A Usage-Based Approach to Recursion in Sentence Processing

Morten H. Christiansen

Cornell University

Maryellen C. MacDonald

University of Wisconsin-Madison

Most current approaches to linguistic structure suggest that language is recursive, that recursion is a fundamental property of grammar, and that independent performance constraints limit recursive abilities that would otherwise be infinite. This article presents a usage-based perspective on recursive sentence processing, in which recursion is construed as an acquired skill and in which limitations on the processing of recursive constructions stem from interactions between linguistic experience and intrinsic constraints on learning and processing. A connectionist model embodying this alternative theory is outlined, along with simulation results showing that the model is capable of constituent-like generalizations and that it can fit human data regarding the differential processing difficulty associated with center-embeddings in German and cross-dependencies in Dutch. Novel predictions are furthermore derived from the model and corroborated by the results of four behavioral experiments, suggesting that acquired recursive abilities are intrinsically bounded not only when processing complex recursive constructions, such as center-embedding and cross-dependency, but also during processing of the simpler, right- and left-recursive structures.

Introduction

Ever since Humboldt (1836/1999, researchers have hypothesized that language makes "infinite use of finite means." Yet the study of language had to wait nearly

a century before the technical devices for adequately expressing the unbound-edness of language became available through the development of recursion theory in the foundations of mathematics (cf. Chomsky, 1965). Recursion has subsequently become a fundamental *property* of grammar, permitting a finite set of rules and principles to process and produce an infinite number of ex-pressions. Thus, recursion has played a central role in the generative approach to language from its very inception. It now forms the core of the Minimalist Program (Boeckx, 2006; Chomsky, 1995) and has been suggested to be the only aspect of the language faculty unique to humans (Hauser, Chomsky, &, Fitch, 2002).

Although generative grammars sanction infinitely complex recursive con-structions, people's ability to deal with such constructions is quite limited. In standard generative models of language processing, the unbounded recursive power of the grammar is therefore typically harnessed by postulating *extrin-sic* memory limitations (e.g., on stack depth; Church, 1982; Marcus, 1980). This article presents an alternative, usage-based view of recursive sentence structure, suggesting that recursion is not an innate property of grammar or an a priori computational property of the neural systems subserving language. Instead, we suggest that the ability to process recursive structure is acquired gradually, in an item-based fashion given experience with specific recursive constructions. In contrast to generative approaches, constraints on recursive regularities do not follow from extrinsic limitations on memory or processing; rather they arise from interactions between linguistic experience and architec-tural constraints on learning and processing (see also Engelmann & Vasishth, 2009; MacDonald & Christiansen, 2002), *intrinsic* to the system in which the knowledge of grammatical regularities is embedded. Constraints specific to particular recursive constructions are acquired as part of the knowledge of the recursive regularities themselves and therefore form an integrated part of the representation of those regularities. As we will see next, recursive constructions come in a variety of forms; but contrary to traditional approaches to recursion, we suggest that intrinsic constraints play a role not only in providing limitations on the processing of complex recursive structures, such as center-embedding, but also in constraining performance on the simpler right- and left-branching recursive structures—albeit to a lesser degree.

Varieties of Recursive Structure

Natural language is typically thought to involve a variety of recursive con-structions.[1] The simplest recursive structures, which also tend to be the most

common in normal speech, are either right-branching as in (1) or left-branching as in (2):

(1) a. John saw the dog that chased the cat.
     b. John saw the dog that chased the cat that bit the mouse.

(2) a. The fat black dog was sleeping.
     b. The big fat black dog was sleeping.

In the above example sentences, (1a) can be seen as incorporating a single level of right-branching recursion in the form of the embedded relative clause *that chased the cat*. Sentence (1b) involves two levels of right-branching recursion because of the two embedded relative clauses *that chased the cat* and *that bit the mouse*. A single level of left-branching recursion is part of (2a) in the form of the adjective *fat* fronting *black dog*. In (2b) two adjectives, *big* and *fat*, iteratively front *black dog*, resulting in a left-branching construction with two levels of recursion. Because right- and left-branching recursion can be captured by iterative processes, we will refer to them together as *iterative* recursion (Christiansen & Chater, 1999).

   Chomsky (1956) showed that iterative recursion of infinite depth can be processed by a finite-state device. However, recursion also exists in more complex forms that cannot be processed in its full, unbounded generality by finite-state devices. The best known type of such *complex* recursion is center-embedding as exemplified in (3):

(3) a. The dog that John saw chased the cat.
     b. The cat that the dog that John saw chased bit the mouse.

These sentences provide center-embedded versions of the right-branching recursive constructions in (1). In (3a), the sentence *John saw the dog* is embedded as a relative clause within the main sentence *the dog chased the cat*, generating one level of center-embedded recursion. Two levels of center-embedded recursion can be observed in (3b), in which *John saw the dog* is embedded within *the dog chased the cat*, which, in turn, is embedded within *the cat bit the mouse*.

   The processing of center-embedded constructions has been studied extensively in psycholinguistics for more than half a century. These studies have shown, for example, that English sentences with more than one center-embedding [e.g., sentence (3b)] are read with the same intonation as a list of random words (Miller, 1962), cannot easily be memorized (Foss & Cairns, 1970; Miller & Isard, 1964), are difficult to paraphrase (Hakes & Foss, 1970; Larkin & Burns, 1977) and comprehend (Blaubergs & Braine, 1974; Hakes,

Evans, & Brannon, 1976; Hamilton & Deese, 1971; Wang, 1970), and are judged to be ungrammatical (Marks, 1968). These processing limitations are not confined to English. Similar patterns have been found in a variety of languages, ranging from French (Peterfalvi & Locatelli, 1971), German (Bach, Brown, & Marslen-Wilson, 1986), and Spanish (Hoover, 1992) to Hebrew (Schlesinger, 1975), Japanese (Uehara & Bradley, 1996) and Korean (Hagstrom & Rhee, 1997). Indeed, corpus analyses of Danish, English, Finnish, French, German, Latin, and Swedish (Karlsson, 2007) indicate that doubly center-embedded sentences are practically absent from spoken language. Moreover, it has been shown that using sentences with a semantic bias or giving people training can improve performance on such structures, but only to a limited extent (Blaubergs & Braine, 1974; Powell & Peters, 1973; Stolz, 1967).

Symbolic models of sentence processing typically embody a rule-based competence grammar that permits unbounded recursion. This means that the models, unlike humans, can process sentences with multiple center-embeddings. Since Miller and Chomsky (1963), the solution to this mismatch has been to impose extrinsic memory limitations exclusively aimed at capturing the human performance limitations on doubly center-embedded constructions. Examples include limits on stack depth (Church, 1982; Marcus, 1980), limits on the number of allowed sentence nodes (Kimball, 1973) or partially complete sentence nodes in a given sentence (Stabler, 1994), limits on the amount of activation available for storing intermediate processing products as well as executing production rules (Just & Carpenter, 1992), the "self-embedding interference constraint" (Gibson & Thomas, 1996), and an upper limit on sentential memory cost (Gibson, 1998).

No comparable limitations are imposed on the processing of iterative recursive constructions in symbolic models. This may due to the fact that even finite-state devices with bounded memory are able to process right- and left-branching recursive structures of infinite length (Chomsky, 1956). It has been widely assumed that depth of recursion does not affect the acceptability (or processability) of iterative recursive structures in any interesting way (e.g., Chomsky, 1965; Church, 1982; Foss & Cairns, 1970; Gibson, 1998; Reich, 1969; Stabler, 1994). Indeed, many studies of center-embedding in English have used right-branching relative clauses as baseline comparisons and found that performance was better relative to the center-embedded stimuli (e.g., Foss & Cairns, 1970; Marks, 1968; Miller & Isard, 1964). A few studies have reported more detailed data on the effect of depth of recursion in right-branching constructions and found that comprehension also decreases as depth of recursion increases in these structures, although not too the same degree as with center-embedded

stimuli (e.g., Bach et al., 1986; Blaubergs & Braine, 1974). However, it is not clear from these results whether the decrease in performance is caused by recursion per se or is merely a byproduct of increased sentence length.

In this article, we investigate four predictions derived from an existing connectionist model of the processing of recursive sentence structure (Christiansen, 1994; Christiansen & Chater, 1994). First, we provide a brief overview of the model and show that it is capable of constituent-based generalizations and that it can fit key human data regarding the processing of complex recursive constructions in the form of center-embedding in German and cross-dependencies in Dutch. The second half of the article describes four online grammaticality judgment experiments testing novel predictions, derived from the model, using a word-by-word self-paced reading task. Experiments 1 and 2 tested two predictions concerning iterative recursion, and Experiments 3 and 4 tested predictions concerning the acceptability of doubly center-embedded sentences using, respectively, semantically biased stimuli from a previous study (Gibson & Thomas, 1999) and semantically neutral stimuli.

A Connectionist Model of Recursive Sentence Processing

Our usage-based approach to recursion builds on a previously developed Simple Recurrent Network (SRN; Elman, 1990) model of recursive sentence processing (Christiansen, 1994; Christiansen & Chater, 1994). The SRN, as illustrated in Figure 1, is essentially a standard feed-forward network equipped with an extra layer of so-called context units. The hidden unit activations from the previous time step are copied back to these context units and paired with the
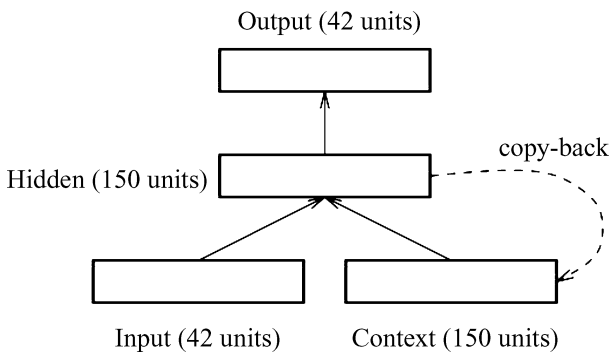

Figure 1 The basic architecture of the SRN used here as well as in Christiansen (1994) and Christiansen and Chater (1994). Arrows with solid lines denote trainable weights, whereas the arrow with the dashed line denotes the copy-back connections.

$$
\begin{array}{rcl}
S & \rightarrow & \text{NP VP .} \\
NP & \rightarrow & \text{N | N PP | N rel | PossP N | N } \textit{and } \text{NP} \\
VP & \rightarrow & V_i \text{ | } V_t \text{ NP | } V_o \text{ (NP) | } V_c \textit{ that } S \\
rel & \rightarrow & \textit{who } \text{VP | } \textit{who } \text{NP } V_{tlo} \\
PP & \rightarrow & \text{prep } N_{loc} \text{ (PP)} \\
PossP & \rightarrow & \text{(PossP) N Poss}
\end{array}
$$

Figure 2 The context-free grammar used to generate training stimuli for the connectionist model of recursive sentence processing developed by Christiansen (1994) and Christiansen and Chater (1994).

current input. This means that the current state of the hidden units can influence the processing of subsequent inputs, providing the SRN with an ability to deal with integrated sequences of input presented successively.

The SRN was trained via a word-by-word prediction task on 50,000 sentences (mean length: 6 words; range: 3–15 words) generated by a context-free grammar (see Figure 2) with a 38-word vocabulary.[2] This grammar involved left-branching recursion in the form of prenominal possessive genitives, right-branching recursion in the form of subject relative clauses, sentential complements, prepositional modifications of NPs, and NP conjunctions, as well as complex recursion in the form of center-embedded relative clauses. The grammar also incorporated subject noun/verb agreement and three additional verb argument structures (transitive, optionally transitive, and intransitive). The generation of sentences was further restricted by probabilistic constraints on the complexity and depth of recursion. Following training, the SRN performed well on a variety of recursive sentence structures, demonstrating that the SRN was able to acquire complex grammatical regularities.[3]

Usage-Based Constituents

A key question for connectionist models of language is whether they are able to acquire knowledge of grammatical regularities going beyond simple co-occurrence statistics from the training corpus. Indeed, Hadley (1994) suggested that connectionist models could not afford the kind of generalization abilities necessary to account for human language processing (see Marcus, 1998, for a similar critique). Christiansen and Chater (1994) addressed this challenge using the SRN from Christiansen (1994). In the training corpus, the noun *boy* had been prevented from ever occurring in a NP conjunction (i.e., NPs such

as *John and boy* and *boy and John* did not occur). During training, the SRN had therefore only seen 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. Additionally, 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*. This suggests that the SRN is able to make nonlocal generalizations based on the structural regularities in the training corpus (see Christiansen & 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.

Here, we provide a more stringent test of the SRN's ability to make appropriate constituent-based generalizations, using the four different types of test sentences shown in (4):

(4) a. Mary says that John and **boy** see.   (known word)
     b. Mary says that John and **zog** see.   (novel word)
     c. *Mary says that John and **near** see.   (illegal word)
     d. Mary says that John and **man** see.   (control word)

Sentence (4a) is similar to what was used by Christiansen and Chater (1994) to demonstrate correct generalization for the known word, *boy*, used in a novel position. In (4b), a completely novel word, *zog*, which the SRN had not seen during training (i.e., the corresponding unit was never activated during training) is activated as part of the NP conjunction. As an ungrammatical contrast, (4c) involves the activation of a known word, *near*, used in a novel but illegal position. Finally, (4d) provides a baseline in which a known word, *man*, is used in a position in which it is likely to have occurred during training (although not in this particular sentence).

Figure 3 shows the summed activation for plural verbs for each of the four sentence types in (4). Strikingly, both the known word in a novel position as well as the completely novel word elicited activations of the plural verbs that were just as high as for the control word. In contrast, the SRN did not activate plural verbs after the illegal word, indicating that it is able to distinguish between known words used in novel positions (which are appropriate given its distributionally defined lexical category) versus known words used in an ungrammatical context. Thus, the network demonstrated sophisticated generalization abilities, ignoring local word co-occurrence constraints while appearing to comply with structural information at the constituent level. It is important to note, however, that SRN is unlikely to have acquired constituency in a categorical form (Christiansen & Chater, 2003) but instead have acquired constituents
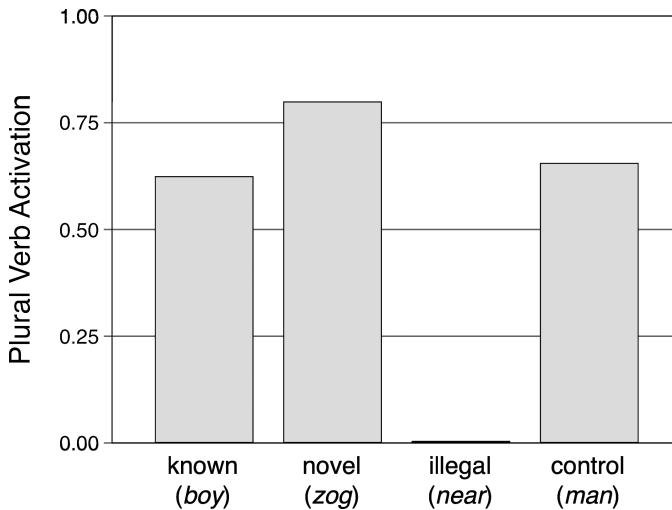
Figure 3 Activation of plural verbs after presentation of the sentence fragment *Mary says that John and N . . .*, where *N* is either a known word in a known position (*boy*), a novel word (*zog*), a known word in an illegal position (*near*), or a control word that have previously occurred in this position (*man*).

that are more in line with the usage-based notion outlined by Beckner and Bybee (this issue).

Deriving Novel Predictions

Simple Recurrent Networks have been employed successfully to model many aspects of psycholinguistic behavior, ranging from speech segmentation (e.g., Christiansen, Allen, & Seidenberg, 1998; Elman, 1990) and word learning (e.g., Sibley, Kello, Plaut, & Elman, 2008) to syntactic processing (e.g., Christiansen, Dale, & Reali, in press; Elman 1993; Rohde, 2002; see also Ellis & Larsen-Freeman, this issue) and reading (e.g., Plaut, 1999). Moreover, SRNs have also been shown to provide good models of nonlinguistic sequence learning (e.g., Botvinick & Plaut, 2004, 2006; Servan-Schreiber, Cleeremans, & McClelland, 1991). The human-like performance of the SRN can be attributed to an interaction between intrinsic architectural constraints (Christiansen & Chater, 1999) and the statistical properties of its input experience (MacDonald & Christiansen, 2002). By analyzing the internal states of SRNs before and after training with right-branching and center-embedded materials, Christiansen and Chater found that this type of network has a basic architectural bias toward locally bounded dependencies similar to those typically found in iterative

recursion. However, in order for the SRN to process multiple instances of iterative recursion, exposure to specific recursive constructions is required. Such exposure is even more crucial for the processing of center-embeddings because the network in this case also has to overcome its architectural bias toward local dependencies. Hence, the SRN does not have a built-in ability for recursion, but instead it develops its human-like processing of different recursive constructions through exposure to repeated instances of such constructions in the input.

In previous analyses, Christiansen (1994) noted certain limitations on the processing of iterative and complex recursive constructions. In the following, we flesh out these results in detail using the Grammatical Prediction Error (GPE) measure of SRN performance (Christiansen & Chater, 1999; MacDonald & Christiansen, 2002). To evaluate the extent to which a network has learned a grammar after training, performance on a test set of sentences is measured. For each word in the test sentences, a trained network should accurately predict the next possible words in the sentence; that is, it should activate all and only the words that produce grammatical continuations of that sentence. Moreover, it is important from a linguistic perspective not only to determine whether the activated words are grammatical given prior context but also which items are not activated despite being sanctioned by the grammar. Thus, the degree of activation of grammatical continuations should correspond to the probability of those continuations in the training set. The GPE assesses all of these facets of SRN performance, taking correct activations of grammatical continuations, correct suppression of ungrammatical continuations, incorrect activations of ungrammatical continuations, and incorrect suppressions of grammatical continuations into account (see Appendix A for details).

The GPE scores range between 0 and 1, providing a very stringent measure of performance. To obtain a perfect GPE score of 0, the SRN must not only predict all and only the next words prescribed by grammar but also be able to scale those predictions according to the lexical frequencies of the legal items. The GPE for an individual word reflects the difficulty that the SRN experienced for that word given the previous sentential context, and it can be mapped qualitatively onto word reading times, with low GPE values reflecting a prediction for short reading times and high values indicating long predicted reading times (MacDonald & Christiansen, 2002). The mean GPE averaged across a sentence expresses the difficulty that the SRN experienced across the sentence as a whole, and such GPE values have been found to correlate with sentence grammaticality ratings (Christiansen & Chater, 1999), with low mean GPE scores predicting low grammatical complexity ratings and high
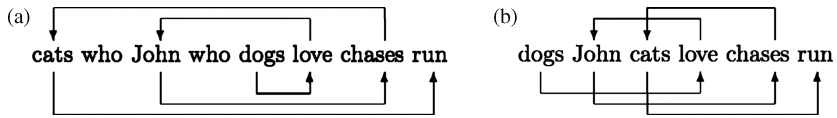
(a)

cats who John who dogs love chases run

(b)

dogs John cats love chases run

Figure 4 An illustration of the dependencies between subject nouns and verbs (arrows below) and between transitive verbs and their objects (arrows above) in sentences with two center-embeddings (a) and two cross-dependencies (b).

scores indicating a prediction for high complexity ratings. Next, we first use mean sentence GPE scores to fit data from human experiments concerning the processing of complex recursive constructions in German and Dutch, after which we derive novel predictions concerning human grammaticality ratings for both iterative and center-embedded recursive constructions in English and present four experiments testing these predictions.

Center-Embedding Versus Cross-Dependency

Center-embeddings and cross-dependencies have played an important role in the theory of language. Whereas center-embedding relations are nested within each other, cross-dependencies cross over one another (see Figure 4). As noted earlier, center-embeddings can be captured by context-free grammars, but cross-dependencies require a more powerful grammar formalism (Shieber, 1985). Perhaps not surprisingly, cross-dependency constructions are quite rare across the languages of the world, but they do occur in Swiss-German and Dutch. An example of a Dutch sentence with two cross-dependencies is shown in (5), with subscripts indicating dependency relations.

(5)   De mannen$_1$ hebben Hans$_2$ Jeanine$_3$ de paarden helpen$_1$ leren$_2$ voeren$_3$

    Literal: The men have Hans Jeanine the horses help teach feed

    Gloss: The men helped Hans teach Jeanine to feed the horses

Although cross-dependencies have been assumed to be more difficult to process than comparable center-embeddings, Bach et al. (1986) found that sentences with two center-embeddings in German were significantly harder to process than comparable sentences with two cross-dependencies in Dutch.

In order to model the comparative difficulty of processing center-embeddings versus cross-dependencies, we trained an SRN on sentences generated by a new grammar in which the center-embedded constructions were replaced by cross-dependency structures (see Figure 5). The iterative

$$S \rightarrow NP\ VP\ .$$
$$S_{cd} \rightarrow N_1\ N_2\ V_{1(tlo)}\ V_{2(i)}\ .$$
$$S_{cd} \rightarrow N_1\ N_2\ N\ V_{1(tlo)}\ V_{2(tlo)}\ .$$
$$S_{cd} \rightarrow N_1\ N_2\ N_3\ V_{1(tlo)}\ V_{2(tlo)}\ V_{3(i)}\ .$$
$$S_{cd} \rightarrow N_1\ N_2\ N_3\ N\ V_{1(tlo)}\ V_{2(tlo)}\ V3_{(tlo)}\ .$$
$$NP \rightarrow N\ |\ N\ PP\ |\ N\ rel\ |\ PossP\ N\ |\ N\ and\ NP$$
$$VP \rightarrow V_i\ |\ V_t\ NP\ |\ V_o\ (NP)\ |\ V_c\ that\ S$$
$$rel \rightarrow who\ VP$$
$$PP \rightarrow prep\ N_{loc}\ (PP)$$
$$PossP \rightarrow (PossP)\ N\ Poss$$

Figure 5 The context-sensitive grammar used to generate training stimuli for the connectionist model of recursive sentence processing developed by Christiansen (1994).

recursive constructions, vocabulary, and other grammar properties remained the same as in the original context-free grammar. Thus, only the complex recursive constructions differed across the two grammars. In addition, all training and network parameters were held constant across the two simulations. After training, the cross-dependency SRN achieved a level of general performance comparable to that of the center-embedding SRN (Christiansen, 1994). Here, we focus on the comparison between the processing of the two complex types of recursion at different depths of embedding.

Bach et al. (1986) asked native German speakers to provide comprehensibility ratings of German sentences involving varying depths of recursion in the form of center-embedded constructions and corresponding right-branching paraphrases with the same meaning. Native Dutch speakers were tested using similar Dutch materials but with the center-embedded constructions replaced by cross-dependency constructions. The left-hand side of Figure 6 shows the Bach et al. results, with the ratings for the right-branching paraphrase sentences subtracted from the matching complex recursive test sentences to remove effects of processing difficulty due to length. The SRN results—the mean sentence GPE scores averaged over 10 novel sentences—are displayed on the right-hand side of Figure 6. For both humans and SRNs, there is no difference in processing difficulty for the two types of complex recursion at one level of embedding. However, for doubly embedded constructions, center-embedded structures
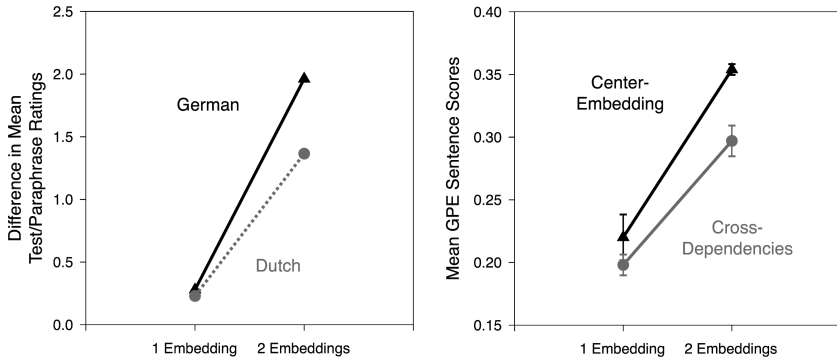
Figure 6 Human performance (from Bach et al., 1986) on center-embedded constructions in German and cross-dependency constructions in Dutch with one or two levels of embedding (left panel). SRN performance on similar complex recursive structures (right panel).

(in German) are harder to process than comparable cross-dependencies (in Dutch). These simulation results thus demonstrate that the SRNs exhibit the same kind of qualitative processing difficulties as humans do on the two types of complex recursive constructions (see also Christiansen & Chater, 1999). Crucially, the networks were able to match human performance without needing complex external memory devices (such as a stack of stacks; Joshi, 1990). Next, we go beyond fitting existing data to explore novel predictions made by the center-embedding SRN for the processing of recursive constructions in English.

Experiment 1: Processing Multiple Right-Branching Prepositional Phrases

In most models of sentence processing, multiple levels of iterative recursion are represented by having the exact same structure occurring several times (e.g., multiple instances of a PP). In contrast, the SRN learns to represent each level of recursion slightly differently from the previous one (Elman, 1991). This leads to increased processing difficulty as the level of recursion grows because the network has to keep track of each level of recursion separately, suggesting that depth of recursion in iterative constructions should affect processing difficulty beyond a mere length effect. Based on Christiansen's (1994) original analyses, we derived specific predictions for sentences involving zero, one, or two levels of right-branching recursion in the form of PP modifications of an NP[4] as shown in (6):

(6) a. The nurse with the vase says that the flowers by the window resemble
        roses. (1 PP)
     b. The nurse says that the flowers in the vase by the window resemble
        roses. (2 PPs)
     c. The blooming flowers in the vase on the table by the window resemble
        roses. (3 PPs)

Predictions were derived from the SRNs for these three types of sentences and
tested with human participants using a variation of the "stop making sense"
sentence-judgment paradigm (Boland, 1997; Boland, Tanenhaus, & Garnsey,
1990; Boland, Tanenhaus, Garnsey, & Carlson, 1995), with a focus on grammat-
ical acceptability rather than semantic sensibility. Following the presentation of
each sentence, participants rated the sentence for grammaticality on a 7-point
scale; these ratings were then compared with the SRN predictions.

Method

Participants

Thirty-six undergraduate students from the University of Southern California
received course credit for participation in this experiment. All participants in
this and subsequent experiments were native speakers of English with normal
or corrected-to-normal vision.

Materials

Nine experimental sentences were constructed with 1 PP, 2 PPs, and 3 PPs ver-
sions as in (6). All items are from this and subsequent experiments are included
in Appendix B. Each sentence version had the same form as (6a)–(6c). The 1
PP sentence type began with a definite NP modified by a single PP (*The nurse
with the vase*), followed by a sentential complement verb and a complementizer
(*says that*), a definite NP modified by a second single PP (*the flowers by the
window*), and a final transitive VP with an indefinite noun (*resemble roses*). The
2 PP sentence type began with the same definite NP as 1 PP stimuli, followed
by the same sentential complement verb and complementizer, a definite NP
modified by a recursive construction with 2 PPs (*the flowers in the vase by the
window*), and the same final transitive VP as 1 PP stimuli. The 3 PP sentence
type began with a definite NP including an adjective (*The blooming flowers*),
modified by a recursive construction with 3 PPs (*in the vase on the table by
the window*), and the same transitive VP as in the other two sentence types.
Each sentence was 14 words long and always ended with the same final NP (*the
window*) and VP (*resemble roses*).

The three conditions were counterbalanced across three lists. In addition, 9 practice sentences and 42 filler sentences were created to incorporate a variety of recursive constructions of equal complexity to the experimental sentences. Two of the practice sentences were ungrammatical as were nine of the fillers. Twenty-one additional stimulus items were sentences from other experiments and 30 additional fillers mixed multiple levels of different kinds of recursive structures.

Procedure

Participants read sentences on a computer monitor, using a word-by-word center presentation paradigm. Each trial started with a fixation cross at the center of the screen. The first press of the space bar removed the fixation cross and displayed the first word of the sentence, and subsequent presses removed the previous word and displayed the next word. For each word, participants decided whether what they had read so far was a grammatical sentence of English. Participants were instructed that both speed and accuracy were important in the experiment and to base their decisions on their first impression about whether a sentence was grammatical. If the sentence read so far was considered grammatical, the participants would press the space bar—if not, they would press a NO key when the sentence became ungrammatical. The presentation of a sentence ceased when the NO was pressed.

When participants finished a sentence, either by reading it all the way through with the space bar or by reading it part way and then pressing the NO key when it became ungrammatical, the screen was cleared and they would be asked to rate how "good" this sentence was.[5] The participants would respond by pressing a number between 1 and 7 on the keyboard, with 1 indicating that the sentence was "perfectly good English" and 7 indicating that it was "really bad English." Participants were encouraged to use the numbers in between for intermediate judgments. The computer recorded the response of the participant.

Participants were assigned randomly to three counterbalanced lists. Each participant saw a different randomization of experimental and filler items.

SRN Testing

The model was tested on three sets of sentences corresponding to the three types shown in (6). The determiner *the* and the adjective in (6c) (*blooming*) could not be included in test sentences because they were not found in the training grammar. Moreover, the actual lexical items used in the network simulations were different from those in the human experiment because of limitations imposed by the training vocabulary, but the lexical categories remained the

same. The three sentence types had the same length as in the experiment, save that (6c) was one word shorter. All sentences involved at least two PPs [although only in (6b) and (6c) were they recursively related]. The crucial factor differentiating the three sentence types is the number of PPs modifying the subject noun (*flowers*) before the final verb (*resemble*). The sentence types were created to include 1, 2, or 3 PPs in this position. In order to ensure that the sentences were equal in length, right-branching sentential complements (*says that* ...) were used in (6a) and (6b) such that the three sentence types are of the same global syntactic complexity. Mean GPE scores were recorded for 10 novel sentences of each type.

Results and Discussion
SRN Predictions
Although the model found the sentences relatively easy to process, there was a significant effect of depth of recursion on GPE scores, $F(2, 18) = 13.41, p < .0001$, independent of sentence length (see Table 1). Thus, the model predicted an effect of sentence type for human ratings, with 3 PPs (6c) rated substantially worse than 2 PPs (6b), which, in turn, should be rated somewhat worse than 1 PP (6a).

Rejection Data
The PP stimuli were generally grammatically acceptable to our participants, with only 6.48% (21 trials) rejected during the reading/judgment task. Only 4.63% of the 1 PP stimuli and 3.70% of the 2 PP stimuli were rejected, and the difference between the two rejection scores was not significant, $\chi^2(1) < 0.1$. In contrast, 11.11% of the items with 3 PPs were rejected—an increase in rejection rate that was significant compared with the 2 PP condition, $\chi^2(1) = 3.51, p < .05$, but only marginally significant in comparison with the 1 PP condition, $\chi^2(1) = 2.43, p = .0595$. Thus, there was a tendency to perceive the 1 PP and 2 PP stimuli as more grammatical than the counterpart with 3 PPs. Figure 7 shows the cumulative profile of rejections across word position in the sentences, starting at the fourth word. Rejections across the three sentence types were more likely to occur toward the end of a sentence, with two thirds of the rejections occurring during the presentation of the last four words, and with only three sentences rejected before the presentation of the 10th word (i.e., *by* in Figure 7). The rejection profile for the 3 PP stimuli suggests that it is the occurrence of the third PP (*by the window*) that makes these stimuli less acceptable than the 1 PP and 2 PP stimuli.
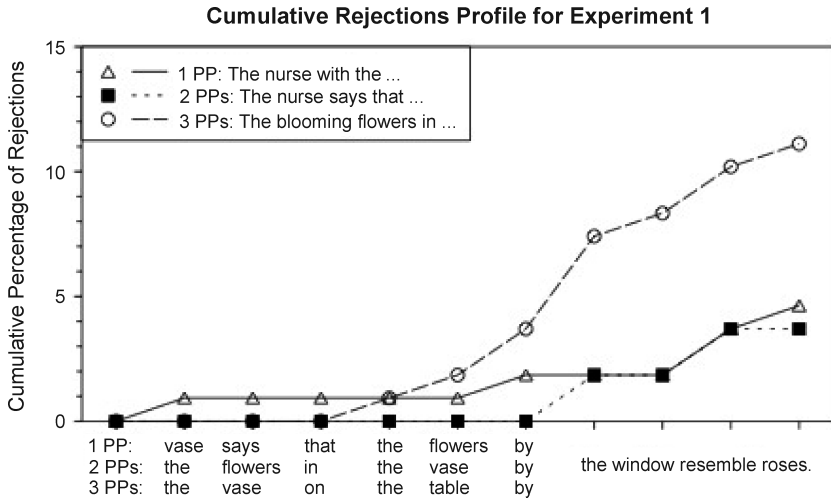
Cumulative Rejections Profile for Experiment 1

Figure 7 The cumulative percentage of rejections for each PP condition at each word position is shown starting from the fourth word.

Table 1 The processing difficulty of multiple PP modifications of NPs

No. of PPs	SRN predictions		Human results	
	Mean GPE	SD	Mean rating	SD
1 PP	0.153	0.008	2.380	0.976
2 PPs	0.161	0.015	2.704	0.846
3 PPs	0.214	0.042	3.269	0.925

Note. NP = noun phrase; PP = prepositional phrase.

Grammaticality Ratings

The ratings for the three sentence types are shown in Table 1. As predicted by the connectionist model, there was a significant effect of sentence type, $F_1(2, 70) = 10.87, p < .0001; F_2(2, 16) = 12.43, p < .001$, such that the deeper the level of recursion, the worse the sentences were rated. The model also predicted that there should be only a small difference between the ratings for the 1 PP and the 2 PP stimuli but a significant difference between the stimuli with the 2 PPs and 3 PPs. The experiment also bears out this prediction t. The stimuli with the 2 PPs were rated only 13.62% worse than the 1 PP stimuli—a difference that was only marginally significant, $F_1(1, 35) = 2.97, p = .094; F_2(1, 8) = 4.56, p = .065$. The items with 3 PPs elicited the worst ratings, which were 37.36% worse than the 1 PP items and 20.89% worse than the 2 PP items. The rating

difference between the sentences with 2 PPs and 3 PPs was significant, $F_1(1, 35) = 5.74, p < .005; F_2(1, 8) = 10.90, p < .02$).

The human ratings thus confirmed the predictions from the connectionist model: Increasing the depth of right-branching recursion has a negative effect on processing difficulty that cannot be attributed to a mere length effect. As predicted by the model, the deeper the level of recursion across the three types of stimuli, the worse the sentences were rated by the participants. This result is not predicted by most other current models of sentence processing, in which right-branching recursion does not cause processing difficulties beyond potential length effects (although see Lewis & Vasishth, 2005).

Experiment 2: Processing Multiple Left-Branching Possessive Genitives

In addition to the effect of multiple instances of right-branching iterative recursion on processing as confirmed by Experiment 1, Christiansen (1994) also observed that the depth of recursion effect in left-branching structures varied in its severity depending on the sentential position in which such recursion occurs. When processing left-branching recursive structures involving multiple prenominal genitives, the SRN learns that it is not crucial to keep track of what occurs before the final noun. This tendency is efficient early in the sentence but creates a problem with recursion toward the end of sentence because the network becomes somewhat uncertain where it is in the sentence. We tested this observation in the context of multiple possessive genitives occurring in either subject (7a) or object (7b) positions in transitive constructions:

(7) a. Jane's dad's colleague's parrot followed the baby all afternoon. (Subject)
     b. The baby followed Jane's dad's colleague's parrot all afternoon. (Object)

Method
SRN Testing
The model was tested as in Experiment 1 on two sets of 10 novel sentences corresponding the two types of sentences in (7).

Participants
Thirty-four undergraduate students from the University of Southern California received course credit for participation in this experiment.

Materials
We constructed 10 experimental items with the same format as (7). As in (7a), the Subject stimuli started with three prenominal genitives, of which the first

always contained a proper name (*Jane's dad's colleague's*), followed by the subject noun (*parrot*), a transitive verb (*followed*), a simple object NP (*the baby*), and a duration adverbial (*all afternoon*). The Object stimuli reversed the order of the two NPs, placing the multiple prenominal genitives in the object position and the simple NP in the subject position, as illustrated by (7b). The conditions were counterbalanced across two lists, each containing five sentences of each type. Additionally, there were 9 practice items (including one ungrammatical), 29 filler items (of which 9 were ungrammatical), and 20 items from other experiments.

Procedure
Experiment 2 involved the same procedure as Experiment 1.

Results and Discussion
SRN Predictions
Comparisons of mean sentence GPE for the two types of sentence materials predicted that having two levels of recursion in an NP involving left-branching prenominal genitives should be significantly less acceptable in an object position compared to a subject position, $F(1, 9) = 110.33, p < .0001$.

Rejection Data
Although the genitive stimuli seemed generally acceptable, participants rejected twice as many sentences (13.24%) as in Experiment 1. The rejection profiles for the two sentence types are illustrated in Figure 8, showing that the rejections are closely associated with the occurrence of the multiple prenominal genitives. However, there was no overall difference in the number of sentences rejected in the Subject (13.53%) and Object (12.94%) conditions, $\chi^2(1) < 1$.

Grammaticality Ratings
As predicted by the SRN model, the results in Table 2 show that multiple prenominal genitives were less acceptable in object position than in subject position, $F_1(1, 33) = 5.76, p < .03; F_2(1, 9) = 3.48, p = .095$. These results suggest that the position of multiple instances of recursion within a sentence affects its acceptability.

Experiment 3: Processing Multiple Semantically Biased Center-Embeddings

In contrast to iterative recursion, complex recursion in the form of center-embedding has often been used as an important source of information about
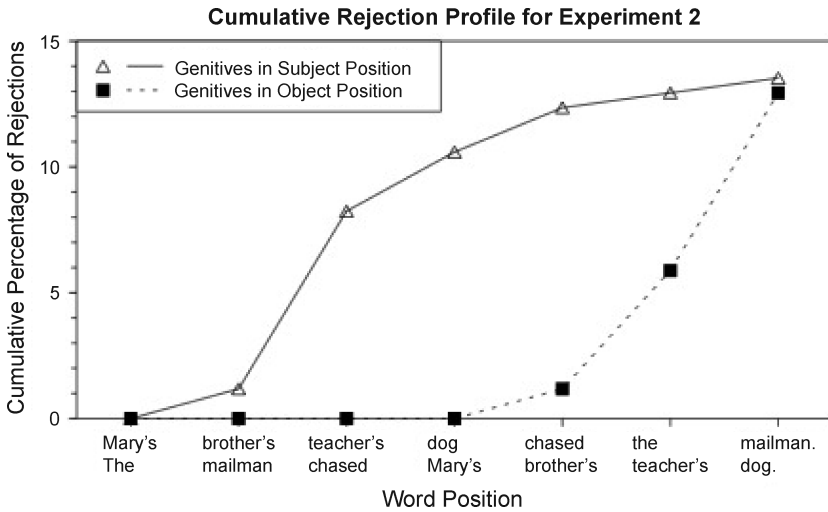
Cumulative Rejection Profile for Experiment 2

Figure 8 The cumulative percentage of rejections for sentences incorporating multiple prenominal genitives in subject or object positions.

complexity effects in human sentence processing (e.g., Blaubergs & Braine, 1974; Foss & Cairns, 1970; Marks, 1968; Miller, 1962; Miller & Isard, 1964; Stolz, 1967). Of particular interest is a study by Gibson and Thomas (1999) investigating the role of memory limitations in the processing of doubly center-embedded object relative clause constructions. Consistent with the external memory limitation account of Gibson (1998), they found that when deleting the middle VP [*was cleaning every week* in (8a)], the resulting ungrammatical sentence (8b) was rated no worse that the original grammatical version.

(8) a. The apartment that the maid who the service had sent over was cleaning every week was well decorated. (3 VPs)

Table 2 The processing difficulty of multiple possessive genitives

Genitive position	SRN predictions		Human results	
	Mean GPE	SD	Mean rating	SD
Sub NP	0.222	0.006	3.606	1.715
Obj NP	0.299	0.025	3.965	1.752

Note. Sub NP = subject noun phrase; Obj NP = object noun phrase.

b. *The apartment that the maid who the service had sent over was well decorated. (2 VPs)

In contrast, Christiansen (1994) noted that the SRN tended to expect that doubly center-embedded sentences would end when it had received two verbs, suggesting that (8b) should actually be rated better than (8a). Christiansen and Chater (1999) further demonstrated that this prediction is primarily due to intrinsic architectural limitations on the processing on doubly center-embedded material rather than insufficient experience with these constructions.[6] Gibson and Thomas' results came from offline ratings, whereas in Experiment 3 we use the online method from the previous two experiments and predict that with this more sensitive measure, sentences such as the ungrammatical (8b) will actually be rated better than grammatical sentences like (8a).

Method

SRN Testing

The model was tested as in Experiments 1 and 2 on two sets of 10 novel sentences corresponding to the sentence types in (8).

Participants

Thirty-six undergraduate students from the University of Southern California received course credit for participation in this experiment.

Materials

Six experimental items were selected from the Gibson and Thomas (1999) stimuli, focusing on the key grammatical (3 VP) versus ungrammatical (2 VPs) version of each sentence as in (8). The two conditions were counterbalanced across two lists, three of each type. In addition, there were 9 practice items (including 2 ungrammatical), 30 filler items (of which 9 were ungrammatical), and 27 items from other experiments.

Procedure

Experiment 3 involved the same procedure as Experiments 1 and 2.

Results and Discussion

SRN Predictions

The mean GPE scores across the two types of sentences followed the preliminary findings by Christiansen and Chater (1999): The grammatical 3-VP sentences were rated significantly worse than the ungrammatical 2-VP sentences, $F(1, 9) = 2892.23, p < .0001$.
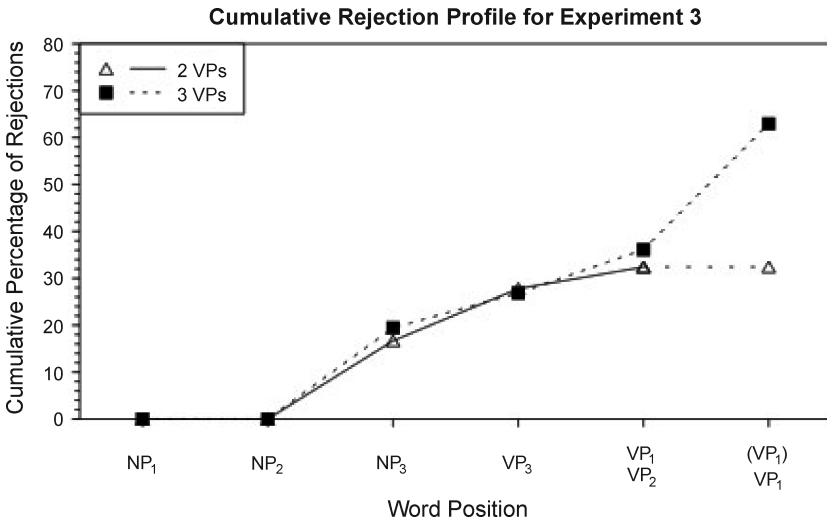
Figure 9 The cumulative percentage of rejections averaged across NP and VP regions in semantically biased center-embedded sentences with 2 VPs or 3 VPs.

Rejection Data

Because the stimuli from Gibson and Thomas (1999) were not equated for length, the number of rejections were averaged for each NP and VP region rather than for each word. The resulting cumulative rejection profile is shown in Figure 9, indicating that significantly more 3 VP sentences were rejected than 2 VP sentences [63% vs. 32.4%; $\chi^2(1) = 20.21, p < .0001$].

Grammaticality Ratings

As predicted by the SRN model (Table 3), the grammatical 3 VP sentences were rated significantly worse than their ungrammatical 2 VP counterparts, $F_1(1, 35) = 15.55, p < .0001; F_2(1, 5) = 6.85, p < .05$. These results suggest

Table 3 The processing difficulty of multiple semantically biased center-embeddings

No. of VPs	SRN predictions		Human results	
	Mean GPE	SD	Mean rating	SD
2 VPs	0.307	0.010	4.778	1.268
3 VPs	0.404	0.005	5.639	1.037

Note. VPs = verb phrases.

humans share the SRN's processing preference for the ungrammatical 2 VP construction over the grammatical 3 VP version (for similar SRN results and additional data for German, see Engelmann & Vasishth, 2009).

Experiment 4: Processing Multiple Semantically Neutral Center-Embeddings

Two potential concerns about the Gibson and Thomas (1999) stimuli used in Experiment 3 are that (a) the results could be an artifact of length because the sentences were not controlled for overall length and (b) the stimuli included semantic biases [e.g., *apartment/decorated*, *service/sent over* in (6b)] that may have increased the plausibility of the 2 VP stimuli. In Experiment 4, we sought to replicate the results from Experiment 3 with semantically neutral stimuli adapted from Stolz (1967), in which adverbs replaced the missing verbs in 2 VP constructions to control for overall length as in (9):

(9) a. The chef who the waiter who the busboy offended appreciated admired the musicians. (3 VPs)
   b. *The chef who the waiter who the busboy offended frequently admired the musicians. (2 VPs)

Method
SRN Testing
The training corpus on which the model was trained did not include semantic constraints (e.g., animacy). Instead, the difference between the center-embedded test items used to make SRN predictions for Experiments 3 and 4 was one of argument structure. The Gibson and Thomas (1999) stimuli in Experiment 3 used optionally transitive verbs, whereas the Experiment 4 stimuli contained transitive verbs. The model was tested as in the previous experiments on two sets of 10 novel sentences matching the structure of the two sentence types in (9).

Participants
Thirty-four undergraduate students from the University of Southern California received course credit for participation in this experiment.

Materials
Ten semantically neutral doubly center-embedded items were adapted from Stolz (1967), each with a 3 VP and 2 VP version as in (9). The conditions were counterbalanced across two lists, each containing five sentences of each
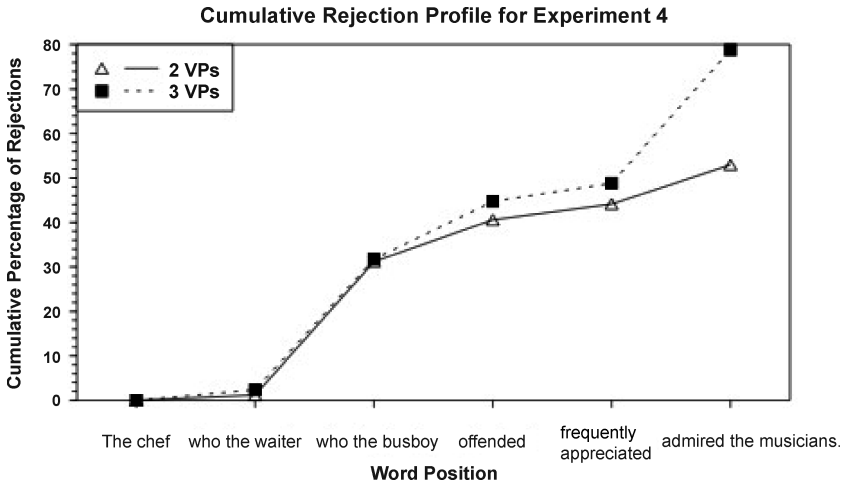
Cumulative Rejection Profile for Experiment 4

Figure 10 The cumulative percentage of rejections averaged across word position in semantically neutral center-embedded sentences with 2 VPs or 3 VPs.

type. Additionally, there were 9 practice items (including one ungrammatical), 29 filler items (of which 9 were ungrammatical), and 20 items from other experiments.

Procedure
Experiment 4 involved the same procedure as Experiments 1–3.

Results and Discussion
SRN Predictions
As in Experiment 3, the mean GPE scores predicted that 3 VP sentences should be rated significantly worse than 2 VP sentences, $F(1, 9) = 43.60, p < .0001$.

Rejection Data
The cumulative rejection profile for the neutral items in the current experiment replicated that for the semantically biased stimuli in the previous experiment (Figure 10): significantly more 3 VP sentences (78.8%) were rejected than the corresponding 2 VP constructions [52.9%; $\chi^2(1) = 25.33, p < .0001$].

Grammaticality Ratings
Again, in line with the SRN predictions (Table 4), the 3 VP sentences were rated significantly worse than their 2 VP counterparts, $F_1(1, 33) = 7.88, p < .01; F_2(1, 9) = 27.46, p < .001$. Thus, the ungrammatical 2 VP constructions

Table 4 The processing difficulty of multiple semantically neutral center-embeddings

No. of VPs	SRN predictions		Human results	
	Mean GPE	*SD*	Mean rating	*SD*
2 VPs	0.360	0.015	5.553	1.251
3 VPs	0.395	0.008	6.165	0.672

Note. VPs = verb phrases.

are preferred over the grammatical 3 VP versions even when controlling for overall length and the influence of semantic bias.

General Discussion

We have presented simulation results from a connectionist implementation of our usage-based approach to recursion, indicating that the model has sophisticated constituent-based generalization abilities and is able to fit human data regarding to the differential processing difficulty of center-embedded and cross-dependency constructions. Novel predictions were then derived from the model and confirmed by the results of four grammaticality judgment experiments. Importantly, this model was not developed for the purpose of fitting these data but was, nevertheless, able to predict the patterns of human grammaticality judgments across three different kinds of recursive structure. Indeed, as illustrated by Figure 11, the SRN predictions not only provide a close fit with the human ratings *within* each experiment but also capture the increased complexity evident *across* Experiments 1–4. Importantly, the remarkably good fit between the model and the human data both within and across the experiments were obtained without changing any parameters across the simulations. In contrast, the present pattern of results provides a challenge for most other accounts of human sentence processing that rely on arbitrary, externally specified limitations on memory or processing to explain patterns of human performance.

Like other implemented computational models, the specific instantiation of our usage-based approach to recursive sentence processing presented here is not without limitations. Although the model covers several key types of recursive sentence constructions, its overall coverage of English is limited in both vocabulary size and range of grammatical regularities. Another limiting factor is that the model predicts only the next word in a sentence. Despite mounting evidence highlighting the importance of prediction to learning, in general (Niv & Schoenbaum, 2008), and language processing, in particular (e.g., Federmeier, 2007; Levy, 2008; Hagoort, in press; Pickering & Garrod, 2007), incorporating
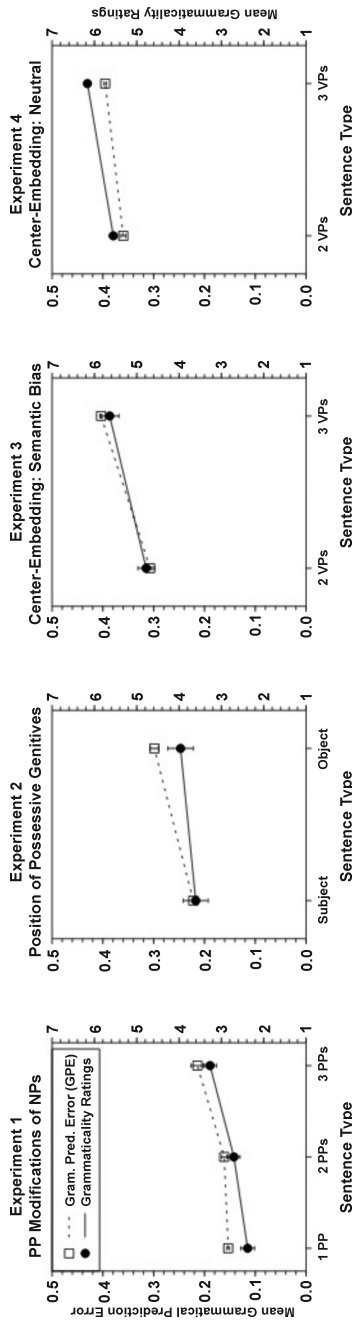
Figure 11 Comparison between SRN predictions of sentence complexity (left *y*-axes: mean GPE) and human ratings (right *y*-axes: mean grammaticality ratings) for four recursive sentence types: (a) multiple PP modifications of NPs (Experiment 1), (b) multiple possessive genitives in subject or object positions (Experiment 2), (c) doubly center-embedded, semantically biased sentences with 2 VPs or 3 VPs (Experiment 3), and (d) doubly center-embedded, semantically neutral sentences with 2 VPs or 3 VPs (Experiment 4).

semantic interpretation of sentences is an important goal for future work along with the scaling up of models to deal with the full complexity of language. Fortunately, there has already been encouraging connectionist work in this direction. For example, SRN models implementing multiple-cue integration have been scaled up to deal with both artificial (Christiansen & Dale, 2001) and full-blown (Reali, Christiansen, & Monaghan, 2003) child-directed speech, the latter even while predicting parts-of-speech information (see Christiansen et al., in press, for a discussion). Other recent work has successfully pushed the boundaries of SRN models of language still further toward increased coverage and semantic interpretation (e.g., Rohde, 2002; Sibley et al., 2008).

Given the characteristics of SRN recursive abilities, our usage-based approach to recursion predicts that substantial individual differences should exist in the processing of specific recursive constructions, not only across development but also across languages. Consistent with this prediction, observational data on preschoolers' language development show that recursive language production emerges gradually over time, structure for structure, rather than wholesale (Dickinson, 1987). In a training study, Roth (1984) further demonstrated that experience with recursive constructions is the key factor in preschoolers' comprehension of recursive structures (and not general cognitive development, including of working memory). In a similar vein, predictions from an SRN model regarding the importance of experience in processing singly embedded subject and object relative clauses (MacDonald & Christiansen, 2002) have been confirmed by a training study with adult participants (Wells et al., 2009). More generally, life experience, in the form of education, has been shown to affect the ability to comprehend complex recursive sentences (Dabrowska, 1997), likely due to differential exposure to language both in terms of the quantity and complexity of the linguistic material. Finally, differences in recursive abilities have also been found across languages. For example, Hoover (1992) showed that doubly center-embedded constructions are more easily processed in Spanish than in English (even when morphological cues were removed in Spanish). Similarly, substantial differences in perceived processing difficulty for the exact same types of recursive structure have been demonstrated across a variety of other languages—including English, German, Japanese, and Persian (Hawkins, 1994)—again highlighting the importance of linguistic experience in explaining recursive processing abilities. Indeed, Evans and Levinson (in press) argue that

> Recursion is not a necessary or defining feature of every language. It is a well-developed feature of some languages, like English or Japanese, rare

but allowed in others (like Bininj Gun-wok), capped at a single level of nesting in others (Kayardild), and in others, like Pirahã, it is completely absent. (p. 28)

In conclusion, we have hypothesized that recursion is not a built-in property of the mechanisms supporting language; rather, as originally suggested by Stolz (1967), the ability to use specific constructions recursively is acquired in an item-based manner. The close fit between our model and the human data provides initial support for this hypothesis (see also Engelmann & Vasishth, 2009). Assuming that future work is able to substantiate the usage-based approach to recursion further, then it would have important theoretical ramifications not only in underscoring the importance of experience in language processing (Wells et al., 2009) but also in suggesting that recursion is unlikely to be key factor in language evolution (Christiansen & Chater, 2008).

<div align="right">Revised version accepted 19 May 2009</div>

Notes

1 Formally, recursion can be defined as follows: If X is a (non-terminal) symbol and α and β are non-empty strings of symbols and terminals, then right-branching recursion occurs when $X \to \alpha\, X$ (i.e., there is a derivation from X to $\alpha\, X$), left-branching recursion when $X \to \beta\, X$, and center-embedding when $X \to \alpha\, X\, \beta$. Recursion can also arise as a consequence of a recursive *set* of rules, none of which need individually be recursive. Thus, rules of the form $X \to \alpha\, Y$ and $Y \to \beta\, X$ together form a recursive rule set because they permit the derivation of a structure, such as $[_X \alpha\ [_Y \beta\ [_X \alpha\ [_Y \beta\ [_X \ldots]]]]]$, involving several levels of right-branching recursion. Similarly, recursive rule sets can be used to derive left-branching and center-embedded constructions.

2 Of the 42 input/output units only 38 were used to represent the vocabulary and other lexical material in a localist format. The remaining four units were saved for other purposes, including for the use as novel, untrained words (as discussed below).

3 All simulations were replicated multiple times (including with variations in network architecture and corpus composition) with qualitatively similar results.

4 Although the sentences with multiple PPs have more attachment ambiguities, our materials always resolve any ambiguity with the most frequent interpretation in English, such that each PP attaches to the NP immediately dominating it. There is no good evidence in the literature that ambiguities resolved with their most favored interpretation are any harder to process than unambiguous sentences, and so the presence of more attachment ambiguity in the 3PP condition should not have affected our results.

5　Obtaining the offline grammaticality ratings after continuous online judgments may have negatively affected the ratings of the sentences that were not read to completion. Future work can address this issue by omitting the online judgments and/or collecting reading times instead.

6　When trained on a corpus that contained only doubly center-embedded constructions, there was only little improvement in performance, similar to the limited effects of training in humans (Blaubergs & Braine, 1974; Powell & Peters, 1973; Stolz, 1967). Christiansen and Chater (1999) further established that the SRN's intrinsic constraints on center-embedding are independent of the size of the hidden unit layer.

References

Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes, 1*, 249–262.

Blaubergs, M. S., & Braine, M. D. S. (1974). Short-term memory limitations on decoding self-embedded sentences. *Journal of Experimental Psychology, 102*, 745–748.

Boeckx, C. (2006). *Linguistic minimalism: Origins, concepts, methods, and aims*. New York: Oxford University Press.

Boland, J. E. (1997). The relationship between syntactic and semantic processes in sentence comprehension. *Language and Cognitive Processes, 12*, 423–484.

Boland, J. E., Tanenhaus, M. K., & Garnsey, S. M. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language, 29*, 413–432.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from *wh*-questions. *Journal of Memory and Language, 34*, 774–806.

Botvinick, M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review, 111*, 395–429.

Botvinick, M., & Plaut, D.C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review, 113*, 201–233.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory, 2*, 113–124.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1995). *The minimalist program.* Cambridge, MA: MIT Press.

Christiansen, M. H. (1994). *Infinite languages, finite minds: Connectionism, learning and linguistic structure*. Unpublished doctoral dissertation, University of Edinburgh.

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. (2003). Constituency and recursion in language. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (2nd ed., pp. 267–271). Cambridge, MA: MIT Press.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31*, 489–558.

Christiansen, M. H., & Dale, R. (2001). Integrating distributional, prosodic and phonological information in a connectionist model of language acquisition. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd annual conference of the Cognitive Science Society* (pp. 220–225). Mahwah, NJ: Lawrence Erlbaum.

Christiansen, M. H., Dale, R., & Reali, F. (in press). Connectionist explorations of multiple-cue integration in syntax acquisition. In S.P. Johnson (Ed.), *Neoconstructivism: The new science of cognitive development*. New York: Oxford University Press.

Church, K. (1982). *On memory limitations in natural language processing*. Bloomington: Indiana University Linguistics Club.

Dabrowska, E. (1997). The LAD goes to school: A cautionary tale for nativists. *Linguistics, 35*, 735–766.

Dickinson, S. (1987). Recursion in development: Support for a biological model of language. *Language and Speech, 30*, 239–249.

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.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*, 71–99.

Engelmann, F., & Vasishth, S. (2009). *Processing grammatical and ungrammatical center embeddings in English and German: A computational model*. Unpublished manuscript.

Evans, N., & Levinson, S. (in press). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*, 491–505.

Foss, D. J., & Cairns, H. S. (1970). Some effects of memory limitations upon sentence comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 9*, 541–547.

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

Gibson, E., & Thomas, J. (1996). The processing complexity of English center-embedded and self-embedded structures. In C. Schütze (Ed.), *Proceedings of the NELS 26 sentence processing workshop* (pp. 45–71). Cambridge, MA: MIT Press.

Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes, 14*, 225–248.

Hadley, R. F. (1994). Systematicity in connectionist language learning. *Mind and Language, 9*, 247–272.

Hagoort, P. (in press). Reflections on the neurobiology of syntax. In D. Bickerton & E. Szathmáry (Eds.), *Biological foundations and origin of syntax. Strüngmann Forum Reports* (Vol. 3). Cambridge, MA: MIT Press.

Hagstrom, P., & Rhee, R. (1997). The Dependency Locality Theory in Korean. *Journal of Psycholinguistic Research, 26*, 189–206.

Hakes, D. T., Evans, J. S., & Brannon, L. L (1976). Understanding sentences with relative clauses. *Memory and Cognition, 4*, 283–290.

Hakes, D. T., & Foss, D. J. (1970). Decision processes during sentence comprehension: Effects of surface structure reconsidered. *Perception and Psychophysics, 8*, 413–416.

Hamilton, H. W., & Deese, J. (1971). Comprehensibility and subject-verb relations in complex sentences. *Journal of Verbal Learning and Verbal Behavior, 10*, 163–170.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hoover, M. L. (1992). Sentence processing strategies in Spanish and English. *Journal of Psycholinguistic Research, 21*, 275–299.

Joshi, A. K. (1990). Processing crossed and nested dependencies: An automaton perspective on the psycholinguistic results. *Language and Cognitive Processes, 5*, 1–27.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122–149.

Karlsson, F. (2007). Constraints on multiple center embedding of clauses. *Journal of Linguistics, 43*, 365–392.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition, 2*, 15–47.

Larkin, W., & Burns, D. (1977). Sentence comprehension and memory for embedded structure. *Memory and Cognition, 5*, 17–22.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126–1177.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*, 375–419.

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.

Marcus, M. (1980). *A theory of syntactic recognition for natural language*. Cambridge, MA: MIT Press.

Marks, L. E. (1968). Scaling of grammaticalness of self-embedded English sentences. *Journal of Verbal Learning and Verbal Behavior, 7*, 965–967.

Miller, G. A. (1962). Some psychological studies of grammar. *American Psychologist, 17*, 748–762.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 419–492). New York: Wiley.

Miller, G. A., & Isard, S. (1964). Free recall of self-embedded English sentences. *Information and Control, 7*, 292–303.

Niv, Y., & Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends in Cognitive Sciences, 12*, 265–272.

Peterfalvi, J. M., & Locatelli, F (1971). L'acceptabilité des phrases [The acceptability of sentences]. *L' Année Psychologique, 71*, 417–427.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*, 105–110.

Plaut, D. C. (1999). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science, 23*, 543–568.

Powell, A., & Peters, R. G. (1973) Semantic clues in comprehension of novel sentences. *Psychological Reports, 32*, 1307–1310.

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 R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th annual conference of the Cognitive Science Society* (pp. 970–975). Mahwah, NJ: Lawrence Erlbaum.

Reich, P. (1969). The finiteness of natural language. *Language, 45*, 831–843.

Rohde, D. L. T. (2002). *A connectionist model of sentence comprehension and production*. Unpublished doctoral dissertation, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Roth, F. P. (1984). Accelerating language learning in young children. *Child Language, 11*, 89–107.

Schlesinger, I. M. (1975). Why a sentence in which a sentence in which a sentence is embedded is embedded is difficult. *Linguistics, 153*, 53–66.

Servan-Schreiber, D., Cleeremans, A., & McClelland, J.L. (1991). Graded state machines: The representation of temporal dependencies in simple recurrent networks. *Machine Learning, 7*, 161–193.

Shieber, S. (1985). Evidence against the context-freeness of natural language. *Linguistics and Philosophy, 8*, 333—343.

Sibley, D. E., Kello, C. T., Plaut, D. C., & Elman, J. L. (2008). Large-scale modeling of wordform learning and representation. *Cognitive Science, 32*, 741–754.

Stabler, E. P. (1994). The finite connectivity of linguistic structure. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 303–336). Hillsdale, NJ: Lawrence Erlbaum.

Stolz, W. S. (1967). A study of the ability to decode grammatically novel sentences. *Journal of Verbal Learning and Verbal Behavior, 6*, 867–873.

Uehara, K., & Bradley, D. (1996). The effect of *-ga* sequences on processing Japanese multiply center-embedded sentences. In *11th Pacific-Asia conference on language, information, and computation* (pp. 187–196). Seoul: Kyung Hee University.

von Humboldt, W. (1999). *On language: On the diversity of human language construction and its influence on the metal development of the human species*. Cambridge: Cambridge University Press. (Original work published 1936)

Wang, M. D. (1970). The role of syntactic complexity as a determiner of comprehensibility. *Journal of Verbal Learning and Verbal Behavior, 9*, 398–404.

Wells, J., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology, 58*, 250–271.

Appendix A
Grammatical Prediction Error

To map SRN predictions onto human performance in our experiments, we use the GPE measure. This SRN performance score has been shown to provide a good model of human behavioral measures, including reading times and grammaticality ratings (Christiansen & Chater, 1999; MacDonald & Christiansen, 2002; Wells et al., 2009). The GPE provides an indication of how well a network is obeying the training grammar when making its predictions. The GPE for predicting a particular word is calculated as:

$$\text{GPE} = 1 - \frac{H}{H + F + M}$$

Hits (*H*) and false positives (*F*) consist of the accumulated activations of the set of units, *G*, that are grammatical and the set of incorrectly activated

ungrammatical units, U, respectively:

$$H = \sum_{i \in G} u_i \quad F = \sum_{i \in U} u_i$$

Missing activation (M; i.e., false negatives) is calculated as a sum over the (positive) discrepancy, m_i, between the target activation for a grammatical unit, t_i, and the actual activation of that unit, u_i:

$$M = \sum_{i \in G} m_i \qquad m_i = \begin{cases} 0 & \text{if } t_i - u_i \leq 0 \\ t_i - u_i & \text{otherwise} \end{cases} \qquad t_i = \frac{(H + F)f_i}{\sum_{j \in G} f_j}$$

The desired activation, t_i, is computed as a proportion of the total activation determined by the lexical frequency, f_i, of the word that u_i designate and weighted by the sum of the lexical frequencies, f_j, of all the grammatical units. Although not an explicit part of the above equations, correct rejections are also taken into account under the assumption that they correspond to zero activation for units that are ungrammatical given previous context.

Appendix B
Stimulus Materials

Experiment 1: Multiple Right-Branching Prepositional Phrases
The student from the city says that the girl from the forest likes pets.
The student says that the girl from the city near the forest likes pets.
The girl from the school in the city near the forest likes pets.

The nurse with the vase says that the flowers by the window resemble roses.
The nurse says that the flowers in the vase by the window resemble roses.
The flowers in the vase on the table by the window resemble roses.

The boss from the office says that the posters across the hall tell lies.
The boss says that the posters in the office across the hall tell lies.
The posters on the desk in the office across the hall tell lies.

The woman at the mall says that the bank behind the airport helps people.
The woman says that the bank at the mall behind the airport helps people.
The bank on the street at the mall behind the airport helps people.
The expert from the town says that the buildings with the flooding need repair.

The expert says that the buildings from the area with the flooding need repair.
The buildings in the town from the area with the flooding need repair.

The man with the drawing says that the story about the river makes sense.
The man says that the story about the castle by the river makes sense.
The story about the drawing of the castle by the river makes sense.

The ranger on the hill says that the deer outside the village avoids hunters.
The ranger says that the deer on the hill outside the village avoids hunters.
The deer on the hill by the lake outside the village avoids hunters.

The chemist on the stairway says that the paint on the balcony contains lead.
The chemist says that the paint in the bedroom with the balcony contains lead.
The paint on the stairway to the bedroom with the balcony contains lead.

The musician by the keyboard says that the guitar from the concert requires
    tuning.
The musician says that the guitar with the broken amp from the concert requires
    tuning.
The guitar by the keyboard with the broken amp from the concert requires
    tuning.

Experiment 2: Multiple Left-Branching Possessive Genitives
Mary's brother's teacher's dog chased the mailman.
The mailman chased Mary's brother's teacher's dog.

Bill's friend's uncle's cat attacked the hamster.
The hamster attacked Bill's friend's uncle's cat.

Jane's dad's colleague's parrot followed the baby.
The baby followed Jane's dad's colleague's parrot.

Ben's band's singer's girlfriend cheated the promoter.
The promoter cheated Ben's band's singer's girlfriend.

Donald's heir's chauffeur's wife avoided the gardener.
The gardener avoided Donald's heir's chauffeur's wife.

John's mom's neighbor's boy visited the doctor.
The doctor visited John's mom's neighbor's boy.

Ann's fiancé's cousin's husband hired the builder.
The builder hired Ann's fiancé's cousin's husband.

Jim's buddy's sister's trainer liked the team.
The team liked Jim's buddy's sister's trainer.

Robert's aunt's company's mechanic offended the trainee.
The trainee offended Robert's aunt's company's mechanic.

Chang's daughter's attorney's partner approached the defendant.
The defendant approached Chang's daughter's attorney's partner.

Experiment 3: Multiple Semantically Biased Center-Embeddings
The lullaby that the famous country singer who the record label had signed to
    a big contract was singing yesterday was written seventy years ago.
The lullaby that the famous country singer who the record label had signed to
    a big contract was written seventy years ago.

The game that the child who the lawnmower had startled in the yard was playing
    in the morning lasted for hours.
The game that the child who the lawnmower had startled in the yard lasted for
    hours.

The crime that the gangster who the story had profiled had planned for weeks
    was quickly solved.
The crime that the gangster who the story had profiled was quickly solved.

The apartment that the maid who the service had sent over was cleaning every
    week was well decorated.
The apartment that the maid who the service had sent over was well decorated.

The shirt that the seamstress who the immigration officer had investigated last
    week was carefully mending needed to be dry cleaned.
The shirt that the seamstress who the immigration officer had investigated last
    week needed to be dry cleaned.

The monologue that the actor who the movie industry had snubbed repeatedly
    was performing last month was extremely well written.
The monologue that the actor who the movie industry had snubbed repeatedly
    was extremely well written.

Experiment 4: Multiple Semantically Neutral Center-Embeddings
The chef who the waiter who the busboy offended appreciated admired the
    musicians.
The chef who the waiter who the busboy offended frequently admired the
    musicians.

The spider that the bullfrog that the turtle followed chased ate the fly.
The spider that the bullfrog that the turtle chased mercilessly ate the fly.

The governor who the secretary who the senator called answered introduced the ambassador.
The governor who the secretary who the senator called often introduced the ambassador.

The chipmunks that the man that the rats feared avoided liked the food.
The chipmunks that the man that the rats feared deeply liked the food.

The baker who the butcher who the shoemaker paid congratulated despised the mayor.
The baker who the butcher who the shoemaker congratulated today despised the mayor.

The dog that the cat that the bird outwitted fought approached the girl.
The dog that the cat that the bird outwitted repeatedly approached the girl.

The detective who the scientist who the lawyer confused questioned told the truth.
The detective who the scientist who the lawyer confused profoundly told the truth.

The clerk that the typist that the receptionist assisted teased thanked the boss.
The clerk that the typist that the receptionist assisted regularly thanked the boss.

The woman who the boy who the policeman found saw hid the evidence.
The woman who the boy who the policeman found yesterday hid the evidence.

The person that the guest that the doorman watched disliked forgot the key.
The person that the guest that the doorman watched briefly forgot the key.