4

Modeling Suppression in Lexical Access

Morton Ann Gernsbacher and Mark F. St. John

Language can be viewed as a specialized skill involving language-specific processes and language-specific mechanisms. In contrast, we view language as drawing on many general cognitive processes and mechanisms. This chapter focuses on one general cognitive mechanism that appears to be crucial for successful language comprehension: the mechanism of suppression.

The mechanism of suppression figures prominently in Gernsbacher's (1990, 1991a, 1995, 1997c) *structure building framework* for understanding language comprehension. According to the structure building framework, the goal of comprehension is to build coherent mental representations, or what we refer to as *mental structures*. These structures represent clauses, sentences, passages, or other meaningful units. In her previous research, Gernsbacher has demonstrated the crucial role that suppression plays in many structure building and language comprehension phenomena. These phenomena include *lexical access* (how comprehenders understand or "access" the appropriate meanings of words; Faust & Gernsbacher, 1996; Gernsbacher & Faust, 1991b), *anaphoric reference* (how comprehenders understand to-whom or to-what anaphors, such as pronouns, refer; Carreiras & Gernsbacher, 1992; Gernsbacher, 1989, 1991b, 1997a; Oakhill, Garnham, Gernsbacher, & Cain, 1992), *cataphoric reference* (how words that are marked by devices, such as spoken stress, gain a privileged status in comprehenders' mental structures; Gernsbacher, 1989; Gernsbacher & Jescheniak, 1995; Gernsbacher & Shroyer, 1989), *surface information loss* (why seemingly superficial information, such as syntactic form, is forgotten more rapidly than seemingly deeper information, such as thematic content; Gernsbacher, 1985), *inferencing* (how comprehenders incorporate information into their mental structures that is only implied by a text or discourse; Beeman, Bowden, & Gernsbacher, 2000; Gernsbacher, 1991b, 1994a; Gernsbacher, Goldsmith, & Robertson, 1992; Gernsbacher, Hallada, & Robertson, 1998; Gernsbacher & Robertson, 1992; Oakhill et al., 1992), *general comprehension skill* (which is skill at comprehending linguistic as well as nonlinguistic media;

Much of the research reviewed in this chapter was supported by grants from the National Institutes of Health (R01 NS 29926, K04 NS01376), the Air Force Office of Scientific Research (AFOSR-89-0258), and the Army Research Institute (DASW0194-K-0004 and DASW01-96-K-0013).

Gernsbacher, 1993, 1997b; Gernsbacher & Faust, 1991a, 1994; Gernsbacher & Robertson, 1995; Gernsbacher, Varner, & Faust, 1990), and the *comprehension challenges* experienced by various subpopulations (Faust, Balota, Duchek, Gernsbacher, & Smith, 1997; Gernsbacher, Tallent, & Bolliger, 1999).

We define suppression as a directed reduction in activation. As we argue below, suppression differs from other mechanisms of activation reduction, such as decay of activation, fixed-sum (compensatory) reduction, or mutual lateral inhibition. Although these are viable means for reducing activation, these mechanisms do not appear to be responsible for the reduction in activation in the comprehension phenomena we have explored. Furthermore, Gernsbacher has demonstrated that suppression plays such a fundamental role in language comprehension in that skilled comprehenders are characterized by efficient suppression mechanisms whereas less-skilled comprehenders are characterized by less-efficient suppression mechanisms. Moreover, this mechanism of suppression appears to be under some level of comprehenders' strategic control, and our most novel proposal is that this mechanism of suppression originates from a configural level representation—a coherent mental structure that directs the reduction in activation at that level and at lower levels of representation.

A number of computational models of lexical access already exist, but none offers a satisfying account of suppression. In this chapter, we present the initial development of a computational model of the mechanism of suppression in lexical access. Our model, which we describe more fully later, is based on St. John's *sentence gestalt* (St. John & McClelland, 1990) and *story gestalt* model (St. John, 1992). These precursors offered a viable architecture for simulating the mental representations (structures) that Gernsbacher describes in her structure building framework; therefore, we used the sentence and story gestalt models as a starting point. We begin by presenting the behavioral data that describe the role of suppression in lexical access; we conclude by presenting our preliminary simulations.

Suppression During Lexical Access: Some Behavioral Data

According to many accounts of word understanding, when comprehenders first hear or read a word, information provided by that word (e.g., its orthography or phonology) activates various potential meanings. Then, constraints provided by lexical, semantic, syntactic, and other sources of information alter those meanings' levels of activation. Eventually, one meaning becomes most strongly activated, and that is the meaning that comprehenders "access" and incorporate into their developing mental structures of the text or discourse (Becker, 1979; Kintsch, 1988; Marslen-Wilson & Welsh, 1978; McClelland & Rumelhart, 1981; Norris, 1986).

Gernsbacher (1990, 1991a; Gernsbacher & Faust, 1991b) argued that one role that the general cognitive mechanism of suppression plays in lexical access is to dampen the activation of the competing meanings. An excellent demonstration of this necessary role is found when one examines how comprehenders access the appropriate meaning of homonyms (i.e., words such as *bugs* that have at least two different meanings). Behavioral data have repeatedly demonstrated that immediately after comprehenders hear or read a homonym, multiple meanings are often activated (see Simpson, 1994, for a review). In fact, multiple meanings are often activated even though one meaning is strongly implied by the sentence context in which the homonym occurs. For example, immediately after comprehenders hear or read the word *bugs* in the sentence *The man was not surprised when he found several spiders, roaches, and other bugs*, both the "insect" meaning and the "covert microphone" meaning of the word *bugs* are activated (Swinney, 1979).

This immediate activation of multiple meanings has been demonstrated in the following experimental task: Participants listen to a series of sentences, and at a critical point during each sentence, they decide rapidly whether a visually presented test word is an English word (i.e., participants make a lexical decision). Presumably, the participants' reaction times and error rates reflect how activated the test words are (and how activated concepts related to those test words are). For example, when the sentence, *The man was not surprised when he found several spiders, roaches, and other bugs,* was presented in Swinney's (1979) experiment, immediately after participants heard the word *bugs,* some participants made lexical decisions to the test word *ANT*; other participants made lexical decisions to the test word *SPY*; and as a control, other participants made lexical decisions to the test word *SEW*. Activation of the appropriate meanings can be estimated by subtracting reaction times to test words like *ANT* from reaction times to test words like *SEW,* and activation of the inappropriate meanings can be estimated by subtracting reaction times to test words like *SPY* from reaction times to test words like *SEW*.

Figure 4.1 presents such an estimate made from Swinney's (1979) data. The two leftmost bars present the data collected immediately after participants heard the homonyms. As these two bars illustrate, when activation was measured immediately after participants heard the homonyms, both the appropriate and inappropriate meanings were more activated than were the unrelated concepts (i.e., bars are reliably above the baseline), and the difference between the two bars was not reliable. In contrast, the two rightmost bars present what happens if participants continue listening to the sentences and are tested only four syllables after hearing the homonyms. After a four-syllable delay, participants still respond more rapidly to test words related to the appropriate meanings than they respond to unrelated test words (i.e., the estimated activation of the appropriate meanings is reliably above the baseline). However, after a four-syllable delay, participants do not respond more rapidly to test words related to the inappropriate meanings than they respond to unrelated test words (i.e., the estimated activation of the inappropriate meanings is not above baseline). Thus, after four syllables, the inappropriate meanings of homonyms are no more activated than are unrelated concepts. Inappropriate meanings become less activated even more quickly than within four syllables, sometimes within only 200 ms (e.g., Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982). This very rapid deactivation of inappropriate meaning is probably why comprehenders are typically consciously aware of activating only one meaning—the contextually appropriate one.

How and why do the inappropriate meanings become less activated? Gernsbacher (1990, 1991a; Gernsbacher & Faust, 1991b) argued that the men-

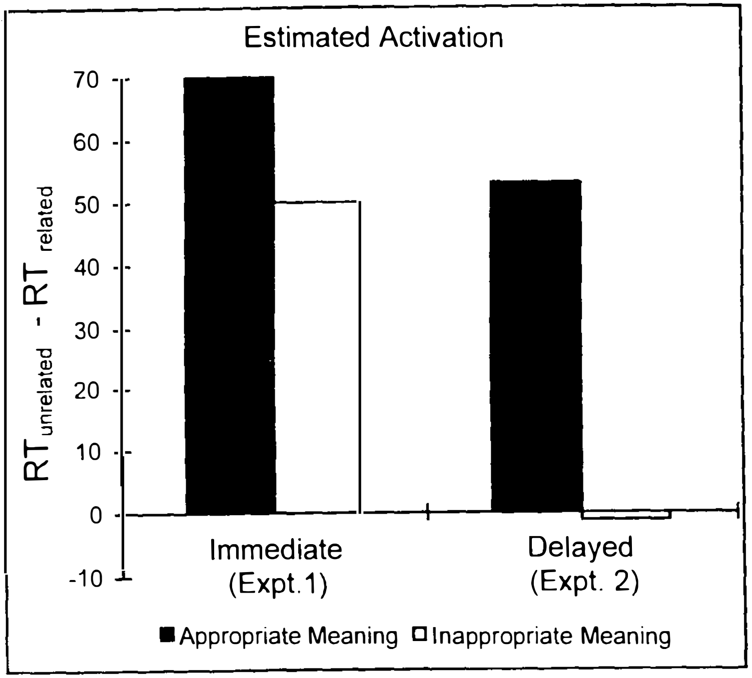

Figure 4.1. Estimated activation of appropriate and inappropriate meaning during sentence comprehension. RT = reaction time; Expt. = experiment. Adapted from tables 2 and 4 of "Lexical Access During Sentence Comprehension: (Re)consideration of Context Effects," by D. A. Swinney, 1979, *Journal of Verbal Learning and Verbal Behavior, 18,* pp. 651, 656. Copyright 1979 by Adademic Press. Adapted with permission.

tal structures that comprehenders build during comprehension can transmit processing signals; these processing signals can dampen (suppress) the activation of other representations (e.g., the inappropriate meanings of homonyms). The first goal in our computational modeling work has been to simulate a sentence-level representation transmitting this type of suppression, as Gernsbacher speculated in her structure building framework. We were particularly interested in modeling this phenomenon because, as we show, our model of lexical access using top-down suppression fits the psychological data better than previous models that have worked by other mechanisms.

The most common way the reduced activation of inappropriate meanings has been computationally modeled is by using an activation-reduction mechanism that Waltz and Pollack (1985) called *lateral inhibition* (also used in Kintsch, 1988). The mechanism is simple: Inhibitory links connect the concepts representing the multiple meanings of a homonym. For example, Kintsch (1988, Table 1) connects the *river* meaning of *bank* to the *money* meaning with a weight of −1.0. The negative (inhibitory) connections between multiple meanings of homonyms act such that when one meaning becomes more activated (because of its support from a biasing context), its greater activation inhibits (and, therefore, reduces) the activation of the inappropriate meaning. Like a seesaw, when the activation of one meaning goes up, the activation of the other comes down. However, if reaction times reflect activation levels, which is what most cognitive psychologists assume (Posner, 1978), the behavioral data do not demonstrate the pattern achieved through lateral inhibition: Both meanings are initially activated, and the appropriate meanings do not further increase in activation as the inappropriate meanings decrease. As demonstrated by Swinney's (1979) data, described above, when the inappropriate meanings' decreased in activation (at the delayed test point), the appropriate meanings did not increase. This is the pattern always observed in the behavioral data.

A related mechanism for reducing the activation of the inappropriate meanings is *compensatory inhibition,* which derives from the assumption that all concepts compete for a fixed amount of activation. When multiple meanings of homonyms are immediately activated, they share this fixed sum. Later, inappropriate meanings must decrease in activation presumably because appropriate meanings take a larger share (of the fixed sum). Like the mechanism of lateral inhibition, the mechanism of compensatory inhibition predicts a seesaw pattern (reduced activation of inappropriate meanings should be accompanied by increased activation of appropriate meanings). As we mentioned above, the behavioral data do not show this seesaw pattern. Perhaps, though, after a delay, appropriate meanings do not increase in activation because after a delay the appropriate meanings must compete with other concepts for the fixed sum of activation. By definition, Swinney's (1979) delayed test point introduced four new syllables; during those four syllables, new concepts might have been introduced, and those new concepts might have consumed some of the fixed amount of activation. However, Gernsbacher and Faust (1991b, Experiment 1) replicated perfectly Swinney's (1979) data in an experimental paradigm in which no additional concepts were introduced during the delay.

An even simpler mechanism for reducing the activation of inappropriate meanings is simply to let them decay. In many models of cognition, mental representations automatically decay when they are not continuously stimulated (J. R. Anderson, 1983). According to a decay explanation, the inappropriate meanings of homonyms become less activated after a delay because they do not receive stimulation from a biasing semantic or syntactic context. In Gernsbacher and Faust (1991b, Experiment 2), we tested the decay explanation. Our stimuli comprised 48 homonyms, such as *quack,* that were just as likely to be thought of as one noun as another (what we refer to as equal-frequency homonyms). For each of the 48 homonyms, we constructed three experimental sentences. One sentence was biased toward one meaning of the homonym, for example, *Pam was diagnosed by a quack.* A second sentence was biased toward a different meaning of the homonym, for example, *Pam heard a sound like a quack.* The third experimental sentence was neutral: Neither its semantic nor its syntactic context was biased toward either meaning, for example, *Pam was annoyed by a* *quack*. We verified the neutrality of these sentences with 50 pilot participants. For each of the 48 homonyms, we selected two test words that were related to the two meanings of the homonyms (e.g., *duck* and *doctor*), and we constructed a control sentence that was identical to the neutral sentence except that the experimental homonym was replaced by another homonym (e.g., *Pam was annoyed by a pupil*). All of the sentences were presented visually, and they continued in meaningful but different ways after the homonyms occurred; however, after participants read the homonyms and before the sentences diverged, we measured the activation of the meanings of the homonyms at two test intervals (without introducing new concepts).

For the biased sentences, the decay and suppression explanations make identical predictions: Immediately, both appropriate and inappropriate meanings should be activated, but after the delay, inappropriate meanings should be less activated. Where the decay and suppression explanations differ is their predictions about the neutral sentences. According to the decay explanation, any meaning will decay if it lacks stimulation from a biasing semantic or syntactic context. Presumably that is why the *doctor* meaning of *quack* loses activation after participants read the sentence *Pam was diagnosed by a quack*; the *doctor* meaning of *quack* lacks stimulation from a biasing context, and, therefore, it decays. Thus, according to the decay explanation, after participants read the neutral sentence *Pam was annoyed by a quack*, the *doctor* and the *duck* meaning of *quack* should decay because both meanings lack stimulation from a biasing context. In contrast, according to the suppression explanation, meanings become less activated because the mental structures representing the sentences' semantic or syntactic context transmit processing signals that suppress the inappropriate meanings. Because neutral sentences provide no bases from which to transmit suppression, the suppression explanation predicts that with a neutral sentence both meanings should be just as activated after the delay as they are immediately.

Figure 4.2 presents the 80 participants' data. As the three leftmost bars indicate, at the immediate test point, the appropriate meanings (of the biased sentences), the inappropriate meanings (of the biased sentences), and both meanings following neutral sentences were reliably activated. As the three rightmost bars indicate, after the delay, the appropriate meanings of the biased sentences were still reliably activated, but the inappropriate meanings of the biased sentences were no longer reliably activated. Indeed, they were statistically no more activated than baseline. This pattern replicates the results of numerous previous experiments (e.g., Swinney, 1979). The novel and critical data are those from the neutral sentences (the checkered bars): After the delay both meanings were still reliably activated; indeed, they were just as activated after the delay as they were immediately. These results confirm the predictions made by the suppression explanation, not the decay explanation. Thus, these results support the hypothesis that inappropriate meanings of homonyms become less activated because suppression is transmitted (top-down) from a sentence-level representation.

Gernsbacher, Robertson, and Werner (see chapter 8, this volume) reported another test of the suppression versus decay hypotheses. We were inspired by the negative priming phenomenon, which demonstrates that selecting against one stimulus in a display renders that stimulus harder to recognize on a subsequent display. If inappropriate meanings become less activated because they are suppressed rather than because they simply decay, then "selecting against" the inappropriate meaning of a homonym—when reading one sentence—should make that meaning harder to access when reading a subsequent sentence. Indeed, this experimental paradigm could show that suppression takes the activation of inappropriate meanings below baseline. In contrast, if inappropriate meanings simply decay, their activation should return only to baseline. Our data strongly support the suppression rather than the decay hypothesis. We (Gernsbacher, 1993; Gernsbacher et al., 1990; Gernsbacher & Faust, 1991a; Gernsbacher & Robertson, 1995) have suggested that less-skilled comprehenders are characterized by less-efficient suppression mechanisms. In Gernsbacher et al. (1990), we discovered that adults' skill in comprehending verbal stories (both written and spoken stories) was highly correlated with their skill in comprehending nonverbal picture stories, providing evidence for a construct we called *general comprehension skill*. In Gernsbacher et al. (1990), we also discovered a critical characteristic of less-skilled adult comprehenders: They are less able to suppress the inappropriate meanings of homonyms. We discovered this phe-

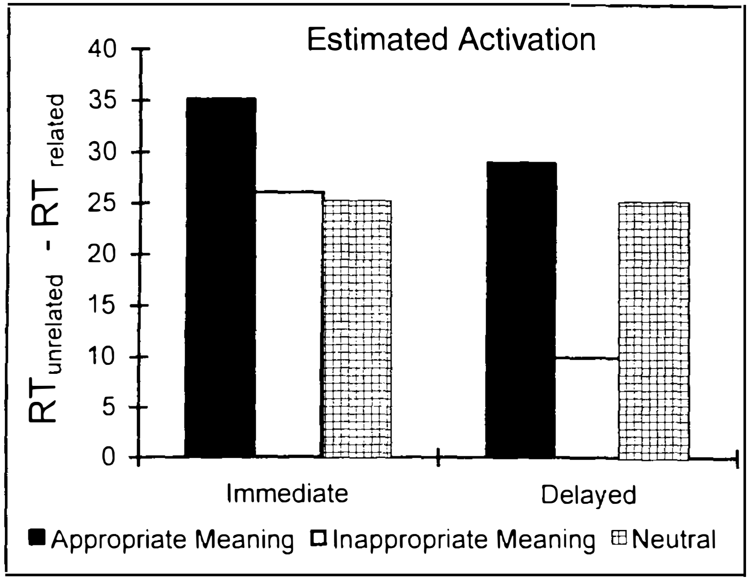

Figure 4.2. Estimated activation of appropriate and inappropriate meaning during sentence comprehension. RT = reaction time. From "The Role of Suppression in Sentence Comprehension," by M. A. Gernsbacher and M. Faust, 1991. In G. B. Simpson (Ed.), *Understanding Word and Sentence*, p. 109. Copyright 1991 by North-Holland. Reprinted with permission.

nomenon in the following way. We began by selecting 64 more- versus less-skilled university-age participants by their performance on a Multi-Media Comprehension Battery (Gernsbacher & Varner, 1988). The Multi-Media Comprehension Battery assesses skill at comprehending written, spoken, and nonverbal picture stories. More-skilled comprehenders were drawn from the upper-third of a distribution of 270 participants, and the less-skilled comprehenders were drawn from the bottom-third.

When these participants returned to the lab, they performed the following task: They read short sentences; after each sentence, they saw a test word. Their task was to decide whether the test word fit the meaning of the sentence they just read. On 80 trials, the test word did indeed fit the sentence, but we were interested in the 80 trials in which the test word did not fit the sentence. On half of those trials, the last word of the sentence was a homonym, for example, the word *spade* in the sentence, *He dug with the spade*. The test word on these trials was related to one meaning of the homonym, but it was a meaning not implied by the context, for example, *ACE*. We measured how long participants took to reject a test word like *ACE* after reading the sentence, *He dug with the spade*, and we compared that latency with how long they took to reject *ACE* after reading the same sentence but with the last word replaced by a nonhomonym, for example, *He dug with the shovel*. From this comparison, we estimated the activation of the inappropriate meanings of the homonyms. For example, the more time participants took to reject *ACE* after the *spade* versus the *shovel* sentence, the more activated the inappropriate meaning of *spade* must have been. We presented the test words (e.g., *ACE*) at two test points: immediately (100 ms) after participants finished reading the sentence-final homonyms or nonhomonyms and at an 850-ms delay.

We measured the inappropriate meanings' estimated activation (i.e., the difference between participants' latencies to reject test words like *ACE* after reading nonhomonyms like *shovel* and their latencies to reject test words like *ACE* after reading homonyms like *spade*). Immediately after both the more- and less-skilled comprehenders read the homonyms, the inappropriate meanings were reliably activated, and the performance difference between the more- and less-skilled comprehenders was not reliable. Thus, for both the more- and less-skilled comprehenders, the inappropriate meanings of the homonyms were immediately activated. In contrast, a different pattern emerged after the 850-ms delay. The inappropriate meanings were no longer reliably activated for the more-skilled comprehenders. But for the less-skilled comprehenders, the inappropriate meanings were still reliably activated; in fact, for the less-skilled comprehenders, the inappropriate meanings were as highly activated after the delay as they were immediately. These data suggest that less-skilled comprehenders are characterized by less-rapid (and therefore less-efficient) suppression mechanisms.

In Gernsbacher and Faust (1991a), we provided more evidence to support the hypothesis that less-skilled comprehenders have less-efficient suppression mechanisms. We discovered that less-skilled comprehenders are also less able to reject the incorrect forms of homophones (e.g., the word *rose* when they read *rows*); less-skilled comprehenders are less able to reject the typical-but-absent members of visual scenes (e.g., a picture of a *tractor* in a farm scene); and less-skilled comprehenders are less able to ignore words superimposed on pictures or pictures surrounding words. However, we also discovered that less-skilled comprehenders are not less appreciative of context; they can accept the contextually appropriate meanings of homonyms as quickly as more-skilled comprehenders do.

In Gernsbacher and Faust (1994), we demonstrated that this type of suppression is susceptible to some forms of strategic control. In an experimental condition in which it behooved participants to suppress inappropriate information, participants appeared to use suppression more rapidly than in a condition in which the need for suppression occurred only rarely. In Gernsbacher and Robertson (1995), we demonstrated that when the task is to suppress the more-frequent but still contextually inappropriate meaning of a homonym (such as the more-frequent meaning of *duck* in the pun *Two men walk into a bar and the third man ducks*), less-skilled comprehenders are also less-efficient at suppression. Finally, in Faust and Gernsbacher (1996), using a split hemifield presentation, we found evidence to suggest that the mechanism of suppression in sentence comprehension is somewhat lateralized. The conclusions we draw from these behavioral data are that suppression is an identifiable mechanism; the mechanism of suppression is responsible for the decreased activation of the inappropriate meanings of homonyms; suppression of the inappropriate meanings of homonyms derives from a higher level representation or structure (be it the representation of a sentence context or even the representation of a word-pair context); inefficient suppression characterizes less-skilled comprehenders; and suppression is under some degree of comprehenders' strategic (although perhaps unconscious) control. We turn now to describe our initial efforts at modeling this mechanism.

Suppression During Lexical Access: A Computational Model

Architecture of the Model

We chose an interactive activation and competition artificial neural network architecture (McClelland & Rumelhart, 1988) as the basis for our model because it exhibits a time course during which it settles into stable representations, and it uses graded and continuous influences (McClelland, 1979, 1991). The time course allows us to simulate moment-by-moment psycholinguistic phenomena (e.g., the immediate multiple activation of word meanings followed by the reduced activation of inappropriate meanings; see Erickson & Allred, chapter 12, this volume) and to investigate moment by moment whether the model accurately captures the behavioral data. The graded influences allow us to simulate graded effects on psychological behavior.

From St. John and McClelland's (1990) sentence gestalt model, we borrow several important features. First, we borrow the idea of a true sentence-level representation (perhaps comparable with the sentence-level structure in Gernsbacher's structure building framework). This representation can be used to simulate top-down sentence contributions to lexical access, anaphoric reference, cataphoric reference, and other sentence comprehension phenomena.

Second, as in the sentence gestalt model, computing sentence meaning is an integral part of processing in our model. By this we mean that sentence meaning is computed by the same parallel, distributed, graded-influence processing that computes lexical access. Whereas other models might use a separate parser for computing sentence meaning (e.g., Kintsch, 1988), we feel that such distinctions pose difficulties for simulating the early contributions of sentence-level processing to other processes. Because our model simultaneously computes multiple levels of representation, we can simulate real-time interactions. Finally, the sentence gestalt model has the capacity to learn to comprehend sentences (i.e., learn through training to compute sentence meaning). In future work, we hope to incorporate that feature into our model.

The network we constructed is organized into four layers of units, each representing a different level of representation in sentence comprehension: an orthographic layer, a lexical layer, a conceptual layer, and a sentence meaning layer that we call the sentence gestalt (SG) layer (St. John & McClelland, 1990). This architecture is graphically illustrated Figure 4.3. Bottom-up and within-layer connections are represented by solid lines; top-down connections are represented by broken lines. The numbers indicate the strengths of the connections.

The model's task is to take in a sequence of words representing a sentence and compute a coherent interpretation for the sentence in the SG layer. The words of a sentence sequentially enter the network at the orthographic layer. Activation flows through the network in a series of processing cycles. On each cycle, the activation of each unit is updated on the basis of its inputs from other units and from external inputs, which occur as each word enters the orthographic layer. Over a series of cycles, the network settles into a configuration of activation values representing the current word and its effects on each level of representation (corresponding to each layer), including the developing sentence meaning in the SG layer. Psychologically, each cycle corresponds to some number of milliseconds. After several cycles, the current word in the orthographic layer is replaced by the next word in the sentence, and processing continues for several more cycles. If the input sentence contains a homonym, the model must interpret it to establish a coherent sentence meaning. We can examine the model's process of interpretation in detail by observing all the units of the model at each level of representation as they slowly settle into a stable configuration of activation values. We can manipulate psycholinguistic variables of the input sentence and the model itself to observe their effects on the settling process.

For simplicity and clarity, the orthographic, lexical, and conceptual representations are local; that is, each orthographic string, word, or concept is represented by one specific unit. Theoretically, however, distributed representations could be used instead to model, for example, the activation of the relevant versus less-relevant associations or features of different concepts (Gernsbacher, 1990; see also Kawamoto, 1993; McKoon & Ratcliff, 1988; Tabossi, 1988b).

The SG layer contains units that represent pairings of thematic roles (e.g., agent, patient, verb) and concepts (e.g., *Pam, doctor, duck*). There is a unit for each possible thematic role–concept pair (e.g., agent–*Pam*, agent–*doctor*, agent–*duck*, patient–*Pam*, patient–*doctor*, patient–*duck*, verb–*diagnosed*). A sentence interpretation corresponds to activating the correct set of thematic role–concept pairs. For example, the interpretation of the sentence *Pam was diagnosed by a quack* would be the set {agent–*doctor*, patient–*Pam*, and verb–*diagnosed*}.

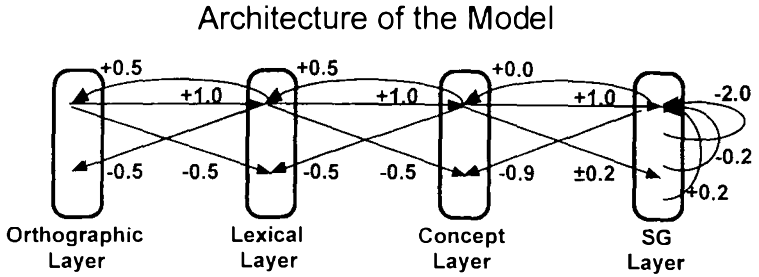

Figure 4.3. Architecture of the comprehension-with-suppression model. SG layer = sentence gestalt layer.

The units in the network are connected in the following way. Orthographic units are connected bottom-up to their equivalent lexical units with weights of +1.0 and to all other lexical units with weights of –0.5. Lexical units are connected top-down to their equivalent orthographic units with weights of +0.5 and to all other orthographic units with weights of –0.5. Nonhomonym lexical units are connected bottom-up to their equivalent conceptual units with weights of +1.0 and to all other conceptual units with weights of –0.5. Each homonym lexical unit is connected to two concept units, one for each meaning. For equal-frequency homonyms, the bottom-up connections are both +1.0. For unequal-frequency homonyms, the bottom-up connections are assigned values proportional to their frequency bias, for example +0.70/+0.30. It is not necessary that these two values sum to 1.0; rather, what is important is that their values encode the assumed difference in frequencies of the meanings of the homonym. Concept units are connected top-down to their equivalent lexical units with weights of +0.5 and to all other lexical units with weights of –0.5.

Each concept unit is connected to several SG units. For example, the concept–*doctor* unit is connected bottom-up to the SG-agent–*doctor* unit and the SG-patient–*doctor* unit with weights of +1.0 and to all other SG-agent and SG-patient units with weights of –0.2. To reflect pragmatic associations (positive, neutral, or negative), the concept–*doctor* unit is connected bottom-up to the moderately related verb–*diagnosed* unit with a weight of +0.2; it is connected bottom-up to the neutral verb–*annoyed* unit with a weight of 0.0; and it is connected to the unrelated verb–*awakened* unit with a weight of –0.2.

We implemented suppression as a top-down contribution from the SG layer to the concept layer. Consequently, in our model, each SG unit is connected top-down to its equivalent concept unit with a weight of 0.0, and, importantly, each SG unit is connected top-down to all other incompatible concept units with a weight of –0.9. For example, the SG-agent–*doctor* unit is connected to the concept–*doctor* unit with a weight of 0.0 but is connected to the concept–*duck* unit with a weight of –0.9.

Each SG unit is also connected within layer to several other SG units. These connections are designed to maintain and complete the activation of a mutually consistent sentence meaning. Connections among agent units, such as SG-agent–*doctor* and SG-agent–*duck*, are strongly inhibitory (–1.0) to reflect their logical inconsistency. All connections among patient units and all connections among verb units are also –1.0. Connections between agent units and verb units or between patient units and verb units reflect their pragmatic associations. Moderate positive associations, such as between agent–*doctor* and verb–*diagnosed*, are +0.2; moderate negative associations, such as between agent–*duck* and verb–*diagnosed*, are –0.2; and neutral associations, such as between agent–*doctor* and verb–*annoyed*, are 0.0.

These SG connections are the only within-layer connections in the model. We are noncommittal about within-layer connections inside the SG layer, but we are theoretically opposed to them at lower layers. We think of these within-layer connections as a shorthand for whatever high-level processing occurs to compute and maintain sentence-level representations. What is important for the current project is to examine the effects of SG activations on lower layers of the model. Finally, a decay parameter, set to 0.1, continuously reduces activations throughout the model by a small amount.

Simulation of Equal-Frequency Homonyms

Our goal in our first simulation was to simulate several key behavioral phenomena that arise during the processing of equal-frequency homonyms, such as *quack*, which has equal-frequency meanings of the concept–*doctor* and the concept–*duck*. These phenomena include the fact that the appropriate and inappropriate meanings are initially activated at nearly the same rate; the maximum activation of the inappropriate meaning is close to the maximum activation of the appropriate meaning; the activation of the inappropriate meaning is reduced to zero (it is suppressed) while the appropriate meaning remains activated; and the appropriate meaning does not gain more activation as the inappropriate loses activation. In this first simulation, we considered only the processing of the sentence-final homonym. The graphs in Figure 4.4 present the time course of activation of select units in the model's concept layer and SG layer, for the sentence, *Pam was diagnosed by a quack*. We assume that the activation of concept units translates directly into priming for lexical decision and naming tasks.

In this simulation, we assumed that the initial words of the sentence had already been computed into an SG representation. (Later in this chapter we present the results of a simulation in which we processed the entire sentence, word by word.) To encode the sentence fragment up to the homonym, we set the external input to the SG-patient–*Pam* unit and the SG-verb–*diagnosed* unit to 0.2. This weak activation of SG units implemented the idea that while a sentence is being processed, there is little time to fully activate an SG representation. To encode the homonym *quack*, we set the external input to the orthographic–*quack* unit to 1.0.

As processing began, activation flowed upward to the lexical layer to activate the word *quack*. From there, activation flowed upward to the concept layer to begin activating both the concept–*doctor* and the concept–*duck*. Simultaneously, activation from the SG layer began to strengthen and flow down to the concept layer. Because the SG-verb–*diagnosed* unit was connected to several appropriate potential agent units, including SG-agent–*doctor*, each began to become activated. Bottom-up activation from the concept–*doctor* and within-layer activation from SG-verb–*diagnosed* activated SG-agent–*doctor* more strongly than its contenders. In turn, SG-agent–*doctor* began to suppress inconsistent concept units, like the concept–*duck*, top-down. Thus, as the SG layer developed (simulating the development of a sentence-level representation), the SG layer suppressed the inappropriate meaning of the homonym.

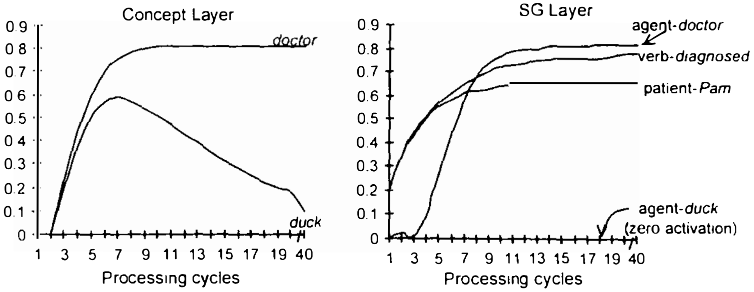

Figure 4.4. Simulation of the processing of an equal-frequency homonym. SG layer = sentence gestalt layer.

As demonstrated in the concept-layer graph, both the concept–*doctor* and the concept–*duck* units were initially activated to a high level, thereby simulating the multiple activation observed in the behavioral data. Indeed, at the peak activation of the inappropriate meaning, the ratio between the inappropriate and appropriate meanings' activation was 0.81. First, note that this ratio is not 1.0. This initially small advantage for the contextually appropriate meaning fits the behavioral data (Lucas, 1999; Simpson, 1994; St. John, 1991). Second, note that the activation of the inappropriate meaning then slowly fell to 0.0, but the activation of the appropriate meaning remained high. Note, however, that the activation of the appropriate meaning did not rise much further, thereby simulating reduced activation of the inappropriate meaning without an accompanying increase in activation of the appropriate meaning. Third, note that the reduction of the activation of the inappropriate meaning—the suppression that we aimed to model—derived solely from the negative top-down connection from the SG layer to the concept layer. Thus, we have simulated the phenomenon of suppression of the inappropriate meaning during the comprehension of an equal-frequency homonym, and we have done so without relying on mutual lateral inhibition, compensatory inhibition, or decay. In this way, our model successfully captures the proposal that suppression emanates from the sentence-level structure, as Gernsbacher (1990, 1991a, Gernsbacher & Faust, 1991b) suggested. This is not to say that lateral inhibition-type models could never be designed to match the data as well. However, the models we have seen in the literature (e.g., Cottrell & Small, 1983; Kawamoto, 1988; Waltz & Pollack, 1985) all show the seesaw phenomenon characteristic of lateral inhibition but uncharacteristic of the behavioral data.

Simulation of Unequal-Frequency Homonyms

What happens when the meanings of a homonym have unequal frequencies? For example, the word *boxer* is more frequently interpreted as the concept–*fighter* than the concept–*canine.* To simulate the processing of an unequal-frequency homonym, such as *boxer,* we set the weight from the word–*boxer* unit to the less-frequent concept–*canine* unit to 0.30 and the weight from the word–*boxer* unit to the more-frequent concept–*fighter* unit to 0.70. Then we examined processing of the sentence, *David was bitten by a boxer*, a sentence in which the less-frequent meaning (*canine*) was contextually appropriate.[1] As in the simulation we described above, we assumed that the initial words of this sentence (*David was bitten by a*) had already been computed into an SG representation; therefore, we set the external inputs to the SG units patient–*David* and verb–*bitten* to 0.2. To encode the homonym *boxer,* we set the external input to the orthographic–*boxer* unit to 1.0. The graphs in Figure 4.5 present the time course of activation of select units in the model's concept layer and SG layer.

As processing began, the activation of concept–*fighter* took an early lead over the activation of the concept–*canine*, because of the concept–*fighter*'s stronger bottom-up weight (simulating its higher experiential frequency). Simultaneously, in the SG layer, activation from the SG-verb–*bitten* began to activate the SG-agent–*canine*. In turn, the SG-agent–*canine* began to suppress the activation of the concept–*fighter*. The final result was that the concept–*fighter* was suppressed to zero activation, whereas the concept–*canine* remained activated. Thus, again we have simulated the phenomenon of suppression of the inappropriate meaning without relying on mutual lateral inhibition, compensatory inhibition, or decay; rather, the inappropriate meaning became less activated due to the suppression emitted from computing the sentence meaning. One of the interesting things that fell out the model but is completely consistent with behavioral data was that the less-frequent, though contextually appropriate, meaning was slower to rise and was never activated as highly as the equal-frequency, contextually appropriate meaning in our first simulation. This result mirrors the behavioral data that have been conducted with "unbalanced" homonyms (Simpson, 1994). The result was produced simply by setting the bottom-up word-to-concept weights at .7 and .3 (to reflect the meanings' frequencies) rather than 1.0 and 1.0.

Consider now a frequency ratio that is particularly lopsided, such as 0.90:0.10, and a sentence context that is only moderately supportive of the very infrequent meaning. In the sentence *Ed was nuzzled by a kid,* the homonym *kid* far more often refers to a child than to a goat. Because of the experiential frequencies, a strong bottom-up influence supports the concept–*child* meaning of *kid* and overrides the moderate sentence context influence supporting the concept–*goat* meaning of *kid*. In fact, the concept–*goat* meaning never gets off the ground. Thus, in simulation, the more-frequent meaning became the more activated meaning, even though it was the contextually inappropriate meaning. This pattern of activation corresponds well with the behavioral data, which demonstrate that comprehenders have difficulty accessing a very infrequent meaning of a homonym without a strong sentence context (Carpenter & Daneman, 1981; Miyake, Just, & Carpenter, 1994), in our terms, stronger pragmatic connections within the SG layer. Moreover, SG-verb–*nuzzled* even begins to show moderate suppression because of its incompatiblitiy with SG-agent–*child.*

---

[1] The preceding sentence was originally developed and modeled before Mike Tyson single-handedly changed the frequencies of the interpretations.

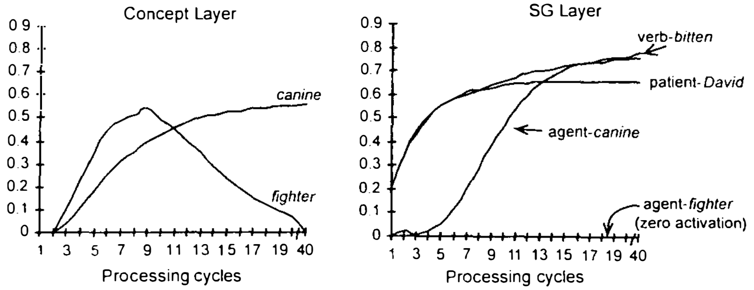

Figure 4.5. Simulation of the processing of an unequal-frequency homonym. SG layer = sentence gestalt layer.

Simulations of Individual Differences in Comprehension Skill

As described in the beginning of this chapter, Gernsbacher and her colleagues have discovered that less-skilled comprehenders are impaired in their ability to quickly suppress inappropriate information, for instance, the inappropriate meanings of homonyms. We can manipulate several variables to investigate where in the model this behavioral result originates, and we can use the model to make predictions for less-skilled comprehenders and other populations. As a preliminary step, we have simulated the inability of less-skilled comprehenders to suppress inappropriate homonym meanings by eliminating the top-down suppression from the SG layer to the concept layer. The graphs in Figure 4.6 present the results of this manipulation for the homonym *quack* in the sentence *Pam was diagnosed by a quack.*

Both meanings at the concept layer were activated bottom-up. These units, in turn, activate their associated SG units. Because we eliminated top-down suppression, both concept meanings remain active. Interestingly, the correct

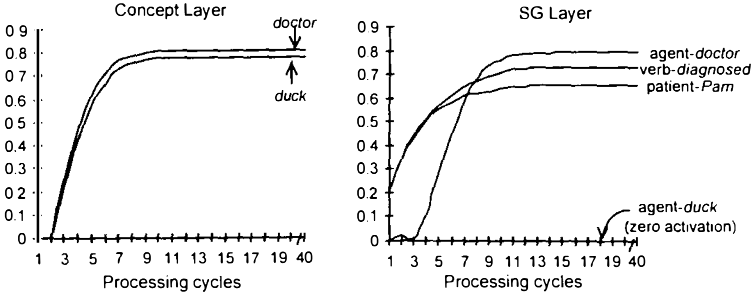

Figure 4.6. Simulation of the processing of an equal-frequency homonym by a less-skilled comprehender. SG layer = sentence gestalt layer.

SG units (agent–*doctor*, verb–*diagnosed*, patient–*Pam*) achieved nearly the same level of activation with or without the top-down suppression weights. This result fits nicely with Gernsbacher's finding that less-skilled comprehenders—despite their inability to suppress contextually inappropriate meanings—do appear to compute the correct sentence meaning (Gernsbacher & Faust, 1991a, Experiment 4).

What happens when less-skilled comprehenders are presented with unequal-frequency homonyms? Although we do not have behavioral data to answer this question, we can use the model to make a prediction. We simulated a less-skilled comprehender (i.e., with no top-down suppression) processing the unequal-frequency homonym *boxer* in the sentence, *David was bitten by a boxer.* Again, at the SG layer, the model computes the correct interpretation (agent–*canine*). However, at the concept layer, both meanings remain active, with the more-frequent concept–*fighter* being more activated than the less-frequent concept–*canine*. We are in the process of investigating whether the behavioral data of less-skilled comprehenders also demonstrate this pattern.

Simulations of the Effects of Sentence Presentation Rate

How does the presentation rate of a sentence affect the lexical access of homonyms? Again we can use the model to make a prediction. The graphs in Figure 4.7 present the results of a simulation of slow sentence presentation for the equal-frequency homonym *quack*, in the sentence, *Pam was diagnosed by a quack.* We simulated the effect of a slow presentation rate by setting the external inputs to the SG units to a higher value, 0.5, instead of the 0.2 used in the simulations we reported above. This increase captures our assumption that at a slow presentation rate, there is more time to compute a strong sentence-level representation before the sentence-final homonym is presented. With this stronger external input, the inappropriate (concept–*duck*) meaning rose more slowly and reached a lower peak activation before getting suppressed. We are also in the process of investigating whether the behavioral data demonstrate this pattern.

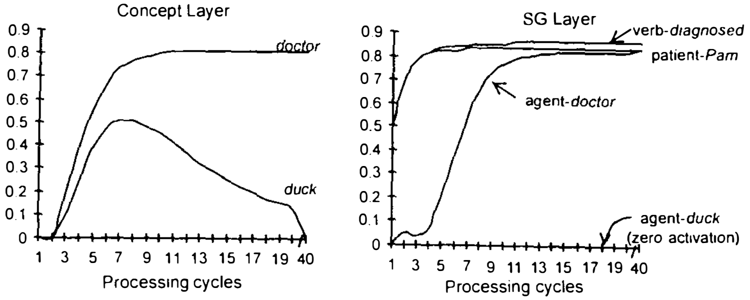

Figure 4.7. Simulation of the processing of a sentence presented at a slow rate. SG layer = sentence gestalt layer.

Simulations of Whole Sentence Processing

The previous simulations examined the processing of the sentence-final word only. We have also begun very preliminary simulations that process a whole sentence, word by word. The graphs in Figure 4.8 chart the early stages of processing (only 13 cycles) of the network as it processed the sentence *Pam was diagnosed by a quack.* Each graph presents the initial time course of activation of select units at each of the four layers. For this simulation, the Interactive Activation and Competition (see McClelland & Rumelhart 1981) decay parameter was increased, to make the model more dynamic and sensitive to each new input. During the first three cycles of processing, the word *Pam* was activated over the orthographic layer. Activation flowed forward and began to settle. On the fourth cycle, *Pam* was replaced by *diagnosed.* Activation began to resettle on this new input, plus the remaining activation from *Pam.* The SG representation became active to represent the meaning of the sentence known so far. On the seventh cycle, when *diagnosed* was replaced by *quack,* the sentence context was in place to suppress the inappropriate (concept–*duck*) meaning. The graphs in Figure 4.8 show initial multiple activation, followed by the beginning of suppression of the inappropriate meaning. Syntax still needs to be computed, and more parameters need to be investigated, but the basic approach appears sound. To develop the model to process syntax, we borrow more ideas from the original sentence gestalt model. The model presented in St. John and Gernsbacher (1998), also based on the sentence gestalt model, successfully learned to process several different syntactic constructions. We will examine the representations that model learned and apply the findings to the current model.

Conclusion

In this chapter, we suggested that suppression is a general cognitive mechanism that plays a prominent role in language comprehension. Our goal was to develop a unified, computational account of this mechanism that displays the

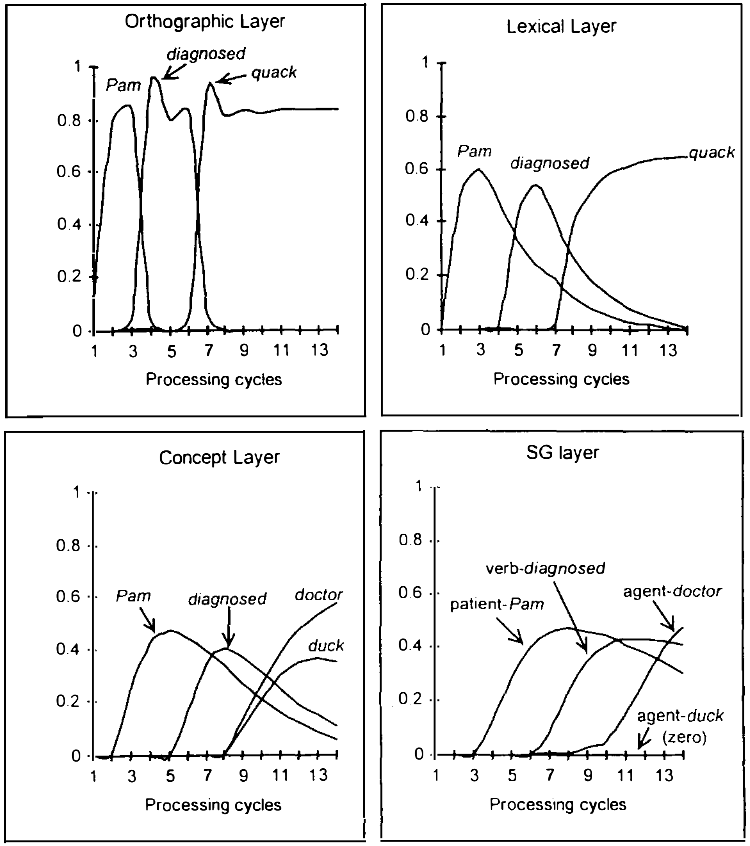

Figure 4.8. Simulation of the processing of an entire sentence, word by word. SG layer = sentence gestalt layer.

time course of suppression in lexical access. Our data indicate, and therefore we designed our model to simulate, suppression as a top-down influence on lexical access from a developing sentence-level representation: the sentence gestalt. Under normal circumstances, the inappropriate meanings of homonyms are actively suppressed because they do not fit with the developing sentence meaning. This suppression is an early and ongoing influence that grows stronger and more effective as the sentence-level representation grows stronger. Thus, top-down influences on lexical access, in the form of suppression, start early and are graded in strength.

The basic phenomenon these graded influences and bottom-up priority produce is rapid bottom-up activation of each meaning of a homonym in accord with its frequency. In the case of equal-frequency homonyms, both meanings are initially strongly activated. Then, weakly at first, but gaining in strength, top-down influences begin to suppress the inappropriate meanings. As the activation of the inappropriate meaning drops, the activation of the appropriate meaning remains constant at its high level. This combination of phenomena is best modeled as a graded, top-down suppression influence.

Beyond the basic phenomenon, a variety of influences and phenomena come together and can be modeled precisely using this interactive, graded-activation framework: the frequency of alternative homonym meanings, the strength of sentence context, the speed of sentence presentation, and the comprehender's skill. Each influence affects the speed and degree of activation of each alternative conceptual meaning. We showed how the model can simulate the processing time course of each of these influences.

Despite the success of this model in accounting for the time course of these phenomena, much remains to be done. First, there are a variety of lexical priming phenomena we have not investigated. Second, the current model's computation of syntactic relations is extremely cursory. Our work with the sentence gestalt model (St. John & Gernsbacher, 1998), however, suggests ways to improve this important aspect of sentence comprehension and integrate it with the lexical access processing. Third, the suppression mechanism applies to a far greater range of phenomena than we have modeled here. In areas related to lexical access, such as the computation of anaphora, cataphora, and even in high-level areas such as discourse-level inferences, suppression plays an important role. Our plan is to investigate these areas, extending the model to address this range of phenomena and thereby provide a unifying and mechanistic account of suppression. It may turn out, of course, on further computational and behavioral investigation, that suppression is actually a family of related mechanisms rather than a single mechanism. Whatever results indicate, a computational approach can help us understand the relations among the mechanisms and how they operate to enable language comprehension.