Generating Structure from Experience: The Role of Memory in Language

Brendan T. Johns[1] & Michael N. Jones[2]
bj23@queensu.ca, jonesmn@indiana.edu
[1]Department of Psychology, Queen's University, Kingston, On.
[2]Department of Psychological and Brain Sciences, Indiana University

Abstract

Theories of language have generally assumed that abstraction of the linguistic input is necessary in order to create higher-level representations of the workings of a language (i.e. a grammar). However, the importance of individual experiences with language has recently been emphasized by many, including usage-based theories (Tomasello, 2003). Based upon this, a formal exemplar model of language is described, which stores instances of sentences across a natural language corpus, using recent advances from models of semantic memory. This memory store is used to generate expectations about the future structure of sentences. The model can successfully capture a variety of different behavioral results. This work provides evidence that much of language processing may be bottom-up in nature, based upon the storage and retrieval of individual experiences with language.

Keywords: Sentence processing; instance memory; semantics

Introduction

Typically, theories of language are abstractionist in nature, where individual experiences are used to create higher-level representations of the workings of a language. This includes the generative viewpoint (Chomsky, 1988), the constraint-based approach (McRae, Spivey-Knowlton, & Tenenhaus, 1998), and the Bayesian perspective (Jurafsky, 1996), among others. All of these approaches differ in terms of the type of information that is required, but all rely upon the abstraction of linguistic input to syntactic categories as the basis of language processing.

An alternative approach to abstractionist theories has been developed in many areas of cognitive psychology, based around the storage of individual experiences, or instances. In terms of language processing, this approach has been championed by the usage-based perspective (Tomasello, 2003; Abbot-Smith & Tomasello, 2006), based on much evidence that language development is largely item-based, and not dependent on acquired syntactic categories One example of this is given by Lieven, Pine, & Baldwin (1997) who found that the majority of a child's utterances are based upon a few experienced lexical patterns.

This research has been backed up by studies with adults that have shown that an increased amount of experience with certain grammatical constructs allows for a greater ease of processing. For instance, Reali & Christiansen (2007) used a corpus analysis to determine the frequency of occurrence of different types of relative clauses. It is a common finding that subject-relative clauses are easier to process than object-relative clauses (e.g. Traxler, Morris, & Seely, 2002), and it was found by Reali & Christiansen (2007) that subject-relatives are the more common construct when personal pronouns are utilized. However, they also found that object-relatives using impersonal pronouns are more frequent than subject-relatives, and a self-paced reading experiment demonstrated that this lead to a processing advantage of object-relative clauses over subject-relative clauses when impersonal pronouns are used. These findings suggest that the amount of experience one has with certain grammatical constructs effects the processing of them.

An instance model, based around Minerva 2 (Hintzman, 1986), has recently been used to understand artificial grammar learning (Jamieson & Mewhort, 2009a, b), thus showing that these models can be readily applied to language-like tasks. In terms of artificial grammar learning, the model proposes that when one sees a probe string (e.g. 'a b c'), whether that string is classified as grammatical or not is based upon the similarity of that string to the other studied strings of that grammar. From this perspective, a grammaticality judgment is not based upon coherence to an implicitly learned grammar, which is what was initially assumed in this task, but instead it is a memory task based on the similarity of a probe string to the exemplars of that language stored in memory. This simple memory model was able to account for classic, and new, experimental results examining implicit memory (Jamieson & Mewhort, 2009a).

The current paper will describe a new computational model, based on an instance theory of memory, which can model many diverse natural language sentence processing findings. It will be based on integrating multiple memory and language models, including BEAGLE (Jones & Mewhort, 2007; Sahlgren, et al., 2007), the semantics retrieval model (Kwantes, 2005), and a classic instance model of memory (Hintzman, 1986). The model will use the storage and retrieval of linguistic experiences as the fundamental operations of language processing. The theoretical foundation of the model will be based in the usage-based view of language (Tomasello, 2003), which rejects the notion that language is solely composed of rules over abstract syntactic categories, but rather it is composed of communicative constructions which emerge through experience with language and the use of it, perhaps as stored in an exemplar memory store (Abbot-Smith & Tomasello, 2006). The proposed model is not meant as a refutation of the importance of higher-level information in language, but instead it will serve as a demonstration of the power that the raw structure of language has.

An Exemplar Model of Sentence Processing

This section will describe the various components of an exemplar model of natural language, as well as the justification for these choices. This will include an examination of both the representation and processing assumptions of the model.

Representation Assumptions

As in Jamieson & Mewhort (2009a, b), the model described here will be loosely based upon the Minerva 2 (Hintzman, 1986) memory model. However, since this model is dealing with real language, some more sophisticated storage assumptions will be required. These storage assumptions will be based upon the fundamentals of distributed memory theory and recent advances in the modeling of semantic memory.

Since we are proposing an exemplar model of language, it is first necessary to establish what an exemplar of language actually is. Most theories of language propose that the word is fundamental, and that many aspects of grammar are encoded in a word's entry in the mental lexicon. A different proposal has been offered by Elman (2009), who suggests, due to the multitude of recent data demonstrating the importance of event knowledge on language processing (e.g. Ferretti, Kutas, & McRae, 2007; Bicknell, et al., 2010), that very little information is actually contained in the lexicon. Instead, Elman (2009) proposes that much of language may be based upon event schemas, which are constructed through the abstraction of event patterns.

Much of what is proposed here is in the same spirit of the proposals of Elman (2009), but within a different formal framework. The current proposal also see event knowledge as central to language, but propose that it is not developed through abstraction at learning, but instead by abstraction at retrieval. That is, given certain linguistic information (e.g. "The farmer grew…"), the system can use this information as a cue to retrieve information about what is likely to occur in this event, similar to a cued recall task in a typical memory experiment. This is a very flexible system as it allows for the combination of multiple memory traces, where the combination is dependent on the current context, allowing for a dynamic language comprehension system. This is in contrast to an abstractionist approach, which would require an individual representation for each event to be stored. However, in order to accomplish this, the model requires complete descriptions of an event, which in this case would be a sentence. Thus, the instance of language that will be used in this model will be a representation of a sentence, taken from a natural language corpus.

In order to construct an instance of a sentence, it is first necessary to determine how word order can be encoded, as this is obviously an essential component of language, and is the main information source that the current model will be dependent upon. The model described here will utilize a new technique to encode order, based off of the distributed memory proposals of Kanerva (1988), and semantic memory modeling (Jones & Mewhort, 2007). Under this proposal, each word is represented with a binary spatter code (called a word's environmental vector), which is a large, sparse vector, and random permutations of these vectors allow for order to be encoded (Sahlgren, et al., 2007). However, instead of binary vectors, more recent models utilize sparse ternary vectors, where non-zero values are either +1 or -1 with equal probability.

To encode order, this approach uses random permutations (RP), which simply takes an environmental vector as input and creates an output vector with a random shuffling of the input values. Word order is encoded by assigning each location within a sentence a unique random permutation. The different permutated environmental vectors are then summed into a composite, giving the representation of a sentence.

In models of semantics (e.g. Sahlgren, et al., 2010) a unique sentential representation is created for each word in the sentence. The representation created for a word is dependent on that word's location within a vector, in order to provide information about that word's role in a sentence. However, for the purposes of the model here that is unnecessary as all that is needed is an encoding of the linear ordering of the total sentence. Based on this, the encoding of a sentence is given with the following equation:

$$Sent = \sum_{i=1}^{n} RP^{i}(word_{i}) \tag{1}$$

Where RPx represents a specific RP for location x in the sentence, and n denotes the total number words in a sentence. The resulting vector serves as an exemplar for a specific sentence and this vector is then stored in memory. Each sentence across a corpus will be stored. Storage of sentence exemplars constructed in this fashion will serve as the basis of this model. Next, how this memory store can be used in sentence processing will be described.

Processing Model

The operation of this model will be based around the concept of expectation generation, or predicting what the upcoming structure of a sentence (or utterance) should be, given the current input. The generation of expectations (or of surprisal to unexpected input), has been a central component of many theories of language processing. There is also a considerable amount of empirical evidence that expectation generation and prediction is a central component of language processing (see Altmann & Mirkovic, 2009 for a review).

The memory model approach to the task of prediction is significantly different from past techniques. It proposes to use the current input (i.e. the cues) to retrieve the structure of the expected future context, based upon one's past experiences with language. This is similar to Simon's (1969) analogy of an ant walking along the beach - much of the complexity in language may not be due to complex internalized representations, but instead it may be due to the structure of the language environment that people are exposed to.

This leads to the question, given the exemplar memory storage previously described, how can past experiences be used to generate expectations about the future states of a sentence? This can be accomplished with the cued retrieval technique described in the Minerva 2 model (Hintzman, 1986), and is similar in nature to work in artificial grammar learning (Jamieson & Mewhort, 2009 a, b) and in retrieval of semantic information about a word (Kwantes, 2005). In

Minerva 2, when a cue is presented, the model activates each trace in memory in parallel. The level of activation of a particular trace is proportional to the similarity between the cue and the trace. The activated traces are then summed into a composite vector (typically referred to as an 'echo'). This represents the aggregated information that is retrieved from memory, in response to the cue. This echo will retrieve information that is attached to the cue, for instance a paired associate in a cued recall task. The Minerva 2 retrieval operation will form the basis of the expectation generation mechanism of the current model, by retrieving the structure that is expected to surround a word in a certain location within a sentence. Using this method, processing of a sentence will take place at two levels: 1) generating expectations to each word in a sentence, and 2) comparison and integration with previously formed expectations.

Expectation generation using words as cues

As discussed above, Minerva 2 retrieves information from memory in response to a cue. Here, the cue will be a word in a sentence, permuted by the position of the word. This will activate the traces in memory where that word occurred in that position (and also traces activated by chance). These traces will then be summed into a composite, which will be referred to as an expectation vector (and represent it with E). This process will generate the words that are expected to surround a given word in a certain position in a sentence. Each word will be summed into the composite based on how similar the probe is to the memory trace. The similarity metric that will be used is a simple vector cosine.

Unlike typical models based on the Minerva framework, here only memory traces that have a positive similarity value will be used, in order to retrieve exemplars that had a similar structure to the inputs. This was done since we want to generate expectations about words that should likely occur, not about words that should not occur. Using only positive values also significantly reduces the amount of computation required, as well as the amount of noise in the resulting vector. The expectation vector is formed in the following manner:

$$E(W) = \sum_{i=1}^{n} (Sim(W, M_i) > 0) \quad M_i * Sim(W, M_i)^{\lambda} \qquad (2)$$

Where n represents the number of traces in memory, W is the word currently being processed, M_x is a trace from the exemplar memory store, and λ is a scaling parameter. The scaling parameter is designed to accentuate the effect of high similarity exemplars over low similarity ones, and by increasing this parameter this difference is enhanced. This also is based on the number of exemplars contained in memory, as the greater the number of exemplars that are contained, the less any single exemplar should contribute. In order to account for sentence processing effects (where expectations are formed in response to multiple words), the expectations across words in a sentence have to be integrated, which will be described next.

Comparison and integration with previously formed expectations

The process described above will retrieve the expectations for what words should surround a word in a sentence. In order to represent the meaning of a sentence these retrieved expectations are summed into a single vector, which will be referred to as the comprehension vector (C). By iteratively constructing expectation vectors, and summing these into a single composite, the meaning of a sentence is 'honed-in' on across the sequence of words. Meaning in this sense refers to a point in multi-dimensional space, similar to the proposals of semantic space models. The comprehension vector will be constructed with the following equation:

$$C_j = C_{j-1} + E\left(RP^j(W_j)\right), j = 1, \dots, \# \, words \qquad (3)$$

Where j is the current position in the sentence, and W_j is the word in that position, and E returns the expectation vector for the word that is currently being processed (the cue is the current word permuted by its location within the sentence). This equation sums the current expectation vector into the comprehension vector, in order to update the expectations about the upcoming words. However, before the expectation vector is summed into the comprehension vector, the expectation vector is normalized so that all values sum to unit length (by dividing each location by the total vector magnitude). This simply ensures that each word adds in the same amount of information into the comprehension vector.

This comprehension vector allows for an expectation value (EV) to be calculated for each word in the sentence, since if a word is expected then its expectation vector should be similar to the comprehension vector. An EV signals how expected the current word was, based on the past words that have been processed. An increase in similarity is assumed to cause an increase in processing efficiency (and hence a decrease in processing time), since the traces in memory that require activation will already be active, due to past processing. The EV for a specific word is calculated by taking the cosine (described in equation 2) between the comprehension vector (C_{j-1}) and the retrieved expectation vector. The EV represents how expected the current word is by determining how much information about that word was previously retrieved. The EV will be the main source of information used to simulate sentence processing results.

Natural Language Simulations

The previous section described a new model of expectation generation in sentence processing, based upon exemplar storage and retrieval. In order to develop increasingly better models of cognition, it is necessary to start training, and testing, computational cognitive models on natural language tasks. This is significantly more challenging, as opposed to artificial languages, as it requires externalized knowledge. That is, it requires information from outside of the specific experimental context.

The corpus that the model will be trained on is the TASA corpus, and 300,000 sentence exemplars were constructed from this corpus. This is less than half of all sentences that are contained in the corpus. No sentences greater than 20 words were included in the analysis. This is a large amount of linguistic data, but it is not too large that it would make the current model implausible. In order to simulate specific empirical results, sentences were taken from the relevant studies. Simulations will be run over a number of different resamples of the environmental vectors for the different conditions in an experiment. This allows for a significance test to be utilized, similar to a typical experiment. The λ parameter will be set at 13, due to the large number of exemplars being used.

Relative Clause Processing

Relative clauses are embedded structures which modify a head noun phrase. It has been found (e.g. Traxler, Morris, & Seely, 2002) that object relative sentences ("The reporter that the senator attacked…") are more difficult to process than subject relative sentences ("The reporter that attacked the senator…"). In a recent study, Reali & Christiansen (2007) conducted a corpus analysis where the relative frequency of different types of relative clauses was measured, in order to determine how the frequency of occurrence effects the processing of this construct. It was found that subject relative clauses are more frequent when personal pronouns are used, but object relative clauses are more frequent when impersonal pronouns are used in the embedded noun phrase, and this pattern was mirrored in the behavioral results. The goal of this simulation is to determine if the model can attain the typical advantage for subject relative clauses, similar to what Traxler, et al. (2002) found, and also the advantage for object relative clauses when impersonal pronouns are used, similar to what Reali & Christiansen (2007) found.

In order to demonstrate the typical subject relative clause processing advantage, 30 clauses from Traxler, et al. (of each type) were attained from the study. Example sentences from this study include:

a) The lawyer that irritated the banker… (SR)
b) The lawyer that the banker irritated… (OR)

The average expectation value for the relative clause region ("irritated the banker" vs. "the banker irritated") was then calculated. In order to simulate the results of Reali & Christiansen, two list sets which demonstrated a processing advantage for object relative clauses were attained. The first list set contained 14 sentences where the noun phrase consisted of second-person pronouns:

a) The consultant that called you… (SR)
b) The consultant that you called… (OR)

The second set also consisted of 14 sentences, but the noun phrase consisted first-person pronouns:

a) The lady that visited me… (SR)
b) The lady that I visited… (OR)

Both of these lists elicited a processing advantage for the two words following the relativizer 'that' (you called/I visited for object relative vs. called you/visited me for subject relative clauses). The average expectancy values were calculated for these same regions. By testing the model across both of these list sets, it simply allows for us to test the model across different types of language, in order to ensure that the difference found is a true one. For all three lists, 20 resamples of the environment vectors were done.

The different expectation values across the three difference sentence types are plotted in Figure 1. This figure demonstrates that for the lists from Traxler, et al. the subject relative clause had higher expectation values, which was highly significant [$F(1,39)=133.74$, $p<0.001$], similar to the behavioral results. However, this pattern reversed itself for both the second-person and first person pronoun sentences, where the object-relative sentences had higher expectation values. Both of these differences were significant, with an $F(1,39)=117.61$, $p<0.001$ for the second-order pronoun sentences and an $F(1,39)=9.485$, $p=0.004$ for the first-order pronoun sentences. This simulation clarifies how this model is generating expectancies: more common structures in language lead to a greater certainty as to what to expect in the upcoming language stream.

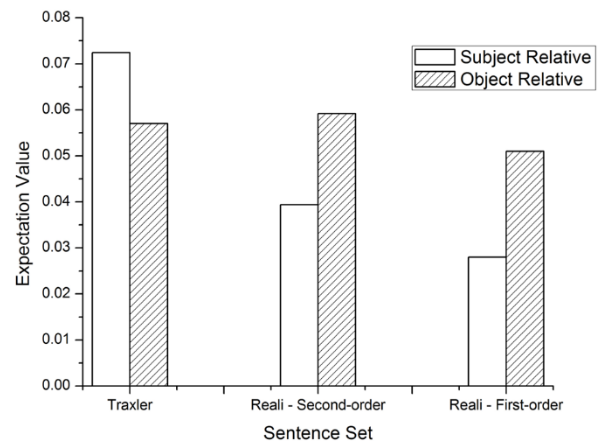

Figure 1. Simulation of the results on relative clause processing from Traxler, et al. (2002) and Reali & Christiansen (2007).

Effects of Contextual Constraint

A number of eye tracking studies of reading have examined the role of contextual constraint on eye movements (e.g. Rayner & Well, 1996). This is typically done by having a target word in a sentence be either congruent or incongruent with the meaning of the sentence. As Rayner & Well (1996) point out, a consistent pattern of results have emerged from these studies: 1) highly constrained words are more often skipped, 2) more regressions are made to unconstrained words, and 3) fixation times are lower for constrained target words. These findings suggest that the sentential context is being used to generate expectancies about what words should occur in the upcoming structure of a sentence.

In order to test whether this model can account for this result, the sentences from Rayner & Well (1996) were tested. These sentences split into three groups of contextual constraint, based on a previously done norming study. Examples of these sentences are:

-High: He mailed the letter without a *stamp*
-Medium: The girl crept slowly towards the *door*

-Low: Jill looked back through the open *curtain*

The set of 72 sentences from Rayner & Well (1996) were used in the following simulation. To fit the data, the expectation value for each target word to the comprehension vector was assessed across the three different conditions. 15 resamples were done for each sentence set.

Figure 2 contains the result of this simulation, which shows a similar pattern to the empirical results: highly constrained words have a higher expectation value (which would manifest themselves in terms of lower reading times, greater probability of skipping, etc…) than medium and low constrained words. This was a significant effect, $F(2, 44) = 69.03$, $p<0.001$, and a planned comparison confirmed that each condition was greater than each condition below it. This is a simple test of this model, but it provides an important basis for its operation: across the words in a sentence, expectations are being generated about the words that are likely to occur in that sentence.

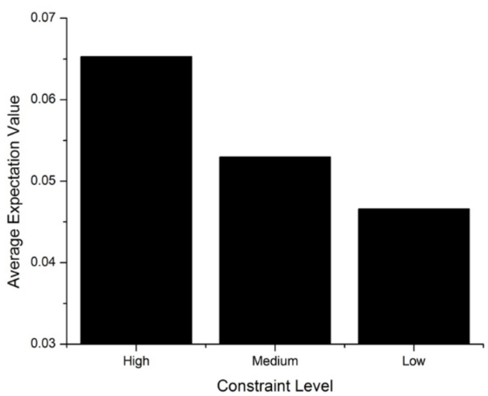

Figure 2. Simulation of Rayner & Well (1996).

Verb Sense and Expectation Generation

Hare, Elman, Tabaczynski, & McRae (2009) conducted a similar study to McRae, et al. (1998) where the sense of a verb was manipulated, in order to determine how this influences expectations about upcoming words. Specifically, they manipulated the transitivity of a verb, where a verb is transitive if it has a direct object and intransitive if it does not. Many verbs can be either, depending on the context. This is in turn related to causation, where causative verbs occur in the transitive (e.g. "he broke the vase") while inchoatives occur in the intransitive (e.g "the vase broke"). If people are sensitive to this type of information, it should lead to expectations about whether a direct object should occur or not. Hare, et al. tested this by manipulating the thematic fit of a subject to be either good theme (e.g. "the glass shattered…") or good cause (e.g. "the brick shattered…") inducing, and measured reading times to the postverb regions of intransitive (e.g. "…into tiny pieces") or transitive ("…the fragile goblet") sentences. It was found that when the sentence were intransitive, the reading times in the postverb region was significantly less for the good theme sentences, and the opposite was true of transitive sentences.

In order to test the model on this result, 15 sentences were attained for each of the four different conditions from Hare, et al. (2009). Expectation values were then calculated at the verb (where no significant difference is expected) and at the first non-function word in the postverb region (e.g. 'tiny' or 'fragile'), where a significant difference is expected. 20 resamples were done for each sentence type.

The results of this simulation (and the data from Hare, et al.) are displayed in Figure 3. This figure demonstrates that the model can approximate the results of this experiment quite well. No significant difference was found at the verb region for either intransitive-biased sentences [$F(1,39)=0.16$, $p>0.1$], or transitive-biased sentences [$F(1,39)=0.775$, $p>0.1$]. However, for intransitive-biased sentences the expectation values for the noun in the good-theme sentences were significantly greater than good-cause sentences [$F(1,39)=29.59$, $p<0.001$]. The opposite was true in transitive-biased sentences, with expectation values for the noun in good-cause sentences being significantly greater than those in good-theme sentences [$F(1,39)=36.51$, $p<0.001$]. This is a very important result for this theory as it demonstrates that it is not generating expectancies simply based on single words, but can generate them also in response to context. Specifically, by summing across expectancy vectors, and combining episodic traces, different expectations about the upcoming structure of a sentence are generated.

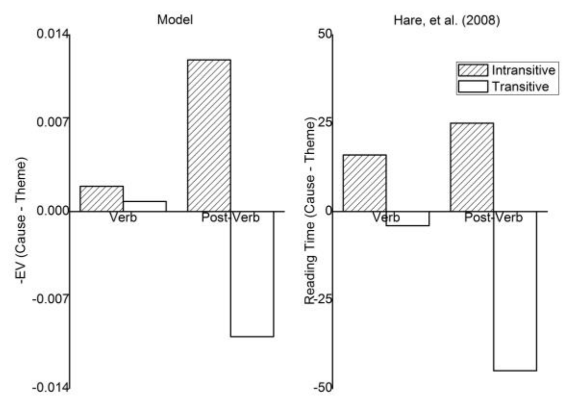

Figure 3. Simulation of the results from Hare, et al. (2009).

Event Knowledge Activation

A recent line of promising research has been to examine how the knowledge of events comes into play during sentence comprehension (Ferretti, et al., 2007). This typically involves manipulating congruent/incongruent or low/high typicality event knowledge that is associated with a particular verb or noun. In particular, we will attempt to simulate two results: 1) Ferretti, et al. (2007) found greater N400 amplitudes to low typicality events vs. high typicality events, suggesting a greater surprisal value to unexpected events, and 2) a recent result by Bicknell, et al. (2010) demonstrating that the reading time (and N400 amplitude) of a certain patient noun (brakes or spelling) depended on the combination of agent and verb (mechanic checked vs. journalist checked).

In order to simulate these results, sentences were taken from the relevant studies (37 sentences for the Ferretti, et al. study and 40 for the Bicknell, et al. study). For the Ferretti, et al. study, expectation values were calculated to the last word of each sentence, where the word was either a high or low typicality word. Two example sentences are (high/low typicality):

a) The girl was skating in the (rink/ring)

b) The king was sitting on the (throne/stage)

For the Bicknell, et al. study expectation values were assessed at the patient noun, for both congruent and incongruent words. Two example sentences of these stimuli are (congruent/incongruent):

a) The (librarian/composer) arranged the shelf

b) The (secretary/speaker) addressed the letter

In order to determine if the model found a difference between the two sets of words, 20 resamples were done for each sentence set.

A significant difference was found for the sentences from the Ferretti, et al. study, with an $F(1, 39) = 73.21$, $p<0.001$. This demonstrates that the model was successfully generating expectations about the event across the structure of the sentence. A significant difference was also found for the sentences from the Bicknell, et al., although this effect was not as large with an $F(1, 39) = 6.97$, $p=0.01$. The smaller difference is not surprising, as it is a more complicated due to requiring both an agent (e.g. secretary) and verb (addressed) to generate the correct event knowledge. However, even this small difference is impressive given the nature of the task. The simulation of the results of Hare, et al. demonstrates that the model is not only able to generate expectations about forthcoming words, but it can generate them about specific types of information, namely event knowledge.

Discussion

Here a new model of expectation generation in sentence processing was tested on natural language sentence processing results. This model is based off of storing exemplars of sentences in memory, and using this memory store to retrieve the expected future structure of a sentence. Unlike most theories of language, this approach is not concerned with learning the rules of a language. Instead the predicted structure of the current language environment is generated based on the previous experiences one has had with language. That is, the current understanding of a sentence is grounded in past experiences with language. This entails that structure in language is not just based upon rules and abstractions of the language input, but different communication patterns used to express different types of ideas. By storing these patterns in memory, this model has demonstrated that sophisticated expectations about forthcoming structures in language can be constructed.

References

Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review*, 23, 275-290.

Altmann, G. T. M., & Mirkovic, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33, 583-609.

Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, 63, 489-505.

Chomsky, N. (1988). Language and the problems of knowledge. *The Managua Lectures*. Cambridge, MA: MIT Press.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, 33, 1-36.

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 182-196.

Hare, M., Elman, J. L., Tabaczynski, T., & McRae, K. (2009). The wind chilled the spectators, but the wine just chilled: Sense, structure, and sentence comprehension. *Cognitive Science*, 33, 610-628.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411-428.

Jamieson, R. K., & Mewhort, D. J. K. (2009a). Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. *Quarterly Journal of Experimental Psychology*, 62, 550-575.

Jamieson, R. K., & Mewhort, D. J. K.. (2009b). Applying an exemplar model to the serial reaction time task: Anticipating from experience. *Quarterly Journal of Experimental Psychology*, 62, 1757-1783.

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114, 1-37.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20, 137-194.

Kanerva, P. (1988). *Sparse distributed memory*. Cambridge, MA: MIT Press.

Kwantes, P. J. (2005). Using context to build semantics. *Psychonomic Bulletin & Review*, 12, 703-710.

Levin, E. V., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24, 187-219.

McRae, K., Spivey-Knowlton, M. J., & Tenenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283-312.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3, 504-509.

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57, 1-23.

Sahlgren, M., Anders, H., & Kanerva, M. (2007). Permutations as a means to encode word order in word space. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, p. 1300-1305.

Simon, H. A. (1969). *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

Tomasello, M. (2003). Constructing a language: A usage-based theory of language acquisition. Cambridge, MA: Harvard University Press.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69-90.