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.

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. Recent work has moved towards a ‘connectionist psycholinguistics’, which captures detailed psychological data.

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 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
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 that the cat chases 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 (N1 N2 V2 V1 versus N1 N2 V1 V2).

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.

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 that the cat chases 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 (N1 N2 V2 V1 versus N1 N2 V1 V2).

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 -endsings 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. -stint pinto).

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).
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 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. 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 of the role of rules in linguistics, or whether it is fatal to connectionist models of language.

Sentence processing
Sentence processing provides a considerable challenge for connectionism. Some connectionists have built symbolic structures directly into the network, whilst others 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.

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. 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. Related models trained on more naturalistic language fragments captured the same data, and provided the basis for novel
into separate context units, because the learning algorithms deal directly with recurrent connections. Future activations via the recurrent links. Recurrent links between hidden units are not viewed as ‘unfolded’ actual and desired output. Information flows bottom-up from input to output units. (c) A simple recurrent architecture as SRNs but are trained using more complex learning algorithms. Current activations affect inputs, providing a limited ability to deal with sequential inputs. (d) Recurrent networks often have the same 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. 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.

Capturing grammaticality ratings in aphasia Some headway has also been made in accounting for data concerning the effects of aphasia on grammaticality judgments. 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. 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. 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 aphasias are errors 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 (Semantically Represented Network) model of grammatical encoding was implemented 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. 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 drówe 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.

**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. 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 argued that the network's non-word reading is well below human performance (although see Ref. 46). A second difficulty is the model's reliance on 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.

**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. 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 is 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. One defense is that current models are too partial (e.g. containing no letter recognition and phonological output components) to model word-level effects. However, this challenge is taken up in a study in which an SRN is trained to pronounce words phoneme-by-phoneme. The network can also reread the input when unable to pronounce parts of a word. The model performs well on words and non-words, and fits empirical data on word length effects. Complementary work using a recurrent network focuses on providing a richer model of phonological knowledge and processing, which may be importantly related to reading development. Finally, it has been shown how a two-route model of reading might emerge naturally from a connectionist learning architecture. 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 eyes, 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, 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, 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, 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, or form a hybrid with symbolic models. The future of connectionist psycholinguistics is therefore likely to have important implications either in overturning, or reaffirming, traditional psychological and linguistic assumptions.