287

ENCYCLOPEDIA OF COGNITIVE SCIENCE

2000

Computational Psycholinguistics

Richard L. Lewis

Associate Professor, Departments of Psychology, Linguistics, and Electrical Engineering & Computer Science

525 E. University, Department of Psychology, Ann Arbor, MI 48109

Phone: (734) 763-1466 Fax: (734) 763-7480 Email: rickl@umich.edu

University of Michigan (Michigan) USA

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 behaviour 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 behaviour 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 modelling more generally. One problem with the label *computational psycholinguistics* is the implication that there is something that can be identified as *non-computational* 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 modelling by its domain (not its techniques), and it is distinguished from other forms of psycholinguistic theorising 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' behaviour fits human data. At some level, there is significant theoretical convergence. All of the models of lexical processing are activation-based: lexical access is modelled 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 & 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 & 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 & Elman (1986) provides an explicit computational realisation 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 bi-directional 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 centred 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 recognise words without clear boundaries known in advance. The bi-directional 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 modelling 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 recognisers. 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 recognisers across multiple time slices. Shortlist is a purely bottom-up model that avoids the duplication of lexical recognisers 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 descendent 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 backpropagation 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 decisions tasks. Because the model exhibits a gradual learning curve, it was also used to simulate the behaviour of children acquiring word recognition skills.

One of the major debates in theories of word naming recognition is whether or not there is a single processing route from print to speech, or dual processing routes—separate lexical and non-lexical 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 non-lexical route implements a general rule-based system that converts letter strings to strings of phonemes. Coltheart et al. also criticise 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 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 realisation 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, Rastle, & Perry, 2001), and the connectionist models have evolved away from feed-forward networks toward recurrent attractor networks that better handle generalisation (Plaut, McClelland, Seidenberg, & Patterson, 1996).

Lexical ambiguity resolution: Processing words in context

One of the key lessons learned from 40 years 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 Kawomoto'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 vs. 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, Pearlmutter, & Seidenberg, 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, *Minimal Attachment* was not implemented until seventeen years after it was introduced (Spivey & Tanenhaus, 1998).) Among the earliest influential computational models were Marcus's (1980) wait-and-see parser, and the Wanner and Maratos (1978) augmented transition network (ATN) grammar, which briefly contended with Minimal Attachment 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 Minimal Attachment 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 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 modelling 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 & Kempen, 2000). Computational modelling 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 recently 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 & 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 (1997); Sturt (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 & Chater, 1999; Tabor, Juliano, & Tanenhaus, 1998). Several hybrid models are also under development, which have the promise of combining some of the strengths of both approaches (Jurafsky, 1996; Just & Carpenter, 1992; Lewis, forthcoming; Stevenson, 1994; Vosse & 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, Weinstein, & Joshi, 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, Grosz, & Gilliom, 1993).

Another influential model is the Construction-Integration (CI) architecture of Kintsch and colleagues on (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, Burger, & Svec, 1997) and Levelt (Levelt, Roelofs, & Meyer, 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, e.g., 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 feed-forward 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, WEAVER++ model accounts primarily for reaction time 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 bi-directional 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 theorising 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 natural-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 modelling of grammar acquisition is currently done using small scale, artificially created grammars or lexicons, in small scale, artificial domains (Feldman, Lakoff, Bailey, Narayanan, & Regier, 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 & 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, Allen, & Seidenberg, 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 centre 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 feed-forward 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 over-generalisation 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 generalised 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 & Prince, 1988). Some of these criticisms have been addressed in revisions to the model (MacWhinney & Leinbach, 1991), but new empirical evidence from adult processing has also accumulated in favour of the dual-route view (Marslen-Wilson & 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 modelling is playing an increasingly important role in theorising 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 will likely continue in the near term. First, there is a gradual move toward 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 toward developing theories that are *jointly constrained by processing and acquisition data* (e.g., Seidenberg & 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 toward *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 likely play out over the longer term (next ten years or so). 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, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition, 63*(2), 121-170.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13*(2-3), 221-268.

Christiansen, M. H., & 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., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*, 589-608.

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

Dell, G. S., Burger, L. K., & Svec, W. R. (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., & Regier, T. (1996). L0: The First Five Years. *Artificial Intelligence Review, 10*, 103-129.

Frazier, L. (1987). Sentence processing: a tutorial review. In M. Coltheart (Ed.), *Attention and Performance XII: The Psychology of Reading* . East Sussex, U.K.: Lawrence Erlbaum Associates, Ltd.

Gibson, E. A. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1-76.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, name, and the centering of attention in discourse. *Cognitive Science, 17*(3), 311-347.

Grosz, B. J., Weinstein, S., & Joshi, A. K. (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, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122-149.

Kawamoto, A. H. (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: Cambridge University Press.

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

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1-38.

Lewis, R. L. (2000). Specifying architectures for language processing: Process, control, and memory in parsing and interpretation. In M. Crocker, M. Pickering, & C. Clifton (Eds.), *Architectures and Mechanisms for Language Processing* . Cambridge: Cambridge University Press.

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

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676-703.

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

Marcus, G. F. (1996). Why do children say "breaked"? *Current Directions in Psychological Science, 5*, 81-85.

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

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*, 1-71.

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

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

McClelland, J. L., & Rumelhart, D. E. (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., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*(1-2), 73-193.

Plaut, D. C., McClelland, J. L., Seidenberg, M., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56-115.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* . Cambridge, MA: MIT Press.

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

Spivey, M. J., & Tanenhaus, M. K. (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., & 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., & Tanenhaus, M. K. (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., & 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., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic Theory and Psychological Reality* . Cambridge, MA: MIT Press.

Additional Reading

Clifton, C., & Duffy, S. A. (2001). Sentence and text comprehension: Roles of linguistic structure. *Annual Review of Psychology, 52*, 167-196.

Crocker, M. W. (1996). *Computational psycholinguistics: An interdisciplinary approach to the study of language*. Dordrecht: Kluwer Academic.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Gernsbacher, M. A. (Ed.). (1994). *Handbook of Psycholinguistics*. San Diego: Academic Press.

Lewis, R. L. (1999). Cognitive Modeling, Symbolic. In R. A. Wilson & F. C. Keil (Eds.), *The MIT Encyclopedia of Cognitive Science* . Cambridge, MA: MIT Press.

McClelland, J. L. (1999). Cognitive Modeling, Connectionist. In R. A. Wilson & F. C. Keil (Eds.), *The MIT Encyclopedia of Cognitive Science* . Cambridge, MA: MIT Press.

Norris, D. (1999). Computational Psycholinguistics. In R. A. Wilson & F. C. Keil (Eds.), *The MIT Encylopedia of Cognitive Science* . Cambridge, MA: MIT Press.

Glossary

Activation A continuously varying quantity associated with elements (usually primitive processing nodes or memory nodes) in a computational architecture. Computation in activation-based architectures happens via the modulation and propagation of activation. The higher the activation of an element, the more available it is to affect further processing.

Ambiguity resolution The process of selecting a preferred interpretation of some locally or globally ambiguous linguistic input (input that is consistent with more than one interpretation). Ambiguity can arise at multiple levels, including syntactic, lexical, semantic, and referential.

Bottom-up processing The processing of some input that is a function of properties of the input only (for example, the orthographic form of a word), not the context in which the input appears.

Cognitive models Theories of cognitive processes that are embodied in explicit working computer programs that generate behaviour constituting predictions to be compared to human data.

Interactive activation architecture Activation-based architectures in which computation emerges from the parallel propagation of activation among a set of highly interconnected neuron-like units.

Symbolic cognitive models Cognitive models that process structured, compositional representations composed of patterns (symbols) that provide distal access to other representations.

Top-down processing That part of the processing of some input that is a function not only of properties of the input but also the specific context in which the input appears (for example, the linguistic context surrounding the occurrence of a word form).