Connectionist psycholinguistics: capturing the empirical data

Morten H. Christiansen and Nick Chater

Connectionist psycholinguistics is an emerging approach to modeling empirical data on human language processing using connectionist computational architectures. For almost 20 years, connectionist models have increasingly been used to model empirical data across many areas of language processing. We critically review four key areas: *speech processing, sentence processing, language production,* and *reading aloud*, and evaluate progress against three criteria: *data contact*, *task veridicality*, and *input representativeness.* Recent connectionist modeling efforts have made considerable headway toward meeting these criteria, although it is by no means clear whether connectionist (or symbolic) psycholinguistics will eventually provide an integrated model of full-scale human language processing.

Morten H. Christiansen*
Depts of Psychology and Linguistics, Southern Illinois University, Carbondale, IL 62901-6502, USA.
*e-mail: morten@siu.edu

Nick Chater
Dept of Psychology, University of Warwick, Coventry, UK CV4 7AL
e-mail: nick.chater@warwick.ac.uk

What is the significance of connectionist models of language processing? Will CONNECTIONISM (see Glossary) ultimately replace, complement or simply implement SYMBOLIC APPROACHES to language? (see Box 1) Early connectionists addressed this issue by attempting to show that connectionism could, in principle, capture aspects of language and language processing. These models showed that connectionist networks could acquire parts of linguistic structure without extensive 'innate' knowledge[1]. Recent work has moved towards a 'connectionist psycholinguistics', which captures detailed psychological data[2].

Criteria for connectionist psycholinguistics

We review progress in connectionist psycholinguistics in four key areas: speech processing, sentence processing, language production, and reading aloud. We suggest that computational models, whether connectionist or symbolic, should meet three criteria: (1) data contact, (2) task veridicality, and (3) input representativeness. Data contact refers to the degree to which a model captures psycholinguistic data. Of course, there is more to capturing the data than simply fitting existing empirical results; for example, a model should also make non-obvious predictions (see Ref. 3 for discussion). Task veridicality refers to the match between the task facing people and the task given to the model. Although a precise match is difficult to obtain, it is important to minimize the discrepancy. For example, many models of the English past tense[4] have low task veridicality because they map verb stems to past tense forms, a task remote from children's language acquisition. Input representativeness refers to the match between the information available to the model and the person. The performance of connectionist models may be impaired by low input representativeness, because the model does not have access to information sources that may be crucial to human performance.

Symbolic computational psycholinguistics

Few symbolic models make direct contact with psycholinguistic data, with the important exception of comprehensive models of word-by-word reading times[5,6] (and see Ref. 7 for a review). Moreover, symbolic models typically do not focus on task veridicality. For example, rule-based theories[8] of the English past tense involve the same stem to past tense mappings as the early connectionist models, and thus suffer from low task veridicality in comparison with more recent connectionist verb morphology models[9]. Input representativeness is typically low in symbolic models[10], where abstract fragments of language are typically modeled, rather than input derived from real corpora. The remainder of the paper considers whether connectionist psycholinguistics is better able to meet these three criteria.

Speech processing

Connectionist modeling of speech processing begins with TRACE, which has an 'interactive activation' architecture, with a sequence of 'layers' of units (see Fig. 1), for phonetic features, phonemes and words[11]. TRACE captured a wide range of empirical data, and, as we shall see, made important novel predictions.

Evidence for interactive models

TRACE is most controversial because it is interactive – the bi-directional links between units mean that information flows TOP-DOWN as well as BOTTOM-UP. Other connectionist models, by contrast, assume purely bottom-up information flow[12]. TRACE provided an impetus to the interactive versus bottom-up debate, with a prediction apparently incompatible with bottom-up models. In natural speech, the pronunciation of a phoneme is affected by surrounding phonemes: this is 'coarticulation'. The speech processor takes account of this via 'compensation for coarticulation' (CFC)[13]. CFC suggests a way of detecting whether lexical information interactively affects the phoneme level when CFC is considered *across* word boundaries; for example, a word-final /s/ influencing a word-initial /k/ as in *Christmas capes*. If the word level influences the phoneme level, the compensation of the /k/ should occur even when the /s/ relies on PHONEME RESTORATION for its identity (i.e. with an ambiguous /s/ in *Christmas*, the /s/ should be restored and thus CFC should

 1364-6613/01/$ – see front matter PII: S1364-6613(00)01600-4

Glossary

Bottom-up process: A process in which representations that are less abstract with respect to perceptual or linguistic input influence more abstract representations (e.g. influence from phonetic to semantic representations; in connectionist terms, influence of layers of units close to the input to layers far from the input).

Center-embedding: The embedding of one sentence within another. For example, the sentence *'the cat chases the mice'* can be embedded in the center of the sentence *'the mice run away'*, yielding the center-embedded sentence *'the mice that the cat chases run away'*.

Connectionism: Computational architecture using networks of processing units, each of which has an output that is a simple numerical function of its inputs. Loosely inspired by neural architecture.

Cross-serial dependency: Syntactic construction similar to center-embedding except that dependencies between nouns and verbs 'cross over' rather than being embedded within each other in an onion-like structure ($N_1N_2V_2V_1$ versus $N_1N_2V_1V_2$).

Distributed representation: Items are represented by a pattern of activation over several connectionist units.

Dynamical system: Approach to cognitive processing that focuses on the way in which systems change, and described in terms of a set of continuously changing, interdependent quantitative variables governed by a set of equations.

Hidden layer: Units in a connectionist network that lie 'between' input and output, and are hence 'hidden'. The invention of backpropagation and other learning algorithms to train networks with hidden units dramatically increased the power of connectionist methods.

Implicit learning: Learning without conscious awareness of or access to what has been learned. What learning (if any) is implicit is highly controversial.

Localist representation: Items are represented by activation of a single connectionist unit.

Phoneme restoration: If the acoustic form of a word is 'doctored' to remove a phoneme (and, for example, replace it with a noise burst), the phoneme is nonetheless sometimes subjectively perceived as present – it has been perceptually 'restored'.

Recursion: A productive feature of language by which, in principle, we can always add to a sentence by embedding new phrases within it.

Regular spelling-to-sound correspondence: A word that has a straightforward rule-like mapping from spelling to sound has regular spelling-to-sound correspondence. For example, the *-int* endings in *tint, lint,* and *mint* are all pronounced in the same way. Exception words by contrast have a more idiosyncratic mapping from spelling to sound (e.g. *-int* in *pint*).

Relative clause: A clause that provides additional information about a preceding noun. In subject relative clauses, such as *'the senator that attacked the reporter admitted the error'*, the first noun (*senator*) is also the subject of the embedded clause. In object relative clauses, such as *'the senator that the reporter attacked admitted the error'*, the first noun is the object of the embedded clause.

Symbolic approach: Computational style in which representations are discrete symbols, and computation involves operations defined on the *form* of those representations. This style of computation is the basis of digital computer technology.

Top-down process: Reverse of bottom-up – a process of influence from more to less abstract representations (e.g. from semantic to phonetic representations; influence of later from earlier layers; in connectionist terms, influence of layers of units far from the input to layers close to the input).

occur as normal). TRACE's novel prediction was experimentally confirmed[14].

Bottom-up connectionist models strike back

Surprisingly, bottom-up connectionist models *can* capture these results. One study used a simple recurrent network (SRN; see Fig. 1) to map phonetic input onto phoneme output[15]. When the net received phonetic input with an ambiguous first word-final phoneme and ambiguous initial segments of the second word, an analog of CFC was observed. The percentages of /t/ and /k/ responses to the first phoneme of the second word depended on the identity of the first word (as in Ref. 14). Importantly, the explanation for this pattern of results cannot be top-down influence from word units, because there *are* no word units. Nonetheless the presence of 'feedback' connections in the HIDDEN LAYER of the SRN might suggest that some form of interactive processing occurs in this model. But this is misleading – the feedback occurs within the hidden layer (i.e. from its previous to its present state), rather than flowing from top to bottom.

This model, although an important demonstration, has poor input representativeness, because it deals with just 12 words. However, a subsequent study scaled-up these results using a similar network trained on phonologically transcribed conversational English[16]. How can bottom-up processes mimic lexical effects? It was argued that restoration depends on local statistical regularities at the phonemic level, rather than depending on access to lexical representations. More recent experiments have since shown that CFC is indeed determined by statistical regularities for non-word stimuli, and that, for word stimuli, there appear to be no residual effects of lexical status, once statistical regularities are taken into account[17]. It is not clear, though, whether bottom-up models can model other evidence that phoneme identification is affected by the lexicon, for example, from signal detection analyses of phoneme restoration[18].

Exploiting distributed representations

A different line of results provides additional evidence that bottom-up models can accommodate apparently top-down effects[19]. An SRN was trained to map a systematically altered featural representation of speech onto a phonemic and semantic representation of the same speech (following a previously established tradition[20]). After training, the network showed evidence of lexical effects in modeling lexical and phonetic decision data[21]. This work was extended by an SRN trained to map sequential phonetic input onto corresponding DISTRIBUTED REPRESENTATIONS of phonological surface forms and semantics[22]. This style of representation contrasts with the LOCALIST REPRESENTATIONS used in TRACE. The ability of the SRN to model the integration of partial cues to phonetic identity and the time course of lexical access provides support for a distributed approach. An important challenge for such distributed models is to model the simultaneous activation of multiple lexical candidates necessitated by the temporal ambiguity of the speech input (e.g. /kæp/ could continue *captain* and *captive*) (see Ref. 23 for a generalization of this phenomenon). The 'coactivation' of several lexical candidates in a distributed model results in a semantic 'blend' vector. Computational explorations[24] of such semantic blends provide explanations of recent empirical results aimed at measuring lexical coactivation[25].

Speech segmentation

Further evidence for the bottom-up approach to speech processing comes from the modeling of speech

Box 1. The debate over connectionist models of language

There are many recurrent themes in the debate over the value of connectionist models of language. Here we list some of the most prominent and enduring.

(1) Learning

Many connectionist networks acquire their 'knowledge' through training on input–output examples, making learning an essential part of these models. By contrast, many symbolic models come with most of their knowledge 'built-in', although some learning may be required to fine-tune this knowledge.

(2) Generalization

People are able to produce and process linguistic forms (words, sentences) that they have never heard before. Generalization to new cases is thus a crucial test[a] for many connectionist models.

(3) Representation

Because most connectionist nets learn, their internal codes are devised by the network to be appropriate for the task. Developing methods for understanding these codes is an important research strand. Whereas internal codes may be learned, the inputs and outputs to a network generally use a code specified by the researcher. The choice of code can be crucial in determining network performance. How these codes relate to standard symbolic representations of language is contentious.

(4) Rules versus exceptions

Many aspects of language exhibit 'quasi-regularities': regularities which usually hold, but which admit exceptions. In a symbolic framework, quasi-regularities may be captured by symbolic rules, associated with explicit lists of exceptions. Symbolic processing models often incorporate this distinction by having separate mechanisms for regular and exceptional cases. In contrast, connectionist nets may provide a *single mechanism* that can learn general rule-like regularities, and their exceptions. The viability of such 'single route' models has been a major point of controversy, although it is not intrinsic to connectionism. One or both separate mechanisms for rules and exceptions could themselves be modeled in connectionist terms[b–d]. A further question is whether networks really learn rules at all, or merely approximate rule-like behavior. Opinions differ on whether the latter is an important positive proposal, which may lead to a revision[e,f] of the role of rules in linguistics, or whether it is fatal[g,h] to connectionist models of language.

References

a Hadley, R. (1994) Systematicity in connectionist language learning. *Mind Lang.* 9, 247–272
b Bullinaria, J.A. (1997) Modelling reading, spelling and past tense learning with artificial neural networks. *Brain Lang.* 59, 236–266
c Pinker, S. (1991) Rules of language. *Science* 253, 530–535
d Zorzi, M. *et al*. (1998) Two routes or one in reading aloud? A connectionist dual-process model. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 1131–1161
e Elman, J.L. (1991) Distributed representation, simple recurrent networks, and grammatical structure. *Machine Learn.* 7, 195–225
f Rumelhart, D.E. and McClelland, J.L. (1987) Learning the past tenses of English verbs: implicit rules or parallel distributed processing? In *Mechanisms of Language Acquisition* (MacWhinney, B. ed.), pp. 195–248, Erlbaum
g Marcus, G.F. (1998) Rethinking eliminative connectionism. *Cognit. Psychol.* 37, 243–282
h Pinker, S. (1999) *Words and Rules: The Ingredients of Language*, Basic Books

segmentation[26]. An SRN was trained to integrate sets of phonetic features with information about lexical stress (strong or weak) and utterance boundary information (encoded as a binary unit) derived from a corpus of child-directed speech. The network was trained to predict the appropriate values of these three cues for the next segment. After training, the network was able to generalize patterns of cue information that occurred at the end of utterances to when the same patterns occurred elsewhere in the utterance. Relying entirely on bottom-up information, the model performed well on the word segmentation task, and captured important aspects of infant speech segmentation.

Summary

Overall, connectionist speech processing models make good contact with psycholinguistic data, and has motivated novel experimental work. Input representativeness is also generally good, with models being trained on large lexicons and sometimes corpora of natural speech. Task veridicality is questionable, however, because the standard abstract representations of the input (e.g. phonetic or phonological representations) may not be computed by the listener[21] and, furthermore, bypass the deep problems involved in handling the physical variability of natural speech.

Sentence processing

Sentence processing provides a considerable challenge for connectionism. Some connectionists[27] have built symbolic structures directly into the network, whilst others[28] have chosen to construct a modular system of networks, each tailored to acquire different aspects of syntactic processing. However, the approach that has made the most contact with psycholinguistic data involves directly training networks to discover syntactic structure from word sequences[29].

Capturing complexity judgment and reading time data

One study has explored the learning of different types of RECURSION by training SRNs on small artificial languages[30]. A measure of grammatical prediction error (GPE) was developed, allowing network output to be mapped onto human performance data. GPE is computed for each word in a sentence and reflects the processing difficulties that a network is experiencing at a given point in a sentence. Averaging GPE across a whole sentence, the model fitted human data concerning the greater perceived difficulty associated with CENTER-EMBEDDING in German compared with CROSS-SERIAL DEPENDENCIES in Dutch[31]. Related models trained on more naturalistic language fragments (M. Christiansen, unpublished results) captured the same data, and provided the basis for novel

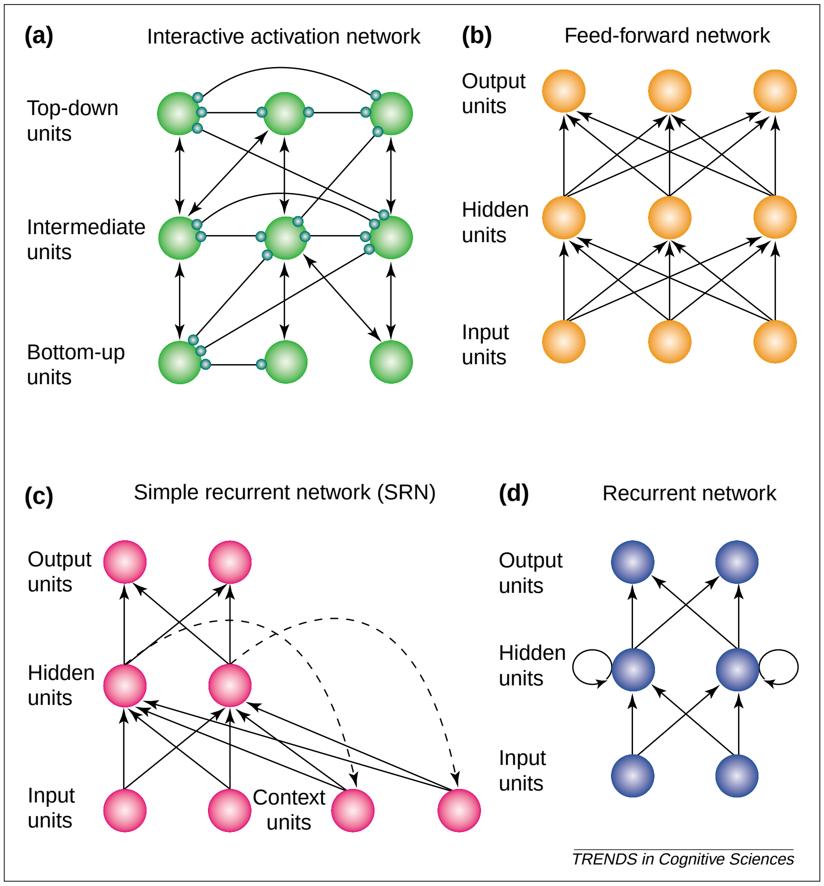

Fig. 1. Network architectures used in connectionist psycholinguistics. (a) An interactive activation network (e.g. TRACE) has bi-directional excitatory (arrows) or inhibitory (filled circles) links. Activation flows bottom-up and top-down, reinforcing mutually consistent states and inhibiting inconsistent states. The weights in an interactive activation network are typically hand-coded rather than learned. (b) A feed-forward network is normally trained using back-propagation[61], which minimizes the discrepancy between the network's actual and desired output. Information flows bottom-up from input to output units. (c) A simple recurrent network (SRN)[1] is essentially a standard feed-forward network equipped with an extra layer of so-called context units. At each time step, input propagate through the hidden units to the output (solid arrows). The hidden unit activation at the previous time step is copied back to the context layer (dashed arrows) and paired with the current input (solid arrows). Thus the hidden units influence the processing of subsequent inputs, providing a limited ability to deal with sequential inputs. (d) Recurrent networks often have the same architecture as SRNs but are trained using more complex learning algorithms[62,63]. Current activations affect future activations via the recurrent links. Recurrent links between hidden units are not viewed as 'unfolded' into separate context units, because the learning algorithms deal directly with recurrent connections.

predictions concerning other types of recursive constructions. These predictions have been confirmed experimentally (M. Christiansen and M. MacDonald, unpublished results). Finally, single-word GPE scores from a related model[32] were mapped directly onto reading times, providing an experience-based account for human data concerning the differential processing of singly center-embedded subject and object RELATIVE CLAUSES by good and poor comprehenders.

Another approach to sentence processing involves a two-component model of ambiguity resolution, combining an SRN with a 'gravitational' mechanism (i.e. a DYNAMICAL SYSTEM)[33]. The SRN was trained in the usual way on sentences derived from a grammar. After training, SRN hidden unit representations for individual words were placed in the gravitational mechanism, and the latter was allowed to settle into a stable state. Settling times were then mapped onto word reading times. The two-component model was able to fit data from several experiments concerning the interaction of lexical and structural constraints on the resolution of temporary syntactic ambiguities (i.e. garden path effects) in sentence comprehension. More recently, the two-component model was extended[34] to account for empirical findings reflecting the influence of semantic role expectations on syntactic ambiguity resolution in sentence processing[35].

Capturing grammaticality ratings in aphasia

Some headway has also been made in accounting for data concerning the effects of aphasia on grammaticality judgments[36]. A recurrent network (see Fig. 1) was trained mutually to associate two input sequences: a sequence of word forms and a corresponding sequence of word meanings. The network was able to learn a small artificial language successfully, enabling it to regenerate the word forms from the meanings and vice versa. Grammaticality judgments were simulated by testing how well the network could recreate a given input sequence, allowing activation to flow from the provided input forms to meaning and then back again. Ungrammatical sentences were recreated less accurately than grammatical sentences, and hence the network was able to distinguish grammatical from ungrammatical sentences. The network was then 'lesioned' by removing 10% of the weights in the network. Grammaticality judgments were then elicited from the impaired network for 10 different sentence types from a classic study of aphasic grammaticality judgments[37]. The aphasic patients had problems with three of these sentence types, and the network fitted this pattern of performance impairment exactly.

Summary

Overall, connectionist models of syntactic processing are at an early stage of development. Current connectionist models of syntax typically use 'toy' fragments of grammar and small vocabularies, and thus have low input representativeness. Nevertheless, the models have good data contact and a reasonable degree of task veridicality. However, more research is required to decide whether promising initial results can be scaled up to deal with the complexities of real language, or whether a purely connectionist approach is beset by fundamental limitations, and can only succeed by incorporating symbolic methods into the models.

Language production

Connectionist models have had a large impact on the field of language production, and played an important role in framing theories of normal and impaired production.

Aphasic word production

One of these models is a paradigm of connectionist psycholinguistics, and *quantitatively* fitted error data from 21 aphasics and 60 normal controls[38]. The network has three layers with bi-directional connections, mapping from semantic features denoting a concept, to a choice of word; and then to the phonemes realizing that word. The model differs from other interactive activation models, such as TRACE, by incorporating a two-step approach to production. First, activation at the semantic features spreads throughout the network for a fixed time. The most active word unit (typically the best match to the semantic features) is 'selected', and its activation boosted. Second, activation again spreads throughout the network for a fixed time, and the most highly activated phonemes are selected, with a phonological frame that specifies the sequential ordering of the phonemes.

Even in normal production, processing sometimes breaks down, leading to semantic errors (*cat* → *dog*), phonological errors (*cat* → *hat*), mixed semantic and phonological errors (*cat* → *rat*), non-word errors (*cat* → *zat*), and unrelated errors (*cat* → *fog*). Normal and aphasic errors are proposed to reflect the same processes, differing only in degree. Therefore, the model's parameters were set by fitting data from controls relating to the five types of errors above. To simulate aphasia, the model was 'damaged' by reducing two global parameters (connection strength and decay rate), leading to more errors. The model fitted the five types of errors found for the aphasics (see Refs 39,40 for discussions). Furthermore, predictions were derived, and subsequently confirmed, concerning the effect of syntactic categories on phonological errors (*dog* → *log*), phonological effects on semantic errors (*cat* → *rat*), naming error patterns after recovery, and errors in word repetition.

Structural priming in syntactic productions

Connectionist models have also been applied to experimental data on sentence production, particularly concerning structural priming. Structural priming arises when the syntactic structure of a previously heard or spoken sentence influences the processing or production of a subsequent sentence. An SRN model of grammatical encoding was implemented[41] to test the suggestion that structural priming may be an instance of IMPLICIT LEARNING. The input to the model was a 'proposition', coded by units for semantic features (e.g. *child*), thematic roles (e.g. *agent*) and action descriptions (e.g. *walking*), and some additional input encoding the internal state of an unimplemented comprehension network. The network outputs a sequence of words expressing the proposition. Structural priming was simulated by allowing learning to occur during testing. This created transient biases in weight space that were sufficiently robust to cause the network to favor (i.e. to be primed by) recently encountered syntactic structures.

The model fitted data concerning the priming, across up to 10 unrelated sentences, of active and passive constructions as well as prepositional ('*The boy gave the guitar to the singer*') and double-object ('*The boy gave the singer the guitar*') dative constructions[42]. The model fitted the passive data well, and showed priming from intransitive locatives ('*The 747 was landing by the control tower*') to passives ('*The 747 was landed by the control tower*'). However, it fitted the dative data less well, and showed no priming from transitive locatives ('*The wealthy woman drove the Mercedes to the church*') to prepositional datives ('*The wealthy woman gave the Mercedes to the church*'). A more recent model with an implemented comprehension network and a less rigid representation of thematic roles provides a better fit with these data[43].

Summary

The connectionist production models make good contact with the data, and have reasonable task veridicality, but suffer from low input representativeness – these models are trained on small fragments of natural language. It seems likely that connectionist models will continue to play a central role in future research on language production. However, scaling up these models to deal with more realistic input is a major challenge for future work.

Reading aloud

Connectionist research on reading aloud has focused on single words. A classic early model used a feed-forward network (see Fig. 1) to map from a distributed orthographic representation to a distributed phonological representation, for monosyllabic English words[44]. The net's performance captured a wide range of experimental data, on the assumption that network error maps onto response time.

This model contrasts with standard views of reading, which assume both a 'phonological route', applying rules of pronunciation, and a 'lexical route', which is a list of words and their pronunciations. Words with a REGULAR SPELLING-TO-SOUND CORRESPONDENCE can be read using either route; exception words by the lexical route; and non-words by the phonological route. It was claimed that, instead, a single connectionist route can pronounce both exception words and non-words.

Critics have responded that the network's non-word reading is well below human performance[45] (although see Ref. 46). Another difficulty is the model's reliance on (log) frequency compression during training (otherwise exception words are not learned successfully). Subsequent research has addressed both limitations, showing that a network trained on actual word frequencies can achieve human levels of performance on both word and non-word pronunciation[47].

Capturing the neuropsychological data

Single and dual route theorists generally agree that there is an additional 'semantic' route, where pronunciation is retrieved via a semantic code – the controversy is whether there are one or two *non-semantic* routes. Some connectionists argue that the division of labor between the phonological and semantic routes can explain diverse neuropsychological syndromes that have been taken to require a dual-route account[47]. On this view, a division of labor emerges between the phonological and the semantic pathway during reading acquisition: the phonological pathway specializes in regular orthography-to-phonology mappings at the expense of exceptions, which are read by the semantic pathway. Damage to the semantic pathway causes 'surface dyslexia' (where exceptions are selectively impaired); damage to the phonological pathway causes 'phonological dyslexia' (where non-words are selectively impaired). According to this viewpoint, 'deep dyslexia' occurs when the phonological route is damaged, and the semantic route is also partially impaired (which leads to semantic errors, such as reading the word *peach* as *apricot*, which are characteristic of the syndrome).

Capturing the experimental data

Moving from neuropsychological to experimental data, connectionist models of reading have been criticized for not modeling effects of specific lexical items[48]. One defense is that current models are too partial (e.g. containing no letter recognition and phonological output components) to model word-level effects[49]. However, this challenge is taken up in a study in which an SRN is trained to pronounce words phoneme-by-phoneme[50]. The network can also refixate the input when unable to pronounce part of a word. The model performs well on words and non-words, and fits empirical data on word length effects[51,52]. Complementary work using a recurrent network focuses on providing a richer model of phonological knowledge and processing[53], which may be importantly related to reading development[54]. Finally, it has been shown how a two-route model of reading might emerge naturally from a connectionist learning architecture[55]. Using backpropagation, direct links between orthographic input and phonological output learn to encode letter-to-phoneme correspondences (a 'phonological route') whereas links via hidden units spontaneously learn to handle exception words (a 'lexical route'). Here, as elsewhere in connectionist psycholinguistics, connectionist models can provide persuasive instantiations of a range of theoretical positions.

Summary

Connectionist research on reading has good data contact and reasonable input representativeness. Task veridicality is questionable: children do not typically associate written and spoken forms for individual words when learning to read (although Ref. 53 partially addresses this issue). A major research challenge is synthesizing insights from accounts of different aspects of reading into a single model.

Conclusion

Current connectionist models involve important simplifications with respect to natural language processing. In some cases, these simplifications are relatively modest. For example, models of reading aloud typically ignore how eye movements are planned, how information is integrated across eye-movements, ignore the sequential character of speech output, and typically deal only with short words. In other cases, the simplifications are more drastic. For example, connectionist models of syntactic processing involve vocabularies and grammars that are vastly simplified. In many cases, these limitations stem from compromises made in order to implement connectionist models as working computational models. Many symbolic models[8], on the other hand, can give the appearance of good data contact simply because they have not yet been implemented and have therefore not been tested in an empirically rigorous way. Nevertheless, we argue that proponents of both connectionist and symbolic models must aim to achieve high degrees of data contact, task veridicality and input representativeness in order to advance computational psycholinguistics.

The present breadth of connectionist psycholinguistics, as outlined above, indicates that the approach has considerable potential. Despite attempts to establish a priori limitations on connectionist language processing[56,57], we suggest that the only way to determine the value of the approach is to pursue it with the greatest possible creativity and vigor. If realistic connectionist models of language processing can be provided, then a radical rethinking of language processing and structure may be required. It might be that the ultimate description of language resides in the structure of complex networks[58], and can only be approximated by symbolic grammatical rules. Conversely, connectionist models might only succeed to the extent that they build in standard linguistic constructs[59], or form a hybrid with symbolic models[60]. The future of connectionist psycholinguistics is therefore likely to have important implications either in overturning, or reaffirming, traditional psychological and linguistic assumptions.

Acknowledgements
We thank the anonymous reviewers for their comments and suggestions regarding this manuscript.

Outstanding questions

- Can connectionist models 'scale-up' successfully to provide more realistic models of language processing, or do they have fundamental computational limitations? And, if connectionist systems can scale up successfully, will the resulting models still provide close fits with the psycholinguistic data?
- To what degree does learning in connectionist networks provide a potential model of human language acquisition? How much structure and knowledge must be 'built' into connectionist networks to deal with real human language?
- Can the connectionist subsystems that have been developed separately to deal with different sub-areas of language processing be integrated? Should connectionist subsystems be considered as separate modules? If so, what are the appropriate modules for connectionist language processing? What implications does this have for the criteria for connectionist psycholinguistics?
- How can we more fully characterize what it means to 'capture the data'? How do we best compare computational models? Should models be required to make non-obvious predictions?

References

1 Elman, J. (1990) Finding structure in time. *Cognit. Sci.* 14, 179–211

2 Christiansen, M.H. and Chater, N. (1999) Connectionist natural language processing: the state of the art. *Cognit. Sci.* 23, 417–437

3 Roberts, S. and Pashler, H. (2000) How persuasive is a good fit? A comment on theory testing. *Psychol. Rev.* 107, 358–367

4 Rumelhart, D.E. and McClelland, J.L. (1986) On learning past tenses of English verbs. In *Parallel Distributed Processing* (Vol. 2) (McClelland, J.L. and. Rumelhart, D.E., eds), pp. 216–271, MIT Press

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

6 Just, M.A. and Carpenter, P.A. (1992) A capacity theory of comprehension: individual differences in working memory. *Psychol. Rev.* 98, 122–149

7 Gibson, E. and Pearlmutter, N.J. (1998) Constraints on sentence comprehension. *Trends Cognit. Sci.* 2, 262–268

8 Pinker, S. (1991) Rules of language. *Science* 253, 530–535

9 Hoeffner, J.H. and McClelland, J.L. (1993) Can a perceptual processing deficit explain the impairment of inflectional morphology in developmental dysphasia? A computational investigation. In *Proceedings of the 25th Annual Stanford Child Language Research Forum* (Clark, E.V., ed.), pp. 39–45, CSLI, Stanford, CA

10 Dresher, B.E. and Kaye, J.D. (1990) A computational learning model for metrical phonology. *Cognition* 34, 137–195

11 McClelland, J.L. and Elman, J.L. (1986) The TRACE model of speech perception. *Cognit. Psychol.* 18, 1–86

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

13 Mann, V.A. and Repp, B.H. (1981) Influence of preceding fricative on stop consonant perception. *J. Acoust. Soc. Am.* 69, 548–558

14 Elman, J.L. and McClelland, J.L. (1988) Cognitive penetration of the mechanisms of perception: compensation for coarticulation of lexically restored phonemes. *J. Mem. Lang.* 27, 143–165

15 Norris, D. (1993) Bottom-up connectionist models of 'interaction'. In *Cognitive Models of Speech Processing* (Altmann, G.T.M. and Shillcock, R., eds), pp. 211–234, Erlbaum

16 Cairns, P. *et al.* (1995) Bottom-up connectionist modelling of speech. In *Connectionist Models of Memory and Language* (Levy, J.P. *et al.*, eds), pp. 289–310, UCL Press

17 Pitt, M.A. and McQueen, J.M. (1998) Is compensation for coarticulation mediated by the lexicon? *J. Mem. Lang.* 39, 347–370

18 Samuel, A.C. (1996) Does lexical information influence the perceptual restoration of phonemes? *J. Exp. Psychol. Gen.* 125, 28–51

19 Gaskell, M.G. and Marslen-Wilson, W.D. (1995) Modeling the perception of spoken words. In *Proceedings of the 17th Annual Cognitive Science Conference*, pp. 19–24, Erlbaum

20 Kawamoto, A.H. (1993) Nonlinear dynamics in the resolution of lexical ambiguity: a parallel distributed processing account. *J. Mem. Lang.* 32, 474–516

21 Marslen-Wilson, W. and Warren, P. (1994) Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychol. Rev.* 101, 653–675

22 Gaskell, M.G. and Marslen-Wilson, W.D. (1997) Integrating form and meaning: a distributed model of speech perception. *Lang. Cognit. Processes* 12, 613–656

23 Allopenna, P.D. *et al.* (1998) Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J. Mem. Lang.* 38, 419–439

24 Gaskell, M.G. and Marslen-Wilson, W.D. (1999) Ambiguity, competition, and blending in spoken word recognition. *Cognit. Sci.* 23, 439–462

25 Gaskell, M.G. and Marslen-Wilson, W.D. (1997) Discriminating local and distributed models of competition in spoken word recognition. In *Proceedings of the 19th Annual Cognitive Science Conference*, pp. 247–252, Erlbaum

26 Christiansen, M.H. *et al.* (1998) Learning to segment speech using multiple cues: a connectionist model. *Lang. Cognit. Processes* 13, 221–268

27 Miyata, Y. *et al.* (1993) Distributed representation and parallel distributed processing of recursive structures. In *Proceedings of the 15th Annual Meeting of the Cognitive Science Society* pp. 759–764, Erlbaum

28 Miikkulainen, R. (1996) Subsymbolic case-role analysis of sentences with embedded clauses. *Cognit. Sci.* 20, 47–73

29 Elman, J.L. (1991) Distributed representation, simple recurrent networks, and grammatical structure. *Machine Learn.* 7, 195–225

30 Christiansen, M.H. and Chater, N. (1999) Toward a connectionist model of recursion in human linguistic performance. *Cognit. Sci.* 23, 157–205

31 Bach, E. *et al.* (1986) Crossed and nested dependencies in German and Dutch: a psycholinguistic study. *Lang. Cognit. Processes* 1, 249–262

32 MacDonald, M.C. and Christiansen, M.H. Reassessing working memory: a comment on Just and Carpenter (1992) and Waters and Caplan (1996) *Psychol. Rev.* (in press)

33 Tabor, W. *et al.* (1997) Parsing in a dynamical system: an attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Lang. Cognit. Processes* 12, 211–271

34 Tabor, W. and Tanenhaus, M.K. (1999) Dynamical models of sentence processing. *Cognit. Sci.* 23, 491–515

35 McRae, K. *et al.* (1998) Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *J. Mem. Lang.* 38, 282–312

36 Allen, J. and Seidenberg, M.S. (1999) The emergence of grammaticality in connectionist networks. In *The Emergence of Language* (MacWhinney, B. ed.), pp. 115–151, Erlbaum

37 Linebarger, M.C. *et al.* (1983) Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition* 13, 361–392

38 Dell, G.S. *et al.* (1997) Lexical access in aphasic and nonaphasic speakers. *Psychol. Rev.* 104, 801–838

39 Ruml, W. and Caramazza, A. (2000) An evaluation of a computational model of lexical access. Comment on Dell *et al.* (1997). *Psychol. Rev.* 107, 609–634

40 Dell, G.S. *et al.* (2000) The role of computational models in neuropsychological investigations of language. Reply to Ruml and Caramazza (2000). *Psychol. Rev.* 107, 635–645

41 Dell, G.S. *et al.* (1999) Connectionist models of language production: lexical access and grammatical encoding. *Cognit. Sci.* 23, 517–542

42 Bock, J.K. and Griffin, M.Z. The persistence of structural priming: transient activation or implicit learning? *J. Exp. Psychol. Gen.* (in press)

43 Chang, F. *et al.* Structural priming as implicit learning: a comparison of models of sentence production. *J. Psycholinguist. Res.* (in press)

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

45 Besner, D. *et al.* (1990) On the connection between connectionism and data: are a few words necessary? *Psychol. Rev.* 97, 432–446

46 Seidenberg, M.S. and McClelland, J.L. (1990) More words but still no lexicon. Reply to Besner *et al.* (1990). *Psychol. Rev.* 97, 447–452

47 Plaut, D. *et al.* (1996) Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychol. Rev.* 103, 56–115

48 Spieler, D.H. and Balota, D.A. (1997) Bringing computational models of word recognition down to the item level. *Psychol. Sci.* 8, 411–416

49 Seidenberg, M.S. and Plaut, D.C. (1998) Evaluating word-reading models at the item level: matching the grain of theory and data. *Psychol. Sci.* 9, 234–237

50 Plaut, D. (1999) A connectionist approach to word reading and acquired dyslexia: extension to sequential processing. *Cognit. Sci.* 23, 543–568

51 Rastle, K. and Coltheart, M. (1998) Whammies and double whammies: the effect of length on nonword reading. *Psychonomic Bull. Rev.* 5, 277–282

52 Weekes, B.S. (1997) Differential effects of number of letters on word and nonword latency. *Q. J. Exp. Psychol.* 50A, 439–456

53 Harm, M.W. and Seidenberg, M.S. (1999) Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychol. Rev.* 106, 491–528

54 Bradley, L. and Bryant, P. (1983) Categorizing sounds and learning to read: a causal connection. *Nature* 301, 419–421

55 Zorzi, M. *et al.* (1998) Two routes or one in reading aloud? A connectionist dual-process model. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 1131–1161

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

57 Marcus, G.F. (1998) Rethinking eliminative connectionism. *Cognit. Psychol.* 37, 243–282

58 Seidenberg, M.S. and MacDonald, M.C. (1999) A probabilistic constraints approach to language acquisition and processing. *Cognit. Sci.* 23, 569–588

59 Smolensky, P. (1999) Grammar-based connectionist approaches to language. *Cognit. Sci.* 23, 589–613

60 Steedman, M. (1999) Connectionist sentence processing in perspective. *Cognit. Sci.* 23, 615–634

61 Rumelhart, D. *et al.* (1986) Learning internal representations by error propagation. In *Parallel Distributed Processing* (Vol. 1) (Rumelhart, D. and McClelland, J., eds), pp. 318–362, MIT Press

62 Williams, R.J. and Peng, J. (1990) An efficient gradient-based algorithm for on-line training of recurrent network trajectories. *Neural Comput.* 2, 490–501

63 Pearlmutter, B.A. (1989) Learning state space trajectories in recurrent neural networks. *Neural Comput.* 1, 263–269