1 COMPUTER MODELS IN PSYCHOLINGUISTICS: AN INTRODUCTION

Ton Dijkstra and Koenraad de Smedt

De Smedt, K. (1996). Computional models of incremental grammatical encoding. In A. Dijkstra & K. de Smedt (Eds.) (1996). *Computational psycholinguistics: AI and connectionist models of human language processing* (pp. 3-23). London: Taylor & Francis, 1996.

1.1 Introduction

In a scientific tradition that is over a century old, psycholinguists have studied human language processing in all its forms, from reading and listening to speaking and writing. Innumerable studies have collected evidence about language perception and production by means of well-established research methods of observation and experimentation. Overviews of this large body of empirical work form the bulk of the material in numerous textbooks and handbooks (a.o. Garman, 1990; Taylor & Taylor, 1990; Miller, 1991; Carroll, 1994; Gernsbacher, 1994).

Over the years, many models have been developed to describe language processing in particular subdomains, giving rise to a rough conception of the architecture of the human language processing system and its components. In the last decades, our understanding of the time course of language processing has grown considerably because of the development of more precise on-line measurement techniques which allowed the collection of more exact data. However, verbal models, which are only informally specified, have not been able to take full advantage of the technological innovations. For example, verbal models can not easily make quantitative predictions, and they allow only marginal checks with respect to their completeness and consistency, which hampers their comparison to empirical data and their theoretical sophistication, respectively.

It is therefore fortunate that the same technological advances that have resulted in better measurement techniques have also stimulated theoretical developments in the form of computer modelling and simulation. Indeed, in recent years, the number of psycholinguistic articles that present computer models in combination with empirical and theoretical work has steadily grown. We believe that due to the increasing complexity of linguistic and psycholinguistic theory and thanks to advances in computer science and technology, we have now reached a point where computer models can no longer be ignored as scientific instruments in the study of human language processing. In this chapter, we want to highlight the general role of computer models in present-day psycholinguistics and give an overview of the models in the remainder of this book.

1.1.1 Computer models

Cognitive science considers human language processing as a form of information processing. Language perception and production are analyzed conceptually into a number of processing steps which recode representations of the incoming information from one form into another, in order to transform the speech signal into an idea or vice versa (cf. Hillgard & Bower, 1975, p. 431). In the context of a psycholinguistic theory, a *model* can be seen as a precise and operationalized rendering of this type of process for a restricted domain of human language processing.

Even though a model may be described *verbally* in ordinary terms, a precise description often requires a *formal* notation borrowed from mathematics or computer science. Insofar as formal models specify a series of computations (algorithm) in the terminology of information processing, they are called *computational*. Well-specified computational models can be implemented as computer programs, written in a particular programming language, that embody the model's algorithm(s) and can be executed on a computer (see also Pylyshyn, 1986, pp. 88–89). We will call such programs *computer models*. In practice, this proposed distinction between computational models (algorithms) and computer models (implemented programs) is not always made, and the two terms are often used interchangeably.

Models are necessary and useful *simplifications* of the real world, which enhance our understanding of the world by revealing the abstract (and perhaps simple) principles underlying its bewildering complexity. In making the world's opaque nature transparent, modelling provides a kind of X-ray vision. Ideally, models abstract away from those aspects of reality that are circumstantial and irrelevant, while highlighting other aspects that are fundamental in explaining what is under investigation. The simplifying assumptions and abstractions can be taken at various levels. It must be decided which aspects of reality should be represented in terms of the model's architecture, which as representational units or their connections, which as steps in the computational process, which as variables or parameters, and which should simply be left out.

Given adequate input and suitable parameter settings, computer models perform computations, the outcomes of which correspond to predictions in accordance with the underlying theory. For example, presenting a language comprehension model with a word or sentence may result in the identification of that word or the interpretation of that sentence. Furthermore, models can also predict error rates and response latencies in specific experimental paradigms.

The model thus *simulates* part of the real world, i.e. the model's behavior is intended to be similar to that observed in real-life and experimental conditions. A comparison of simulation results and experimental data can be expected to lead to a revision or further refinement of theoretical insights. This, in turn, leads to further adaptations in the model and perhaps to more experiments, etc. As a result of the introduction of models, the classical empirical cycle is extended to the form as shown in Figure 1.1.

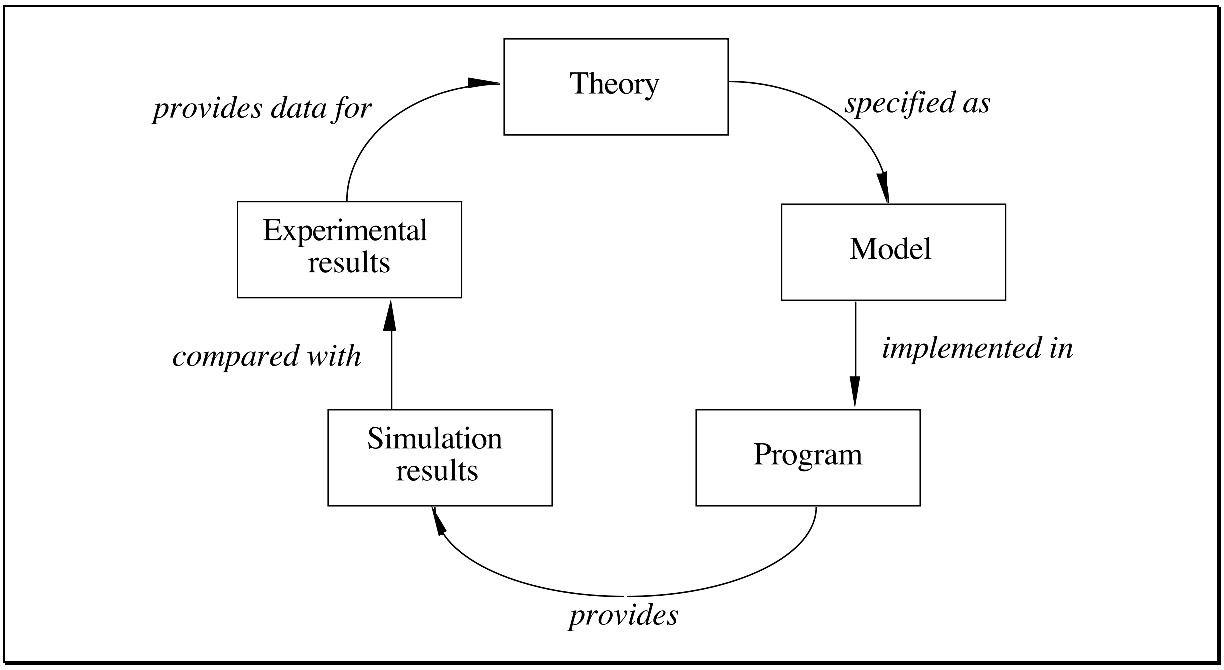

Figure 1.1 The empirical cycle including computer modelling

Both computer models and humans can be considered as (more or less complex) information processing systems. Marr (1982) distinguished three levels at which an information processing device can be understood: the level of computational theory, that of representation and algorithm, and that of the hardware implementation. The most general level is a *computational theory* in Marr's sense, specifying what is the goal of the computation, why it is appropriate, and what is the general strategy. With respect to natural language processing, Marr assigns theories of linguistic competence at this level, for example Chomsky's transformational grammar, because transformations in this theory define abstract mappings between linguistic descriptions, not operations of cognitive processes. Such operations are specified as *algorithms* at the second level. Distinctions such as the relative speed of various algorithms, their sensitivity to inaccuracies in the data, or whether they run in parallel or serially play a role here. We will use the term *computational* to refer to the operations at this algorithmic level (as witnessed by the title of this book), rather than to refer to the first level of Marr. The third level specifies the *physical realization* of the machine executing the algorithm. The wiring and other physical properties of this machine may put severe restrictions on, for instance, speed and parallelism of realizable algorithms, as well as on memory capacity.

The computational models in this book are mainly described at the theoretical level of representation and algorithm. To be interesting for psycholinguistics, such models may abstract away from the human hardware (or rather, wetware) implementation, but they must still be correct and specific about the algorithm underlying a particular cognitive process and the representations involved in it. Such computational models must also give accounts of cognitive errors and response latencies. In this respect, psycholinguistic models are different from models that are product-oriented and based on technological solutions. Such models (perhaps "programs" is a better term) do not care about the particular algorithm underlying a linguistic process, as long as input and output representations are optimal. For instance, in reading aloud the newspaper to a blind person, a machine does not need to use the same series of algorithms as a human reader, it just has to perform the task optimally and efficiently.

1.1.2 Simulation

Implementing a model on a computer has practical advantages over only formally specifying a computational model. The automaticity, speed, and precision of computers make it possible to run fast and accurate simulations with the implemented model. In the most general sense, a psycholinguistic computer model is designed to reach an outcome similar to that of human language processing. A sentence parser is meant to process those types of sentences that a human can process under various conditions. For example, whether a sentence can be parsed or not may depend on the number of embedded clauses (see Chapter 8). Going further, the model could be made to predict not only qualitative aspects of parsing behavior, but also quantitative aspects, for example, that one parsing is preferred over another (ordinal scale) and how strong this preference is (interval scale). Other quantitative simulations concern the success rate of parsing for different sentences and the time needed to parse these (interval scale).

Due to the increasing complexity of scientific theories, simulation is not only practical, but even indispensable in many fields, including psycholinguistics. We will attempt to list some important uses of computer simulation and their advantages.

First, without computer simulation it is practically impossible to check whether the model is complete and the different parts of the model are internally consistent (in the sense that they do not produce contradictions). It is often impossible to test a verbal model by manually computing its results when it is applied to sample situations, let alone by performing exhaustive tests on a comprehensive set of data. This is even so for formally specified or mathematical (*closed-form*) models, for example when these predict probabilistic behavior or yield complex response distributions.

Second, simulation may sometimes be essential in the interpretation of empirical results. Simulations can reveal that data which *prima facie* seem to contradict verbal theory may in fact be consistent with it. This case was illustrated in language production by Levelt et al. (1991), whose computer implementation of the so-called Standard Theory (assuming two autonomous serial processing stages in production) could be reconciled with what appeared to be conflicting empirical data (see Chapter 13 for a discussion of this issue from a different viewpoint).

Moreover, predictions made on the basis of a verbal theory for different experimental conditions may be fine-tuned by the computer model to the actually used stimulus material. As has been observed already by Cutler (1981), it is becoming more and more difficult for experimenters to run experiments using stimuli that have been controlled for the increasing number of characteristics that are considered relevant. For example, in the domain of word recognition, relevant factors include word frequency, bigram frequency, number of neighbors, frequency of the neighbors, familiarity, concreteness, etc. (see Chapters 5 and 6). If one of these factors is the topic of investigation, the experimenter wants to match the stimulus material with respect to the other factors. However, given the large number of relevant factors, the experimenter must often use less stringent criteria on stimulus matching than she would have liked to, simply because the optimal word material does not exist in the language. This results in stimulus material that may to some extent be suboptimal or noisy, and in experimental data that will deviate to an unknown extent from what would have occurred under ideal circumstances.

In this case, computer simulation can be applied as a tool to evaluate the effects of variability in the material that is uncontrolled for or unavoidable. Suppose the experimenter adheres to the assumptions on the architecture and process of visual word recognition that are incorporated in the Interactive Activation (IA) model (discussed in Chapter 6), and that the model is run with the stimuli selected for the test conditions. If the assumptions underlying the model are correct, then different stimulus conditions should result in the differential behavior that was expected on the basis of the verbal theory, since the model is assumed to be a specification of this theory. For example, in a very precise way the model takes into account how bigram frequency, target frequency, and frequency and number of neighbors interact. The actual pattern of simulation results can therefore be considered to reflect more precise predictions *tuned to this particular stimulus material* than those made on the basis of the verbal theory (cf. Van Heuven, Dijkstra, & Grainger, in prep.). If only a verbal model were available, it would have been impossible to obtain predictions that are corrected for the inherent biases in the stimulus material.

Even if the model's predictions indicate that the stimulus material deviates in some respects from the theoretically motivated criteria, later collected empirical data can still be compared to the model's performance. Though the material does not reflect the theoretically distinct dimensions completely, the results of the simulation may still be considered to be predictions based on the cognitive architecture and processes assumed by the theory and implemented in the model.

Third, modelling not only results in predictions of linguistic behavior under empirically known conditions, taking into account many different aspects of the stimulus material simultaneously, but even under previously unknown conditions. This may lead to ideas for new experiments which can then check the predictions made by the simulations. As an example, consider a semantic priming situation in which a target word (e.g., *nurse*) is presented at different temporal intervals (Stimulus Onset Asynchronies or SOAs) after a prime stimulus (e.g., *doctor*). When fit to empirical data for a particular range of SOAs, a computer model can predict target latencies for an as yet untested SOA between prime and target. The heuristic value of computer models is further enhanced by the use of computer models as practical tools that support research by visualizing processes and representations (cf. Chandrasekaran, Hari Narayanan, & Iwasaki, 1993; Gentner & Stevens, 1983; Lakoff & Johnson, 1980). In the earlier example, a suitable new SOA might be suggested by examining a visual display of the simulation results for already tested or simulated SOAs.

Fourth, even manipulations which cannot be performed in a direct way on human subjects are amenable to simulation. For example, one may simulate successive degrees of lesions to a computer model to invoke effects of aphasia (e.g., Patterson, Seidenberg, & McClelland, 1989; Haarmann & Kolk, 1991). Although obvious ethical considerations make the replication of such manipulations in an experimental way impossible, simulation results can nevertheless be compared to observable facts ("nature's own experiments") in the real world. A similar role for simulation is common in other scientific fields, for instance in astronomy, where it is impossible to experiment by manipulating real galaxies, but where computer simulations of galaxy formation are compared to observed galaxies in different stages.

These advantages of computer modelling are widely discussed and are obviously shared by the authors and editors of this book. We hope that, by studying the various computer models described in this book, the reader will be infected by our enthusiasm and become convinced of the exciting possibilities offered by simulation studies. However, to warn for potential pitfalls in modelling, we will now address the difficult issue of *how to specify* a computational model.

1.1.3 Model specification

Even though computational models are becoming increasingly complex, with many different parameters and constructs at different hierarchical levels, they still simplify human behavior to a tremendous degree, not in the least because simulation tests of the model are subject to practical and theoretical limitations. These simplifications become apparent if we compare the architecture and behavior of a particular model to what is known about the cognitive subsystems serving a similar purpose in humans. The organisation of the model reflects the choices and restrictions made by the modeller with respect to the representation of reality in terms of (at least) structure, process, task, and resources and strategies (cf. McCloskey, 1991).

Structural choices in computer models are made with respect to input and output *representations* of messages, as well as to internal representations within the processing system. Some of the choices are willingly made simplifications that help to avoid irrelevant complexity ("nuisance factors") in the model. It may not be very useful, for example, to incorporate letter features within a model that focuses on general aspects of text representation, despite the observation that changing a single letter feature could make a difference to the meaning of the word, sentence, and even text it belongs to.

Sometimes structural simplifications are forced because not enough is known about human cognitive architecture to provide the model with empirically motivated and thus "realistic" representations. Some representational choices cannot be avoided because either the output or the input of the model cannot be directly observed. For example, we cannot directly observe the semantic representations that are the product of language comprehension or that lie at the origin of language production, so that models at these levels must use abstract, "invented" (even if plausible) semantic representations.

More generally, in order to be able to operate at all, computer models require ad hoc solutions to deal with the underspecification of less central parts of the model. Such underspecifications may become painfully apparent especially during the model's implementation phase, and remedies to this problem may to some extent depend on characteristics of the chosen computer environment. One must beware of theoretically unmotivated "patchwork" since the implemented solutions may affect the model's performance.

To summarize this point, most current computational models can be criticized in that they incorporate only a crude and possibly disfigured replica of some central part of human cognitive architecture, while they disregard less salient aspects.

In this respect, *process* simplifications are even more apparent. Most available computer models are static in the sense that they tend to focus at the end products of processing (Parisi & Burani, 1988). Even when models have an explicit architecture, many are still like black boxes in the sense that they do not allow the *on-line* examination of the developing linguistic process they are meant to mimick. Moreover, since, as described earlier, we often cannot precisely know the input or output of human processing, we do not even know where certain processes start and end.

This indicates how difficult – if not impossible – it is to distinguish cognitively motivated from technologically motivated (e.g., pure AI) or behaviorist accounts of input-output relationships (cf. Pylyshyn, 1989; Anderson, 1983, 1991). However, the currently available experimental and clinical evidence from different disciplines concerning human information processing constrains the number of possible human mental architectures in many more ways than product-oriented approaches take into account. Also, the cognitive relationship between input and output representations is so complex that modelling is indispensable as a powerful heuristic means to chart potential intermediate mental processes, representations, and knowledge sources.

Furthermore, the time-course of the mental process and the informational pathways taken depend on the representational units in the model and the way they are linked up. Consider a network model in which two parameters regulate the amount of the forward and backward spreading activation between representations of a letter and the word in which it occurs (cf. Chapter 6). When a model includes these or other parameters with a variable strength, a particular implemented version of a verbally or formally specified model is in fact but one possible realization of a general class of models. It is very well possible that some parameter settings may result in model behavior that is not only quantitatively but also qualitatively different from that emerging with other settings. For several models found in the literature, this has led to the wise decision to restrict empirical investigations to one particular model with a fixed set of parameters and parameter settings, or alternatively to experimentally vary these parameter settings (cf. Jacobs & Grainger, 1992).

Task specifications are unfortunately often lacking in computer models. The input-output associations established by computer models are seldom explicitly linked to particular tasks. Little is known about which task simplifications can be allowed in a computer model without seriously deforming the similarity between model and human processing. Whereas the human mind is characterized by flexibility and seems to have at its disposal complex task-dependent identification and decision processes, most models at best embody a simple decision process that operates in the same rigid manner in different tasks on a common task-independent identification process (cf. Jacobs & Grainger, 1994). Admittedly, the task dependence of human language processing is still an open issue. The question is whether, for instance, human syntactic parsing operates differently depending on the task, which may vary from 'deep' understanding to skimming, correcting and reading aloud. It would be useful if modelers would standardly specify for which particular tasks their models are appropriate and to which extent they believe the models are generalizable to other task situations.

Human flexibility is also prominent in issues relating to *mental resources* and *strategic decisions*. Very few of the current models account for attentional constraints on information processing, e.g., limitations in working memory capacity in relation to mental processing load. In addition, a subject's resort to particular *strategies* in a particular task may sometimes be related to such resource limitations as well, or may in other cases be related to attempts to obtain an over-all benefit in task performance (Stone & Van Orden, 1993). However, such extraordinary flexibility of subjects is unaccounted for by practically all models. Recent approaches to language processing have just started to pay attention to these issues, sometimes in relation to interindividual differences in task performance (e.g., Just & Carpenter, 1992; King & Just, 1991). One interesting idea that has been brought forward is to implement strategic control as a form of control over parameter settings. Even though paying attention to strategic factors may currently seem to require a considerable investment of modelling efforts, "rigorous treatments of strategy may, in the long run, simplify rather than complicate the big picture" (Stone & Van Orden, 1993, p. 771).

In the foregoing, we have elaborated on the choices the modelling researcher must make during implementation for several reasons. Each choice with respect to structure, process, and other characteristics of the model not only constrains and co-determines the actual behavior of a complete and operative model, but also necessarily implies a simplification of the real world. Furthermore, as we shall see in more detail below, each choice can be used in the process of evaluating the model. If the model's behavior is not sufficiently similar to the empirical data, this could be due to incorrect assumptions with respect to fundamental model characteristics such as the model's architecture (e.g. interactivity, or lack of it) and the representations it uses (cf. Lachter & Bever's, 1988, and Besner et al.'s, 1990, criticisms concerning the use of Wickeltriples in models of language acquisition and word recognition). Trying to improve the fit between the model's behavior and empirical data by fiddling with parameter settings may not be as illuminative or useful as examining the simplifying choices that are made with respect to different model components.

Furthermore, the modeller will ideally not build the model just to account for one data set, but will have an open eye for the research context and potential future developments. Keeping this broader perspective in mind, a modeller should from the very beginning take into account those aspects that will allow generalisation to different materials, tasks, and subjects in future research. A model with many simplifications may be very useful on a small-scale, but in order to generalize it while keeping the same underlying explanatory constructs it may be necessary to improve the description of structural, process, task, resource, or strategic aspects.

In our opinion the advantages of computer models far outweigh the model design problems we have just signalled. Computer models are tools, and tools can be useful even if they are not perfect, as long as we remain aware of their limitations. The remainder of this book will therefore pay attention not only to the achievements, but also the shortcomings of computational psycholinguistic models.

1.1.4 Model Evaluation

There are several criteria by which to judge the adequacy of a model. One obviously essential step of model evaluation lies in the comparison of the model's results with appropriate *empirical data*. But as already hinted at above, such a comparison can only be considered as a rather weak test of the model if the empirical results were previously known and the model was constructed to account for those data in the first place. Especially models that try to fit their results with the help of many parameters may amount to rather trivial views of mental processing. Researchers in the field of simulation are familiar with the saying that 'a good scientist can draw an elephant with three parameters, and with four he can tie a knot in its tail' (Miller, Galanter, & Pribram, 1960, p. 182). A stronger test consists in having an already existing model produce predictions for untested conditions. Subsequent empirical tests can then falsify or verify these predictions. Though pure model predictions for new stimulus material are seldom made (see Chapter 12 for a notable exception in the domain of lemma retrieval), researchers do often check if their models generalize to data sets which were not taken into account during model construction.

We would like to point out that in an optimal situation, the stimulus material tested in the experiments would be nominally identical to that used in the computer simulations. In this case, the results of the experiments can be related to the model behavior without making the implicit assumption that differences in simulation and experiment are not due to unknown stimulus differences. So, in practice, models and experiments impose restrictions on each other. We recommend building new models in combination with planning experiments, while making sure the models can handle the stimulus material to be used in experiments. This approach seems to ensure an optimal coordination between, on the one hand, the top-down, hypothetical-deductive theory development which is typically enforced by the modelling effort and, on the other hand, the bottom-up development of theory on the basis of experimental data.

Clearly, a model may try to account for *multiple* aspects of human language behavior at once, and may thus be evaluated on several dimensions. For instance, learning connectionist models typically include a training phase before testing the output. If the language learning process is assumed to be similar to that in human subjects, this provides an extra criterion to evaluate the model, in addition to the evaluation of the learned behavior. Therefore, when models become more complex, there may on the one hand be more ways to 'solve' deviations with respect to the data, but on the other hand, the more complex models may offer more dimensions to be tested and evaluated upon. When the model's performance range is enlarged, the chance that there is real convergence between the computational model and human processing may also be enlarged.

In scientific practice, models are often evaluated also in a *model to model comparison*. But the comparison of several models for the same domain should be done carefully, taking into account the specific simplifications and assumptions in each of the models. In addition, models often cut the cognitive pie in very different pieces, even when they relate to the same subdomain, so that even input and output are often hardly comparable.

A general set of evaluation criteria is discussed by Jacobs and Grainger (1994) under the headings of *descriptive and explanatory adequacy, simplicity, generality,* and *falsifiability.* These criteria, originally conceived for evaluating models of visual word recognition, can be usefully applied to computer models of psycholinguistic processes at large. In addition to these criteria which relate models to the empirical data they wish to simulate, Jacobs and Grainger also mention other criteria such as *modifiability, equivalence class, completeness*, and *research generativity* of models. The main motivation behind this last set of criteria is that models that can more easily be tested or changed, or that are based on more abstract principles, should be preferable to others. We refer the reader to this article for detailed discussion and confine ourselves to a few remarks, starting with descriptive adequacy.

We consider a model to be *descriptively adequate* in a qualitative way, if the model, despite several simplifications and abstractions, nevertheless retains essential properties of the human cognitive processing system and the representations it uses. This suggests a model evaluation in terms of the implementation choices described in 1.1.3. However, this reveals an implicit conflict: On the one hand, a simple model is preferable to a complex one in order to be more easily controllable and to avoid Bonini's paradox ('the simulation turns out to be no easier to understand than the real-world processes it is supposed to illuminate'). On the other hand, a good model must retain enough of the essential complexity of human behavior. We would consider a model's input representation to be sufficient and acceptable if it incorporates just the right amount and kind of simplification compared to the signals a human use as input, while the model still results in the same behavior as that of the humans. However, in the current state of psycholinguistics, it is very hard to define “the right amount and kind of simplification” or even “the same behavior”. Still, as a rough guideline, models that are based upon widely accepted empirical assumptions are preferable to models that make ad hoc assumptions.

However, Jacobs and Grainger use the term *descriptive adequacy* of a model in a more quantitative way, referring to the degree of accuracy with which a model can predict a data set. This, by the way, is unfortunately often the main or only criterion used in practice. Jacobs and Grainger point out that while verbal models should be formulated so as to allow predictions at the level of an ordinal scale, algorithmic models should enable us to compute goodness-of-fit indices at an interval scale. These predictions of course depend on how well units on a model scale correspond to those in reality. For instance, activation values (see Chapter 3) in the model may be transformed into response probabilities, or time measurements in terms of the model’s processing cycles may be transformed into reaction times.

While descriptive adequacy amounts to an “in depth” criterion, models can also be judged with respect to their “breadth” in terms of *generalizability* to other stimulus sets, tasks, and response measures. Sure enough, models that are more generally applicable are often more useful than models operating in a restricted domain, but benefits in generality often bring about costs in terms of depth of analysis or in terms of model complexity. A tough problem in model comparison here is that, even though we sometimes count the number of free parameters in determining a model’s goodness of fit, we do not yet have a good means to evaluate the implicit “theoretical degrees of freedom” which a model consumes in terms of aspects of its architecture (cf. Newell et al., 1989, p. 126; Massaro & Cohen, 1991). Some first suggestions on how to obtain a “simplicity rating” for computer models are given by Jacobs and Grainger.

Approaching model evaluation from a practical perspective, we suggested to the authors of Parts II and III of the book to keep in mind a number of aspects of the computer models in their discussion. First, relating to section 1.1.3 on model specification are the following aspects: the quality of the model input compared to that of human language; the psychological validity of the processing assumptions; the types of internal representations used during processing; the quality of the model output compared to that of humans; the assumed types of decision processes. Second, relating to the current section on model evaluation are the following: the coverage of the available experimental data, e.g., the range of data sets accounted for (generalisability); and the goodness of fit (descriptive adequacy). Third, models should also be evaluated in terms of the types of problems that remain as yet unsolved, which may lead to suggestions for improvements in future research.

1.2 Organization of the book

The book is organized in three parts. Part I is about computational modelling as a new approach to psycholinguistics in general. Part II discusses models in the various subdomains of language comprehension and Part III does the same for language production.

1.2.1 Overview of Part I: Modelling paradigms

Part I consists of three chapters of which the current introductory chapter is the first. The two remaining ones each deal with a broad group of paradigms for cognitive modelling.

In Chapter 2, Daelemans and De Smedt introduce the paradigms of mainstream Artificial Intelligence, based on symbolic representations of mental states and concepts. Computation in an AI model is driven by an algorithm that manipulates symbols and is itself also symbolically specified. The chapter describes a number of formalisms for the representation of knowledge, including semantic or associative networks, frames, inheritance mechanisms, marker passing, conceptual dependency structures and conceptual graphs, production systems and logic, as well as various types of grammars.

Chapter 3 by Murre and Goebel introduces the connectionist approach, which is inspired by ideas about how the human brain works, even though the majority of the resulting neural network models in psycholinguistics do not make neurophysiological claims. The chapter distinguishes between *localist* and *distributed* neural networks. In the former, also referred to as Interactive Activation (IA) networks, concepts to be modelled (including linguistic notions) are represented by single network nodes, as in associative networks, and the weights on connections are usually preprogrammed. Distributed models, also called Parallel Distributed Processing (PDP) models, are based on networks in which concepts are associated with activation patterns over several nodes simultaneously and the system itself can often learn to categorize and generalize. Connectionist models in general are characterized by the fact that regular behavior arises from not from hard rules, but from the interaction of many small influences. Among other things, Chapter 3 describes various network architectures and system dynamics, and a number of widely-used learning rules, such as the Hebb-rule, the delta rule, and backpropagation.

1.2.2 The language user framework

Parts II and III discuss prominent computer models, organized by the subdomains in Psycholinguistics to which they belong. Given the limited scope of this book, no attempt at completeness has been made, neither with respect to the subdomains nor with respect to modelling techniques. In particular, accounts of language learning and language disorders are limited in this book. Nevertheless, we hope to present a sufficiently wide range of representative computational models to give an indication of their importance for psycholinguistics. Furthermore, each chapter in these parts has roughly the same structure, proceeding from an overview of basic concepts and phenomena to empirical and theoretical work and then to a number of computer models, which are explained in some detail and evaluated.

As an overall framework to subdivide Psycholinguistics in convenient sections, we use the language user framework proposed by Dijkstra and Kempen (1993; 1984; see also Levelt, 1989) and depicted in Figure 1.2. Like the human body which incorporates a number of organs that each have a different function, the language user system can be conceived as consisting of functionally different mechanisms (cf. Chomsky, 1980, pp. 38–39). Thus, to understand human language processing, we distinguish a number of subsystems or (in a loose terminology) *modules*, each of which deals with a specific subtask (see also Stone & Van

Orden, 1993). For example, the task of the Word Recognition subsystem is to identify morphemes or words that can be used by the Parser to construct a syntactic representation of a sentence. The Word Recognition system and the Parser make use of specialized linguistic knowledge stored in a specific part of the language user system (e.g., the Mental Lexicon and the Grammar). Figure 1.2 shows the various components or mechanisms of the language user framework grouped around the knowledge sources containing representations of linguistic units and rules.

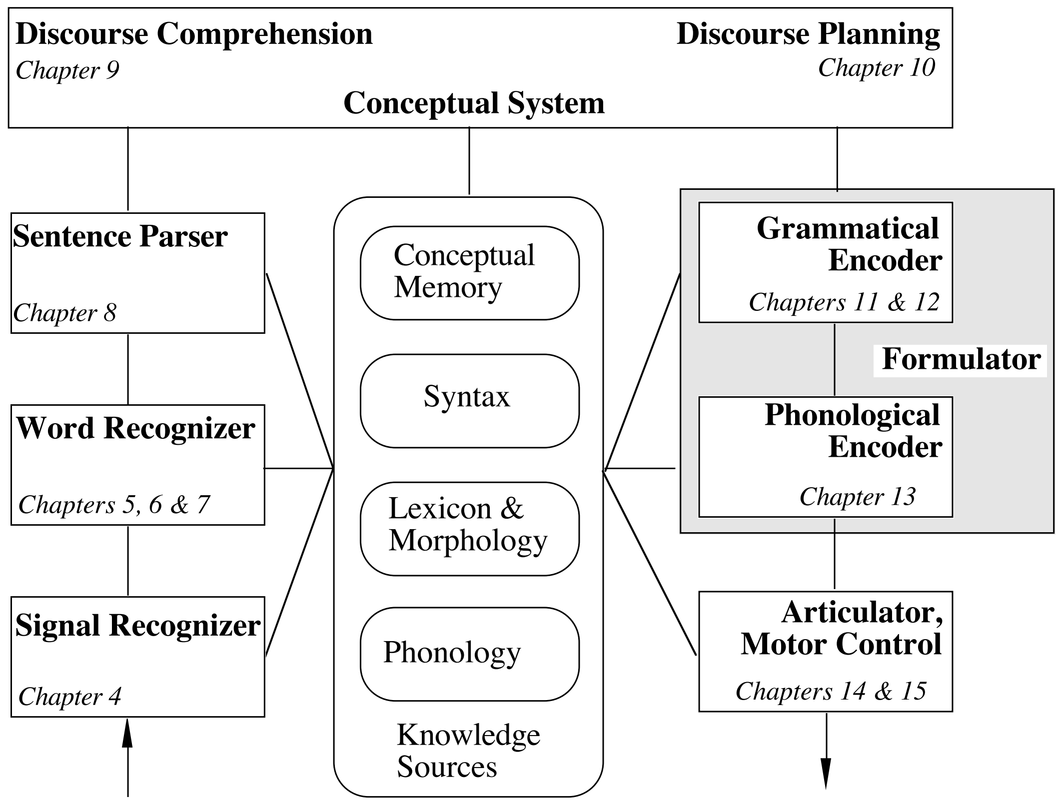

Figure 1.2 The language user framework which serves to organize the book

The language user framework can be subdivided in several ways. First, a distinction can be made between the linguistic knowledge that is stored in long-term memory (middle of the model), and the mechanisms (situated at the edges) that use that knowledge to transform the incoming or outgoing representations into new formats. To some extent, this distinction corresponds to the traditional subdivision of linguistics and psycholinguistics.

Second, the model contains modules for both language perception (left) and language production (right), and makes the simplifying assumption that these employ representations from a common database, even though the processes underlying production and perception may necessarily be very different (see Zwitserlood, 1994, for a recent analysis of the viability of this position with respect to phonological representations of lexical form; also see Foreword).

Third, the framework contains (horizontal) layers that roughly correspond to linguistic units of different sizes: letter–phoneme, morpheme–word, sentence, and discourse. To these units correspond different linguistic subdisciplines: phonetics and phonology, morphology and lexicology, syntax, and semantics and pragmatics.

The framework purposely leaves open the complicated issue of whether information between modules flows unidirectionally or bidirectionally between modules, e.g., only from speech signal to concept in language comprehension or with feedback. In fact, some researchers may stress the interactions between different components to such an extent that they subdivide the language user system in a different way. Some may even find the whole enterprise of specifying modules within the linguistic system futile. However, the language user framework offers a convenient classification to structure Parts II and III of this book. For this reason, Figure 1.2 indicates the mapping between modules and chapters.

The models to be discussed in Parts II and III are for the most part psychologically inspired models that are not only well-known in their domain but describe a special, restricted set of phenomena that have attracted attention from many researchers. For example, the IA model for visual word recognition (Chapter 6) is discussed not only because it is a “classic” model in its subdomain, but also because it makes psychologically interesting predictions with respect to the influence of competing word candidates on target recognition. Other thematic choices include the discussion of syntactic preferences in sentence understanding (Chapter 8), incremental processing in sentence production (Chapter 11), and speech errors in phonological encoding (Chapter 13). In other cases, the authors chose particular models in a domain because they were the only models available, even though these models currently have little psychological motivation. This was the case, for instance, in the area of text planning (Chapter 10), where most available models are mainly product-oriented. In still other cases, the authors chose the only implemented models that are currently available for a specific task, and which often happened to be their own models. A case in point is the modelling of handwriting (Chapter 15).

1.2.3 Overview of Part II: Language comprehension

In Chapter 4, Massaro presents two well-known computer models in the area of speech perception and evaluates these with respect to a number of prominent factors that affect processing. These models are his own Fuzzy Logic Model of Perception (FLMP) and the IA model TRACE. The models are discussed in the light of the question which perceptual units should be considered most important in speech perception. Massaro argues that the syllable is a likely candidate representation.

While the TRACE model is considered at the sublexical level in Chapter 4, it is reconsidered by Frauenfelder in Chapter 5 from the perspective of word recognition. Frauenfelder systematically compares the architecture and (simulation) behavior of TRACE with that of two influential verbal models of auditory word recognition, COHORT I and COHORT II, and with SHORTLIST, a hybrid connectionist model. Empirical evidence and theoretical considerations are presented to evaluate the different ways the models handle the positive and negative effects of competing lexical candidates in the time-course of auditory word recognition and the way they select the final candidate.

Many of the same issues with respect to representational units, interactivity, and competitor effects return in Chapter 6 on visual word recognition. Grainger and Dijkstra give an overview of the empirical literature in search for relevant constraints on computer models. Two implemented computer models are described next, the IA model of McClelland and Rumelhart and Seidenberg and McClelland's more recent PDP-model. Even though both are connectionist models, comparing them is not easy, since they differ on several dimensions, particularly with respect to the experimental tasks to which they have been applied.

While the models described by Grainger and Dijkstra can handle only the recognition of morphologically monomorphemic words, Baayen and Schreuder in Chapter 7 consider models that deal with morphologically complex words. The chapter gives a brief but useful overview of morphology, and proceeds to the discussion of linguistically oriented models, taking into account linguistic considerations, as well as psycholinguistic models that take empirical evidence into account. Apart from a discussion of verbal psycholinguistic models (such as the AAM and MRM), the chapter pays attention to the recently implemented connectionist PDP-model of Gasser.

In Chapter 8, the book's attention moves up to the sentence level. Considering parsing from a more general psychological point of view than is usually done, Kempen discusses the need for parallel, incremental, and interactive sentence processing. Several linguistic and psycholinguistic parsers that have been proposed in the last decades are described and compared, among which Augmented Transition Networks (ATNs), shift-reduce parsers, Marcus's PARSIFAL, race-based parsers and, finally, Kempen and Vosse's Unification Space model. Kempen concludes his chapter by stating that he expects a future perspective for dynamically oriented parsers, perhaps in the guise of connectionist models.

Competition between AI-based linguistic models and connectionist based psycholinguistics models can also be observed in Chapter 9, in which Garnham contrasts Kintsch's construction-integration model of discourse comprehension with the connectionist model of text comprehension by Sharkey. The models are discussed after a review of theoretical constraints and empirical findings about the interrelationship of sentences with respect to inference processes, anaphorical expressions, and nonliteral meaning. Both implemented models are discussed in the light of the mental models theory of discourse.

1.2.4 Overview of Part III: Language production

Both Chapter 9 on discourse understanding and Chapter 10 on discourse planning are oriented more towards the visual modality (written text) than towards the spoken modality (conversation). However, in the area of discourse planning, the existing computer models are to a much larger extent application-oriented, as witnessed, for instance, by a system that answers questions about ships in a naval data base. In Chapter 10, Andriessen, De Smedt, and Zock review a number of psycholinguistic constraints on models of discourse planning, and then discuss three models that are application-oriented but show an increasingly psycholinguistic orientation: the TEXT schema-based model of McKeown, Hovy's RST-based text structurer, and a recent model of Moore and Paris on planning text for advisory dialogues.

In Chapter 11, De Smedt discusses how speakers compose their sentences in language production, a process technically called “grammatical encoding”, paying special attention to the incremental (piecemeal) syntactic construction of sentences. Interesting theoretical relationships can be found between this chapter and those on parsing and discourse planning. De Smedt first considers conceptual, lexical, and syntactic factors that affect the construction of sentence frames, and then discusses a number of computational models for incremental sentence generation: the AI-based models IPG, IPF, and POPEL, and a connectionist model called FIG.

The construction of sentence frames that De Smedt describes goes hand in hand with the selection of word material at an abstract level called the *lemma*. In Chapter 12, Roelofs gives an account on lemma retrieval in which he discusses a number of different views on how particular lexical items can be found for the conceptual content that needs to be uttered. He describes his own implemented computer model, a hybrid model which incorporates symbolic notions (such as flagging of current nodes), together with mathematical notions (application of a hazard rate) and connectionist notions (spreading activation). Support for the model is provided by reviewing empirical data.

In Chapter 13, Dell and Juliano describe two computational approaches that account for the process of phonological encoding, which is the attribution of a speech form to retrieved lemmas. Giving the verbal symbolic standard theory as background, Dell’s well-known spreading activation model is presented and compared with a recent learning distributed connectionist model. In this chapter, the authors focus on speech error evidence as a major empirical data source for modelling phonological encoding.

In Chapter 14, Boves and Cranen consider the last stage in language production, that of articulation. After explaining the basic concepts of speech articulation and speech acoustics, they describe how the actual execution of articulation commands by the speech apparatus can be captured by Articulatory Synthesis (as implemented by the *task-dynamics* model) and Terminal Analog Synthesis. The chapter makes clear we still have only rather limited knowledge about the intricacies of the speech production system. Furthermore, this chapter is to some extent complementary to that on Speech Perception by Massaro. For example, the authors also discuss the issue of what the basic unit(s) in speech are, but approach it from a more phonetically oriented angle. While Massaro argues for an important role of the syllable in speech perception, Boves and Cranen pay much attention to phonemic and subphonemic types of representations in speech production.

In Chapter 15, the last chapter of the book, Schomaker and van Galen also consider the last stage of language production, but in the written mode, dealing with the hand movements in the process of connected, cursive handwriting. The major empirical issues in this domain, often neglected in textbooks on Psycholinguistics, are first discussed within a general framework that is remarkably compatible with the language user framework presented earlier in the current chapter. In the light of this framework, the chapter explains an implemented symbolic model (named the Cursive Connections Grammar) that simulates the process of concatenating cursive characters. Next comes a discussion of two variants of a connectionist model that focuses on the actual trajectory formation (articulation) of individual letters. Though these models differ to some extent in focus and in content from the other psycholinguistic models presented in earlier chapters, similar issues arise with respect to the modelling of discrete or continuous time and value dimensions.

To conclude, the models in this book are samples stemming from different traditions. Traditional AI models have stimulated computational modelling from the sixties onwards. Linguistic information processing is specified in AI in terms of symbol manipulation (Chapter 2). From the eighties onward, the AI models have been getting tough competition from connectionist models, which consider information processing in terms of a brain-based metaphor (Chapter 3). The clash between these two types of models is evident in sometimes fierce discussions (cf. Seidenberg & McClelland, 1989, 1990, vs. Besner et al., 1990, and Coltheart et al., 1993; Levelt et al., 1991a, 1991b, vs. Dell & O'Seaghda, 1991; Fodor & Pylyshyn, 1988, Fodor & McLaughlin, 1990, vs. McClelland, 1993, Rumelhart, 1989, Seidenberg, 1994, Smolensky, 1988; Rumelhart & McClelland, 1986, MacWhinney & Leinbach, 1991, Plunkett & Marchman, 1991, vs. Lachter & Bever, 1988, Pinker & Mehler, 1989, Pinker & Prince, 1988; and, to some extent, Massaro, 1988, 1989, Massaro & Cohen, 1991, vs. McClelland, 1991, McClelland & Elman, 1986). In this book we do not take sides, but merely report on the achievements of individual models in particular subdomains.

As a consequence of the confrontation of radically different paradigms for human cognitive processing, the field has become sensitized to new creative modelling approaches that go beyond the established ones. Hybrid models are starting to appear that combine advantages and characteristics of two or more approaches (Gutknecht, 1992; Lehnert, 1991; Stone, in press). In this book, the number of such models is still rather limited. For example, the non-decompositional spreading activation model in Chapter 12 combines features from the classical symbolic, mathematical, and connectionist paradigms. Another example is the Unification Space model in Chapter 8, which incorporates a symbolic grammar, but also makes use of a new connectionist optimization method (simulated annealing) within the framework of an unexpected metaphor (that of chemistry and physics). In the literature at large, more and more models start to appear that deviate from the mainstream, describing cognitive and psycholinguistic models from as yet less familiar viewpoints such as adaptive dynamic systems theory (see, e.g., Smolensky, 1986, on harmony theory; Grossberg, 1980, on resonance; and Van Orden & Goldinger, in press, on covariant learning) and genetic algorithms (Holland, 1975).

Almost a century ago, brave pioneers contrived all sorts of amazing contraptions in an attempt to fulfill man's desire to fly. Perhaps, in our age, at the brink of the twenty-first century, computer models are the devices that stimulate man's imagination. And as before, of all the amazing models that are devised, only time can tell which will fall and which will fly.

1.3 References

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1991). *The adaptive character of thought.* Hillsdale, NJ: Lawrence Erlbaum.

Besner, D., Twilley, L., McCann, R. S., & Seergobin, K. (1990). On the association between connectionism and data: Are a few words necessary? *Psychological Review, 97,* 432–446.

Carroll, D. W. (1994). *Psychology of language*. Belmont, CA: Brooks/Cole.

Chandrasekaran, B., Hari Narayanan, N., & Iwasaki, Y. (1993). Reasoning with diagrammatic representations. *AI Magazine, 14 (2),* 49–56.

Chomsky, N. (1980). *Rules and representations*. Oxford: Blackwell.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100,* 589–608.

Cutler, A. (1981). Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990? *Cognition, 10,* 65–70.

Dell, G. S., & O'Seaghda, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review, 98,* 604–614.

Dijkstra, A., & Kempen, G. (1984). *Taal in uitvoering*. Groningen, The Netherlands: Wolters-Noordhoff.

Dijkstra, A., & Kempen, G. (1993). *Taalpsychologie*. Groningen, The Netherlands: Wolters-Noordhoff.

Fodor, J. A., & McLaughlin, B. (1990). Connectionism and the problem of systematicity: why Smolensky's solution doesn't work. *Cognition, 35,* 183-204.

Fodor, J. A., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28,* 3–71.

Garman, M. (1990). *Psycholinguistics*. Cambridge: Cambridge University Press.

Gentner, D., & Stevens, A. L. (Eds.), (1983). *Mental models*. Hillsdale, NJ: Lawrence Erlbaum.

Gernsbacher, M. A. (1994). *Handbook of Psycholinguistics*. Orlando: Academic Press.

Grossberg, S. (1980). How does the brain build a cognitive code? *Psychological Review, 87,* 1–51.

Gutknecht, M. (1992). The 'Postmodern Mind': Hybrid models of cognition. *Connection Science, 4,* 339–364.

Haarmann, H. J., & Kolk, H. H. J. (1991). A computer model of the temporal course of agrammatic sentence understanding: The effects of variation in severity and sentence complexity. *Cognitive Science, 15,* 49–87.

Hillgard, E. R., & Bower, G. H. (1975). *Theories of learning*. Englewood Cliffs, NJ: Prentice Hall.

Holland, J. H. (1975). *Adaptation in natural and artificial systems.* Ann Arbor, MI: University of Michigan Press.

Jacobs, A. M., & Grainger, J. (1992). Testing a semi-stochastic variant of the interactive activation model in different word recognition experiments. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1174–1188.

Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance, 20.*

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30,* 580–602.

Lachter, J., & Bever, T. (1988). The relation between linguistic structure and associative theories of language learning: A constructive critique of some connectionist learning models. *Cognition, 28,* 195–247.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lehnert, W. (1991). Symbolic/subsymbolic sentence analysis – Exploiting the best of two worlds. In J. A. Barnden & J. B. Pollack (Eds.) *High-level connectionist models.* Norwood, NJ: Ablex.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

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

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, Th., & Havinga, J. (1991b). Normal and deviant lexical processing: Reply to Dell and O'Seaghda (1991). *Psychological Review, 98,* 615–618.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Freeman.

Massaro, D. W. (1988). Some criticisms of connectionist models of human performance. *Journal of Memory and Language, 27,* 213–234.

Massaro. D. W. (1989). Testing between the TRACE model and the Fuzzy Logical Model of Perception. *Cognitive Psychology, 21,* 398–421.

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology, 23,* 558–614.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition, 40,* 121–157.

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology, 23*, 1–44.

McClelland, J. L. (1993). Toward a theory of information processing in graded, random, and interactive networks. In D. E. Meyer & S. Kornblum (Eds.), *Attention & performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 655–688). Cambridge, MA: MIT Press.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

McCloskey, M. (1991). Networks and theories: The place of connectionism in cognitive science. *Psychological Science, 2,* 387–395.

Miller, G. (1991). *The science of words*. New York: Freeman & Company.

Miller, G.A., Galanter, E., & Pribram, K.H. (1960). Plans and the structure of behavior. London: Holt, Rinehart, & Winston.

Newell, A., Rosenbloom, P. S., & Laird, J. E. (1989). Symbolic architectures for cognition. In M. I. Posner, (Ed.), *Foundations of Cognitive Science*. Cambridge MA: MIT Press. (pp. 93–131)

Patterson, K. E., Seidenberg, M. S., & McClelland, J. L. (1989). Connections and disconnections: Acquired dyslexia in a computational model of reading processes. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 131–181). Oxford: Clarendon.

Parisi, D., & Burani, C. (1988). Observations on theoretical models in neuropsychology of language. In G. Denes, C. Semenza, & P. Bisacchi (Eds.), *Perspectives on Cognitive Neuropsychology*. London: Lawrence Erlbaum.

Pinker, S., & Mehler, J. (1989). *Connections and symbols*. Cambridge, MA: MIT Press.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a Parallel Distributed Processing model of language acquisition. *Cognition, 28,* 73–193.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multilayered perceptron: Implications for child language acquisition. *Cognition, 38,* 43–102.

Posner, M. I. (Ed.). (1989). *Foundations of Cognitive Science*. Cambridge MA: MIT Press.

Pylyshyn, Z. W. (1986). *Computation and cognition: Toward a foundation of cognitive science*. Cambridge, MA: Bradford Books.

Pylyshyn, Z. W. (1989). Computing in cognitive science. In M. I. Posner (Ed.), *Foundations of Cognitive Science* (pp. 49–91). Cambridge, MA: MIT Press.

Rumelhart, D. E. & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, and the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition: Vol 2. Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press.

Rumelhart, D. E. (1989). The architecture of mind: A connectionist approach. In M. I. Posner (Ed.), *Foundations of Cognitive Science* (pp. 133–159). Cambridge, MA: MIT Press.

Seidenberg, M. S. (1994). Language and connectionism: the developing interface. *Cognition, 50,* 385-441.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Seidenberg, M. S., & McClelland, J. L. (1990). More words but still no lexicon. Reply to Besner et al. (1990). *Psychological Review, 97,* 447–452.

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing* (Vol. 1, pp. 194–281). Cambridge, MA: MIT Press.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11*, 1-23.

Stone, G. O. (in press). Combining connectionist and symbolic properties in a single process. In R. Corregan, G. Iverson, & S. Lima (Eds.). *The reality of linguistic rules.*

Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 744–774.

Taylor, I., & Taylor, M. M. (1990). *Psycholinguistics: Learning and using language.* Englewood Cliffs, NJ: Prentice-Hall.

Van Heuven, W., Dijkstra, A., & Grainger, J. (in preparation). Neighborhood effects in bilingual word recognition: The BIA-model and experiments.

Van Orden, G. C., & Goldinger, S. D. (in press). The interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance*.

Zwitserlood, P. (1994). Access to phonological-form representations in language comprehension and production. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 83–106). Hillsdale, NJ: Erlbaum.