

# 5

## The Skill of Speaking

Willem J.M. Levelt

*Max-Planck-Institut für Psycholinguistik, Nijmegen, Netherlands*

After an introductory note on the general neglect that this universal, species-specific and existential skill has met in the history of psychology, an outline is sketched of its internal organisation. Various relatively autonomous processing components interact to map a speaker's communicative intention into a fluent sequence of articulatory gestures. The chapter then proceeds to discuss each component's functioning in some detail on the hand of ongoing empirical, especially experimental, research in the Max-Planck-Institut für Psycholinguistik. The first component process discussed is the conceptual preparation of speech, the retrieval of communicatively relevant information, its linearisation and its propositional encoding. The second component process is grammatical encoding, the retrieval of appropriate words for the concepts to be expressed, and the creation of a syntactic plan. The next component deals with phonological encoding, i.e. the retrieval of phonological information for each word, and the computation of a phonetic plan for each word and the utterance as a whole. This plan is executed by the articulatory component, with overt speech as output. Finally, this output is fed back through the speaker's speech comprehension system, where it is monitored for problems of delivery. In case of necessity, the speaker will initiate a self-repair. It is shown that self-repairing is a highly systematic process.

### INTRODUCTION

The theory of how we speak has traditionally been a stepchild of psychology. Although most people spend substantial parts of the day talking, the topic of speaking has been systematically evaded in just about all traditional textbooks of psychology. And this goes back to the beginnings of scientific psychology. Of the 26 chapters in William James' *Textbook of Psychology* (1892), for

instance, not a single one deals with language or speech. In his *The Principles of Psychology* (1890) he squibs at language as something utterly un-psychological: "Language", he writes, "was originally made by men that were not psychologists". And the American tradition in psychology followed suit. In Europe, the Gestalt psychologists gave positive lip-service to the importance of speech, but factually ignored it completely. There were also, however, some towering exceptions. Both Wilhelm Wundt (1900) and Karl Bühler (1934) placed language and speech quite central in their theories of psychology, although they disagreed between them on about every detail. But even this tradition definitely died with Bühler's flight from the Nazi's in 1938. Otherwise, the study of language and speech was left to linguists, phoneticians and neurologists.

All this is very surprising because there can be no doubt that speaking is our most complex, and our most species-specific, cognitive-motor skill. Psychologists are masters in studying the trivial, and in ignoring that which makes us biologically human. But the tide has been turning recently, and the study of how we speak is becoming fashionable again among psychologists. In the following, I will report on some work from my own laboratory—the Max-Planck-Institut für Psycholinguistik—but much of it is based on insights we owe to a small number of pioneers and friends who had the guts to take pity on the stepchild, to feed and to fondle it.

## GENERAL OUTLINE OF THE SKILL

As any skill, speaking involves the interaction of several processing components. Even pre-theoretically, it is obvious that speaking at least comprises a level of intentions and ideas, a level of words and sentences, and a level of sound production or articulation. These three levels of processing have their own characteristic speeds of operation. In normal fluent speech, new elementary ideas (roughly corresponding to the occurrence of main verbs) come at the (average) speed of about one every 4–5 sec. The generation of ideas to be expressed in speech is a relatively slow strategic process. Individual words are produced at a rate of about two or three per second. This is already quite fast; retrieving words is a largely automatic process. Individual speech sounds, finally, are produced at machine gun rate, some 10–15 per second. The process is fully automatic; there is no way for the speaker to ponder about choosing an /a/ or an /o/, a /k/ or a /p/.

Figure 5.1 is a blueprint or working model of the main processing components that cooperate in the production of fluent speech. The slow strategic component is labelled *conceptualiser*. Speaking is a form of goal-directed behaviour. The speaker conceives of some communicative intention, for instance to inform the interlocutor about something, to commit the interlocutor to some course of action (as in requesting), to share irritation or elation about something, etc. In all these cases, the speaker must select information whose expression will reveal the

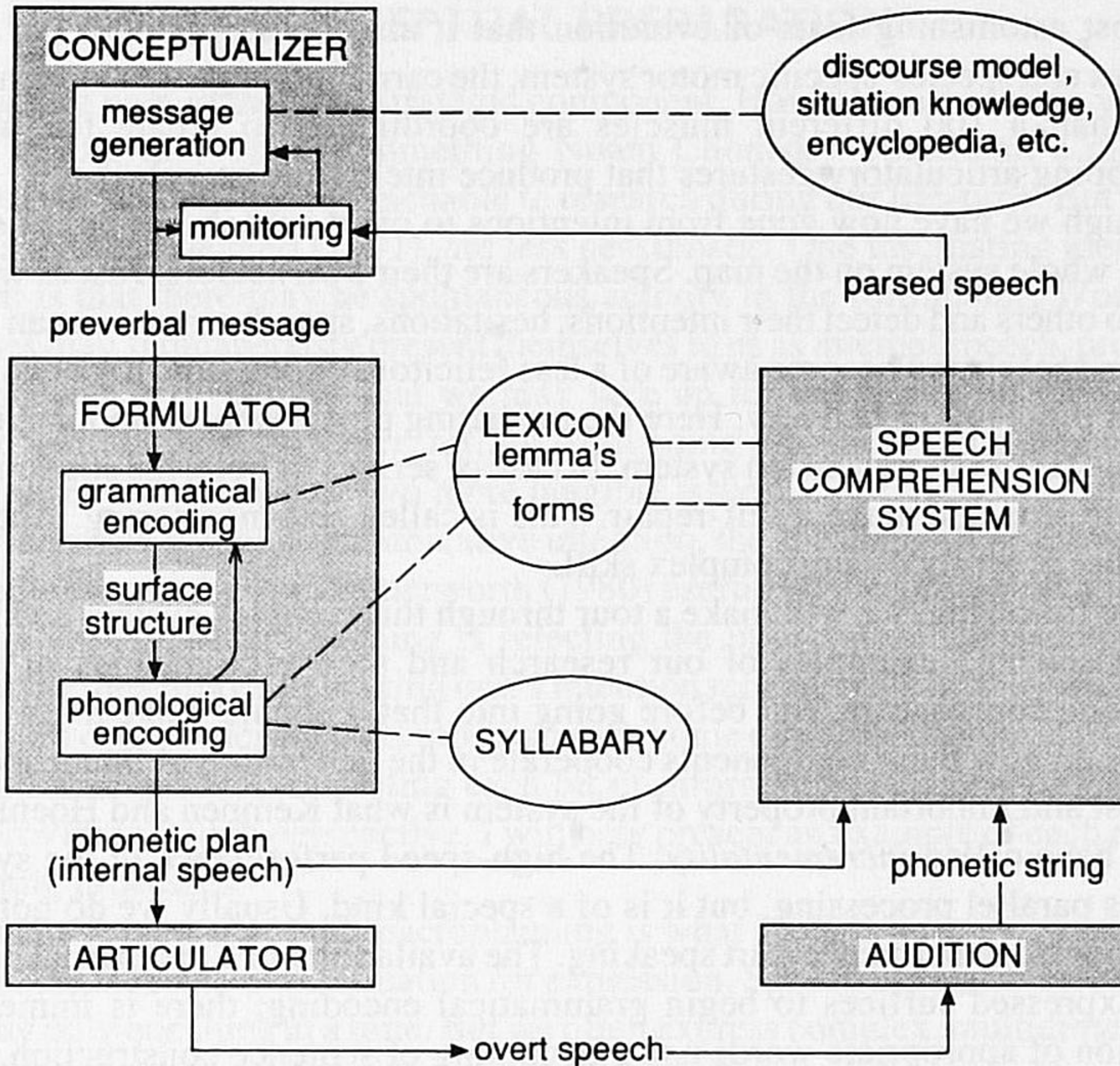


FIG. 5.1 Blueprint of the speaker. Boxes represent processing components; circle and ellipses represent knowledge stores.

speaker's intention to the partner in conversation. This information is technically called the speaker's "message". Which message will do the job depends on myriad circumstances, on what was said before in the ongoing conversation, on what the partners in speech share in terms of beliefs and expectations, on their shared perceptions, on their status relation, and so on.

But whatever the message is going to be, it has to be cast in linguistic form. This is the task of the *formulator*. In fact, it performs two quite distinct operations, the first of which is *grammatical encoding*. It consists of selecting the appropriate words from the lexicon and putting them in some appropriate syntactic order. This developing syntactic structure is called a "surface structure". The second one is to generate an articulatory or phonetic shape for all words and for the utterance as a whole. This is called *phonological encoding*. The final result is an articulatory plan in which the speech sounds to be articulated and the prosody of the utterance are fully specified. Phenomenologically, this phonetic plan presents itself to us as "internal speech".

This articulatory plan, finally, can be executed by our articulatory system. This is really our apparatus for breathing and for the ingestion of food. It is one of

the most astonishing feats of evolution that it also developed into our most complex and species-specific motor system, the carrier of language. In speaking, more than a 100 different muscles are coordinated to create the highly overlapping articulatory gestures that produce intelligible speech.

Though we have now gone from intentions to overt speech, we have not yet put the whole system on the map. Speakers are their own hearers. Just as we can listen to others and detect their intentions, hesitations, speech errors, we can parse our own speech and become aware of a less felicitous expression, a speech error or other problems of delivery. Here the mediating processing component is our own language comprehension system. In case of serious trouble, the speaker may decide to stop and make a self-repair. This is called self-monitoring. And that is another property of any complex skill.

In the following, we will make a tour through this complex system and I will present selected examples of our research and theory on various of these processing components. But before going into that, I should make a few short remarks on how these components cooperate in the generation of fluent speech.

A first and important property of the system is what Kempen and Hoenkamp (1987) have called *incrementality*. The high-speed performance of the system requires parallel processing, but it is of a special kind. Usually we do not plan the full message before we start speaking. The availability of a first partial notion to be expressed suffices to begin grammatical encoding; there is immediate activation of appropriate words and a beginning of sentence construction. And as soon as there is a sentence initial word, phonological encoding is initiated. Also, we need only one or a few syllables and overt articulation can begin. Self-monitoring can begin as soon as any stretch of internal speech has become available. This roofing tile style of parallel processing is called "incremental processing".

Parallel processing is only possible because of the *automaticity* of the various processing components. That is a second major property of the system. We spend most of our attention on *what* to say, the topic of discourse, our communicative intention. But *how* we say it largely takes care of itself.

And, finally, the system has a high degree of *modularity*. The various processing components do their own work largely independent of what the other components are doing at the same time. There is surprisingly little interaction or feedback in the system, in spite of repeated claims to the contrary—especially in the connectionist literature. We have spent major efforts on testing such claims experimentally and time and again the result has been negative (see the section on Lexical Access below). And this quite limited interactiveness makes good functional sense. The processing components have to perform wildly different tasks. If they were to affect one another during processing, the system would become highly error-prone. But it is not. Speech errors are exceedingly rare events. Modularity is nature's protection against processing failure.

## CONCEPTUAL PREPARATION

Let us turn now to the first, strategic component. How do we conceive what we want to express? This is something Noam Chomsky would call a mystery, something that will not be amenable to research during our life-time. But others, such as Daniel Dennett (1991), are less pessimistic. One fascinating idea I owe to him is that there may be spontaneous activity in the formulator. Words and phrases may spontaneously present themselves to us as internal speech, providing us with tatters of notions that we may take up for expression in the ongoing conversation. I will, however, refrain from going into these intriguing matters, and concentrate on one or two more tangible issues.

Given some sort of communicative intention, the speaker will go through two procedures, which Brian Butterworth (1980) usefully called *macroplanning* and *microplanning*. Macroplanning is selecting the information whose expression will serve the purpose of making one's intention recognisable by the interlocutor. It also involves ordering that information in some effective way for the listener. Microplanning involves casting each bit of information in propositional form, and providing it with perspective. I will now present an example of each of these planning processes.

One important aspect of macroplanning is what I called *linearisation* (Levelt, 1981), the ordering of information for expression. Speech is a linear medium; we can say only one thing at a time. But we often express complex, multidimensional information. In those cases, we have to decide what to say first, what to say next, etc. We are not very good at that. Try to explain the game of chess and you drown in the trouble of linearisation. Solving the linearisation problem means finding a path through the complex information structure to be expressed. How speakers do this we studied by means of patterns such as those in Fig. 5.2.

We asked our subjects to describe the patterns, starting at the arrow, in such a way that a listener should be able to draw them from the tape-recorded description. What we found is that speakers follow three principles in linearising such networks. The first is a principle of connectivity:

*Principle of connectivity:* Wherever possible, choose as the next node to be described one that has a direct connection to the current node.

Without exception, subjects go from dot to dot, following the connectivities in the pattern; in pattern (a) they will go from green to blue to orange to yellow to pink, never from green to yellow, skipping blue and orange. But connectivity has its limits. The stumbling block is choice nodes. In describing pattern (b), subjects again start out in a connected fashion. They may, for instance, go from blue to black to orange to grey and then to red. But all nodes they pass between blue and red are choice nodes to which they must return in order to complete the pattern's description. How do they do this? Here they follow the stack principle:

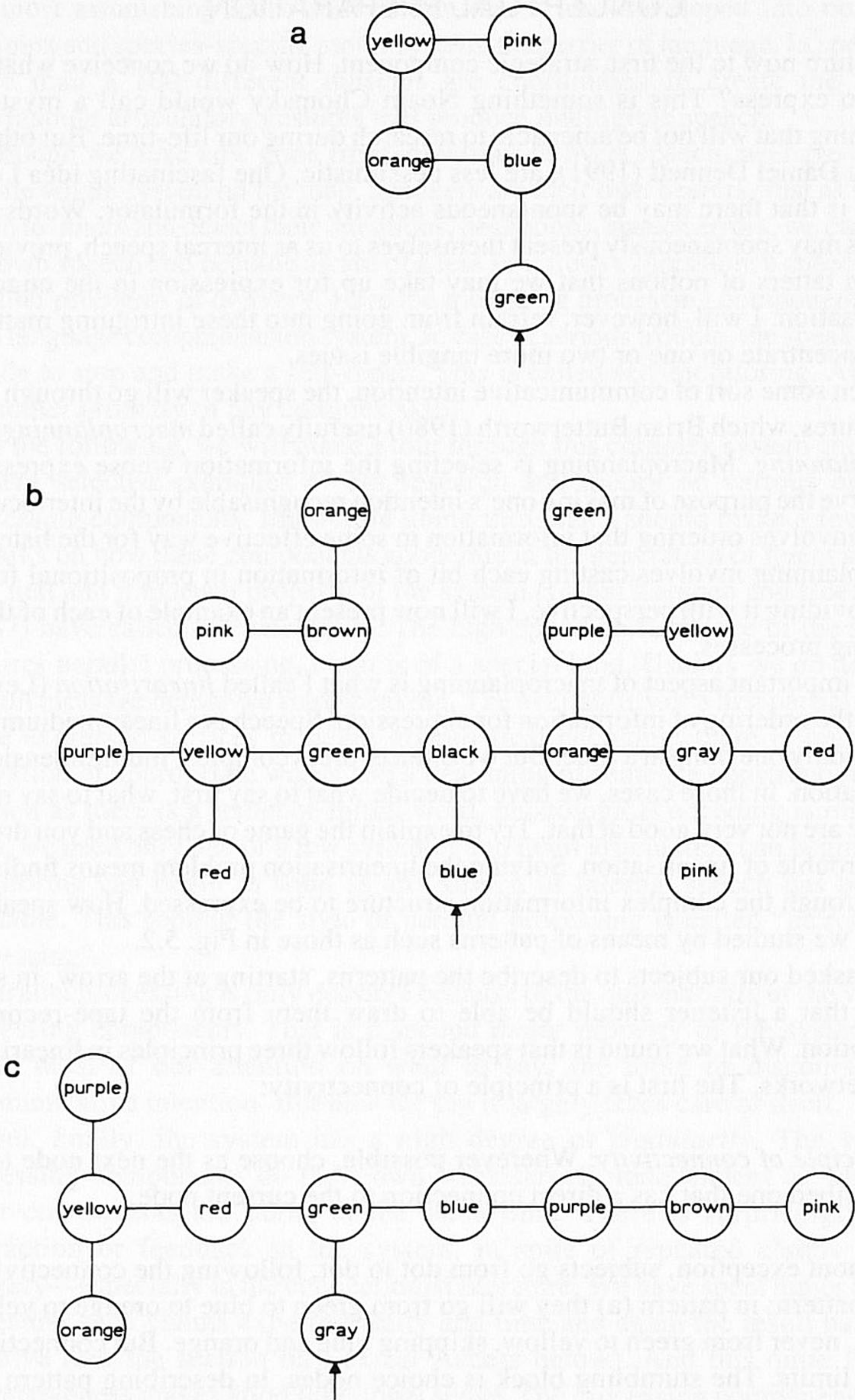


FIG. 5.2 Patterns used to study speakers' linearisation strategies.

*Stack principle:* Return to the last choice node in the waiting line.

The principle says that one should always return to the last choice node one passed. One should make a memory stack of the choice nodes passed, and return to the top node on the stack. And that is indeed what subjects do. From red they return to grey and mention the connection to pink; they then return to orange and describe the whole structure above it. They then return to the last choice node on the stack, black, and describe the left side of the pattern.

The third principle tells us which branch to take first from a choice node. Do you go left first, or right, or straight?

*Minimal load principle:* Order alternative continuations in such a way that the resulting memory load for return addresses is minimal.

The principle says "take the least memory loading branch first". From the green choice node in pattern (c), for instance, it is the branch to the right; contrary to the branch to the left, it does not contain any further choice nodes.

These three principles could be implemented in an augmented transition network with a push-down memory (Levelt, 1981). That model adequately describes what speakers do in linearising such patterns. We found that these principles also hold for quite different information structures. And that should not surprise us, as they are based on quite general principles of memory management.

Turning now to microplanning, it can be demonstrated from the same kind of examples. Microplanning involves the propositionalisation of information. Take the sub-pattern of the orange (O) and yellow (Y) nodes in pattern (a). Dependent on their direction of linearisations, speakers will say something like "above O is Y", or "below Y is O". One object (the theme) is located with respect to the other object (the referent). In doing this, the speaker tacitly takes perspective. When starting at the green node, two-thirds of the speakers indeed locate yellow *above* orange. They tell you how their gaze moves over the pattern, namely upward from the yellow node. But some speakers take a different type of perspective. From green they go "straight" to blue, then "left" to orange and "right" to yellow, as if walking or driving a car through the pattern. They make a body tour, not a gaze tour. The same visual direction, therefore, can be propositionalised in different ways (involving the concept ABOVE or the concept RIGHT OF), dependent on the tacit perspective chosen. Listeners are demonstrably confused when a speaker takes a body tour perspective instead of the more common gaze tour perspective.

The example shows that the alignment of perspective is not a trivial matter. It is, moreover, a pervasive problem in conversation, not limited to spatial discourse. When we talk, we always choose a topic and a perspective on that topic. And in some way interlocutors must manage to align their perspectives.

I will now leave this intriguing issue and turn to the next component, which is the linguistic heart of the system.

## FORMULATING

### Overview

Once the speaker has a message, or an initial part of it available, formulating begins, and it involves two phases, grammatical encoding and phonological encoding. A first step in *grammatical* encoding is the retrieval of appropriate words for the concepts to be expressed. This is a fascinating process. In normal fluent speech, we retrieve words at a rate of two or three per second. At this rate we hit the appropriate words, sure as a gun, in a huge lexicon containing several tens of thousands of items. The occasional errors of selection, such as

Don't burn your toes (*intended*: fingers)

show that the lexicon has a semantic organisation; misselections often involve semantically related items. In addition, the words that we retrieve, or *lemmas* as the technical term goes, each have their own syntactic properties. They are nouns, verbs (transitive or intransitive) or other categories, each requiring its own specific syntactic environment. Grammatical encoding is somewhat like solving a set of simultaneous equations. The syntactic requirements of all words retrieved have to be simultaneously realised (the technical term here is *unification*). The occasional syntactic misordering, such as

Seymour sliced the knife with a *salami*

shows that a word's syntactic requirements are almost always met. Nouns only exchange with nouns, verbs only with verbs, etc. Though syntactic planning is an essential part of grammatical encoding, I will not discuss it here (but see Levelt, 1989). I will rather concentrate on lexical selection, to which I will presently turn (after having introduced phonological encoding).

The second phase in formulating is *phonological* encoding. For each selected lemma, the speaker retrieves a phonological code, a sound form specified in the lexicon. It is also called a *lexeme*. And this is used to build up a phonetic program for the words and for the utterance as a whole. The occasional phonological speech error at this level, such as

With this wing I thee red

shows that there are no phonetic whole word templates. Rather, a word's phonological units—roughly *phonemes*—are retrieved one by one.

The two-phase process of selecting a lemma and computing its phonetic shape is called *lexical access*. In recent years, we have been working on a model of lexical access, the outlines of which I would like now to present.

## Lexical Access

The model is a network through which activation spreads. It is not a connectionist model, but rather a model in the tradition of Collins, Quillian and Loftus, with labelled nodes and arcs and various conditions on the gating of information. Figure 5.3 is a small fragment of the network.

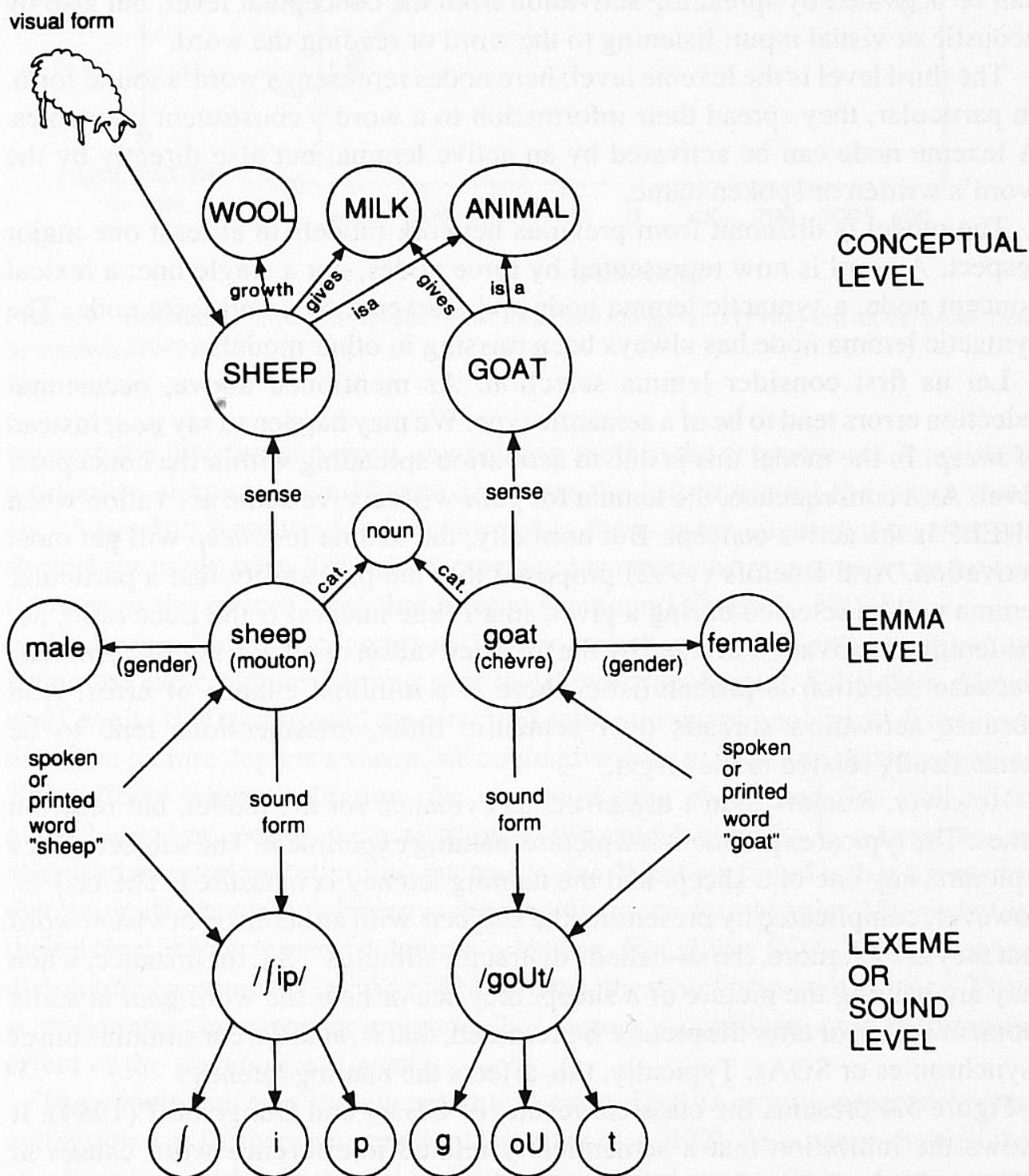


FIG. 5.3 Fragment of lexical network (from Bock & Levelt, 1994). Reproduced by permission.

There are three levels of strata to be distinguished. The top, conceptual level contains concept nodes. A concept's semantic properties are represented by its network of labelled connections to other concepts. For instance, a sheep is a domestic animal, has wool as growth, and produces milk. Conceptual nodes can be activated by various means. The sheep node can, for instance, be activated by visual information, such as a visual image of a sheep. Some concepts, such as the one for SHEEP are *lexical* concepts; they have a direct link to a node on the second level.

This second level is the lemma level. Lemma nodes have syntactic properties. *Sheep* is of the category nouns and (in French) *mouton* is of male gender. *Goat* is also a noun lemma. In French (*chèvre*) it has female gender. A lemma node can be activated by spreading activation from the conceptual level, but also by acoustic or visual input: listening to the word or reading the word.

The third level is the lexeme level; here nodes represent a word's sound form. In particular, they spread their information to a word's constituent phonemes. A lexeme node can be activated by an active lemma, but also directly by the word's written or spoken name.

The model is different from previous network models in at least one major respect. A word is now represented by three nodes, not a single one: a lexical concept node, a syntactic lemma node and a lexeme or sound form node. The syntactic lemma node has always been missing in other models.

Let us first consider lemma selection. As mentioned above, occasional selection errors tend to be of a semantic type. We may happen to say *goat* instead of *sheep*. In the model this is due to activation spreading within the conceptual level. As a consequence, the lemma for *goat* will receive some activation when SHEEP is the active concept. But normally, the lemma for *sheep* will get most activation. Ardi Roelofs (1992) proposed that the probability that a particular lemma will be selected during a given small time interval is the Luce ratio, i.e. the lemma's activation divided by the total activation of all other active lemmas. Because selection is probabilistic, there is a minimal chance of error. And because activation spreads over semantic links, misselections tend to be semantically related to the target.

However, Roelofs didn't use errors as evidence for his model, but reaction times. The typical experiment is a picture-naming experiment. The subject names a picture, say one of a sheep, and the naming latency is measured. The task is, however, complicated by presenting the subjects with an acoustic or visual word that they are to ignore, the so-called "distractor stimulus". So, for instance, when they are naming the picture of a sheep, they see or hear the word *goat* at some moment before or after the picture is presented, that is, at different stimulus onset asynchronies or SOAs. Typically, this affects the naming latency.

Figure 5.4 presents the classical results of Glaser and Dünghoff (1984). It shows the inhibition that a semantically related interference word causes at different SOAs. It also shows how Roelofs' model fits the data. You get inhibition

## PICTURE NAMING INTERFERENCE

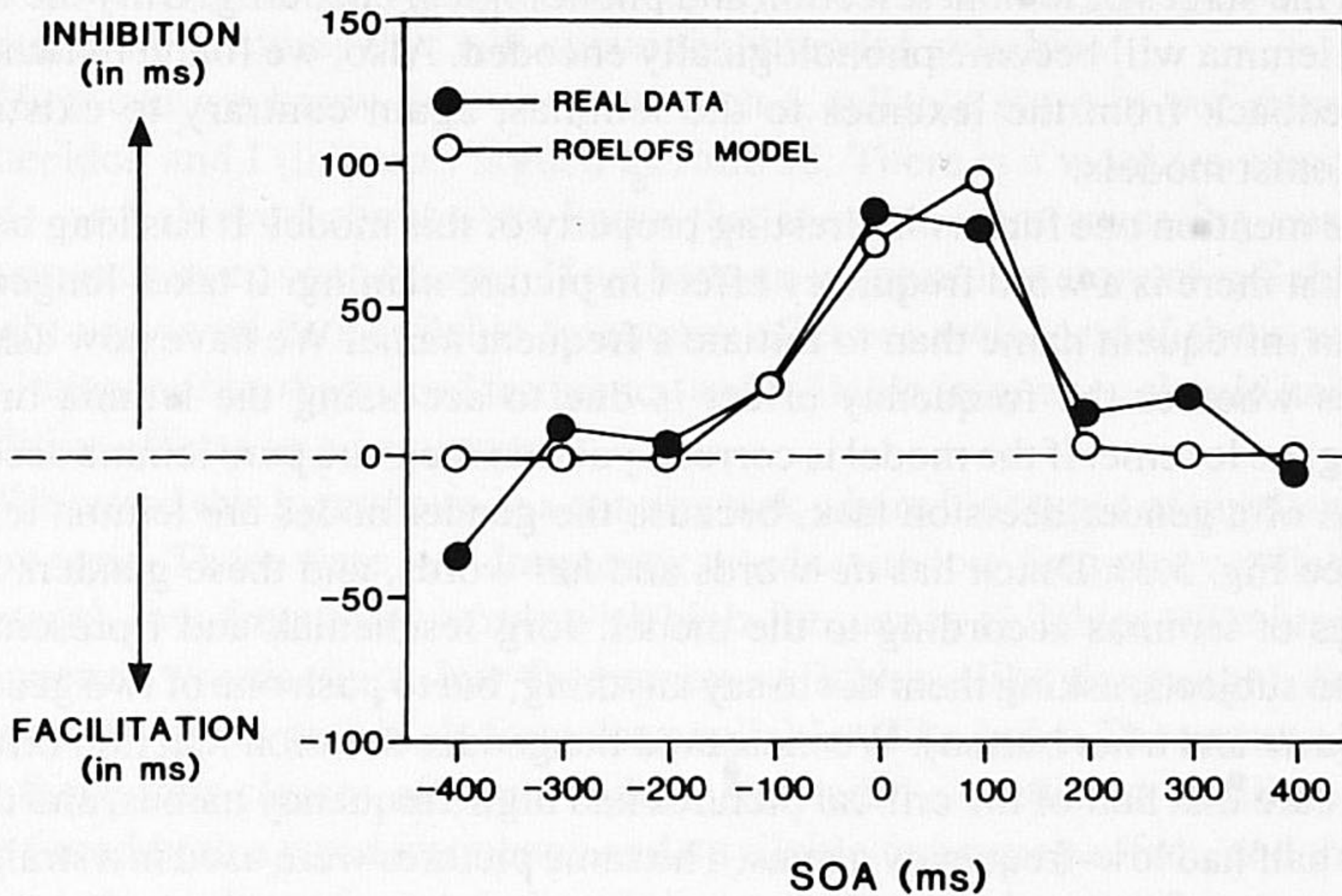


FIG. 5.4 Semantic interference results from Glaser and Döngelhoff (1984) and model predictions by Roelofs (1992).

because the interference word, say *goat*, is itself in the response set. It activates a response alternative, and hence decreases the Luce ratio for the target word, i.e. *sheep*. But Roelofs' model also makes the non-trivial prediction that you should get facilitation if the distractor word is not in the response set and when it precedes the picture. And that is what was found (Roelofs, 1992).

We further found that lemma activation strictly precedes lexeme activation. Using the same picture-naming plus interference technique, Schriefers, Meyer and Levelt (1990) compared the effects of semantic and phonological distractors. So, if the picture depicts a sheep, we could give *goat* as a spoken distractor word. This affects lemma selection. Or we could give *sheet*, and this will affect phonological encoding, the activation of the speech sounds. As a baseline, we also used an unrelated stimulus, such as *house*. What we found is that a semantic distractor affects naming latency when the distractor word begins 150 ms before the picture; it interferes with lemma selection. But at that SOA, a phonological distractor is without effect. At a SOA of 0 ms, the distractors switch roles. There is no semantic interference any more, but there is a significant (and facilitatory) effect of the phonological prime.

The conclusion that lemma activation and selection strictly precede lexeme activation was further confirmed by Levelt et al. (1992). This paper showed that an active lemma does not send any activation to its lexeme. Only when a lemma

is selected can the activation pass through to the lexeme stratum. This contradicts all published connectionist and cascading models. They all predict that activation will spread all the way down through a network. We found a strict separation between the stages of lemma selection and phonological encoding. Only the one selected lemma will become phonologically encoded. Also, we found that there is no feedback from the lexemes to the lemmas, again contrary to existing connectionist models.

Let me mention one further interesting property of this model. It has long been known that there is a word frequency effect in picture naming. It takes longer to initiate an infrequent name than to initiate a frequent name. We have now asked ourselves whether the frequency effect is due to accessing the lemma or to accessing the lexeme. If the model is correct, you can measure pure lemma access by means of a gender decision task, because the gender nodes are lemma level nodes (see Fig. 5.3). Dutch has *de*-words and *het*-words, and these genders are properties of lemmas according to the model. Jörg Jescheniak and I presented pictures to subjects, asking them not to say anything, but to push one of two gender buttons (a *de* and a *het* button). We measured the gender decision reaction times. We took care that half of the critical pictures had high-frequency names, and that the other half had low-frequency names. The same pictures were used in a straight naming task, where we measured naming latencies. After some training with the pictures and tasks, subjects produced the usual frequency effect in the naming task, but no frequency effect in the gender decision task. This shows that the frequency effect is not due to accessing the word's lemma, but to accessing its lexeme or phonological code. And indeed, a low-frequency homophone like *bank* (riverside) behaves like its high-frequency twin *bank* (financial institution) in terms of naming latency; that is because they share the same lexeme. It is, in short, the lexeme thresholds that are frequency dependent, not the lemma thresholds.

So much for lexical access. Let us now assume that we have retrieved the word's phonological form. How then is a phonetic program constructed? Here the central process is *syllabification*.

## Syllabification

When lexemes are activated, their sound units, usually phonemes, are spelled out one after another from beginning to end (see Levelt, 1992b, for discussion and references). And this is the beginning of syllabification. Syllabification is the chunking of successive speech sounds into pronounceable syllables. Syllabification ignores word boundaries. When you say *Peter gave him it*, then the three words *gave*, *him* and *it*, will cluster together, creating one so-called "phonological word". The phonological word is the domain of syllabification. Following the rules of the language, we get *ga-vi-mit* here, which violates all lexical word boundaries. This incremental syllable production follows both universal and language-specific rules (see Levelt, 1992b).

The main unanswered question now is: How does the speaker get from such composed syllables to detailed articulatory motor programs? Our hypothesis is that speakers have access to a mental *syllabary* (see Fig. 5.1). For each syllable the syllabary has an almost fully specified articulatory program—a set of articulatory gestures that will create the intended syllable.

How can we know that we have such a syllable store in our mind? Linda Wheeldon and I (in prep.) argued as follows: There is a word frequency effect, and—as I showed above—we know that it stems from accessing our store of lexemes or word sound forms. If we have an independent store of syllables, there might very well be a syllable frequency effect as well. And if these two stores are independent, then word frequency and syllable frequency should have purely additive effects in a naming task.

We tested this hypothesis in a naming task where four types of bisyllabic words were used. There were low-frequency words with low-frequency syllables (like *lantern*), low-frequency words with high-frequency syllables (like *litter*), high-frequency words with low-frequency syllables (like *language*) and high-frequency words with high-frequency syllables (like *lady*). The naming latencies for these four classes of items indeed showed the predicted additive relation: We found both a word frequency and a syllable frequency effect, and the effects were independent of one another. It is, to my knowledge, the first experimental result to testify to the existence of an independent syllabary in the speaker's mind, but the theory needs further corroboration.

Returning now to my “blueprint” of the speaker (Fig. 5.1), I have discussed some aspects of grammatical and phonological encoding, but by no means all aspects. In particular, I skipped the absolutely essential mechanisms of syntactic and prosodic planning. I will also pass the Articulator's operation, the way in which syllable programs are executed by laryngeal and supralaryngeal musculature (but see Levelt, 1989), and finally turn to the speaker's mechanisms of self-monitoring and self-repair.

## SELF-MONITORING

Speakers listen to their own internal and overt speech. When they notice trouble of some kind, they may decide to interrupt themselves and make a repair, as in the following example:

We can go straight to the ye- to the orange node

In general, a speaker stops immediately upon detecting trouble, but not every bit of trouble is detected. Remember that speakers spend most of their attention on planning what to say. Monitoring requires attention. One should, therefore, predict that the detection of trouble is best for words that appear towards the end of phrases. Those are the places where the planning of an idea is just over and

the speaker's attention can be momentarily free. And this is what I found in an analysis of spontaneous self-repairs (Levelt, 1983).

Although speakers can interrupt themselves at any place, interrupting any linguistic constituent, like in the middle of the word (see the above example), the way they restart after self-interruption is surprisingly systematic. Restarting is governed by the language's rules of coordination. Consider these two repairs:

1. Is the nurse—er—the doctor interviewing patients?
2. Is the doctor seeing—er—the doctor interviewing patients?

The first repair above we experience as well-formed, as a possible repair. The second repair, however, is odd or ill-formed. And, in fact, the latter kind of repair is quite rare in our data. But notice that the repair proper, i.e. the part after self-interruption, is the same in the two cases. Hence, the acceptability of a repair depends on something else than on the way the speaker restarts. It rather depends on the repair's relation to the interrupted stretch of speech, which is different in the two cases. How?

The first, well-formed repair (1) corresponds to an equally well-formed coordination (3). The second, ill-formed repair (2), however, corresponds to a coordination that we experience as ill-formed (4):

3. Is the nurse or the doctor interviewing patients?
4. Is the doctor seeing patients or the doctor interviewing patients?

When we repair, we keep the interrupted syntax in abeyance, and graft the repair onto it in a linguistically principled manner, just as we do in coordination.

There are many other fascinating aspects of self-monitoring that I must leave undiscussed; the reader is referred to Levelt (1989).

## CONCLUSION

We have made a tour through the main processing components involved in speaking. But it could not do justice to all the complexities of the skill at hand. What I call the "blueprint" of the speaker (Fig. 5.1) is, on the one hand, a frame of reference for summarising two or three decades of research in speaking by several colleagues and by my own Institute. But, on the other hand, it is also a theoretical framework that inspires a research programme. The reader can find more on both in Levelt (1989, 1992a).

Much more important, though, than the details of the architecture proposed should be the conclusion that the most complex and most ignored human cognitive-motor skill is not beyond scientific analysis, whether experimentally or theoretically. We can expect major progress here in the near future.

## REFERENCES

- Bock, J.K., & Levelt, W.J.M. (1994). Language production: Phonological encoding. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. New York: Academic Press.
- Bühler, K. (1934). *Sprachtheorie*. Stuttgart: Fisher.
- Butterworth, B. (1980). Evidence from pauses in speech. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk*. London: Academic Press.
- Dennett, D. (1991). *Consciousness explained*. Boston: Little, Brown and Co.
- Glaser, W.R., & Dünghoff, F.J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 640–654.
- James, W. (1890). *The principles of psychology*. New York: Henry Holt.
- James, W. (1892). *Textbook of psychology*. London: Macmillan.
- Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11, 210–258.
- Levelt, W.J.M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.
- Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W.J.M. (Ed.) (1992a). Special issue on *Lexical access in speech production*. *Cognition*, 42, 1–314.
- Levelt, W.J.M. (1992b). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42, 1–22.
- 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 naming. *Psychological Review*, 98, 122–142.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107–142.
- Schriefers, H., Meyer, A.S., & Levelt, W.J.M. (1990). Exploring the time course of lexical access in production: Picture-word interference studies. *Journal of Memory and Language*, 29, 86–102.
- Wundt, W. (1900). *Die Sprache*, 2 Vols. Leipzig: Körner.