Psycholinguistics, Computational

Advanced article

Richard L Lewis, University of Michigan, Ann Arbor, Michigan, USA

CONTENTS

Introduction
Models of lexical processing
Models of comprehension
Models of production
Models of acquisition
Current directions

Computational psycholinguistics seeks to build theories of human linguistic processes that take the form of working computational models. These models address processes ranging from word recognition to discourse comprehension, and produce behavior that constitutes predictions to be compared to human data.

INTRODUCTION

Computational psycholinguistics seeks to build theories of human linguistic processes that take the form of implemented computational models. These models are intended to explain how some psycholinguistic function is accomplished by a set of primitive computational processes. The models perform a psycholinguistic task and produce behavior that can be interpreted as a set of predictions to be compared to human data. As such, computational psycholinguistics is a paradigmatic example of cognitive modeling more generally. One problem with the label *computational psycholinguistics* is the implication that there is something that can be identified as *noncomputational* psycholinguistics. This is not presently the case: all psycholinguistic theories are, at some level, assertions about computational processes. Computational psycholinguistics is distinguished from other forms of cognitive modeling by its domain (not its techniques), and it is distinguished from other forms of psycholinguistic theorizing by its focus on producing functioning computational mechanisms that embody an explicit process model. The remainder of this article is devoted to reviewing the state of computational modeling in several of the major subfields of psycholinguistics.

MODELS OF LEXICAL PROCESSING

The most influential computational models in psycholinguistics have been those focused on word-level processes, in particular, spoken and visual word recognition. In fact, there are currently no major psycholinguistic theories of word recognition that do not take the form of a computational model. Competing theories are routinely tested by running the corresponding computational models to determine how well the models' behavior fits human data. At some level, there is significant theoretical convergence. All of the models of lexical processing are activation-based: lexical access is modeled as a dynamic process of modulating the activation of patterns of representation that encode information associated with specific lexical (or morphological) items. However, the models differ dramatically along many important architectural dimensions, such as the degree of top-down feedback and the nature of the computational principles determining the dynamic activation patterns.

Spoken Word Recognition

Models of spoken word recognition must satisfy a number of challenging functional and empirical constraints. These include: speech occurs in time, with no clear boundaries between words or phonemes, which may in fact overlap; there are effects of both left and right context on word recognition; lower-level phoneme identification may depend on higher-level lexical information; and there may be considerable noise in the environment (McClelland and Elman, 1986).

Current computational models of word recognition are extensions of ideas first put forward explicitly in the COHORT theory of speech perception (Marslen-Wilson and Tyler, 1980). The key principles in COHORT are that the initial sound of a word establishes a *cohort* or candidate set of possible words beginning with that sound, and this candidate set is incrementally narrowed down in real time as subsequent acoustic input arrives. Word recognition is achieved when the candidate set is narrowed to one, which may occur before the end of the word.

The TRACE model of McClelland and Elman (1986) provides an explicit computational realization of these basic ideas in COHORT, while addressing some of its most critical shortcomings. In particular, COHORT had no clear account of how word boundaries were identified in the continuous speech stream, and it assumed accurate bottom-up identification of phonemes. TRACE is an interactive-activation architecture with bidirectional excitatory connections between nodes representing acoustic features, phonemes, and words. Each time slice of input occupies a separate part of the input vector, and there are multiple copies of phoneme and word detectors centered over every three time slices. There are also inhibitory links within levels between mutually incompatible words or phonemes; thus, word and phoneme recognition is a competitive process. This competition and the distribution of multiple detectors across the network permits the model to recognize words without clear boundaries known in advance. The bidirectional nature of the within-level connections provides a way for the lexicon to directly influence the perception of lower-level phonemic and acoustic features.

TRACE has been used to account for a wide range of psycholinguistic data on word recognition, including the signature data originally used to motivate COHORT. Among these phenomena are: the effect of lexical context on phoneme recognition and its modulation by factors such as ambiguity; phonotactic rule effects on phoneme recognition, and their modulation by specific lexical items (phonotactic rules determine what sequences of phonemes are possible in a language); and the categorical nature of phoneme perception. TRACE was one of the prominent early successes of the PDP (parallel distributed processing) approach to modeling cognition and perception, and played a significant role in establishing the viability of the PDP paradigm.

TRACE has been challenged on both empirical and theoretical grounds, most notably by the Shortlist model of Norris (1994). A number of empirical studies have directly tested the assumption of top-down feedback in TRACE and yielded results more consistent with a purely bottom-up architecture in which phoneme recognition is autonomous and receives no feedback from lexical recognizers. For example, certain top-down lexical influences are dependent on using degraded stimuli, though TRACE should predict the effects in undegraded stimuli as well. Norris also argued that the TRACE architecture is implausible because it assumes the duplication of the entire network of lexical recognizers across multiple time slices. Shortlist is a purely bottom-up model that avoids the duplication of lexical recognizers by separating the process of generating candidate words (the 'shortlist') and the process of resolving identification via lexical competition.

Visual Word Recognition: Lexical Naming and Decision

Current prominent models of visual word recognition also take the form of computational models. One of the most influential of these models, the connectionist model of Seidenberg and McClelland (1989) (henceforth SM89), is a descendant of the McClelland and Rumelhart (1981) interactive activation model of word perception, which used localist word, letter, and feature units with hand-coded connections. SM89 builds on this earlier model but adopts distributed representations of both orthographic and phonological information. The model is a feedforward network with one hidden layer interposed between orthographic and phonological units. The connections between units were trained by back propagation on a word-naming task. The model accounts for several phenomena in word-naming, including differences among regular and exception words and differences in word-naming and lexical decision tasks. Because the model exhibits a gradual learning curve, it was also used to simulate the behavior of children acquiring word recognition skills.

One of the major debates in theories of word recognition is whether or not there is a single processing route from print to speech, or dual processing routes – separate lexical and nonlexical routes. The SM89 model is a clear example of a single-route architecture, and has come under sharp criticism from proponents of dual-route architectures. For example, Coltheart *et al.* (1993) note that the SM89 model actually performs more poorly on nonwords than humans do. Dual-route architectures are well suited to handing nonwords because the nonlexical route implements a general rule-based system that converts letter strings to strings of phonemes. Coltheart *et al.* also criticize the SM89 model for its inability to account for the dissociations evident in pure developmental surface dyslexia: normal nonword reading accuracy accompanied by gross impairments in reading exception words. Coltheart *et al.* offer a modular dual-route computational model, the Dual-Route Cascaded Model, which incorporates a learning algorithm for inducing the general pronunciation rules from examples (it was tested on the same letter-string/phone-string pairs used by SM89). Although Coltheart *et al*. did not commit to the details of the lexical route, they suggest that something like the original McClelland and Rumelhart (1981) model may be an appropriate realization of that part of the word-naming system.

The debate surrounding dual-route and single-route architectures continues, with data from various forms of dyslexia playing an increasingly important role. The dual-route models have evolved to include explicit accounts of both reading aloud and lexical decision (Coltheart *et al*., 2001), and the connectionist models have evolved away from feedforward networks towards recurrent attractor networks that better handle generalization (Plaut *et al*., 1996).

Lexical Ambiguity Resolution: Processing Words in Context

One of the key lessons learned from several decades of attempting to program computers to process natural language is that massive local ambiguity is pervasive at all levels of linguistic representation. This is clearly evident in lexical processing, in which individual words are often associated with multiple syntactic and semantic senses, some mutually inconsistent, some partially inconsistent. Many of the theoretical themes noted above in word recognition are important in ambiguity resolution as well, in particular, the degree of autonomy or interaction present in initial lexical access. Differing positions on this issue distinguish the major theories of ambiguity resolution: *selective access models*, most closely associated with interactive theories, assume that contextual information provides direct top-down influence on initial sense activation; *ordered access models* assume that different senses are accessed in order of frequency of use; *exhaustive access* models, most closely associated with modular theories, assume that all senses are autonomously and exhaustively accessed in parallel; and *hybrid* models assume some combined effects of context and frequency.

In contrast to word recognition, the major theories of lexical ambiguity resolution are not strongly identified with specific implemented computational models (for reasons discussed below). However, there have been attempts to build detailed comprehensive computational models. One of the most successful is Kawamoto's (1993) recurrent connectionist model of ambiguity resolution. In this model, each lexical entry is represented by a pattern of activity over a 216-bit vector divided into separate subvectors representing a word's spelling, pronunciation, part of speech, and meaning. The network is trained with a simple error-correction algorithm by presenting it with the lexical patterns to be learned. The result is that these patterns become *attractors* in the 216-dimensional representational space. The network is tested by presenting it with just *part* of a lexical entry (e.g. its spelling pattern) and noting how long various parts of the network take to settle into a coherent pattern corresponding to a particular lexical entry. Kawomoto used these settling times to predict reading times, lexical decision times, and semantic access times. The model accounts for a wide range of phenomena, including frequency effects on processing of unambiguous and ambiguous words, context interactions with frequency, and the effect of task on the relative difficulty of processing ambiguous versus unambiguous words.

MODELS OF COMPREHENSION

Language comprehension involves more than the identification and disambiguation of words; the meanings of these parts must be pieced together in real time to yield the meanings of the sentences and the discourse. The state-of-the-art in computational linguistics and artificial intelligence places an upper bound on the field's ability to develop functional theories of comprehension processes. The best understood of these processes computationally and psychologically is syntactic parsing, the incremental assignment of grammatical structure to a string of words. Syntactic parsing is often assumed (though not universally) to be a necessary precursor to assigning a semantic interpretation.

Parsing

The major computational problem in parsing is how to handle local ambiguity. In fact, the prominent theories of sentence processing are actually theories of ambiguity resolution, and are distinguished by the positions they take on the key architectural questions surrounding ambiguity resolution. These include: are multiple structures computed and maintained in parallel at ambiguous points, or does the parser commit to a single structure immediately? What determines what structures the parser prefers when faced with ambiguity (e.g. referential discourse context, structural complexity, frequency of usage)? How do syntactic and lexical ambiguity resolution interact?

Two of the most influential models of sentence processing take opposing positions on most of these issues (though many of the issues are orthogonal).

Frazier's (1987) Garden Path Model asserts that the parser computes and pursues a single structure at ambiguous points, and that this initial structure is computed on the basis of general phrase structure rules without appeal to frequency, context, or detailed lexical information. Instead, structural simplicity is the principle that determines which structure is pursued in the case of local ambiguity. In contrast, the Constraint-based Lexicalist approach (MacDonald *et al*., 1994) claims that parsing is a constraint-satisfaction process that uses multiple information sources (or constraints), including context and detailed lexical information, without special architectural priority given to any particular constraint.

In sharp contrast to theories of word recognition, the dominant theories of sentence processing have not been strongly identified with specific computational models. (For example, the Garden Path Model was not implemented until 17 years after it was introduced (Spivey and Tanenhaus, 1998).) Among the earliest influential computational models were Marcus's (1980) wait-and-see parser, and the Wanner and Maratsos (1978) augmented transition network (ATN) grammar, which briefly contended with the Garden Path Model as a framework for understanding ambiguity resolution. Nevertheless, implemented computational models of sentence processing largely dropped from the scene in the 1980s.

Understanding why this happened will help place current parsing models in context. First, the early success of the Garden Path Model and the rise of modularity as a central theoretical theme in cognitive science jointly led the field to focus on modularity as the key architectural issue in sentence processing, and on ambiguity resolution as the key phenomenon providing insight into that issue. Second, Minimal Attachment is an extremely simple and practical theory – it can be stated in a few sentences and easily used to derive predictions cross-linguistically (once the underlying syntactic structures have been agreed upon). Computational models offered little advantage over such a theory, given this relatively narrow empirical and theoretical focus.

Two developments in the field are now leading researchers to develop more computational models. One is the need to provide more comprehensive, integrated accounts of sentence processing. Modularity is but one of several important architectural issues (Lewis, 2000), and computational modeling provides a way to develop and test interactions among components in a more functionally complete architecture. For example, computational models figure prominently among recent attempts to provide integrated accounts of both garden-path effects and working memory complexity effects in unambiguous constructions (Gibson, 1998; Lewis, 2000; Vosse and Kempen, 2000). Computational modeling also provides a way to import theoretical constraints from other areas of cognitive psychology, as in the Just and Carpenter (1992) working memory-constrained model.

A second development leading to more computational models is the rise of the constraint-based theories of sentence processing noted above. While these theories were initially proposed without associated computational models, it has become clear that the nature of these theories demands that they be formulated and tested as precise computational models. Several activation-based/connectionist models (e.g. Spivey and Tanenhaus, 1998) have been developed in the constraint-based framework.

Unlike computational models of word-level processes, which are almost exclusively the domain of connectionism, current computational theories of sentence processing are a mix of symbolic, connectionist, probabilistic, and hybrid models. As a class, the symbolic models tend to account for more complex cross-linguistic data, such as phenomena in head-final languages (e.g. Konieczny *et al*., 1997; Sturt and Crocker, 1996). However, recent models based on recurrent networks are attempting to push connectionist models in the direction of handling more complex syntactic structures, including difficult center-embeddings (Christiansen and Chater, 1999; Tabor *et al*., 1998). Several hybrid models are also under development, which have the promise of combining some of the strengths of both approaches (Jurafsky, 1996; Just and Carpenter, 1992; Lewis, forthcoming; Stevenson, 1994; Vosse and Kempen, 2000).

Discourse Processing

Processing running discourses of sentences in a text or verbal exchanges between interlocutors requires keeping track of multiple related levels of information (including, at least, the linguistic structure of the utterances, the goals and intentions of the participants, and the content of what is being discussed). Several major discourse processing theories have long been associated with implemented computational models. These include the Centering theory of Grosz and colleagues (Grosz *et al*., 1995), which provides an explicit algorithm for keeping track of attentional shifts among discourse entities and binding referring expressions to these entities. The theory makes predictions about preferential patterns of pronominal reference that have been tested in reading time experiments (Gordon *et al*., 1993).

Another influential model is the Construction-Integration (CI) architecture of Kintsch and colleagues (Kintsch, 1998). Comprehension in the CI architecture is an activation-based process that proceeds in two phases. The *construction* phase produces local sentence-level propositions using simple, context-independent rules. The *integration* phase uses a constraint satisfaction process to integrate the possibly incoherent set of local propositions into a coherent whole organized by higher-level macropropositions. Many of the CI model's predictions about anaphora resolution, word identification, and the generation and retrieval of macropropositions have been empirically confirmed (Kintsch, 1998).

MODELS OF PRODUCTION

The dominant psycholinguistic theories of production are now associated with implemented computational models. Most psycholinguistic theories of production focus on the final stages of production: producing an ordered set of phonemes corresponding to some (given) intended utterance. (In contrast, much work on production in computational linguistics and artificial intelligence is focused on the functionally more difficult processes of higher-order discourse and speech act planning.) The theoretical landscape is quite similar to theories of lexical processing: all the models are activation-based, but differ in their assumptions about the nature of interaction between independent levels of representation. Among the best-known models are those of Dell (Dell *et al*., 1997) and Levelt (Levelt *et al*., 1999), which take opposing positions along this dimension. The Dell model is an interactive-activation-based theory that takes an ordered set of word units as input and generates a string of phonemes. Most of the important phenomena accounted for by the model are speech errors, including perseverations (e.g. *beef needle soup*) and anticipations (e.g. *cuff of coffee*). Dell's model consists of a network of word units (lemmas) and phoneme units and bidirectional links between word units and their constituent phonemes. The signature phenomenon accounted for by the feedback from phonemes to words is the statistical overrepresentation of mixed errors, such as saying *rat* when the intention is *cat*. When the word node for *cat* is active, the phoneme segments /k/, /æ/, and /t/ are activated. The latter two segments then feed activation to *rat*, which may already be above baseline due to a semantic association.

The WEAVER++ model (Levelt *et al*., 1999) is also activation-based, but eliminates bidirectional connections. Processing is staged in strictly feedforward fashion, starting with conceptual preparation (not implemented), and proceeding to lexical selection, morphological and phonological encoding, phonetic encoding, and finally articulation. Unlike most other production theories, the WEAVER++ model accounts primarily for reaction time (RT) data, and was developed exclusively on the basis of RT data from simple production paradigms such as picture naming. However, Levelt and colleagues have also shown that the model can account for some speech errors as well, including those used to motivate the bidirectional connectivity in the strongly interactionist models.

MODELS OF ACQUISITION

With one prominent exception noted below, computational models have only recently begun to play an important role in theorizing about language acquisition. A fundamental difficulty facing the development of serious computational models of acquisition is that the input to such models must generally be a large corpus of utterances *in context*. Although large computer databases of naturally occurring text and speech are now readily available, such databases currently lack a component that nearly all acquisition theories assume is necessary: some representation of the context in which the utterance occurs. For this reason, much computational modeling of grammar acquisition is currently done using small-scale, artificially created grammars or lexicons, in small-scale, artificial domains (Feldman *et al*., 1996).

However, current speech and text databases are well suited to exploring *distributional* theories of acquisition. For example, certain kinds of lexical and syntactic information can be determined from purely distributional analyses (Cartwright and Brent, 1997). One important example is specific verb subcategorization frames, which play a critical role in all modern syntactic theories and sentence comprehension theories. Computational models of speech segmentation have also been developed that learn to identify word boundaries from exposure to continuous speech (Christiansen *et al*., 1998).

By far the most controversial and influential computational acquisition model is the Rumelhart and McClelland (1986) (henceforth RM86) connectionist model of the acquisition of the past tense form of English verbs. Past tense inflection acquisition has served as a kind of *Drosophila* for research on the mechanisms underlying apparently rule-governed linguistic behavior, and lies at the center of a much broader debate on connectionism and language. The RM86 model was proposed as an alternative account to the traditional view that the past tense form of English verbs is formed by dual routes: an abstract rule that handles all regular forms by adding *–ed* to a stem, and a memory that contains a list of irregular exception words (such as *ran*). The connectionist model instead proposed a single processing route, implemented as a feedforward network with a single hidden layer, and no explicit representation of a rule. The network was trained on 460 pairs of root and inflected forms. The network reproduced the well-known U-shaped performance curve often taken as prima facie evidence for the formation of a general *–ed* rule: children initially do not make overgeneralization errors (e.g. saying *runned* for *ran*), but then go through a period of apparently over-applying the general rule, and finally recover to adult levels of performance. Crucially, the network also generalized and transferred appropriately to novel low-frequency verbs (e.g. the network correctly produced *wept* as the past tense of *weep*), capturing subregularities among the irregular words in the corpus.

Every aspect of this work has come under sharp criticism, including the content of the artificial database on which RM86 trained their original network, the empirical robustness of the U-shaped curve itself, and the use of connectionist architectures more generally as accounts of human linguistic and cognitive performance (Marcus, 1996; Pinker and Prince, 1988). Some of these criticisms have been addressed in revisions to the model (MacWhinney and Leinbach, 1991), but new empirical evidence from adult processing has also accumulated in favor of the dual-route view (Marslen-Wilson and Tyler, 1998).

CURRENT DIRECTIONS

A number of short-term and long-term theoretical directions are evident in this review. One overarching trend is clear: computational modeling is playing an increasingly important role in theorizing in all subfields of psycholinguistics. There are several reasons for this, all related to theoretical trends in psycholinguistics more generally. There are four trends in particular that are likely to continue in the near term. First, there is a gradual move towards providing more *integrated accounts* of multiple components of linguistic processing. For example, several computational models now combine theories of lexical ambiguity resolution and sentence processing, or ambiguity resolution and working memory (e.g. Kintsch, 1998). Second, there is an increasing move towards developing theories that are *jointly constrained by processing and acquisition data* (e.g. Seidenberg and McClelland, 1989). Accompanying this trend is a growing reliance on large machine-readable corpora to test models that have some role for linguistic experience. Third, theories of normal linguistic performance are increasingly constrained by *neuropsychological data* from patients with linguistic deficits due to brain damage. Computational models of intact performance can be ‘lesioned’ and tested against both normal and patient data (e.g. Plaut *et al.*, 1996). Fourth, there is increasing convergence in all subfields of psycholinguistics towards *continuous activation-based* models of processing. These include parallel distributed processing approaches, but also many activation-based symbolic models.

There are also some emerging trends that will most likely play out over the longer term. These include increasing attempts to integrate psycholinguistic models with other process theories in cognitive psychology, such as detailed models of memory and skill, and increasing convergence with efforts in computational linguistics as both fields attempt to tackle functionally difficult areas such as word sense disambiguation and robust parsing. These latter efforts will naturally result in greater contact with linguistic theory. In particular, linguistic theories which prove to be important in the development of scalable and robust speech and natural language systems will be incorporated in psycholinguistic models that place a premium on functionality and scalability.

References

Cartwright TA and Brent MR (1997) Syntactic categorization in early language acquisition: formalizing the role of distributional analysis. *Cognition* **63**(2): 121–170.

Christiansen MH, Allen J and Seidenberg MS (1998) Learning to segment speech using multiple cues: a connectionist model. *Language and Cognitive Processes* **13**(2–3): 221–268.

Christiansen MH and Chater N (1999) Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science* **23**(2): 157–205.

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

Coltheart M, Rastle K and Perry C (2001) DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review* **108**(1): 204–256.

Dell GS, Burger LK and Svec WR (1997) Language production and serial order: A functional analysis and a model. *Psychological Review* **104**(1): 123–147.

Feldman J, Lakoff G, Bailey D, Narayanan S and Regier T (1996) L0: the first five years. *Artificial Intelligence Review* **10**: 103–129.

Frazier L (1987) Sentence processing: a tutorial review. In: Coltheart M (ed.) *Attention and Performance XII: The Psychology of Reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gibson EA (1998) Linguistic complexity: locality of syntactic dependencies. *Cognition* **68**: 1–76.

Gordon PC, Grosz BJ and Gilliom LA (1993) Pronouns, name, and the centering of attention in discourse. *Cognitive Science* **17**(3): 311–347.

Grosz BJ, Weinstein S and Joshi AK (1995) Centering: a Framework for modeling the local coherence of discourse. *Computational Linguistics* **21**(2): 203–225.

Jurafsky D (1996) A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science* **20**(2): 137–194.

Just MA and Carpenter PA (1992) A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* **99**(1): 122–149.

Kawamoto AH (1993) Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language* **32**: 474–516.

Kintsch W (1998) *Comprehension: A Paradigm for Cognition*. Cambridge, UK: Cambridge University Press.

Konieczny L, Hemforth B, Scheepers C and Strube G (1997) The role of lexical heads in parsing: evidence from German. *Language and Cognitive Processes* **12**(2/3): 307–348.

Levelt WJM, Roelofs A and Meyer AS (1999) A theory of lexical access in speech production. *Behavioral and Brain Sciences* **22**(1): 1–38.

Lewis RL (2000) Specifying architectures for language processing: process, control, and memory in parsing and interpretation. In: Crocker M, Pickering M and Clifton C (eds) *Architectures and Mechanisms for Language Processing*. Cambridge, UK: Cambridge University Press.

Lewis RL (forthcoming) *Cognitive and Computational Foundations of Sentence Processing*. Oxford, UK: Oxford University Press.

MacDonald MC, Pearlmutter NJ and Seidenberg MS (1994) The lexical nature of syntactic ambiguity resolution. *Psychological Review* **101**: 676–703.

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

Marcus GF (1996) Why do children say 'breaked'? *Current Directions in Psychological Science* **5**: 81–85.

Marcus MP (1980) *A Theory of Syntactic Recognition for Natural Language*. Cambridge, MA: MIT Press.

Marslen-Wilson W and Tyler LK (1980) The temporal structure of spoken language understanding. *Cognition* **8**: 1–71.

Marslen-Wilson W and Tyler LK (1998) Rules, representations, and the English past tense. *Trends in Cognitive Sciences* **2**(11): 428–435.

McClelland JL and Elman JL (1986) The TRACE model of speech perception. *Cognitive Psychology* **18**: 1–86.

McClelland JL and Rumelhart DE (1981) An interactive activation model of context effects in letter perception: 1. An account of the basic findings. *Psychological Review* **88**: 375–407.

Norris D (1994) Shortlist: a connectionist model of continuous speech recognition. *Cognition* **52**: 189–234.

Pinker S and Prince A (1988) On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition* **28**(1–2): 73–193.

Plaut DC, McClelland JL, Seidenberg M and Patterson KE (1996) Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review* **103**: 56–115.

Rumelhart DE and McClelland JL (1986) On learning the past tense of English verbs. In: McClelland JL and Rumelhart DE (eds) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT Press.

Seidenberg M and McClelland JL (1989) A distributed developmental model of word recognition and naming. *Psychological Review* **96**: 523–568.

Spivey MJ and Tanenhaus MK (1998) Syntactic ambiguity resolution in discourse: modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition* **24**: 1521–1543.

Stevenson S (1994) Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research* **23**(4): 295–322.

Sturt P and Crocker M (1996) Monotonic syntactic processing: a cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes* **11**: 449–494.

Tabor W, Juliano C and Tanenhaus MK (1998) Parsing in a dynamical system: an attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes* **12**: 211–272.

Vosse T and Kempen G (2000) Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition* **75**: 105–143.

Wanner E and Maratsos M (1978) An ATN approach to comprehension. In: Halle M, Bresnan J and Miller GA (eds) *Linguistic Theory and Psychological Reality*. Cambridge, MA: MIT Press.

Further Reading

Clifton C and Duffy SA (2001) Sentence and text comprehension: roles of linguistic structure. *Annual Review of Psychology* **52**: 167–196.

Crocker MW (1996) *Computational Psycholinguistics: An Interdisciplinary Approach to the Study of Language*. Dordrecht: Kluwer Academic.

Elman JL (1990) Finding structure in time. *Cognitive Science* **14**: 179–211.
Gernsbacher MA (ed.) (1994) *Handbook of Psycholinguistics*. San Diego, CA: Academic Press.
Lewis RL (1999) Cognitive modeling, symbolic. In: Wilson RA and Keil FC (eds) *The MIT Encyclopedia of Cognitive Science*. Cambridge, MA: MIT Press.
McClelland JL (1999) Cognitive modeling, connectionist. In: Wilson RA and Keil FC (eds) *The MIT Encyclopedia of Cognitive Science*. Cambridge, MA: MIT Press.
Norris D (1999) Computational psycholinguistics. In: Wilson RA and Keil FC (eds) *The MIT Encyclopedia of Cognitive Science*. Cambridge, MA: MIT Press.

Psychology: Experimental Methods

Introductory article

Robert W Proctor, Purdue University, West Lafayette, Indiana, USA
EJ Capaldi, Purdue University, West Lafayette, Indiana, USA
Kim-Phuong L Vu, Purdue University, West Lafayette, Indiana, USA

CONTENTS

Use of experiments in psychology
Control in psychology experiments
Between-subjects and within-subjects designs
Complex designs and interactions
Statistical significance testing
Varieties of scientific methods

Experimental methods in psychology are the procedures used to isolate the effects of manipulations on behavioral measures.

USE OF EXPERIMENTS IN PSYCHOLOGY

Experimental methods have been associated with psychology since the field's beginnings. The founding of psychology as a science is usually dated to 1879, when Wilhelm Wundt established the first laboratory devoted to experimental investigation of psychological phenomena. This association of psychology with the experimental method and its emphasis on control of the environment is what enabled the discipline to make the claim of being a science. The scientific approach provides a more objective method for establishing facts and evaluating alternative possible explanations. Throughout its history, experimentation has remained the central method of psychology, although it has not been without its critics, and non-experimental methods such as naturalistic observation and survey research have come into increasingly wide use.

Experiments can be conducted with humans or animals. The specific population that is studied will depend on several factors, including the topic with which the research is concerned, the theoretical predispositions of the researcher, the specific methods that are feasible with a particular population (e.g. humans cannot be lesioned, but nonhumans cannot provide verbal reports), and the fact that more control can be exerted over a laboratory animal's history and environment than can be extended over a human's. Much psychological research in the late nineteenth century used human subjects, in part because researchers had an interest in the subjective experience of perceptual events. Beginning in the early twentieth century with the behaviorist movement, the use of animals increased. Because the learning and conditioning principles studied by the behaviorists were considered to be generalizable across species, much of the research focused on rats and pigeons, animals that can be studied easily.

Research on humans continued to be conducted throughout this period, but a major renewal of interest occurred with the advent of contemporary cognitive psychology in the 1950s. Most experimental research on humans is conducted in laboratory settings with undergraduate psychology students. One concern with such research is the extent to which the principles derived from it