have evidence that one or more later lemmas have been retrieved. This is, probably, a safeguard for fluency of delivery.

TEMPORAL ALIGNMENT

Clearly, there must be temporal alignment when there is dependency, and we discussed a few such cases. But if two or more processes of lexical access run independently, they may still overlap in time. This would, in fact, serve fluency. If we encoded each next word from scratch after completing the current word's articulation, our speech would get disconnected. This is prevented when successive access operations are telescoped to some degree. How much telescoping is done?

Eye tracking

Recently, we have been pursuing this issue by means of an eye tracking set-up. When subjects describe a scene like the one in Figure 6(a), they almost always focus the objects in left-to-right order and this is also the order of mention (*the scooter and the ball*). An important dependent variable in this project is *viewing time*. It is the duration of looking at an object. That interval begins with the first fixation on the object; it ends when the saccade to the other object is made. The other dependent variable is, of course, speech onset latency.

Out initial conception of how speakers align access procedures here was that they would attend to the left object (the scooter in Figure 6a), just long enough to recognise it. This should suffice to trigger all further encoding operations for the phrase *the scooter*. They could then shift attention to the second object (the ball in the example scene) in order to recognise it. This would allow the linguistic encoding procedures for the two noun phrases to overlap maximally, such serving fluency. In other words, we applied the principle of incrementality, which has been used so profitably in modelling speech production.

The subjects' gaze patterns can be revealing in this matter. When the subject shifts gaze from the first to the second object, we can be sure the subject shifts attention to that object. So, when will the gaze be shifted? On our initial conception, viewing time on the left object would be just long enough to do object recognition. This would be in agreement with current human performance theory. Sanders (1998), in his comprehensive treatment of human performance theory, discusses the two-object scanning case in great depth. His conclusion from the experimental evidence is that the saccade out of the left-hand object is triggered when the perceptual stage has been completed. The viewing time is uncontaminated by response selection. It is, therefore, a challenge to test whether, in our two-object naming case, the left object viewing time is affected by perceptual factors (it should) and by "response selection" factors (it should not).

Meyer, Sleiderink, and Levelt (1998) used contour deletion (see Figure 6b) as a perceptual factor. It should take longer to recognise the object in the degraded case; hence viewing time should be longer. How to affect "response selection"? We decided to use word frequency as an independent variable. As discussed earlier, accessing a word's phonological code (i.e., the "response code") is faster for high-frequency words than for low-frequency ones. Hence, pictures were chosen that had either high- or low-frequency names. (We determined for our experimental pictures that the time needed to *recognise* them was uncorrelated with word frequency; this was done in a pre-experiment where subjects made object/non-object decisions). Viewing times should not be affected by word frequency.

The data told us otherwise. As predicted, viewing times were significantly longer for degraded pictures than for full ones, by 15 ms. However, viewing times were also dependent on word frequency. Viewing times were, on average, 34 ms longer for pictures with low-frequency names

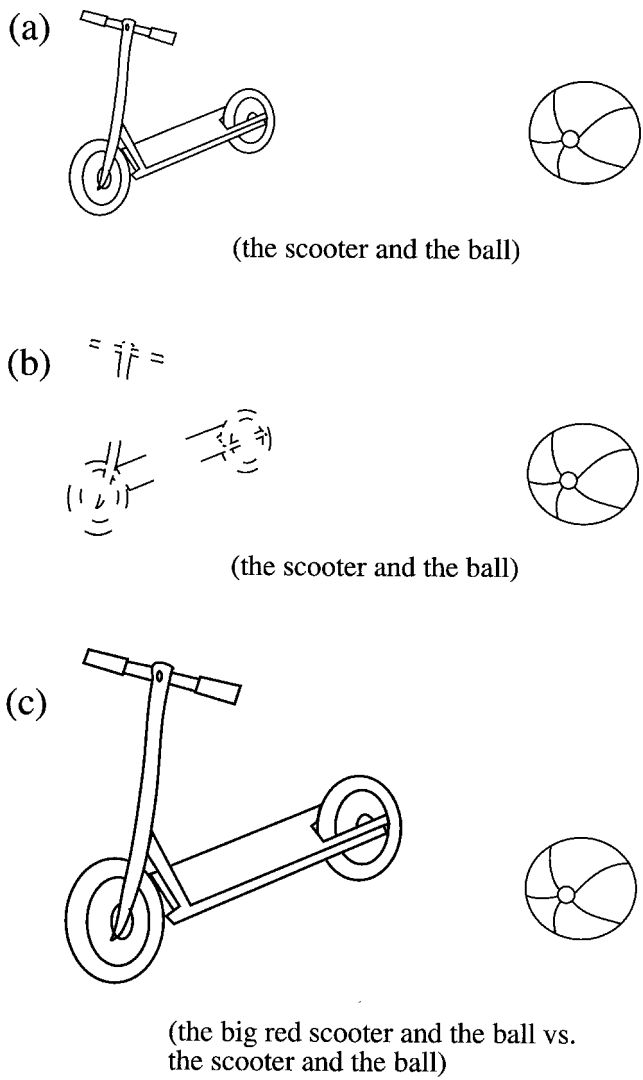

Figure 6. Some stimuli used in eye tracking experiments: (a) a two object scene; (b) visual degradation of the left object; (c) manipulation of object size.

than for pictures with high-frequency names. Hence, we had a genuine surprise. The subject apparently does not move out the gaze before at least the word's phonological code has been accessed. Our initial conception was wrong.

Before giving in, however, we tested the issue once more in a different manner. There is another way to affect the duration of phonological encoding, as we have seen, namely by presenting the subject with an auditory phonological prime. So, for instance, when the subject describes a scene by the utterance *the book and the pipe*, you can prime the phonological encoding of "book" by giving either *hook* or *boot* as an auditory stimulus, as compared to an unrelated control stimulus, such as *sail*. The prime affects selection of the phonological code for *book*. If the subject's gaze moves out before phonological encoding begins, viewing time should be insensitive to this type of auditory priming. But it is not. Meyer and van der Meulen (in press) report that, under auditory priming, viewing time on the left object is diminished by 50 ms on average. This reconfirms that the speaker's attention stays on the first object at least till the phonological code has been retrieved.

Does it await just retrieval of the code, or rather completion of phonological encoding, in particular full syllabification of the noun phrase? Earlier we mentioned that syllabification is an incremental process. The more syllables there are to be encoded, the longer it takes. Is viewing time dependent on the length of the referring expression prepared by the speaker? If so, the shift of attention not only awaits accessing the phonological code, but the full completion of phonological encoding. In order to test this, we had scenes such as the one in Figure 6(c) be described in one of two ways. The subject was either instructed to say *the scooter and the ball* or to say *the big red scooter and the ball*. Hence, the first noun phrase is either simple and short or complex and long. In the Dutch equivalent we used, they differ by four syllables. In this case, the left object viewing times turned out to be hugely different; 559 ms for the simple utterance, 1229 ms for the complex utterance—a difference of 670 ms. Still, the difference in speech onset time was small and non-significant (713 ms for the simple utterance, 755 for the complex one). It is apparently not the case that speech onset triggers the saccade to the second object. The saccade is triggered 154 ms *before* speech onset for the simple utterance, but as much as 474 ms *after* speech onset for the complex utterance. Our best guess is that the gaze shift is triggered by the completion of the phonological encoding, i.e., the syllabification of the referential phrase, *the scooter* or *the big red scooter*. That would predict that, measured from the saccade to the right object, the latency of initiating the second phrase (*and the ball*) should be the same in the simple and complex cases, and our estimates suggest that it is: about 580 ms in both cases.

These results show that, different from what we expected and from what human performance theory suggests, the time course of double lexical access is not maximally telescoped. The encoding of referent 2 begins only after the phonological encoding of reference 1 has been completed. It may even be the case that temporal overlap is not maximised, but rather minimised. Under the conditions of the present experiments, a still later gaze shift would probably have created speech dysfluencies. After the gaze shift the subjects had, on average, no more than 584 ms to go before fluently initiating the second phrase (*and the ball*), that is about 700 ms before initiating the noun phrase *the ball*. Remember that, for the simple utterance, the subjects needed 713 ms from seeing the left object to initiating the first noun phrase, *the scooter*. Hence, some 700 ms were probably needed for planning the second noun phrase (*the ball*) as well. In other words, a still later saccade to the right object would have led to dysfluency.

CONCLUSION

It is fair to say that we found less evidence for dependency and less evidence for temporal overlap than we had expected. Our speakers preferred to begin their access to the first content word without having to bother about later content words. Systematic dependency occurred when grammatical features of the later word were relevant for the encoding of the current word, as is the case for gender congruency in Dutch adjective-noun phrases. It also occurred when lemma retrieval of two (or more) words is jointly dependent on the selection of a single "superlemma" as is the case in producing an idiom. But without such linguistic ties between successive encoding operations, accessing the first content word usually proceeded without regard to later lexical material.

What purpose can be served by this strategy of shifting attention as late as possible? The main function may well be to minimise processing load. Staggering same-kind processes often goes at the expense of buffering resources. This is apparently also the case for phonological encoding processes. By spreading them thin, we keep processing load at a comfortable level. What speakers apparently do tolerate in these experiments is temporal overlap of the articulation of one item with the encoding of a following one. That is in agreement with earlier findings by Wheeldon and Levelt (1995), where phonological encoding of a word was observed to proceed under concurrent articulation of another word.

Temporal overlap of successive lexical access procedures was minimised to the level of brinkmanship. Our speakers seemed to rather risk dysfluency than having to cope with extra cognitive load or interference

in the encoding of subsequent content words. Hence, one may paradoxically conclude that a speaker's encoding of a subsequent lexical item is made highly dependent on completing the encoding of the current one. This context-dependency guarantees undisturbed, "modular" execution of each subsequent access procedure.

Manuscript received December 1999
Revised manuscript received April 2000

REFERENCES

Bachoud-Lévi, A.-C., Dupoux, E., Cohen, L., & Mehler, J. (1998). Where is the length effect? A cross-linguistic study of speech production. *Journal of Memory and Language*, *39*, 331–346.

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

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

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

Jescheniak, J.D., & Levelt, W.J.M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 824–843.

Levelt, W.J.M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104.

Levelt, W.J.M. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*, 223–232.

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

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

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

Luce, R.D. (1959). *Individual choice behavior*. New York: John Wiley.

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

Meringer, R., & Mayer, K. (1978). *Versprechen und Verlesen*. Benjamins. (Original work published 1895)

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. (1996). Lexical access in phrase and sentence production. *Journal of Memory and Language*, *35*, 477–496.

Meyer, A.S. (1997). Conceptual influences on grammatical planning units. *Language and Cognitive Processes*, *12*, 859–863.

Meyer, A.S., & van der Meulen, F.F. (in press). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychological Bulletin and Review*.

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*, 1146–1160.

Meyer, A.S., Sleiderink, A.M., & Levelt, W.J.M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*, B25–B33.

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

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

Rizzolatti, G., & Gentillucci, M. (1988). Motor and visual-motor functions in premotor cortex. In P. Rakic & W. Singer (Eds.), *Neurobiology of motor cortex*. Chichester, UK: Wiley.

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

Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition*, *47*, 59–87.

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

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

Sanders, A.F. (1998). *Elements of human performance: Reaction processes and attention in human skill*. Mahwah, NJ: Lawrence Erlbaum Associates.

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*, 841–850.

Schriefers, H., & Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. *European Journal of Cognitive Psychology*, *11*, 17–50.

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

Sprenger, S., Levelt, W.J.M., & Kempen, G. (1999). *Producing idiomatic expressions: Idiom representation and access*. Poster presented at AMLaP-99, Edinburgh.

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

Wheeldon, L.R., & Levelt, W.J.M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, *34*, 311–334.