the beginning of the list, followed by more complex, probably less common, characters. Within stroke count groups, the order is, essentially traditional.

Hsu Shen's radical groups numbered as high as 800. Today, there are around 200 in use, depending on the dictionary. Over the centuries, Hsu Shen's indexing system has been simplified for ease of use. Most modern dictionaries use the simplified radical list developed in the Ming Dynasty (c. 1615). Several dictionaries prior to the Ming had reduced the radical count from Hsu Shen's approximately 800.

Hsu Shen also included etymological information in the belief that this would be a useful memory aid. A current dictionary that provides this type of information is *Chinese characters – a genealogy and dictionary.*

Hsu Shen's method has proved useful but raises its own problems. Many Chinese characters contain phonetic information, but the phonetic information may refer to pronunciations that are no longer current. Using stroke counts depends heavily on knowledge of traditional writing methods and on the fonts used. Small font sizes often make accurate stroke counts difficult, if not impossible, for complex characters. The Chinese language has prompted hosts of reform methods, none of which have achieved universal acceptance. Simplification schemes run into cultural opposition as well as the flexibility of the language itself. There is no theoretical limit to the number of characters. In a sense, all dictionaries are special-purpose works.

See also: Chinese.

Bibliography

Harbaugh R (1998). *Chinese characters – a genealogy and dictionary.* Taipei: Han Lu Book & Publishing Co.

Relevant Websites

http://www.zhongwen.com/ – Distributor of *Chinese characters – a genealogy and dictionary.*

Huilliche *See:* Mapudungan.

Human Language Processing: Connectionist Models

L Onnis, Cornell University, Ithaca, NY, USA
M Christiansen, Cornell University, Ithaca, NY, USA
N Chater, University of Warwick, Coventry, UK

Connectionist psycholinguistics is an emerging approach to modeling empirical data on human language processing using connectionist computational architectures. Over the last 20 years, a wide range of psycholinguistic phenomena have been modeled, such as speech processing, impaired and normal reading, aphasic word production, and structural priming in sentence production. In this article, we will focus on syntax, and especially on the theoretical relationship between traditional notions in syntax (constituency, structure dependency, recursion) and connectionist networks, which do not appear to make reference to these notions intrinsically. For an overview of other aspects of connectionist language processing, see Christiansen and Chater (2001) and Rohde and Plaut (2003).

Connectionist models, or neural networks, are neurally inspired devices based on numerical computation, rather than symbolic manipulation. They are formed by several simple processing units called nodes, which mimic the activity of single neurons. Each node is excited or inhibited by information coming from other units to which it is connected. A set of homogeneous nodes that collectively represents some information is often called a layer. For instance, there could be a phoneme layer, a word layer, and so on. A popular architecture in connectionist psycholinguistics is the feed-forward network (Figure 1), where an input layer receives signals (e.g., raw words) from outside the network, an output layer sends signals to outside the network, and an intermediate layer of so-called hidden units does not directly connect to the outside of the network. Hidden unit layers are often considered the internal states that

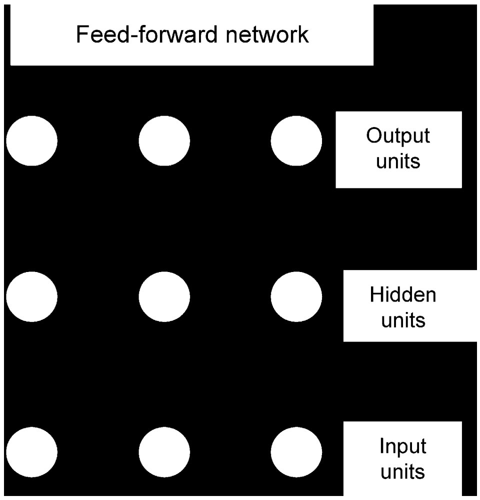

Figure 1 A standard feed-forward network containing three layers of units. Information flows bottom-up from input to output units. These types of networks are normally trained using the back-propagation algorithm, which minimizes the discrepancy between the network's actual and desired output.

represent the knowledge acquired by the network at the end of training. Information flows from input to output and training proceeds by adjusting the weights between single units until the network produces a desired output (for instance, the production of the correct word syntactic category in a given context).

Two aspects of neural networks are particularly interesting for syntactic processing: first, networks **learn** to perform some specific task, e.g., process an English sentence, from being exposed to relatively unanalyzed raw input, such as a sequence of words. This feature makes them suitable for explaining a great deal of language learning and language processing in a psychologically plausible manner. Secondly, the knowledge of the networks is not represented in any explicit way, but rather emerges as a pattern of neuronal activity between the interconnected nodes, in a parallel distributed processing fashion. This is particularly interesting because it contrasts with more traditional symbolic views in cognitive science, suggesting that humans process sentences by transforming representations according to sets of rules. For instance, a representation of the sentence *the girl liked a boy*, requires the constituents *the*, *a*, *girl*, *boy*, *liked*, and an explicit representation of the syntactic relationships between these constituents. The rules of syntax govern how constituents can be combined together, allowing, for instance, *the boy liked a girl* as a legal sentence, but not *boy girl a liked the*. Phrase structure rules describe the relationship between constituents in the above sentence:

S → NP VP
NP → (det) N
VP → V (NP)

Besides such constituent structure, to capture the full generativity of human language, recursion needs to be introduced, for instance by adding a new rule that adds a potential prepositional phrase (PP) to the NP:

NP → (det) N (PP)
PP → prep NP

These rules are recursive because the expansion of the right-hand sides of each can call the other. For example, the complex NP *the flowers in the vase* has the simple NP *the vase* recursively embedded within it. Because this process can be applied arbitrarily often, constructions of arbitrary complexity can in principle be generated. Constituency and recursion are some of the most fundamental concepts in linguistics. Since both are defined in terms of relations between symbols, symbolic models of language processing therefore incorporate these properties by fiat. In this article, we discuss how constituency, structure dependency, and recursion may fit into a connectionist framework, and the possible implications this work may have for linguistics and psycholinguistics.

Constituency

Connectionist models can address constituency in three increasingly radical ways. First, some models – especially early ones – are implementations of symbolic language processing models in neural hardware; for example, Fanty's (1986) network implementation of a context-free grammar contains explicit representations of the constituent structure of a sentence in just the same way as a nonconnectionist implementation of the same model would have. Connectionist implementations of this kind provide feasibility proofs that traditional symbolic models of language processing are compatible with a 'brain-style' computational architecture, although they add little that is new to the treatment of constituency. The remaining two classes of connectionist models actually **learn** to process constituent structure, rather than having this ability hardwired. One approach is to have a network learn from input 'tagged' with information about constituent structure. Kim *et al.* (2002) train a network to map a combination of orthographic and cooccurrence-based semantic information about a word onto a structured representation encoding the minimal syntactic environment for that word. With an input vocabulary consisting of 20 000 words, this model has an impressive coverage, and can account for certain results from the psycholinguistic literature concerning ambiguity resolution in sentence processing. Henderson and Lane (1998) provide a model known as a simple synchrony network trained to parse sentences preencoded as parts of speech. The network takes the part of speech tags for the sentence constituents as input and is trained to output the parse tree fragment of a given constituent when that constituent is queried. The network learns to parse a corpus of written English reasonably well. However, because in these models constituent structure is compiled either into the input or the output representations, this style of model does not offer any fresh insight into how the linguistic constituency might operate, based on connectionist principles.

The third class of connectionist models addresses the more ambitious problem of learning the constituent structure of a language from untagged linguistic input. Such models have the potential to develop a new or unexpected notion of constituency, and hence may have substantial implications for theories of constituency in linguistics and psycholinguistics.

To understand how the more radical connectionist models address constituency, we can divide the problem of finding constituent structure into two interrelated parts: segmenting the sentence into chunks that correspond, to some extent, to linguistic constituents; and categorizing these units appropriately. The first problem is an aspect of the general problem of segmenting speech into appropriate units (e.g., phonemes, words, etc.) and more generally is an aspect of perceptual grouping. The second problem regards the general problem of classifying linguistic units, for instance, recognizing different classes of phonemes or establishing the parts of speech of individual lexical items. The segmentation and classification problems need not be solved sequentially. Indeed, there may be mutual influence between the decision to segment a particular chunk of language and the decision that it can be classified in a particular way. Nonetheless, it is useful to keep the two aspects of the analysis of constituency conceptually separate.

It is also important to stress the difference between the problem of assigning constituent structure to novel sentences, where the language is known, and the problem of acquiring the constituent structure of an unknown language. Statistical symbolic parsers are able to make some inroads into the first problem (Charniak, 1993). For highly stylized language input, and given a prestored grammar, they can apply grammatical knowledge to establish one or more possible constituent structures for novel sentences. But symbolic methods are much less advanced in acquiring the constituent structure of language, because this requires solving the hard problem of learning a grammar from a set of sentences generated by that grammar. It is therefore in relation to the **acquisition** of constituency that connectionist methods, with their well-developed learning methods, have attracted the most interest.

We begin by focusing on the problem of classifying, rather than segmenting, the linguistic input. One connectionist model (Finch and Chater, 1993) learns parts of speech of individual words by clustering words together on the basis of the immediate linguistic contexts in which they occur. The rationale is based on the replacement test: if two words are observed to occur in highly similar immediate contexts in a corpus, they probably belong to the same syntactic category. Finch and Chater used a single-layer network with Hebbian learning to store cooccurrences between target words and their immediate neighbors, such that each target word was associated with a vector representing the contexts in which it typically occurs. A competitive learning network classified these vectors, grouping together words with similar syntactic categories. This method is able to operate over unrestricted natural language, in contrast to most symbolic and connectionist models. From a linguistic perspective, the model slices lexical categories too finely, producing, for example, many word classes that correspond to nouns or verbs. On the other hand, the words within a class tend to be semantically related, which is useful from a cognitive perspective. The same method can be extended to classify sequences of words as NPs, VPs, etc. An initial classification of words is used to recode the input as a sequence of lexical constituents. Then short sequences of lexical constituents are classified by their context, as before. The resulting groups of phrases (e.g., Determiner-Adjective-Noun) are readily interpretable as NPs, VPs, PPs, and so on, but again these groupings are too linguistically restrictive (i.e., only a small number of NPs are included in any particular cluster). Moreover, this phrasal level classification has not yet been implemented in a connectionist network.

A different attack on the problem of constituency involves training simple recurrent networks (SRNs) on linguistic input (Elman, 1990). An SRN involves a crucial modification to a feed-forward network: the current set of hidden unit values is 'copied back' to a set of additional input units, and paired with the next input to the network (Figure 2). The current hidden unit values can thus directly affect the next hidden unit values, providing the network with a memory for past inputs. This enables it to tackle online sentence processing, where the input is revealed sequentially over time.

Segmentation into constituents can be achieved in two ways by an SRN trained to **predict** the next input.

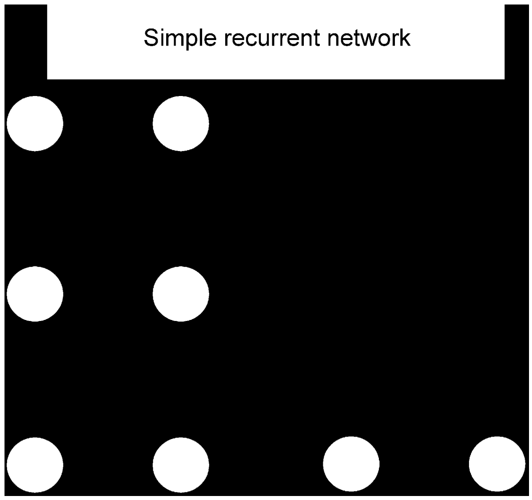

Figure 2 A simple recurrent network (SRN) is essentially a standard feed-forward network equipped with an extra layer of so-called context units. At each time step, an input propagates through the hidden units to the outputs (solid arrows). The hidden unit activation at the previous time step is copied 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.

One is based on the assumption that predictability is higher within constituents than across constituent boundaries, and hence that high prediction error indicates a boundary. This method has been advocated as potentially applicable at a range of linguistic levels (Elman, 1990), but in practice has only been successfully applied on corpora of unrestricted natural language input in finding word boundaries (Cairns *et al.*, 1997). Even here, the prediction strategy is a very partial cue to segmentation. If the network is provided with information about naturally occurring pauses between utterances (or parts of utterances), an alternative method is to assume that constituent boundaries occur where the network has an unusually high expectation of an utterance boundary. The rationale is that pauses tend to occur at constituent boundaries and hence the prediction of a possible utterance boundary suggests a constituent boundary may have occurred. This approach seems highly applicable to segmenting sentences into phrases, but has also primarily been used for finding word boundaries in real corpora of language, when combined with other cues (Christiansen *et al.*, 1998).

So far, we have considered how SRNs might find constituents. But how well do they classify constituents? At the word level, cluster analysis of hidden unit activations shows that, to some extent, the hidden unit patterns associated with different word classes group naturally into syntactic categories, for SRNs trained on simple artificial grammars (Elman, 1990). These results are important because they show that even though the SRN may not learn to classify constituents **explicitly**, it is nevertheless able to use this information to process constituents appropriately (see Fodor and Pylyshyn, 1988 versus van Gelder, 1998). Another way of assessing constituency learning is to see if networks generalize to predict novel sentences of a language. The logic is that to predict successfully, the SRN must exploit linguistic regularities that are defined across constituents, and hence develop a notion of constituency to do so. However, Hadley (1994) points out that this type of evidence is not compelling if the novel sentences are extremely similar to the network's training sentences. He suggested that, to show substantial evidence for generalization across constituents, the network should be able to handle novel sentences in which words appear in sentence locations where they have **not** previously occurred. For example, a novel sentence might involve a particular noun in object position, where it has previously occurred only in subject position. This property of the human brain, which Fodor and Pylyshyn (1988) call systematicity, captures the idea that if one can understand the sentence *Bob loves Mary* one can also understand *Mary loves Bob*; the underlying knowledge that supports understanding of the first sentence enables understanding the second. Hence, to generalize effectively, the network must presumably develop some abstract category of nouns.

Christiansen and Chater (1994) addressed Hadley's challenge presenting evidence that connectionist models are able to attain strong generalization. In their training corpus, the noun *boy* was prevented from ever occurring in a noun phrase conjunction (i.e., noun phrases such as *John and boy* and *boy and John* did not occur). During training, the SRN had therefore only been presented with singular verbs following *boy*. Nonetheless, the network was able to correctly predict that a plural verb must follow *John and boy* as prescribed by the grammar. In addition, the network was still able to correctly predict a plural verb when a prepositional phrase was attached to *boy* as in *John and boy from town*, providing even stronger evidence for strong generalization. In contrast, when the SRN was presented with ungrammatical lexical items in the second noun position, as in *John and near*, it did not activate the plural nouns. Instead, it activated lexical items that were not grammatical given the previous context. This suggests that the SRN is able to make nonlocal generalizations based on the structural regularities in the training corpus (see Christiansen and Chater, 1994 for further details). If the SRN relied solely on local information it would not have been able to make correct predictions in either case. Thus, the network demonstrated sophisticated generalization abilities, ignoring local word cooccurrence constraints while appearing to comply with structural information at the constituent level. Additional evidence of strong generalization in connectionist nets are found in Niklasson and van Gelder (1994) (but see Hadley, 1994 for a rebuttal).

Another strong test of generalizations that connectionist models seem to fail concerns equivalence relations: Marcus (1998) trained an SRN on sentences like *A rose is a rose* or *A tulip is a tulip*, but when given a novel sentence fragment like *A blicket is a*..., the SRN could not predict that *blicket* was going to be the next word because it activated all the words that it had seen in this sentence position (e.g., *rose*, *tulip*, etc.). This suggests that it did not develop abstract variable-based frames like *a X is a X*, where *X* is a variable that can be bound to any word. In reply to Marcus, Chang (2002) argued that a model of sentence production could generalize well if the input message was encoded in several separate representations that were linked together. Chang's architecture had one pathway for representing the mapping of object semantics to word forms and another for representing and mapping objects (and the words that describe them) into appropriate sentence positions. The model contained a bank of units representing event roles (agent, patient, etc.) that feeds onto a layer of lexical units (e.g., DOG, CAT, etc.). This way the model could represent the different roles of a word, say *dog*, in various events while activating a common semantics that all dogs have. Semantic units connected to word label units that represented the phonetics of each word, allowing the learning of a word label for each meaning. Since there was only one set of semantic units for different event role units, learning the mapping of the semantic feature DOG to the word label *dog* allowed the model to generalize this word to other event roles. Hence, even if the model had only experienced the sentence *the dog chases the cat*, it was able to generalize and produce a novel sentence *the cat chases the dog*.

Rohde (2002) developed a model (CSCP) that performs both comprehension and production of complex, multiclausal sentences using a larger corpus than previous models. In this respect, neural networks promise to scale up to real samples of natural language. The model deals with several sentence comprehension and production phenomena that involve discovery of constituent structure such as multiple verb tenses and voices, adverbs and adjectives, prepositional phrases, relative and subordinate clauses, and sentential complements. As in Chang (2002), the network has a semantic component, which is used to help the net predict the next word. The semantic component is also used to produce a word. When a word is selected for production, it is fed back into the comprehension input and the model proceeds to produce the next word, and so on. Therefore the model also claims that language production is learned primarily by formulating implicit predictions while attempting to comprehend one's language.

In this discussion, it must be noted that connectionist models do not mirror classical constituency precisely. That is, they do not derive rigid classes of words and phrases that are interchangeable across contexts. Rather, they divide words and phrases into clusters without precisely defined boundaries and treat them differently depending on the linguistic contexts in which they occur. This context-sensitive constituency can be viewed either as the undoing of connectionist approaches to language, or their radical contribution. The potential problem with context-sensitive constituency is the productivity of language: to take Chomsky's famous example, how do we know that the sentence *colorless green ideas sleep furiously* is syntactically correct, except by reference to a context-**insensitive** representation of the relevant word classes? This seems necessary, because each word occurs in a context where it has rarely been encountered before. But Allen and Seidenberg (1999) argue that this problem may not be fatal for context-sensitive notions of constituency. They trained a network to mutually 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: it was able to regenerate the word forms from the meanings and vice versa. Allen and Seidenberg then tested whether the network could recreate a sequence of word forms presented to it, by passing information from form to meaning and back. Ungrammatical sentences were recreated much less accurately than grammatical sentences, and the network was thus able to distinguish grammatical from ungrammatical sentences. Importantly, this was true for sentences in which words appeared in novel grammatical combinations, as specified by the systematicity criterion, and as exampled in Chomsky's famous sentence. Thus, the context-sensitivity of connectionist constituency may not rule out the possibility of highly creative and novel use of language, because abstract relations may be encoded at a semantic level, as well as at the level of word forms.

If the apparent linguistic limitations of context-sensitive constituency can be overcome, then the potential psychological contribution of this notion is enormous. First, context-sensitivity seems to be the norm, throughout human classification. Second, much data on sentence processing can be explained naturally by assuming that constituents are represented in a fuzzy and context-bound manner.

Structure Dependency

A linguistic concept intimately tied to constituency is structure dependency, the notion that grammatical knowledge relies on the structural relationships between constituents, rather than on the linear sequence of items. One of the fiercest arguments leveled at distributional models of learning (of which neural networks can be seen as an instantiation) concerns the uninformativeness of such mechanisms for detecting linguistically relevant properties (Pinker, 1984). Pinker argued that the most relevant properties of language are abstract, such as phrase structure configurations, grammatical relations, and syntactic categories, whereas information contained in raw input pertains to serial position, adjacency, and co-occurrence relations among words. To the extent that connectionist models, and in particular SRNs, process information simply relying on the linear sequencing of linguistic elements (at any level of analysis, phonemes, morphemes, words, etc.), they would seem doomed to fail to capture structure dependency. Reali and Christiansen (in press) demonstrated that SRNs could master a classic example of structure dependency in English, auxiliary fronting. In the generative linguistics framework, declaratives are turned into questions by fronting the correct auxiliary. For example, to turn the declarative form *The man who is hungry is ordering dinner*, it is correct to front the main clause auxiliary as in (1a), but fronting the subordinate clause auxiliary produces an ungrammatical sentence as in (1b) (Chomsky, 1980).

(1a) *Is the man who is hungry ordering dinner?*
(1b) **Is the man who hungry is ordering dinner?*

A structure-independent rule would move the first *is*, but the correct structure-dependent rule involves only the movement of the *is* from the main clause. Because children do not erroneously move the first *is* to the front of the sentence (Crain and Nakayama, 1987), and because they receive too little direct evidence for inferring correct auxiliary fronting (e.g., Chomsky, 1980), it has been claimed that structure dependency could not be learned from positive evidence by an associative mechanism. Reali and Christiansen trained SRNs to predict the next lexical category from a corpus of child-directed speech that contained no relevant examples of correct auxiliary fronting. The networks, tested on completely novel sentences, produced more accurate predictions for grammatical test sentences such as *Is the boy who is hungry nearby?* than for ungrammatical sentences, such as **Is the boy who hungry is nearby?* The prediction of the well-formed relative clause continuation, V (i.e., *is*), was highly preferred over the ill-formed version, ADJ (i.e., *hungry*). This pattern of predictions reflected the networks' sensitivity to the statistical properties of the corpus. The networks distinguished chunks of lexical categories that were more frequent in the training input from less frequent ones (i.e., the lexical categories corresponding to PRON V ADJ [*who is hungry*] vs. PRON ADJ V [*who hungry is*]). This work suggests that the networks were sensitive to the indirect statistical evidence present in the corpus, developing an appropriate bias toward the correct forms of AUX-questions without having a built-in rule for question formation.

Recursion

As with constituency, connectionist models have dealt with recursion in three increasingly radical ways. The least radical approach is to hardwire recursion into the network (e.g., as in Fanty's [1986] implementation of phrase structure rules) or to add an external symbolic (first-in-last-out) stack to the model (e.g., as in Kwasny and Faisal's [1990] deterministic connectionist parser). In both cases, recursive generativity is achieved entirely through standard symbolic means, and although this is a perfectly reasonable approach to recursion, it adds nothing new to symbolic accounts of natural language recursion. The more radical connectionist approaches to recursion aim for networks to *learn* to deal with recursive structure. One approach is to construct a modular system of networks, each of which is trained to acquire different aspects of syntactic processing. For example, Miikkulainen's (1996) system consists of three different networks: one trained to map words onto case-role assignments, another trained to function as a stack, and a third trained to segment the input into constituent-like units. Although the model displays complex recursive abilities, the basis for these abilities and their generalization to novel sentence structures derive from the configuration of the stack network combined with the modular architecture of the system rather than being discovered by the model.

The most radical connectionist approaches to recursion attempt to learn recursive abilities with minimal prior knowledge built into the system. In this type of model, the network is most often required to discover both the constituent structure of the input as well as how these constituents can be recursively assembled into sentences. As with the similar approach to constituency described above, such models may provide new insights into the notion of recursion in human language processing.

Before discussing these modeling efforts, we need to assess to what extent recursion is observed in human language behavior. It is useful to distinguish simple and complex recursion. Simple recursion consists in recursively adding new material to the left (e.g., the adjective phrases (AP) in *the grey cat → the fat grey cat → the ugly fat grey cat*) or right (the PPs in *the flowers in the vase → the flowers in the vase on the table → the flowers in the vase on the table by the window*) of existing phrase material. In complex recursion, new material is added in more complicated ways; for example, through center-embedding of sentences (*the chef admired the musicians → the chef who the waiter appreciated admired the musicians*). Psycholinguistic evidence shows that people find simple recursion relatively easy to process, whereas complex recursion is almost impossible to process with more than one level of recursion. For instance, the following sentence with two levels of simple (right-branching) recursion, *The busboy offended the waiter who appreciated the chef who admired the musicians* is much easier to comprehend than the comparable sentence with two levels of complex recursion, *The chef who the waiter who the busboy offended appreciated admired the musicians*. Because recursion is built into symbolic models, there are no intrinsic limitations on how many levels of recursion can be processed. Instead, such models must invoke extrinsic constraints, such as the competence/performance distinction, to accommodate the human performance asymmetry on simple and complex constructions. The radical connectionist approach models human performance directly without the need for extrinsic performance constraints.

The SRN model by Elman (1991) was perhaps the first connectionist attempt to simulate human behavior on recursive constructions. This network was trained on sentences generated by a small context-free grammar incorporating center-embedding and a single kind of right-branching recursive structure. In related work, Christiansen and Chater (1994) trained SRNs on a recursive artificial language incorporating four kinds of right-branching structures, a left-branching structure, and center-embedding. The behavior of these networks was qualitatively comparable with human performance in that the SRN predictions for right-branching structures were more accurate than on sentences of the same length involving center-embedding, and performance degraded appropriately when depth of center-embedding increased. Weckerly and Elman (1992) further corroborated these results, suggesting that semantic bias (incorporated via co-occurrence restrictions on the verbs) can facilitate network performance in center-embedded constructions similarly to the semantic facilitation effects found in human processing. Using abstract artificial languages, Christiansen and Chater (1999) show that the SRN's general pattern of performance is relatively invariant across network size and training corpus, and conclude that the human-like pattern of performance derives from intrinsic constraints inherent in the SRN architecture.

Connectionist models of recursive syntax typically use 'toy' fragments of grammar and small vocabularies. Aside from raising concerns over scaling up, this makes it difficult to provide detailed fits with empirical data. Nonetheless, some attempts have recently been made toward fitting existing data and deriving new empirical predictions from the models. For example, the Christiansen and Chater (1999) SRN model fits grammaticality ratings data from several behavioral experiments, including an account of the relative processing difficulty associated with processing center-embeddings (with the following relationship between nouns and verbs: $N_1\_N_2\_N_3\_V_3\_V_2\_V_1$) versus cross-dependencies (with the following relationship between nouns and verbs: $N_1\_N_2\_N_3\_V_1\_V_2\_V_3$). Human data have shown that sentences with two center-embeddings (in German) were significantly harder to process than comparable sentences with two cross-dependencies (in Dutch). Christiansen and Chater (1999) developed a measure of grammatical prediction error (GPE) that allowed 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, Christiansen and Chater (1999) fitted human data concerning the greater perceived difficulty associated with center-embedding in German compared to cross-serial dependencies in Dutch (Bach *et al.*, 1986). The simulation results demonstrated that the SRNs exhibited the same kind of qualitative processing difficulties as humans on these two types of complex recursive constructions. MacDonald and Christiansen (2002) derived novel predictions concerning other types of recursive constructions, and these predictions were later confirmed experimentally. They mapped single-word GPE scores 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 in human participants with different levels of reading comprehension ability.

Just as the radical connectionist approach to constituency deviates from classical constituency, the above approach to recursion deviates from the classical notion of recursion. The radical models of recursion do not acquire true recursion because they are unable to process infinitely complex recursive constructions. However, the classic notion of recursion may be ill-suited for capturing human recursive abilities. Indeed, the psycholinguistic data suggest that people's performance may be better construed as being only quasi-recursive. The above-mentioned semantic facilitation of recursive processing further suggests that human recursive performance may be partially context-sensitive; for example, the semantically biased *The bees that the hive that the farmer built housed stung the children* is easier to comprehend than neutral *The chef that the waiter that the busboy offended appreciated admired the musicians* even though both sentences contain two center-embeddings. This dovetails with the context-sensitive notion of constituency, and suggests that context-sensitivity may be a more pervasive feature of language processing than typically assumed by symbolic approaches.

Connectionist Models of Language Processing

What is the significance of connectionist models of language processing? Will connectionism ultimately replace, complement or simply implement symbolic approaches to language? Early connectionists addressed this issue by attempting to show that connectionism could, in principle, capture aspects of language and language processing (Christiansen and Chater, 2001). These models showed that connectionist networks could in principle acquire parts of linguistic structure without extensive innate knowledge. Recent work has moved toward a connectionist psycholinguistics that captures detailed psychological data. This article has outlined several ways in which constituency and recursion – two fundamental properties of linguistic knowledge – may be accommodated within a connectionist framework, ranging from direct implementation of symbolic systems to the acquisition of constituency and recursion from untagged input. We have focused on the radical approach because this has the greatest potential impact on psycholinguistics and linguistic theory.

However, much of this research is still preliminary. Future work is required to decide whether promising, but limited, initial results can eventually be scaled up to deal with the complexities of real language input, or whether a radical connectionist approach is beset by fundamental limitations. Another challenge is to find ways – theoretically and practically – to interface models, which have been proposed at different levels of linguistic analyses, with one another (e.g., interfacing models of morphology with models of sentence processing).

Nevertheless, the connectionist models described in this article have already influenced the study of language processing. First, connectionism has helped promote a general change toward replacing box-and-arrow diagrams with explicit computational models. Second, connectionism has reinvigorated the interest in computational models of learning, by including learning properties, such as recursion and constituent structure, which were previously assumed to be present *a priori* in humans, and therefore taken for granted. Finally, connectionism tends to discard the separation between competence and performance as artificially construed, and possibly misleading. From this perspective, linguistic recursion is a conceptual artifact of the competence/performance distinction instead of a necessary characteristic of the underlying computational mechanism. In this light, the problem facing connectionist models of language processing is not whether they can implement some kind of recursive mechanism, but whether they will be able to account for the (limited) recursive structure found in natural language behavior purely in terms of nonsymbolic computation.

Connectionism has thus already had a considerable impact on the psychology of language. But the final extent of this influence depends on the degree to which practical connectionist models can be developed and extended to deal with complex aspects of language processing in a psychologically realistic way. Recent models have already started to scale up to more realistic corpora and sentence processing behaviors (Kim *et al.*, 2002; Reali *et al.*, 2003; Rohde, 2002). If realistic connectionist models of language processing can be provided, then the possibility of a radical rethinking not just of the nature of language processing, but of the structure of language itself, may be required.

See also: Associationism and Connectionism; Computational Approaches to Language Acquisition; Constituent Structure; Formal Models and Language Acquisition; Generalization; Generative Grammar; Human Language Processing: Symbolic Models; Language Processing: Statistical Methods; Morphology in Parallel Distributed Processing; Philosophy of Linguistics; Systematicity.

Bibliography

Allen J & Seidenberg M S (1999). 'The emergence of grammaticality in connectionist networks.' In MacWhinney B

(ed.) *The emergence of language.* Mahwah, NJ: Lawrence Erlbaum Associates. 115–151.
Bach E, Brown C & Marslen-Wilson W (1986). 'Crossed and nested dependencies in Dutch and German.' *Language and Cognitive Processes 1,* 249–262.
Cairns P, Shillcock R C, Chater N & Levy J (1997). 'Bootstrapping word boundaries: a bottom-up corpus-based approach to speech segmentation.' *Cognitive Psychology 33,* 111–153.
Chang F (2002). 'Symbolically speaking: a connectionist model of sentence production.' *Cognitive Science 26,* 609–651.
Charniak E (1993). *Statistical language learning.* Cambridge: MIT Press.
Chomsky N (1980). *Rules and representation.* Cambridge: MIT Press.
Christiansen M H, Allen J & Seidenberg M S (1998). 'Learning to segment speech using multiple cues: a connectionist model.' *Language and Cognitive Processes 13,* 221–268.
Christiansen M H & Chater N (1994). 'Generalization and connectionist language learning.' *Mind and Language 9,* 273–287.
Christiansen M H & Chater N (1999). 'Toward a connectionist model of recursion in human linguistic performance.' *Cognitive Science 23,* 157–205.
Christiansen M H & Chater N (2001). 'Connectionist psycholinguistics: capturing the empirical data.' *Trends in Cognitive Sciences 5,* 82–88.
Crain S & Nakayama M (1987). 'Structure dependence in grammar formation.' *Language 63,* 522–543.
Elman J L (1990). 'Finding structure in time.' *Cognitive Science 14,* 179–211.
Elman J L (1991). 'Distributed representation, simple recurrent networks, and grammatical structure.' *Machine Learning 7,* 195–225.
Fanty M A (1986). 'Context-free parsing with connectionist networks.' In Denker J S (ed.) *Neural networks for computing.* New York: American Institute of Physics. 140–145.
Finch S & Chater N (1993). 'Learning syntactic categories: a statistical approach.' In Oaksford M & Brown G D A (eds.) *Neurodynamics and psychology.* New York: Academic Press. 295–321.
Fodor J & Pylyshyn Z (1988). 'Connectionism and cognitive architecture.' *Cognition 28,* 3–71.
Hadley R F (1994). 'Systematicity in connectionist language learning.' *Mind and Language 9,* 247–272.
Henderson J & Lane P (1998). 'A connectionist architecture for learning to parse.' In *Proceedings of 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics (COLING-ACL '98)*, University of Montreal, Canada. 531–537.
Kim A E, Srinivas B & Trueswell J C (2002). 'The convergence of lexicalist perspectives in psycholinguistics and computational linguistics.' In Merlo P & Stevenson S (eds.) *Sentence processing and the lexicon: formal, computational and experimental perspectives.* Philadelphia: John Benjamins. 109–135.
Kwasny S C & Faisal K A (1990). 'Connectionism and determinism in a syntactic parser.' *Connection Science 2,* 63–82.
MacDonald M C & Christiansen M H (2002). 'Reassessing working memory: a comment on Just & Carpenter (1992) and Waters & Caplan (1996).' *Psychological Review 109,* 35–54.
Marcus G F (1998). 'Rethinking eliminative connectionism.' *Cognitive Psychology 37,* 243–282.
Miikkulainen R (1996). 'Subsymbolic case-role analysis of sentences with embedded clauses.' *Cognitive Science 20,* 47–73.
Niklasson L & van Gelder T (1994). 'On being systematically connectionist.' *Mind and Language 9(3),* 288–302.
Pinker S (1984). *Language learnability and language development.* Cambridge: Harvard University Press.
Reali F, Christiansen M H & Monaghan P (2003). 'Phonological and distributional cues in syntax acquisition: scaling up the connectionist approach to multiple-cue integration.' In *Proceedings of the 25th Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum. 970–975.
Reali F & Christiansen M H (in press). 'Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence.' *Cognitive Science*.
Rohde D L T (2002). 'A connectionist model of sentence comprehension and production.' Ph.D. diss. Carnegie Mellon University.
Rohde D L T & Plaut D C (2003). 'Connectionist models of language processing.' *Cognitive Studies 10(1),* 10–28.
Van Gelder T J (1998). 'The dynamical hypothesis in cognitive science.' *Behavioral and Brain Sciences 21,* 1–14.
Weckerly J & Elman J L (1992). 'A PDP approach to processing center-embedded sentences.' In *Proceedings of the Fourteenth Annual Meeting of the Cognitive Science Society.* Hillsdale: Lawrence Erlbaum Associates. 414–419.