Chapter 12

Computational Models of Lemma Retrieval

Ardi Roelofs

Introduction

Suppose a speaker wants to produce the utterance *These birds drink*. This requires, among other things, accessing in memory the words *bird* and *drink*. Psycholinguists assume that the process of lexical access consists of two major steps, called lemma retrieval and phonological encoding (e.g., Garrett, 1975; Kempen and Huijbers, 1983; Dell, 1986; Kempen and Hoenkamp, 1987; Levelt, 1989, 1992). Lemma retrieval is the process by which a message concept (e.g., BIRD) is mapped onto a lemma. A lemma is a representation of the meaning and the syntactic properties of a word. For instance, the lemma of the word *bird* specifies the conceptual conditions for the appropriate use of the word, and indicates, among other things, that the word is a noun. Importantly, lemmas do not make explicit phonological information. The retrieval of lemmas from memory is a crucial step in the process of grammatical encoding (see Chapter 11). The building of a phrasal, clausal, or sentential structure (e.g., making the noun *bird* the head of the noun phrase *these birds*) requires the syntactic information that lemmas contain. Furthermore, lemma retrieval mediates between message encoding and the second stage of lexical access, phonological encoding, both in the production of connected speech and in the production of single words. For example, in object naming, a word's lemma is used in mapping a lexical concept (e.g., BIRD) onto the appropriate word form (i.e., /bɜrd/). Phonological encoding is the process by which phonological information about a word is recovered from memory and used to compute the phonetic shape of the word (see Chapter 13).

The distinction between lemma retrieval and phonological encoding is based on a large variety of empirical findings (for an overview, see e.g., Levelt, 1989). Some classical evidence comes from errors in spontaneous speech. The distinction explains, for example, the difference in distribution between word and sound exchanges. Word exchanges such as *The tall <u>dog</u> beats the <u>boy</u>* typically involve items of the same lexical category and occur across phrasal boundaries. By contrast, sound errors such as *The ball toy beats the dog* typically do not respect lexical category and occur within a phrase. Other evidence comes from controlled speech production experiments. For example, different frequency effects are obtained from accessing lemma information (such as grammatical gender) and from accessing word-form information (Jescheniak and Levelt, 1994). Furthermore, during lexical access, lemma information is available before word-form information (Schriefers, Meyer and Levelt, 1990).

Lemma retrieval is an extremely efficient process. First, lemmas are retrieved very fast. In normal conversation, a speaker can easily retrieve up to two lemmas per second, but speeding up to five lemmas poses no difficulty (Levelt, 1989). Of course, in producing connected speech, lemmas may be retrieved in parallel. Thus, the number of words per second does not reveal the speed of retrieval of a single lemma *per se*. Studies of object naming, however, suggest that the retrieval time for a lemma can be as small as 100–150 ms (Levelt *et al.*, 1991). Second, lemmas are retrieved very accurately. If a speaker wants to express a concept, and the mental lexicon contains a lemma for the concept, then typically that lemma is retrieved and no other. In an object naming experiment, a subject accesses a wrong lemma on only a small percentage of the trials. In normal conversation, a speaker fails to retrieve the right lemma roughly once per 2,000 words (Levelt, 1989). This is a great achievement given the vastness of the mental lexicon — it is conjectured that a speaker has an active vocabulary of some 30,000 words (Levelt, 1989). The efficiency of the retrieval of lemmas poses a challenge to a model builder. Any computational model proposed for lemma retrieval should be able to explain the speed and accuracy of retrieval, referred to by Levelt and Flores d'Arcais (1987) as the criteria of *speed* and *convergence*.

Convergence in lemma retrieval cannot be taken for granted, as the following examples demonstrate. The word *parent* is called a hyperonym of *father*, and *father* is called a hyponym of *parent*. If the conceptual conditions for the application of a word such as *father* are met, then those of its hyperonyms such as *parent* are automatically satisfied as well. A father may be referred to as a *parent*. Why, then, does the retrieval process not recover both the intended word and its hyperonyms (Levelt, 1989)? Or consider words that express disjunctive concepts. For example, *sibling* means BROTHER OR SISTER. If the conceptual conditions for *sibling* are satisfied, then those of *brother* or those of *sister* are satisfied too, and vice versa. Why does the lemma retrieval process not deliver both the word and its hyponyms? As a final example, take the word *father* and the phrase ... *male parent*, which express the same underlying conceptual content. How does the retrieval process know whether a speaker wants to produce a single word instead of a phrase, or vice versa?

In this chapter, I briefly review the major computational models of lemma retrieval in the literature and highlight some important issues concerning the memory representation and the retrieval algorithm for lemmas. For instance, what is the nature of message concepts, the input representations to the retrieval process? In particular, are message concepts conceptually decomposed or not? If messages are decomposed, do lemmas contain semantic tests which, in lemma retrieval, are applied to the input representations testing for these conceptual components? If such tests are involved, are they applied in series or in parallel? It is indicated where the models stand with respect to these issues. I evaluate the models with respect to the criteria of speed and convergence, and discuss whether they are in accordance with relevant empirical findings.

As we will see, all models except one fail to meet the convergence criterion. As a consequence, the comparison between models and experimental results in this chapter is to a large extent carried out for one (i.e., the converging) model only. In this respect, the chapter is somewhat biased. However, this reflects the state of the art. At present, the best one can do is to evaluate empirically the converging model together with the suggestions that have been put forward to solve the convergence problems for the other models. Also, I restrict myself to existing models and attempted solutions — no new models and solutions are advanced. Furthermore, the chapter concentrates on the task domain of the retrieval of single lemmas, as it occurs in typical experimental tasks in language production research. I do not address such important issues as, for example, the role of conceptual context in lemma retrieval or the retrieval of lemmas in the production of connected speech (for an extensive discussion of such issues, see e.g., Levelt, 1989; Bierwisch and Schreuder, 1992; Roelofs, 1994).

Before describing the models, it is useful to go into the structure of lemmas in somewhat more detail (for an extensive description, see Levelt, 1989). As indicated, a lemma is a representation of the meaning and the syntactic properties of a word. For example, the lemmas of the words *bird* and *drink* describe the conceptual conditions for the appropriate use of these words. In addition, the lemma of *bird* indicates that this word is a noun, and the lemma of *drink* says that the word is a verb. A verb's lemma also makes explicit the word's functional structure, that is, the way it maps conceptual arguments (e.g., actor, theme) onto syntactic functions (e.g., subject, object). Typically, several morphosyntactic parameters of a lemma can be set. For a noun, there is a parameter for its number (i.e., singular and plural). For a verb, there are, in addition, parameters for the word's tense (e.g., past, present), person (i.e., first, second, third), mood (e.g., indicative), and so forth. The parameter values play an important role in phonological encoding, the construction of an articulatory program for the word. For example, in encoding the word form of *drink* with the morphosyntactic parameters set to third person, plural, and present tense, the corresponding morpheme and the speech segments /d/, /r/, /ɪ/, /ŋ/, and /k/ have to be retrieved (see Chapter 13). The morphological and phonological composition of a word is specified in the form lexicon, which can be accessed via the word's lemma. Also, the form inventory is used in accessing lemmas during spoken language comprehension (see Chapter 5). Via a specification of the orthography of the word, a lemma can be accessed in visual language comprehension (see Chapter 6). Thus, a lemma can be accessed by meaning as well as by form (for a detailed description, see Roelofs, in preparation). In language production, lemma access is based on meaning.

Major Computational Models

The process of lexical access has not received as much attention in the study of language production as it has in the study of language comprehension. Furthermore, models of lexical access in speaking primarily address the process of phonological encoding. Although typically some assumptions are made about lemma retrieval, only a few models address this process in depth. The most detailed computational models of conceptually driven lemma retrieval are discrimination nets (Goldman, 1975), decision tables (Miller and Johnson-Laird, 1976), logogens (Morton, 1969),[1] and spreading-activation networks (Dell and O'Seaghdha, 1991, 1992; Roelofs, 1992a, 1992b, 1993, in preparation). The basic idea behind these models is simple. Below, I briefly describe the models. Some of their shortcomings will immediately become clear.

Discrimination Nets

The discrimination nets proposed by Goldman (1975) are binary trees with non-terminal nodes that represent semantic tests and terminal nodes that represent lemmas. To retrieve a lemma for a message concept, semantic tests are applied to the concept, starting with the test at the root of the tree. If the concept passes the test, control moves to the left daughter node; if not, control moves to the right one. Tests are run until a terminal node is reached.

Figure 12.1: Illustration of a discrimination net

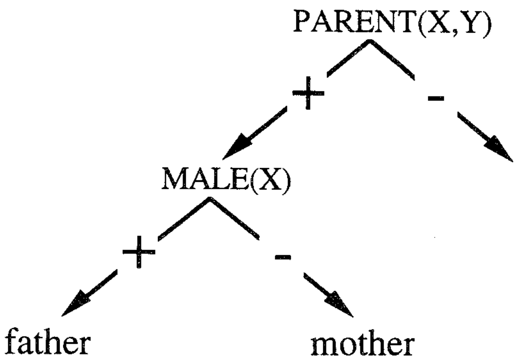


Figure 12.1 illustrates the way a binary discrimination net operates. The figure shows a discrimination net with a test for PARENT(X,Y) at the root, followed by a test for MALE(X), which leads to the lemma nodes of *father* and *mother*. Assume a speaker wants to express the concept FATHER, which is presumed to be

represented within the message by the conceptual features MALE(X) and PARENT(X,Y). First the test for PARENT(X,Y) at the root of the tree will be applied to the message concept. This test will evaluate to True (+), thus the left branch will be taken. Next, the test for MALE(X) will be applied. This test will also evaluate to True, and therefore the lemma node of *father* will be reached. Note that if the outcome of the last test had been False (−), then the lemma of *mother* would have been reached. But what if the speaker wants to express the concept PARENT? How is the lemma of *parent* represented and recovered? This poses a difficulty to this model. For someone to be a parent it is irrelevant whether that person is a male or female — both a mother and a father can be referred to as *parent*. Therefore, the MALE(X) test cannot be appropriately used for *parent*. What, then, is the test for *parent*? Perhaps, the inclusion of a feature that indicates whether gender is relevant would enable convergence to take place. Then, when the tests for PARENT(X,Y) and this feature evaluate to True, *parent* would be reached. Of course, such a proposal is *ad hoc* because for the problem of disjunctive terms and the word-to-phrase synonymy problem other solutions have to be found.

Decision Tables

Decision tables (Miller and Johnson-Laird, 1976) are access matrices where the row margins represent semantic tests, the matrix columns stand for outcome patterns on a series of such tests, and the column margins represent lemmas. An entry of a matrix can contain the value True (+), False (−), or can be blank. The latter indicates that the outcome of a test is irrelevant. Semantic tests are applied in parallel to each concept a speaker wants to express, and the pattern of outcome triggers a particular lemma.

Figure 12.2: Illustration of a decision table

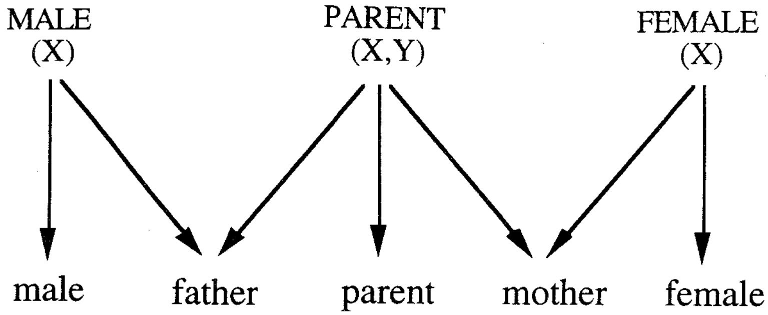
PARENT(X,Y)	+	+	+		
MALE(X)		+	-	+	-
	parent	father	mother	male	female

Figure 12.2 illustrates the way a decision table works. Horizontally, the semantic tests for PARENT(X,Y) and MALE(X) are depicted, and vertically the lemmas for *parent*, *father*, *mother*, *male*, and *female* are shown. The table specifies, for example, that triggering the lemma of *father* requires both the test for PARENT(X,Y) and the test for MALE(X) to evaluate to True (+). Note that in contrast to a binary discrimination net, the representation of the meaning of *parent* is straightforward. The lemma of *parent* has simply a blank for the outcome of the MALE(X) test, indicating the irrelevance of the result of this test. Assume again that a speaker wants to express the concept FATHER. Then, both the MALE(X) test and the PARENT(X,Y) test will evaluate to True. The outcome pattern for *father* will thus be satisfied, but the condition for the retrieval of, for example, *parent* will simultaneously be satisfied as well. The only requirement for triggering this lemma is that the PARENT(X,Y) test evaluates to True, the outcome of the MALE(X) test is irrelevant. Consequently, if a speaker wants to verbalize FATHER, both the lemma of *father* and the lemma of *parent* will be retrieved, thereby violating the convergence criterion. Thus, although decision tables can easily represent hyperonyms, they cannot cope with hyperonymy in the actual retrieval of words (Levelt, 1989). For the same reasons, they cannot handle disjunctive concepts and word-to-phrase synonymy. The major problem posed by disjunctive terms (e.g., *sibling*) is that the satisfaction of a disjunction of conceptual conditions (e.g., BROTHER OR SISTER) implies that the condition of at least one of the disjuncts is met. A decision table cannot deliver a disjunctive term without producing its hyponyms, and it cannot deliver a word that has a disjunctive hyperonym without producing the hyperonym. The problem posed by word-to-phrase synonymy is that a decision table cannot know whether to deliver *male* and *parent* instead of *father*, or vice versa, when the tests for MALE(X) and PARENT(X,Y) evaluate to True. Therefore, a decision table fails as a model of lemma retrieval.

Logogens

A logogen (Morton, 1969) is a device counting how many of a word's conceptual features are present in the message. When the count surpasses a critical threshold the logogen will fire, thereby making the word available. In lemma retrieval a set of conceptual features is switched on, and the logogen that fires first will be selected. Figure 12.3 illustrates the working of a system of logogens for lemma retrieval. The conceptual features MALE(X), PARENT(X,Y), and FEMALE(X) are depicted at the top, the logogens for the lemmas of *male*, *father*, *parent*, *mother*, and *female* are shown at the bottom. The arrows from conceptual features to logogens indicate which features are relevant to a particular logogen. For example, the conceptual features PARENT(X,Y) and MALE(X) are the features relevant to *father*. Note that to represent the values True, False, and Irrelevant, the feature

Figure 12.3: Illustration of a system of logogens and a featural spreading-activation net



FEMALE(X) is needed in addition to the feature MALE(X), because turning MALE(X) *on* or leaving it *off* cannot encode three values. In retrieving the lemma of *father*, the features MALE(X) and PARENT(X,Y) are turned on. Consequently, the logogen of *father* will exceed threshold, because all the relevant features are present. However, similar to a decision table, hyperonymy poses difficulty (Levelt, 1989). The set of features relevant to a word includes those relevant to its hyperonyms. Therefore, in switching the features MALE(X) and PARENT(X,Y) on, the logogen of *parent* will also exceed threshold and fire too (*parent* requires PARENT(X,Y) to be on and nothing else). For the same reasons, a system of logogens cannot handle word-to-phrase synonymy, nor can it cope with disjunctive concepts. Thus, similar to a decision table, a system of logogens cannot explain how the retrieval process converges onto the appropriate words in memory.

Spreading-Activation Nets

Spreading-activation networks come in two varieties: featural nets and non-decompositional nets. Featural nets have much in common with logogens. In such a net, conceptual feature nodes are directly connected to lemma nodes (cf., Dell and O'Seaghdha, 1991, 1992).[2] By contrast, in a non-decompositional net, conceptual features are indirectly connected to lemma nodes, via independent nodes for lexical concepts (Collins and Loftus, 1975; Roelofs, 1992a, 1993). Lemma retrieval begins by activating the set of conceptual feature nodes (in a featural net) or the lexical-concept node (in a non-decompositional net) making up the to-be-verbalized concept. The activation is not all-or-none (as was the case in the logogen model), but is real valued. Nodes then spread a proportion of their activation towards the associated lemma nodes. This proportion is the same for all nodes, that is, there is a general spreading rate. Finally, after a certain period of time (e.g., depending on the speaking rate), the highest activated lemma node (or the first node exceeding a response threshold) is selected.

Featural Spreading-Activation Networks

I illustrate the working of a featural spreading-activation net by using Figure 12.3 again. The top layer represents conceptual feature nodes, the bottom layer consists of lemma nodes, and the arrows indicate the connections between the feature nodes and the lemma nodes. For example, the nodes for PARENT(X,Y) and MALE(X) are connected to the lemma node of *father*. The connections are usually taken to be bi-directional (i.e., from features to lemmas and from lemmas back to features or 'interactive') so that the network can also be used for language comprehension (cf., Dell and O'Seaghdha, 1991, 1992).

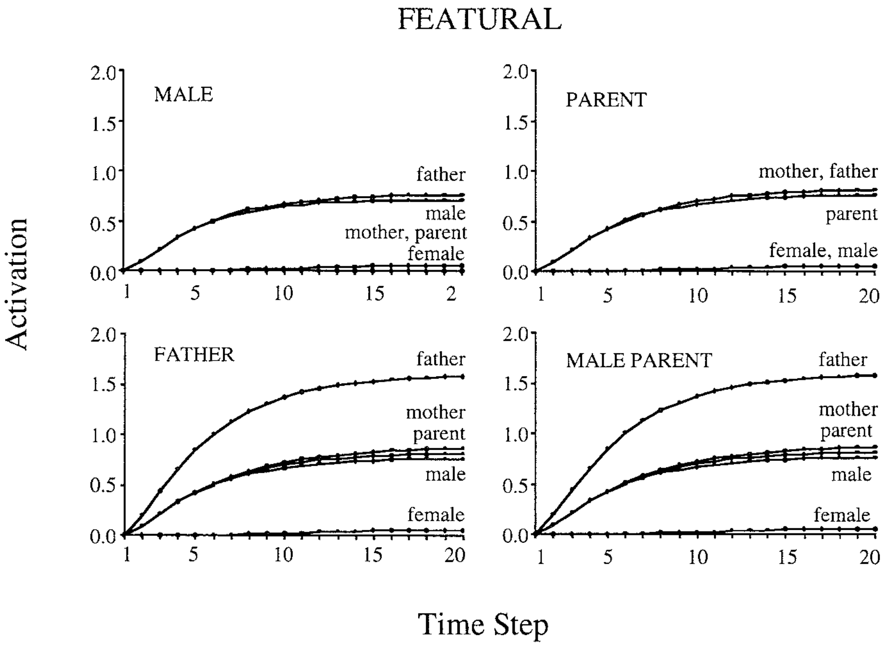
In verbalizing the concept FATHER, the feature nodes MALE(X) and PARENT(X,Y) are activated. The activation then spreads towards the lemma nodes (and back again), activating the lemma node of *father*, among other things. Often, activation is assumed to spread according to

$$a(m,s+1) = a(m,s)(1-d) + \sum_n r\, a(n,s) \qquad (1)$$

where $a(m,s)$ is the activation level of node m at time step s, and d is a decay rate with $0 < d < 1$ (e.g., Dell and O'Seaghdha, 1991, 1992; Roelofs, 1992a, 1993). The rightmost term denotes the amount of activation m receives between s and $s+1$, where $a(n,s)$ is the output of neighbour n (equal to its level of activation). The factor r indicates the spreading rate. Figure 12.4 illustrates the behaviour of the network by showing activation curves for the lemma nodes. The curves were obtained by using the spreading equation with parameter values (taken from Dell and O'Seaghdha, 1991, 1992) $d = 0.4$ and $r = 0.1$, and an external input to the feature nodes of size 1.0.

In contrast to a decision table and a system of logogens, there is no hyperonym problem, as shown by the activation plots of Figure 12.4. In activating the conceptual features MALE(X) and PARENT(X,Y) to express FATHER, the lemma of the word *parent* will not be selected. Although the lemma nodes of the hyperonyms of *father* such as *parent* will be activated, they will not be activated as high as the node of *father*. Both conceptual features for FATHER (i.e., MALE(X) and PARENT(X,Y)) will send a proportion of their activation to *father*, but only one of these features (i.e., PARENT(X,Y)) will activate *parent*. Thus, *father* will have a higher level of activation than its hyperonyms. According to this model, lemmas are selected on the basis of their level of activation (e.g., Dell, 1986). The retrieval process suffers, however, from a hyponym problem, because all features that activate a word will also activate the word's hyponyms. As can be seen, the network produces incorrectly *mother* and *father* for PARENT, and *father* for MALE. In verbalizing PARENT, *mother* and *father* will end up with a higher level of activation than *parent* due to a reverberation of activation from FEMALE(X) and MALE(X) via the backward links. The same holds for the verbalization of MALE, where the level of activation of *father* exceeds that of *male* due to the feedback of activation from PARENT(X,Y). Finally, consider the situation with two message concepts, MALE and PARENT, requiring that two lemma nodes are selected to

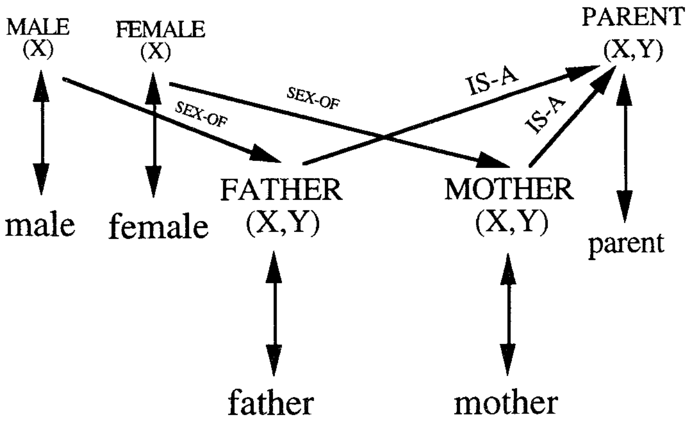
Figure 12.4: Activation curves for the interactive featural spreading-activation net



produce the phrase ... *male parent*. Figure 12.4 shows that the system will incorrectly select the lemmas of *father* and *mother*, since they are the two lemmas with the highest activation.

Of course, there is nothing in the featural approach that requires a general spreading rate *r* (i.e., equal weights on the links between feature nodes and lemma nodes). Thus, perhaps the hyponym problem can be solved by abandoning the assumption of a general spreading rate. Instead, one could put appropriately tuned weights on the links between conceptual feature nodes and lemma nodes. To prevent the problem, the weights have to be set (e.g., during a learning process, see Chapter 3) such that in activating the conceptual feature nodes of a word, its lemma node will receive more activation than the nodes of the word's hyperonyms and hyponyms. This ought to hold for all words in the lexical network. However, tuning weights is insufficient to solve all convergence problems. Again, the problem of disjunctive terms remains unresolved. Also, the word-to-phrase synonymy problem cannot be handled by the same mechanism. For example, if MALE(X) and PARENT(X,Y) are connected to the lemma nodes of *father*, *male*, and *parent* such that activating these components can retrieve *father* as well as *male* and *parent*, then activating the components will retrieve all these terms. To solve this problem, one might perhaps assume that a single verbalization and building the phrase are accomplished on the basis of different patterns of activation over the feature nodes. That is, by assigning different activation patterns to MALE PARENT and FATHER. Alternatively, one might assume

Figure 12.5: Illustration of a non-decompositional spreading-activation net



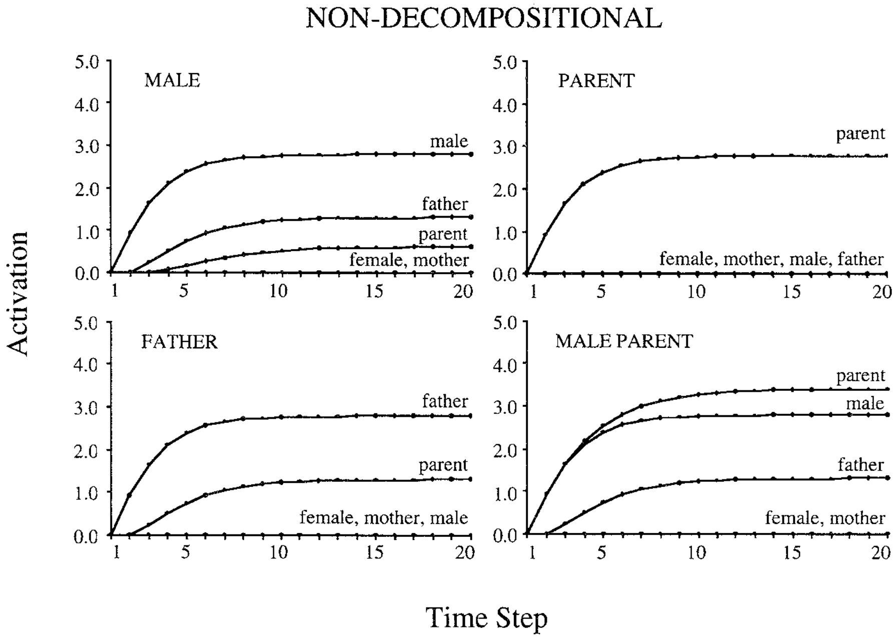
that retrieving *male parent* without accessing *father*, and vice versa, is achieved by nodes that stand for combinations of components (i.e., a node for MALE PARENT). Of course, both solutions would amount to giving up decomposition. To conclude, a general solution to the convergence problems requires more than tuning weights.

Non-Decompositional Spreading-Activation Networks

Figure 12.5 illustrates how a non-decompositional spreading-activation net works (Roelofs, 1992a, 1993). In contrast to a featural network, each lexical concept (e.g., FATHER, MOTHER, PARENT) is represented by an independent node in the net. The conceptual properties of the concept FATHER are specified by labelled links (pointers) between its lexical concept node FATHER(X,Y) and other concept nodes in the network. For example, FATHER(X,Y) is linked to PARENT(X,Y) by an IS-A link, and MALE(X) is connected to FATHER(X,Y) by a SEX-OF link. In expressing the concept FATHER, the activation level of the lexical-concept node FATHER(X,Y) is enhanced. Thus, the crucial difference with a featural network is that lemma retrieval starts with the activation of FATHER(X,Y), and not with activation of the features MALE(X) and PARENT(X,Y). These features provide background information about FATHER, but are not directly engaged in the lemma retrieval process.

Figure 12.6 illustrates the behaviour of the network by showing the activation curves for the lemma nodes in the verbalization of FATHER, PARENT, MALE, and MALE PARENT. Similar curves are obtained with other connectivity patterns at the conceptual level. The parameter values (taken from Roelofs, 1992a, 1993) were $d = 0.60$, $r = 0.25$ for the conceptual connections, and $r = 0.19$ for the connections between lexical-concept nodes and lemma nodes. The external input to the lexical-concept nodes was of size 4.91. Consider the verbalization of FATHER. After activating FATHER(X,Y), activation will spread towards the lemma nodes, and the first node whose activation level exceeds that of the other nodes by some critical

Figure 12.6: Activation curves for the non-decompositional spreading-activation net



amount (e.g., 1.5) is selected. As can be seen in Figure 12.6, the first node reaching the response threshold will be the node of *father*. Although the lemma node of *parent* will be co-activated due to the IS-A link between FATHER(X,Y) and PARENT(X,Y), the *parent* node will receive only a proportion of a proportion of the activation of FATHER(X,Y), whereas the lemma node of *father* will get a full proportion. For the same reasons, the process correctly delivers *parent* for PARENT, *male* for MALE, and *male* and *parent* instead of *father* when MALE and PARENT are the intended concepts (and MALE(X) and PARENT(X,Y) are made part of the message). This example demonstrates that a non-decompositional spreading-activation net can cope with the hyperonym problem, the hyponym problem, and the word-to-phrase synonymy problem (for mathematical proof, see Roelofs, 1994).

The decompositional models suffer from the convergence problems in lemma retrieval because the set of conceptual features of a concept (e.g., BIRD, FATHER) contains that of its superordinates (e.g., ANIMAL, PARENT) as a proper part. The non-decompositional model seems to be confronted with a convergence problem too, albeit in message encoding rather than in lemma retrieval. In a non-decompositional approach, set inclusion may be prevented by adhering to some form of economy in organizing the conceptual network. In particular, there should be no full redundancy in the storage of features with concepts in memory.

Then, some properties shared by a concept and its superordinate are not stored with the concept but only with the superordinate. The concept inherits these properties from its superordinate via an IS-A link (cf., Collins and Loftus, 1975). In this way, set inclusion is prevented and convergence in the selection of lexical concepts in message encoding may be achieved (for details, see Roelofs, 1992b).

Evaluation of the Models

The models of lemma retrieval can be characterized along a number of dimensions. A first dimension concerns the nature of the input to the retrieval process. As indicated, with respect to the input, the models take either a *decompositional* or a *non-decompositional* stance. A non-decompositional approach assumes that the lemma of a word such as *father* is retrieved on the basis of the abstract representation FATHER(X,Y). For a lemma of a word in the language there is a corresponding element in the vocabulary for the specification of messages (cf., Fodor *et al.*, 1980; Roelofs, 1994). By contrast, decompositional models assume that the lemma of a semantically complex word (i.e., a word whose meaning can be further analyzed into more elementary concepts) is retrieved on the basis of a number of primitive concepts. For example, the lemma of *father* is retrieved on the basis of MALE(X) and PARENT(X,Y). Thus, each element for the specification of messages is typically linked to several lemmas (cf., Bierwisch and Schreuder, 1992). PARENT(X,Y) is associated to both *father* and *parent*. All models described above are decompositional except the one proposed by Roelofs (1992a, 1993), in which lemmas are accessed from independent lexical-concept nodes in a lexical network by means of the spreading of activation.

The models can be further classified on the basis of whether they distinguish between *conceptual representations* on the one hand and *semantic representations* (word meanings) on the other (cf., Jackendoff, 1987). Models that make this distinction assume that the input representation (i.e., the message concept) and the semantic part of a lemma are mentally realized by separate, independent representations. This holds for discrimination nets and decision tables. Here, lemmas contain semantic tests (making up a semantic representation) to be applied to the message concept (making up a conceptual representation). By contrast, in a logogen system and in a spreading-activation net, the message concept and the word meaning are embodied by one and the same representation. The representations of the conceptual features or the lexical-concept node make up both the message concept *and* the word meaning.

Discrimination nets and decision tables are *symbolic* models (for a definition, see Chapter 2). They involve mental operations (i.e., the semantic tests) on structured mental representations (i.e., the conceptual structures making up the messages). Logogens are *connectionistic* in that no structure sensitive operation is involved but only a process of counting (cf., Chapter 3). That is, they involve the accumulation of discrete inputs until a selection threshold is exceeded. The same holds for the featural spreading-activation model of Dell and O'Seaghdha (1991, 1992), except that the accumulation involves real-valued inputs. The general framework adopted by Dell (1986), however, is a *hybrid* one. For example, in grammatical encoding a selected lemma is associated to the first empty slot of a syntactic frame. This is clearly a structure sensitive process, involving such notions as *first*, *slot*, and *frame*. The non-decompositional spreading-activation model of Roelofs (1992a, 1993) is also a hybrid one for the same reasons. In addition, in that model the links between the nodes are labelled, following a computational tradition starting with the work of Quillian in the mid-1960s (for a review, see Collins and Loftus, 1975). The lexical network is made up by structured representations, such as IS-A(FATHER, PARENT) and LEXICAL_CATEGORY(father, NOUN). The labels on the links do not play a direct role in lemma retrieval, but are relevant for grammatical encoding processes. For example, in selecting a noun as opposed to a verb in the construction of a phrase, the formulation processes have to access the information specified by the category link and node. Also, the labels play a role in the retrieval of conceptual information from semantic memory in language comprehension and in message encoding (e.g., Collins and Loftus, 1975).

In the logogen model and the featural spreading-activation model, lemma nodes are *local* computational units. Each lemma node represents a single lemma. A lexical concept such as FATHER is represented by its component conceptual features, in the connectionist literature often referred to as *microfeatures* and in the symbolic literature as *semantic markers*. Each microfeature or marker is assumed to be represented by a single node in the network. Thus, similar to the lemma nodes, conceptual features are represented in a local way. Instead, the conceptual features may be represented in a *distributed* manner (see Chapter 3). Then, each node would be involved in the representation of several conceptual features, and each conceptual feature would be represented by several nodes. A conceptual feature is then represented by an activation pattern over a set of nodes (i.e., an activation vector). The unit of computation of the system at the representational level is the activation pattern.

Finally, the models differ in whether they assume that semantic tests are applied (or conceptual features are activated) in *series* or in *parallel*. The discrimination nets assume a serial ordering of the tests. Each test is applied to the message concept after the test has received the control of its immediate predecessor. The decision tables presume parallel testing. Similarly, in a logogen system and in a featural spreading-activation net, conceptual features will be activated simultaneously.

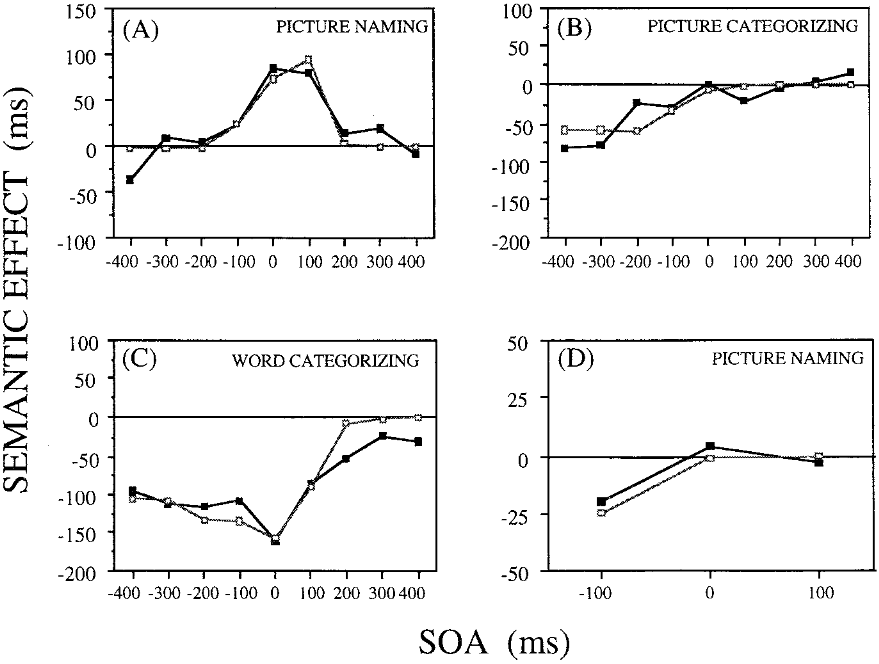
The speed of access makes a serial accessing mechanism an unattractive option (Levelt, 1989). In a discrimination net, the lemma access time will be equal to the sum of the latencies of the semantic tests leading to the lemma. Thus, the speed depends on the number of tests. A much more attractive option is a parallel accessing mechanism such as a decision table. Here, the slowest semantic test determines the retrieval speed. The lemma retrieval latency will be equal to the maximum of the latencies of the tests to be satisfied for the lemma. In case each lexical concept is linked to a single lemma, the retrieval latency will be short too.

Empirical Tests of the Models: Retrieval Speed

At first sight, one can derive strong empirical predictions from the models about the speed of lemma retrieval (for an extensive discussion of latency predictions from models, see McGill, 1963; Townsend and Ashby, 1983; Luce, 1986). For example, if the tests for a word are a proper superset of the tests for the word's hyperonyms, and the time to retrieve a lemma equals the sum of the latencies of the semantic tests, then words should be retrieved more slowly than their hyperonyms. Unfortunately, however, the failure of the decompositional models to handle hyperonymy makes such empirical tests problematic in that the retrieval of the correct word on an experimental trial is a problem for these models right from the start.

By contrast, the non-decompositional spreading-activation model does not suffer from the convergence problems. Consequently, it has been possible to put the model to a number of empirical tests. In testing this model, the so-called picture-word interference paradigm was used (for a review of this paradigm, see Glaser, 1992). In recent years, this experimental paradigm has become one of the major tools in the study of lexical access in language production (e.g., Schriefers, Meyer and Levelt, 1990). Subjects have to name pictured objects (e.g., they have to say *bird* to a pictured bird) while trying to ignore so-called distractor words. A distractor word is a written word superimposed on the picture or a spoken word presented via headphones. The speed of retrieval of the target lemma is affected depending on, among other things, the temporal relationship (the stimulus onset asynchrony or SOA) and the content relationship between picture and word. Usually, the distractor is presented just before (e.g., −400, −300, −200, or −100 ms), simultaneously with, or right after (e.g., +100, +200, +300, or +400 ms) picture onset, and is either semantically related to the pictured object (e.g., distractor *fish* in a pictured bird) or semantically unrelated (e.g., distractor *shoe*). Alternatively, the subjects are asked to refer to the picture or to the word printed in the picture by producing a hyperonym — called, respectively, picture categorization and word categorization. For example, they have to say *animal* to a depicted bird, while trying to ignore the word printed in the picture. Or they have to say *animal* to the word *bird* and ignore the picture. Typically, one observes semantic inhibition (i.e., naming latencies are slower for the related condition than for the unrelated one) at SOAs between −100 and +100 ms for picture naming, semantic facilitation (i.e., naming latencies are faster for the related condition than for the unrelated one) at SOAs between −400 and −100 ms for picture categorization, and even more semantic facilitation at SOAs between −400 and +200 ms for word categorization. The panels A, B, and C of Figure

Figure 12.7: SOA functions of the semantic effect from distractors in picture naming, picture categorization, and word categorization. Real data (filled squares) and the predictions by the non-decompositional spreading-activation model (empty squares). [Real data in (A), (B), and (C) are from Glaser and Düngelhoff, 1984; real data in (D) are from Roelofs, 1992]. A positive value means an inhibitory effect from a distractor word, and a negative value indicates a facilitatory effect



12.7 show the empirical SOA curves of the semantic effect for picture naming, picture categorization, and word categorization obtained by Glaser and Düngelhoff (1984). The panels depict for each SOA (in ms) the mean latency in the semantically related condition minus that in the unrelated condition (in ms). Thus, a positive value indicates semantic inhibition, and a negative value indicates semantic facilitation.

The panels also show the SOA curves for the three tasks as derived from the non-decompositional spreading-activation model by computer simulation. The simulations used the activation equation (with the parameter values) given above and networks consisting of lemma nodes and lexical-concepts nodes as shown in Figure 12.5. There were two different semantic fields, each typically consisting of a superordinate (e.g., ANIMAL(X)) and two subordinates (e.g., BIRD(X) and FISH(X)), which were connected to each other and to their lemma nodes in a bi-directional manner. The presentation of a picture and a word was simulated by assigning external activation to the corresponding concept node and lemma node, respectively, representing the output from picture and word perception. The external input was provided using the appropriate SOA, with the duration of a time step taken to be 25 ms. The probability of actual selection of the target node in any time step was assumed to be equal to a ratio. The numerator of the ratio was formed by the activation level of the target node. The denominator was equal to the sum of the activation levels of the target and all the other lemma nodes. On the basis of the spreading equation and the selection probability, the mathematical expectation of the retrieval time of the target lemma was computed for each SOA and distractor type. Figure 12.7 depicts for each SOA the theoretically expected retrieval time in the semantically related condition minus that in the unrelated condition. For details of the simulations, I refer to the original publications (i.e., Roelofs, 1992a, 1993). As can be seen in Figure 12.7 (and has been shown by a statistical χ^2 measure of fit), the fit of the model to the data is good.

The non-decompositional spreading-activation model explains the semantic effects as follows. If *fish* is superimposed as distractor on a pictured bird, the activation from the picture and from the distractor word will converge in the network on the lemma node of the distractor word *fish*. This is because BIRD(X) and FISH(X) are connected in the network through links which are bi-directional (unlike in the *father* example). If *shoe* is superimposed, there will be no such convergence of activation — BIRD(X) and SHOE(X) are not connected. As a result, *fish* will be a stronger competitor to the target lemma than *shoe*. This is what the SOA curve in panel A of Figure 12.7 shows. According to the model, distractor lemmas will only be competitors if they are potential responses in the experiment (this was the case in the picture-naming experiment just described). Thus, in a categorization experiment where subjects respond by using hyperonyms such as *animal*, the distractors *fish* and *shoe* should be no competitors because they are no longer potential responses. Instead, a facilitatory effect from *fish* should be obtained, because *fish* will prime the lemma of *animal* via the conceptual links in the network, whereas *shoe* will not. This prediction is supported by the experimental data, as shown in panel B and C of Figure 12.7. A prediction along the same line is that distractors such as the hyperonym *animal*, the cohyponym *fish*, and the hyponym *robin* should facilitate the naming response (*bird*) if the pictured bird is the only animal to be named in the experiment. Then, the distractors are not permitted responses, and therefore they will not be a competitor to the target lemma. Thus, they should lead to facilitation if they are semantically related, as we saw for the categorization experiments. This prediction is confirmed by the empirical data, shown in panel D of Figure 12.7. The panel depicts the mean semantic effect per SOA for distractor words such as *animal*, *fish*, and *robin* (there was no difference in effect between hyperonym, cohyponym, and hyponym distractors). To conclude, the non-decompositional spreading-activation model gives a good account for several important findings on the speed of lemma retrieval. Importantly, these findings were obtained with a number of different experimental tasks, which cross-validates the model.

Recently, Bierwisch and Schreuder (1992) suggested that the hyperonym problem may be solved within a decompositional model by an inhibitory link between a word and its hyperonyms. For example, in the logogen model, the logogen of a word may have inhibitory connections to the logogens of its hyperonyms. Then, if the logogen of a word exceeds threshold, it inhibits the logogens of its hyperonyms (e.g., *father* inhibits *parent*). Consequently, only one logogen will fire, as required by the convergence criterion.

Of course, such a proposal is *ad hoc* since for the solution of the word-to-phrase synonymy problem and the problem of disjunctive terms other mechanisms have to be found. Furthermore, an inhibitory link may cause additional problems. In retrieving *male parent* by MALE(X) and PARENT(X,Y), *father* (activated by these conceptual features) will inhibit its hyperonyms *male* and *parent* (the targets of retrieval) too.

Nevertheless, it is an empirical question whether such inhibitory links exist. An inhibitory link between a word and its hyperonyms has clear consequences. It implies, for instance, that the naming of a pictured object should be inhibited by a distractor that is a hyponym of the picture name. For example, *robin* should inhibit the retrieval of *bird* (relative to an unrelated hyponym as distractor). Similarly, *bird* as distractor should inhibit the retrieval of *animal*. It is irrelevant whether the distractor is a potential response in the experiment, because the inhibitory link is assumed to be hard-wired in the mental lexicon. As we saw earlier, picture-word interference experiments using these types of distractor have been conducted. The results were clear-cut: No inhibitory effect from hyponyms on the retrieval of a word is obtained. Instead, the retrieval of the lemma of *bird* is facilitated by the distractor word *robin*, and the retrieval of the lemma of *animal* is facilitated by the word *bird*. This holds for nouns (see panels B, C, and D of Figure 12.7) as well as for verbs (Roelofs, 1993). For example, the retrieval of *laugh* is facilitated by *chuckle*. Clearly, an inhibitory link between a word and its hyperonyms is not the answer to the hyperonym problem.

Alternatively, one may implement a principle of *specificity*, as proposed by Levelt (1989, 1992). This principle says that of all lemmas whose conceptual conditions are satisfied, the most specific one is selected. For example, if BIRD triggers *bird*, *animal*, and so forth, then *bird* will be selected because it is the most specific term. The principle might be incorporated in a decision table. Then, of all the lemmas that match the concept, the one with the most specific condition would be given preference. Similar to an inhibitory channel, however, a specificity principle predicts an inhibitory effect from hyponym distractors on the retrieval of the picture name. In case of a distractor hyponym (e.g., *robin*), the distractor word would be preferred to the name of the picture (*bird*) because it is more specific. Note, however, that the principle may be reconciled with the empirical findings if its domain of application is restricted to the lemmas that are permitted responses in the experiment. Then, although *robin* is more specific than *bird*, it is not considered for selection because it is not a permitted response in the experiment.

There are, however, other problems with a specificity principle. Although the principle solves the hyperonym problem, it fails on the word-to-phrase synonymy problem. Also, it cannot handle the retrieval of words that express disjunctive concepts. For example, if a speaker wants to verbalize SIBLING (i.e., BROTHER OR SISTER), the conceptual conditions of *brother* or of *sister* will be satisfied too. Consequently, the latter terms will be selected because they are more specific. And if a speaker wants to say . . . *male parent*, the principle of specificity leads to retrieval of *father* and not to retrieval of *male* and *parent*.

Conclusion

In this chapter, I have reviewed the major existing computational models of lemma retrieval: discrimination nets, decision tables, logogens, and featural and non-decompositional spreading-activation nets. I have shown that, first, the decompositional proposals for lemma retrieval fail to solve the convergence problem, which is composed of several subproblems — how to avoid retrieving hyperonyms and hyponyms along with or instead of the intended words, and how to correctly retrieve either a word or a synonymous phrase. Second, I have shown that the non-decompositional model satisfies the speed and convergence criteria and accounts for several major empirical findings on conceptually driven lemma retrieval. These findings concerned the time course of the inhibitory and facilitatory influences of semantic relatedness in the picture-word interference paradigm. This poses a challenge to a builder of computational models — to develop a decompositional model for lemma retrieval that meets the computational criteria and accounts for the data in a more parsimonious way.

Notes

1 Originally, discrimination nets, decision tables, and logogens were proposed as models for lexical access *per se*, without making the distinction between lemma retrieval and phonological encoding. This distinction originates from more recent research.

2 Dell and O'Seaghdha do not make a case for a featural representation of word meaning; their featural approach was one of convenience. I refer to their model because it is a computational model and gives a good description of feature-based lemma retrieval within a spreading-activation framework.

References

BIERWISCH, M. and SCHREUDER, R. (1992) 'From concepts to lexical items', *Cognition*, 42, pp. 23–60.

COLLINS, A.M. and LOFTUS, E.F. (1975) 'A spreading-activation theory of semantic processing', *Psychological Review*, 82, pp. 407–28.

DELL, G.S. (1986) 'A spreading-activation theory of retrieval in sentence production', *Psychological Review*, 93, pp. 283–321.

DELL, G.S. and O'SEAGHDHA, P.G. (1991) 'Mediated and convergent lexical priming in language production: A comment on Levelt *et al.* (1991) 'Psychological Review, 98, pp. 604–14.

DELL, G.S. and O'SEAGHDHA, P.G. (1992) 'Stages of lexical access in language production', *Cognition*, 42, pp. 287–314.

FODOR, J.A., GARRETT, M.F., WALKER, E.C.T. and PARKES, C.H. (1980) 'Against definitions', *Cognition*, 8, pp. 263–367.

GARRETT, M.F. (1975) 'The analysis of sentence production', in BOWER, G.H. (Ed) *The Psychology of Learning and Motivation*, New York: Academic Press.

GLASER, W.R. (1992) 'Picture naming', *Cognition*, 42, pp. 61–105.

GLASER, W.R. and DÜNGELHOFF, F.-J. (1984) 'The time course of picture-word interference', *Journal of Experimental Psychology: Human Perception and Performance*, 10, pp. 640–54.

GOLDMAN, N. (1975) 'Conceptual generation', in SCHANK, R. (Ed) *Conceptual Information Processing*, Amsterdam: North-Holland, pp. 289–371.

JACKENDOFF, R. (1987) *Semantics and Cognition*, Cambridge, MA: MIT Press.

JESCHENIAK, J.-D. and LEVELT, W.J.M. (1994) 'Word frequency effects in speech production: Retrieval of syntactic information and phonological form', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, pp. 824–43.

KEMPEN, G. and HOENKAMP, E. (1987) 'An incremental procedural grammar for sentence formulation', *Cognitive Science*, 11, pp. 201–58.

LEVELT, W.J.M. (1989) *Speaking: From Intention to Articulation*, Cambridge, MA: MIT Press.

LEVELT, W.J.M. (1992) 'Accessing words in speech production: Stages, processes and representations', *Cognition*, 42, pp. 1–22.

LEVELT, W.J.M. and FLORES D'ARCAIS, G.B. (1987) 'Snelheid en uniciteit bij lexicale toegang [Speed and uniqueness in lexical access], in CROMBAG, H.F.M., VAN DER KAMP, L.J. Th. and VLEK, C.A.J. (Eds) *De Psychologie Voorbij*, Lisse: Swets & Zeitlinger, pp. 55–68.

KEMPEN, G. and HUIJBERS, P. (1983) 'The lexicalization process in sentence production and naming: Indirect election of words', *Cognition*, 14, pp. 185–209.

LEVELT, W.J.M., SCHRIEFERS, H., VORBERG, D., MEYER, A.S., PECHMANN, T. and HAVINGA, J. (1991) 'The time course of lexical access in speech production: A study of picture naming', *Psychological Review*, 98, pp. 122–42.

LUCE, R.D. (1986) *Response Times: Their Role in Inferring Elementary Mental Organization*, New York: Oxford University Press.

MCGILL, W.J. (1963) 'Stochastic latency mechanisms', in LUCE, R.D., BUSH, R.R. and GALANTER, E. (Eds) *Handbook of Mathematical Psychology, Vol. 1*, New York: Wiley, pp. 309–60.

MILLER, G.A. and JOHNSON-LAIRD, P.N. (1976) *Language and Perception*, Cambridge, MA: Harvard University Press.

MORTON, J. (1969) 'The interaction of information in word recognition', *Psychological Review*, 76, pp. 165–78.

ROELOFS, A. (1992a) 'A spreading-activation theory of lemma retrieval in speaking', *Cognition*, 42, pp. 107–42.

ROELOFS, A. (1992b) *Lemma Retrieval in Speaking: A Theory, Computer Simulations, and Empirical Data*, Doctoral dissertation, NICI Technical Report 92–08, University of Nijmegen.

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

ROELOFS, A. (1994) 'Word retrieval in speaking: A case for conceptually non-decomposed access', Manuscript submitted for publication.

ROELOFS, A. (in preparation) 'From lexical concept to articulatory program: A theory of lexical access in spoken word production'.

SCHRIEFERS, H., MEYER, A. S. and LEVELT, W. J. M. (1990) 'Exploring the time course of lexical access in language production: Picture-word interference studies', *Journal of Memory and Language*, 29, pp. 86–102.

TOWNSEND, J. T. and ASHBY, F. G. (1983) *Stochastic modeling of elementary psychological processes*, Cambridge: Cambridge University Press.