

Generating Structure From Experience: A Retrieval-Based Model of Language Processing

Brendan T. Johns
Queen's University

Michael N. Jones
Indiana University Bloomington

Standard theories of language generally assume that some abstraction of linguistic input is necessary to create higher level representations of linguistic structures (e.g., a grammar). However, the importance of individual experiences with language has recently been emphasized by both usage-based theories (Tomasello, 2003) and grounded and situated theories (e.g., Zwaan & Madden, 2005). Following the usage-based approach, we present a formal exemplar model that stores instances of sentences across a natural language corpus, applying recent advances from models of semantic memory. In this model, an exemplar memory is used to generate expectations about the future structure of sentences, using a mechanism for prediction in language processing (Altmann & Mirković, 2009). The model successfully captures a broad range of behavioral effects—reduced relative clause processing (Reali & Christiansen, 2007), the role of contextual constraint (Rayner & Well, 1996), and event knowledge activation (Ferretti, Kutas, & McRae, 2007), among others. We further demonstrate how perceptual knowledge could be integrated into this exemplar-based framework, with the goal of grounding language processing in perception. Finally, we illustrate how an exemplar memory system could have been used in the cultural evolution of language. The model provides evidence that an impressive amount of language processing may be bottom-up in nature, built on the storage and retrieval of individual linguistic experiences.

Keywords: language processing, exemplar memory, sentence processing, semantic memory, grounded cognition

Typically, theories of language are abstractionist in nature, with individual experiences being used to create higher level representations of the workings of a language. An alternative approach to abstractionist theories has been explored in many areas of cognitive psychology—one based on the storage of individual experiences, or instances. Different instance-based theories have been developed to explain categorization (Nosofsky, 1986), task automatization (Logan, 1988), recognition memory (Hintzman, 1988), attention (Kruschke, 1992), and schema abstraction (Hintzman, 1986), among many other phenomena (for an integrated view, see Logan, 2002).

In the realm of language processing, the exemplar approach has been championed from the usage-based perspective (Abbot-Smith & Tomasello, 2006; Tomasello, 2003) on the basis of much evidence that language development is largely item based and not dependent on acquired syntactic categories (Tomasello, 2000). One example of this is Lieven, Pine, and Baldwin (1997), who found that the majority of a child's utterances are based on a few experienced lexical patterns. As Tomasello (2003) noted, this type of research demonstrates that perhaps the best way to generate

grammatical utterances is to use examples of previously heard communications. This is also similar to the viewpoint of the distributed approach to language (e.g., Dale, 2012), which proposes that language is culturally constructed.

The usage-based perspective has been supported by adult studies demonstrating that an increased amount of experience with certain grammatical constructs allows for a greater ease of processing. For example, Reali and 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 Reali and Christiansen found that subject-relative clauses are the more common construct when impersonal pronouns are used. However, they also found that object-relative clauses using personal pronouns are more frequent than subject-relative ones, and a self-paced reading experiment demonstrated that this leads to a processing advantage for object-relative clauses over subject-relative clauses when personal pronouns are used. These findings suggest that the amount of experience one has with certain grammatical constructs affects one's processing of them. Wells, Christiansen, Race, Acheson, and MacDonald (2009) further examined this question by manipulating the number of object-relative clauses that participants experienced during reading over a period of several weeks. An increase in processing efficiency for object-relative clauses emerged, suggesting that the greater experience participants have with a construct, the easier it is to process. Similar findings have been found across a range of different language tasks, demonstrating the importance of experience in language processing.

This article was published Online First May 11, 2015.

Brendan T. Johns, Department of Psychology, Queen's University;
Michael N. Jones, Department of Psychological and Brain Sciences, Indiana University Bloomington.

Correspondence concerning this article should be addressed to Brendan T. Johns, Department of Psychology, Queen's University, 62 Arch Street, Kingston, Ontario K7L 3N6, Canada. E-mail: bjohns11@gmail.com

Situated, grounded, and embodied theories of language have also made similar claims about the importance of individual experiences. Situation models (Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998) propose that language is a set of instructions used to create a mental representation of a context that is described in language, with storage and retrieval of previously experienced situations playing a central role in this construction process (Zwaan & Radavansky, 1998). A related, and more general, theory of cognition is the perceptual symbols systems (PSS) approach of Barsalou (1999). PSS proposes that the symbols used in cognition are not amodal but are, instead, based in the sensory and motor modalities. In terms of language processing, this theory suggests that our perceptual experiences lay down traces within memory, which can then be retrieved to be used to generate a perceptual simulation of a described event in language. Specifically, language can act as a cue to retrieve the correct situational, perceptual, and action information that the language stream describes (Zwaan, 2004; Zwaan & Madden, 2005). The importance of both usage-based and grounded information was first proposed by Wittgenstein (1953), with the current article agreeing with many of the principles of this approach.

An instance model of memory—specifically, Hintzman’s (1986, 1988) MINERVA 2 model—has recently been adopted to understand artificial grammar learning (Jamieson & Mewhort, 2009a, 2009b, 2010, 2011), showing that such a model can be readily applied to language-like tasks. In the artificial grammar-learning task, 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 dependent on the similarity of the string to the other studied strings of the language. From this perspective, a grammaticality judgment is not based on coherence to an implicitly learned grammar, which is what was initially assumed in this task (see Reber, 1976), but, instead, it is a memory task based on the similarity of the probe string to exemplars of the language stored in memory. The model generates a grammaticality judgment by determining whether the current probe string is similar enough to the studied strings to be considered grammatical, similar in principle to an episodic-recognition judgment. This simple memory model was able to account for both classic and new experimental results examining grammaticality judgments with artificial languages, even though the model uses no explicit grammatical representations (Jamieson & Mewhort, 2009a). This model has been able to extend to a number of different tasks, such as the serial reaction time task (Jamieson & Mewhort, 2009b) and stem completion (Jamieson & Mewhort, 2010). If one accepts that artificial language experiments provide important information about the mechanisms underlying language processing, this leads to the question of what role memory processes may play in natural language. Similar models have also been used to explore semantic learning (Kwantes, 2005), lexical access (Goldinger, 1998), and perceptual inference (Johns & Jones, 2012), among other phenomena, suggesting that this approach to examining language is impressively robust.

A number of other models have been proposed that are based on similar principles. Dennis (2004, 2005) used an exemplar memory model as one aspect of a rather sophisticated model of verbal cognition, with much more intricate encoding and processing assumptions than the approach adopted here. Nonetheless, given the similarity between the assumptions of these two models, it is

likely that they are complementary to one another, and the success of the approach of Dennis (2004; 2005) also provides solid evidence that this type of model is useful in the modeling of language processing. Bannard, Lieven, and Tomasello (2009) used a Bayesian selection model to extract item-based grammars from a corpus of child-directed speech and found that these grammars were highly predictive of the utterances that children made. In addition, Thiessen and Pavlik (2013) used an adapted MINERVA model as a general theory of distributed learning, including different aspects of language acquisition, complementing the approach taken here. These previous models provide evidence that an item-based approach can provide a solid basis for a general model of language processing.

The current article describes a new computational model, based on an instance theory of memory, that can model many diverse natural language sentence-processing findings. It is based on integration of multiple memory and language models, including the bound encoding of the aggregate language environment (BEAGLE) model of semantic memory (Jones, Kintsch, & Mewhort, 2006; Jones & Mewhort, 2007; Recchia, Jones, Sahlgren, & Kanerva, 2010) and a classic instance model of memory (Hintzman, 1986, 1988). The model uses the storage and retrieval of linguistic experiences as the fundamental operations of language processing. The theoretical foundation of the model is the usage-based view of language (Tomasello, 2003), which rejects the notion that language is composed of rules over abstract syntactic categories, contending that, rather, it is composed of communicative constructions that emerge through experience with language and the use of it, perhaps as stored in an exemplar memory store (Abbot-Smith & Tomasello, 2006).

In addition, given some of the similarities between the usage-based view and the grounded cognition perspective (Barsalou, 1999), we demonstrate how an exemplar model can be integrated with perceptual information to ground the operation of the model in the perceptual environment. This is then used to explain a number of different results from the visual world paradigm. Finally, in addition to learning linguistic structure, we show how the exemplar-based approach can generate structured linguistic utterances, tested in an iterative-learning cultural evolution framework (Kirby, 2002). In the next part, we sketch out the theoretical basis and formal implementation of the model.

An Exemplar Model of Sentence Processing

This part describes the various components of an exemplar model of language learning as well as the theoretical justification for the modeling framework and choice of mechanisms. This includes an examination of both the representation and the processing assumptions of the model.

Representation Assumptions

As in Jamieson & Mewhort (2009a, 2009b, 2010, 2011), the model described here is loosely based on the MINERVA 2 (Hintzman, 1986, 1988) memory model. However, because our model deals with real language learned from a corpus, some more sophisticated storage assumptions are required. These storage assumptions are based on the fundamentals of distributed memory theory and recent advances in the modeling of semantic memory.

The current view agrees with other recent proposals about the importance of event knowledge information (e.g., Elman, 2009); however, the mechanisms proposed are different. Specifically, the current model is based on the view that language can be construed as a retrieval cue. That is, given certain linguistic information (e.g., “The farmer grew . . .”), the system can use the 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. The result is a very flexible system, because it allows for the combination of multiple memory traces (similar to the proposals of Zwaan & Madden, 2005), with this combination being dependent on the current context, allowing for a dynamic language-comprehension system. However, to accomplish this, the model requires a complete description of an event, which in this case would be a sentence. Thus, the instance of language used in this model is a representation of a sentence taken from a natural language corpus.

To construct an instance of a sentence, it is first necessary to determine how word order can be encoded, because this is obviously an essential component of language. In the memory-modeling literature, a number of different methods have been developed to encode order, such as the use of convolution in the theory of distributed associative memory, or TODAM, class of models (Murdock, 1982, 1995). Convolution is a technique that takes two vectors and creates a third, unique vector that represents the connection between the two items. This technique has been used in the BEAGLE model of semantics (Jones & Mewhort, 2007), in which it has been shown that by taking into account sentential order, a better semantic representation can be created than is possible by simply relying on pure co-occurrence information.

However, a new technique to encode order has recently been developed on the basis of the sparse distributed memory theory of Kanerva (1988, 2009) and used to approximate BEAGLE, which is significantly more computationally complex (Recchia et al., 2010). Under this proposal, each word is represented with a binary spatter code (called a word’s *environmental vector*)—which is essentially a large, sparse vector—and random permutations of such vectors allow for order to be encoded (Sahlgren, Holst, & Kanerva, 2008). These models use binary spatter codes (sparse ternary vectors), where nonzero values are either +1 or –1 with equal probability. These vectors are typically very sparse, with less than 1% of their values being nonzero.

To encode order, this approach uses random permutation (RP), which simply takes an environmental vector as input and creates an output vector that is a random shuffling of the input values. Word order is encoded by assigning each location within a sentence a unique RP. The different permuted environmental vectors are then summed into a composite, giving the representation of a sentence. Recchia et al. (2010) created a modified version of BEAGLE using RPs and found that the model was both more scalable to large corpora and gave better fits to semantic similarity norms than did circular convolution.

In models of semantics (e.g., Recchia et al., 2010), a unique sentential representation is created for each word in a sentence. The representation created for a word is dependent on that word’s location within a vector so as to provide information about that word’s role in the sentence. However, for the purposes of the present model, that is unnecessary, because all that is needed is an

encoding of the linear ordering of the total sentence. On the basis of this, the encoding of a sentence is given with the following equation:

$$\text{Sent} = \sum_{i=1}^n RP^i(\text{word}_i) \quad (1)$$

where RP^i represents a specific RP for location i 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 is stored. Storage of sentence exemplars constructed in this fashion serves as the basis of this model. Next, we describe how this memory store is used in sentence processing.

Processing Model

The operation of this model is based on 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, including constraint-based models and simple recurrent network models, because of the incremental nature of these models (see Altmann & Mirković, 2009; Elman, 2009; Levy, 2008; McRae & Matsuki, 2013). There is also a considerable amount of empirical evidence that expectation generation and prediction is a central component of language processing, both in behavioral studies (Altmann & Kamide, 1999, 2007; Kamide, Altmann, & Haywood, 2003) and in event-related potential studies (DeLong, Urbach, & Kutas, 2005; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005).

The exemplar model approach to predication attempts to use the current input (i.e., the cues) to retrieve the structure of the expected future context on the basis of 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 attributable to complex internalized representations but, instead, may be a result of the structure of the language environment that people are exposed to. This is also similar to the proposals of theories of constructive memory (e.g., Schacter, Addis, & Buckner, 2007; Schacter, Norman, & Koutstaal, 1998), in which it is suggested that processing of the current environment and predictions of the future are based on the structure of past experience. The present model proposes that the current structure of the environment is used as a retrieval cue to activate similar past experiences, an action that forms the underlying representation of the meaning of the current context. That is, the meaning of the current context is grounded in the past. What occurs in this model is not simply the prediction of future words but, instead, active construction of the meaning of the current linguistic input on the basis of past experience.

This leads to a question: Given the exemplar memory storage previously described, how can past experiences be used to generate expectations about the future states of a sentence? The model accomplishes expectation with the cued-retrieval technique described in the MINERVA 2 model (Hintzman, 1986, 1988) and is similar in nature to work in artificial grammar learning (Jamieson & Mewhort, 2009a, 2009b, 2010).

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*), which represents the aggregated information that is retrieved from memory, in response to the cue. The echo retrieves information that is attached to the cue—for instance, a paired associate in a cued-recall task. Using this method, processing of a sentence takes place at two levels: (a) generation of expectations to each word in the sentence and (b) comparison and integration with previously formed expectations.

Expectation generation using words as cues. MINERVA 2 retrieves information from memory in response to a cue. Here, the cue is a word in a sentence permuted by the position of that word. This activates the traces in memory where that word occurred in that position (and also traces activated by chance). These traces are then summed into a composite, referred to as an *expectation vector* (and represented with \mathbf{E}). This process generates the words that are expected to surround a given word in a certain position in a sentence. Each word is summed into the composite on the basis of how similar the probe is to the memory trace. The similarity metric used is a simple vector cosine (a length-normalized dot product):

$$\text{Sim}(\mathbf{P}, \mathbf{T}) = \frac{\sum_{i=1}^n \mathbf{P}_i \mathbf{T}_i}{\sqrt{\sum_{i=1}^n \mathbf{P}_i^2} \sqrt{\sum_{i=1}^n \mathbf{T}_i^2}} \quad (2)$$

where n represents the dimensionality of the vectors, \mathbf{P} is the probe vector, and \mathbf{T} is a memory trace.

Unlike typical models based on the MINERVA framework, we use only memory traces that have a positive similarity value so as to retrieve exemplars that have a similar structure to the inputs, because we want to generate expectations about words that should likely occur, not about words that should not occur. The expectation vector is formed with the following equation:

$$\mathbf{E}(\mathbf{W}) = \sum_{i=1}^n (\text{Sim}(\mathbf{W}, \mathbf{M}_i) > 0) \text{Sim}(\mathbf{W}, \mathbf{M}_i)^\lambda \times \mathbf{M}_i, \quad (3)$$

where n represents the number of traces in memory, \mathbf{W} is the word currently being processed, \mathbf{M}_i 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, because the greater the number of such exemplars, the less any single exemplar should contribute. This equation sums all positively activated traces into a single composite, proportional to the level of activation of the word. Again, we believe that this is similar in principle to a memory-retrieval task, in which one uses a word (in a certain position) to generate expectancies about what other words should surround it on the basis of one's past episodic experiences. The result of this computation is a high-dimensional semantic representation of the meaning of the word in that position, similar to the results of a semantic-space model.

In terms of sentence comprehension, this allows for expectations to be generated about the upcoming words in a sentence. However, it only retrieves what is expected in response to a single word (in a specific position). To account for sentence-processing effects (in which expectations are formed in response to multiple words), the expectations across words in a sentence have to be integrated, and this is described next.

Comparison and integration with previously formed expectations. The process just described will retrieve the expectations for what words should surround a word in a sentence. To represent the meaning of a sentence, these retrieved expectations are summed into a single vector, referred to as the *comprehension vector* (\mathbf{C}). By iteratively constructing expectation vectors, and summing these into a single composite, the meaning of a sentence is homed in on across the sequence of words. The comprehension vector is constructed as follows:

$$\mathbf{C}_j = \mathbf{C}_{j-1} + \mathbf{E}(\text{RP}^j(\mathbf{W}_j)) \quad j = 1, \dots, \# \text{ words} \quad (4)$$

where j is the current position in the sentence, \mathbf{W}_j is the word in that position, and \mathbf{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 to update expectations about the upcoming words. However, before the expectation vector is summed into the comprehension vector, it is normalized so that all values sum to unit length (by dividing each location by the total vector magnitude). The use of a normalization procedure ensures that each word adds the same amount of information to the comprehension vector (i.e., each word contributes equally to the predictions being formed). This simplifying assumption is used mainly to mitigate the contributions of the frequency of different words, because without normalization, very high-frequency words (e.g., function words) would overwhelm the prediction process, even though it is not obvious that this should be the case (indeed, it could be argued that low-frequency words provide more information about the meaning of a sentence because of the sparseness of their usage). Thus, the model uses a simplifying assumption that each word contributes equally to the comprehension process, even though this is likely not the case. More research is needed to determine formal mechanisms that can measure the relative importance of different words in a sentence, although the semantic deviation calculations used in Jones, Johns, and Recchia (2012) and Johns, Gruenenfelder, Pisoni, and Jones (2012) to explore lexical organization offer a promising solution.

The comprehension vector allows for an *expectation value* (EV) to be calculated for each word in the sentence, because 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 on the basis of the past words that had been processed. An increase in similarity is assumed to cause an increase in processing efficiency (and, hence, a decrease in processing time), because the traces in memory that require activation will already be active because of past processing. The EV for a specific word is calculated by taking the cosine (described in Equation 2) between the comprehension vector (\mathbf{C}_{j-1}) and the retrieved expectation vector. This value represents how expected the current word is by determining how much information about that word was previously retrieved. In terms of the semantic space, this signals how similar the overall sentence representation at time $j - 1$ is to

the retrieved representation of the word at position j . The EV is the main source of information used to simulate reading times in sentence-processing results.

Discussion

This part outlined a new model of sentence processing that is based in storage and retrieval from an exemplar memory system and is consistent with a variety of previous proposals (Barsalou, 1999; Hintzman, 1986; Jamieson & Mewhort, 2009a; Tomasello, 2003). This model proposes that language input is heavily structured and that this structure can be used to guide our comprehension process. *Comprehension* in this case refers to the active prediction of future words (there are obviously a number of different aspects to comprehension), which is an important aspect of this cognitive process (Elman, 2009; Levy, 2008). This model also allows for the idea of an exemplar memory explanation of language (Abbot-Smith & Tomasello, 2006) to be tested objectively. The model is also both mathematically and conceptually clear, which allows for a good understanding of both its successes and failures to be attained. Next, we examine how accurate the model's predictions of natural language processing are across a variety of sentence-processing effects.

Natural Language Simulations

The previous part described a new model of expectation generation in sentence processing based on exemplar storage and retrieval. To develop increasingly better models of cognition, it is necessary to start training, and testing, computational cognitive models on natural language. This is significantly more challenging than training and testing on artificial languages, because it requires externalized knowledge. That is, it requires information from outside of the specific experimental context (i.e., knowledge about the meanings of words).

We trained the model on 300,000 sentence exemplars extracted from the Touchstone Applied Science Associates (TASA) corpus (Landauer & Dumais, 1997)—slightly less than half of all sentences contained in the corpus. No sentences greater than 20 words in length were included in the analysis. The environmental vectors had a length of 25,000 elements, with each having five nonzero values, meaning that the vectors were very sparse. It is worth pointing out that the model is fairly resistant to differing levels of sparsity, but both the storage and processing complexity are reduced as one reduces the sparsity level.

To simulate behavioral results, the specific sentences used were taken from the particular studies under question. Simulations were run over a number of different resamples of the environmental vectors for the different conditions in an experiment. This allowed for a significance test to be used, similar to in a typical experiment. The λ parameter was fixed at 13 because of the large number of exemplars being used.

Relative Clause Processing

Relative clauses are embedded structures that modify a head noun phrase. It has been found (e.g., Holmes & O'Regan, 1981; Traxler et al., 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 . . .”). This has also been one of the areas of research that has been used to clarify the role of experience in language (see Wells et al., 2009). In a recent study, Reali and Christiansen (2007) conducted a corpus analysis in which the relative frequency of different types of relative clauses was measured to determine how frequency of occurrence affects the processing of this construct. They found that subject-relative clauses are more frequent when impersonal pronouns are used, but object-relative clauses are more frequent when personal pronouns are used, in the embedded noun phrase. In a series of self-paced reading tasks, Reali and Christiansen demonstrated that the ease of processing of the different clause types was related to frequency of occurrence in the language environment. Specifically, they found that when personal pronouns are used, object-relative clauses are easier to process than subject-relative clauses. The goal of the current simulation was to determine whether the model can produce the typical advantage for subject-relative clauses, similar to what Traxler et al. found for subject-relative clauses, and also the advantage for object-relative clauses when personal pronouns are used, similar to what Reali and Christiansen found. This was a straightforward simulation for this model because of the demonstration of the role of experience in the processing of this syntactic construct.

To demonstrate the typical subject-relative clause processing advantage, 30 clauses from Traxler et al. (2000 [of each type]) were used. Example sentences from this study are the following:

1. The lawyer that irritated the banker . . . (subject relative)
2. The lawyer that the banker irritated . . . (object relative)

The average EV for the relative clause region (“irritated the banker” vs. “the banker irritated”) was then calculated. To simulate the results of Reali and Christiansen (2007), two list sets that demonstrated a processing advantage for object-relative clauses were attained. The first list set contained 14 sentences in which the noun phrase consisted of second-person pronouns, as in these examples:

1. The consultant that called you . . . (subject relative)
2. The consultant that you called . . . (object relative)

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

1. The lady that visited me . . . (subject relative)
2. The lady that I visited . . . (object relative)

Both of these lists elicited a processing advantage for the two words following the relativizer “that” (“you called”/“I visited” for object-relative clauses vs. “called you”/“visited me” for subject-relative clauses). The average EVs were computed for these same regions. Testing the model across both of these list sets simply allowed us to test the model across different types of language to ensure that the difference found was a true one. For all three lists, 20 resamples of the environment vectors were conducted.

The different EVs across the three difference sentence types are plotted in Figure 1. This figure demonstrates that for the lists from

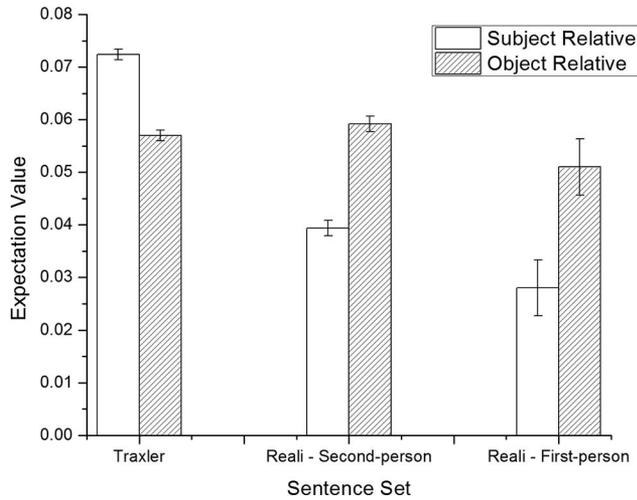


Figure 1. Simulation of the results for relative clause processing from Traxler et al. (2002) and Reali and Christiansen (2007). Error bars represent standard error.

Traxler et al. (2002), subject-relative clauses had higher EVs, which was statistically reliable, $F(1, 39) = 133.74, p < .001$, similar to the behavioral results. However, this pattern reversed itself for both the second-person and first-person pronoun sentences, in which the object-relative clauses had higher EVs. These differences were significant for both the second-person pronoun sentences, $F(1, 39) = 117.61, p < .001$, and the first-person pronoun sentences, $F(1, 39) = 9.485, p = .004$, replicating the results of Reali and Christiansen (2007) that the more common a syntactic construction is, the more predictable it becomes.

Effects of Contextual Constraint

A number of eye-tracking studies of reading have examined the role of contextual constraint on eye movements (e.g., Ehrlich & Rayner, 1981; Rayner & Well, 1996). This is typically done by making a target word in a sentence be either congruent or incongruent with the meaning of the sentence. As Rayner and Well pointed out, a consistent pattern of results has emerged from these studies: (a) Highly constrained words are more often skipped, (b) more regressions are made to unconstrained words, and (c) fixation times are shorter for constrained target words. These findings suggest that sentential context is used to generate expectancies about what words should occur in the upcoming structure of a sentence.

To determine whether this behavioral pattern would emerge from the model, we tested sentences from Rayner and Well (1996). These sentences were taken from Schwanenflugel (1986), who gave participants a sentence and asked them to produce a word that most fit with that context. This resulted in a production-probability value for different target words. Rayner and Well constructed 24 sentences for three conditions of contextual constraint: (a) high (range of production: 73%–100%), (b) medium (range of production: 13%–68%), and (c) low (range of production: 3%–8%). Examples of these sentences are as follows:

1. He mailed the letter without a stamp. (high)

2. The girl crept slowly toward the door. (medium)

3. Jill looked back through the open curtain. (low)

The set of 72 sentences from Rayner and Well was used in this simulation. To fit the data, the EV for each target word to the comprehension vector was assessed across the three different conditions. Fifteen resamples were conducted for each sentence set.

The results of this simulation are shown in Figure 2, demonstrating a similar pattern to the empirical results: Highly constrained words had higher EVs (manifesting in shorter reading times, greater probabilities of skipping, etc.) than medium- and low-constraint words. This was a significant effect, $F(2, 44) = 69.03, p < .001$, and a planned comparison confirmed that each condition was greater than each condition below it. This was a simple test of this model, but it provided an important basis for its operation: Across the words in a sentence, expectations are generated about the words that are likely to occur in the sentence.

Verb Sense and Expectation Generation

Hare, Elman, Tabaczynski, and McRae (2009) conducted a study in which the sense of a verb was manipulated to determine how it influenced expectations about upcoming words. Specifically, Hare, Elman, et al. manipulated the transitivity of a verb (with a verb being transitive if it has a direct object and intransitive if it does not). Many verbs can be either transitive or intransitive, depending on context. This is, in turn, related to causation, with causative verbs occurring in the transitive form (e.g., “he broke the vase”) and inchoatives occurring in the intransitive form (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, Elman, et al. tested this by manipulating the thematic fit of a subject to be inducing of either good theme (e.g., “the glass shattered . . .”) or good cause (e.g., “the brick shattered

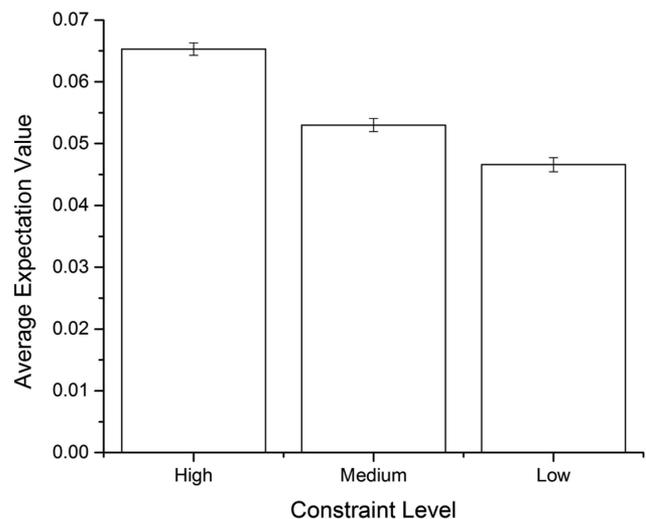


Figure 2. Simulation of the results from Rayner and Well (1996), demonstrating that the more contextually constrained a sentence is, the easier it is for the model to form predictions about upcoming words in the sentence. Error bars represent standard error.

. . .”), and they measured reading times to the postverb regions of intransitive (e.g., “. . . into tiny pieces”) and transitive (“. . . the fragile goblet”) sentences. They found that when a sentence was intransitive, reading times in the postverb region were significantly shorter for good-theme sentences, and the opposite was true for transitive sentences. This experiment demonstrated that verb sense manipulates the expectations that are generated, and, hence, it provided a nice test to determine whether an exemplar memory model could use this same verb-based information to generate expectancies.

To test the model on this result, 15 sentences were attained for each of the four different conditions from Hare, Elman, et al. (2009). EVs were then calculated at the verb (where no significant difference would be expected) and at the first nonfunction word in the postverb region (e.g., “tiny,” “fragile”), where a significant difference would be expected. Twenty resamples were simulated for each sentence type.

The results of this simulation (and the data from Hare, Elman, et al., 2009) are displayed in Figure 3. This figure demonstrates that the model could approximate the results of this experiment quite well. No significant difference was found at the verb region for either intransitive-biased sentences, $F(1, 39) = 0.16, p > .10$, or transitive-biased sentences, $F(1, 39) = 0.775, p > .10$. However, for intransitive-biased sentences, the EVs for nouns in good-theme sentences were significantly greater than those for nouns in good-cause sentences, $F(1, 39) = 29.59, p < .001$. The opposite was true for transitive-biased sentences, with EVs for nouns in good-cause sentences being significantly greater than those for nouns in good-theme sentences, $F(1, 39) = 36.51, p < .001$. This is a very important result for this theory as it demonstrates that the model does not simply generate expectancies on the basis of 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. These expectancies are dependent on the combination of words and not simply the sequential order of the sentence.

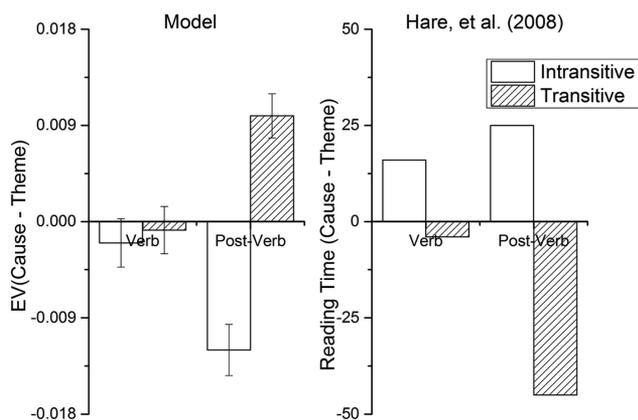


Figure 3. Simulation of the results from Hare Elman, Tabaczynski, and McRae (2009). When a verb is intransitive, it is theme inducing, whereas a transitive verb is cause inducing, and this leads to processing differences in the postverb region. The y-axis is reversed for the model because of lower expectation values (EVs) resulting in slower processing, and vice versa. Error bars represent standard error.

Event-Knowledge Activation

A recent line of promising research has examined how knowledge of events comes into play during sentence comprehension (Bicknell, Elman, Hare, McRae, & Kutas, 2010; Ferretti, Kutas, & McRae, 2007) and the role it plays in semantic memory in general (Hare, Jones, Thomson, Kelly, & McRae, 2009). This typically involves manipulating congruent/incongruent or low-/high-typicality event knowledge that is associated with a particular verb or noun. In particular, we attempted to simulate two results: (a) a finding by Ferretti et al. that greater N400 amplitudes are exhibited to low-typicality events than to high-typicality events, suggesting a greater surprisal value to unexpected events, and (b) a recent result obtained by Bicknell et al. 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”).

To simulate these results, sentences were taken from the relevant studies (37 sentences from Ferretti et al., 2007, and 40 from Bicknell et al., 2010). For the Ferretti et al. study, EVs were calculated to the last word of each sentence, where the word was either a high- or low-typicality word. Two example sentences (high/low typicality) are these:

1. The girl was skating in the (rink/ring).
2. The king was sitting on the (throne/stage).

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

1. The (librarian/composer) arranged the shelf.
2. The (secretary/speaker) addressed the letter.

To determine whether the model would find a difference between the two sets of words, 20 resamples were simulated for each sentence set.

A significant difference was found for the sentences from the Ferretti et al. (2007) study, $F(1, 39) = 73.21, p < .001$, demonstrating 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 Bicknell et al. (2010), although this effect was not as large, $F(1, 39) = 6.97, p = .01$. The smaller difference is not surprising, because the task was more complicated, 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 Bicknell et al. (2010) and Ferretti et al. (2007) demonstrated 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.

Internal/External Change-of-State Verbs

A recent influential theory of sentence processing is the meaning through syntax (MTS) approach (McKoon & Macfarland, 2002; McKoon & Ratcliff, 2003). This approach proposes that the syntactic information about a verb is stored as a template in the

lexicon for that word. Such a template describes how the subject and objects (in the case of transitive sentences) are linked by the verb. The linking rules in turn describe how the arguments are encoded in sentence order. If this is the case, the model described here should be able to approximate this representation, because during retrieval, the arguments that surround a verb should be contained within a word's expectation vector.

One of the main points of analysis in the MTS framework is based on the difference between internal and external causation verbs, because it is presumed that the event structures for these verbs would differ in their complexity. For internal causative verbs, there is only a change of state involved, whereas for external causative verbs, there is both a cause and a change of state. Under the computational framework proposed here, this should manifest in terms of less determinacy about the expectations for external verbs.

To test this, two sets of sentence sets were taken from [McKoon and Macfarland \(2002\)](#). One of the sets was composed of transitive sentences (14 sentences for both internal and external verbs), and the other was composed of intransitive sentences (26 sentences for both internal and external verbs). Examples of these sentences are as follows:

1. The flowers bloomed. (internal/intransitive)
2. The concrete crumbled. (external/intransitive)
3. The intense heat wilted the crowd. (internal/transitive)
4. The telephone call awoke the residents. (external/transitive)

Because in the intransitive sentences, the verb occurred in the final position, it was not possible to calculate EVs in the same manner as done previously. Instead, the similarity between the expectation vector for the verb and the preceding noun was calculated to determine how predictive the verb was of the previous word. For the transitive sentences, the EVs were generated in the same manner as in the previous simulations, with EVs averaged from the verb to the end of the second noun phrase. This was done across 20 resamples for each sentence set.

It was found that the EVs for internal verbs were significantly greater than those for external verbs, for both intransitive, $F(1, 39) = 10.70, p = .002$, and transitive, $F(1, 39) = 32.40, p < .001$, sentences. This demonstrates that the complexity of a verb's event structure can also manifest in terms of the predictability of the surrounding structure of that verb, suggesting that retrieval from memory may provide a plausible mechanism for generating a word's event template, in coherence with the proposals of MTS theory.

Discussion

We tested a simple exemplar-based model of expectation generation on standard sentence-processing tasks from the literature. The model is based on storage of exemplars of sentences in memory (based on models of semantic memory) and usage of this memory store to retrieve the expected future structure of a sentence (given limited input). 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 environ-

ment is generated on the basis of the previous experiences one has had with language, which allows for a prediction to be formed about future input. That is, the current understanding of a sentence is grounded in past experiences with language.

Grounded models of cognition ([Barsalou, 1999, 2008; Zwaan, 2004; Zwaan & Madden, 2005](#)) make similar claims about the importance of exemplar memories in language processing. They propose that language can be used to retrieve attached referential information ([Zwaan & Madden, 2005](#)) to construct a perceptual simulation of the current language context. How this can be formalized within the current framework is described next.

Integration of Linguistic and Perceptual Information

Language is not simply an amodal communication mechanism; instead, it is used to refer to the physical and embodied environment that one is a part of ([Barsalou, 2008](#)), and perceptual-motor information is in turn used to ground the meaning of words. The importance of grounded and embodied information in language processing has been demonstrated conclusively across a variety of different tasks, including priming studies ([Myung, Blumstein, & Sedivy, 2006](#)), conceptual processing ([Connell & Lynott, 2011](#)), and word recognition ([Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008](#)), among many others. This research suggests that linguistic and perceptual-motor information are intricately connected to one another, so any proposed theory of language must somehow include this information. To accomplish this, we attempt to formalize aspects of PSS ([Barsalou, 1999](#)) and the interconnected trace hypothesis ([Zwaan, 2004; Zwaan & Madden, 2005](#)) within our exemplar-based framework. Although these theories make strong assumptions about the role of grounded and motor information in language comprehension, multiple studies have questioned the strictness of this assumption (e.g., [Garcea, Dombovy, & Mahon, 2013; Stasenko, Garcea, & Mahon, 2013](#)). We are neutral on exactly how integrated language and perception are; instead, we simply recognize the importance of grounded and perceptual information in the language-comprehension process and offer a simple formal mechanism for how this integration is possible.

PSS theory ([Barsalou, 1999](#)), the cornerstone of the grounded cognition movement ([Barsalou, 2008](#)), has been proposed as a competitor to pure linguistic models as an explanatory theory for language comprehension. The basis of PSS is the dismissal of amodal symbols as the central component underlying human mental representations. Rather, the PSS approach proposes that the symbols used in reasoning, memory, language, and learning are grounded in sensory modalities.

In the realm of language, PSS proposes that the mental representation of a word is based on the perceptual and motor states that underlie experiences with the word's physical referent ([Barsalou, 1999](#)). Across many experiences with words, the underlying neural states tend to stabilize and create an accurate perceptual-motor representation of a word that is grounded across sensory areas in the cortex ([Barsalou, Simmons, Barbey, & Wilson, 2003](#)). These perceptual simulators can then be combined, guided by the structure of the linguistic input, to construct simulations of the meanings of sentences and discourse.

The specific method of grounding language that we espouse here is very similar to the proposals of the interconnected experi-

ential trace hypothesis (Zwaan & Madden, 2005). This framework proposes that there are two main types of traces laid down in memory during experiences with language: (a) linguistic traces (i.e., traces of linguistic structure created by receiving or producing language) and (b) referential traces (i.e., perceptual information experienced during the linguistic experience). Language input then acts as a retrieval cue by activating linguistic traces stored in memory, which in turn activates referential traces, and this leads to a mental simulation of the described events in the language stream in the sensory–motor areas of the brain.

To store referential traces in a formal model, one must first have a representation of the perceptual referents of words. There have been a number of different methods developed to estimate perceptual representation (e.g., Andrews, Vigliocco, & Vinson, 2009; see McRae & Jones, 2013). The technique we use here is the generating perceptual representations (GPR) model of Johns and Jones (2012), which uses global lexical similarity to construct perceptual representations about words that have no perceptual information in their lexical representations. The bases of the representation are the feature norms from McRae, Cree, Seidenberg, and McNorgan (2005). These norms only contain representations for around 500 words, but Johns and Jones used the global lexical similarity among words to generate perceptual representations. The result of this process was that all words had inferred perceptual representations that were cross-validated with reasonable precision.

Representational Assumptions

The GPR model is able to generate the expected perceptual representations for a wide variety of words, which we use to create referential traces. However, this is a very simple approximation compared with what we believe is actually stored in these traces, and it ignores many of the complexities contained in the physical environment. This is a necessary evil as it is currently unclear how to create more complex traces on a wide scale, and the proposals described here should be seen as simple jumping off points and as ways to formalize some of the claims of grounded theories.

To construct perceptual representations, we first used the GPR model to construct inferred perceptual representations for the 24,000 most frequent words in the TASA corpus (for the specific method and parameter values used, see Johns & Jones, 2012). These individual word representations were then used to construct a perceptual representation of a sentence. To create a referential trace for a specific sentence, all the words that had a generated perceptual representation were summed into a single trace:

$$\mathbf{R} = \sum_{j=1}^n \mathbf{P}(\mathbf{W}_j) \quad (5)$$

where \mathbf{R} is the created referential trace, \mathbf{P} retrieves the perceptual representation of a particular word, \mathbf{W} contains the words in the sentence for which a referential trace is currently being constructed, and n is the length (in words) of the sentence. This referential trace was then concatenated to the linguistic exemplar and stored in memory. However, because not all words experienced have perceptual referents, only 25% of linguistic traces will have a referential trace attached to them. These perceptual traces are used to generate expectations about upcoming perceptual information on the basis of what is contained in the linguistic stream.

Retrieval of Referential Information

The previous section described how referential information is stored in the linguistic memory store. As studies in the visual world paradigm have demonstrated (e.g., Altmann & Kamide, 1999), people are capable of constructing predictions about perceptual information on the basis of the sentence they are currently experiencing. The exemplar model attempts to explain these types of results by constructing a perceptual simulation of the linguistic stream by combining past referential experiences through a memory-retrieval operation in a very similar manner to the process described in Part 1.

Expectations about the objects that a specific sentence is referring to are constructed by using the words in a sentence to retrieve the correct perceptual information. This is accomplished by cuing the exemplar memory system. As was done in the pure language iteration model, this is accomplished by first generating expectations regarding a word in a particular position. This is done in the same manner as the creation of an expectation vector in Equation 3, but referential traces are summed instead of linguistic ones. The cue is still a word, permuted by the position of that word in a sentence. A referential expectation (\mathbf{RE}) vector is constructed as follows:

$$\mathbf{RE}(\mathbf{W}) = \sum_{i=1}^n (\text{Sim}(\mathbf{W}, \mathbf{LM}_i) > 0) \text{Sim}(\mathbf{W}, \mathbf{LM}_i)^\alpha \times \mathbf{R}_i, \quad (6)$$

where \mathbf{W} is the current word being processed (permuted by its location within the sentence), \mathbf{LM} is the linguistic memory store, α is a scaling parameter that determines how much information one referential trace adds into the resulting \mathbf{RE} vector, and \mathbf{R} is the referential trace attached to the \mathbf{LM} vector. This vector retrieves the perceptual information that the word is expected to be referring to. Again, traces with only positive similarity values are used so as to only retrieve expectations about references that should occur.

However, it has been demonstrated that individuals generate predictions in response to multiple words, which means that \mathbf{RE} vectors need to be summed into a single vector, similar to the use of the comprehension vector in Part 1. The composite vector formed by summing the \mathbf{RE} vectors is called a perceptual simulation (\mathbf{PS}) vector, giving the constructed perceptual representation for a specific sentence:

$$\mathbf{PS}_j = \mathbf{PS}_{j-1} + \mathbf{RE}(\mathbf{R}^j(\mathbf{W}_j)) \quad j = 1, \dots, \# \text{ words} \quad (7)$$

This vector represents the total expected perceptual information that the language stream is referring to. These two vectors (\mathbf{RE} and \mathbf{PS}) are the main sources used to simulate a number of different behavioral results. Specifically, by taking the similarity between some environmental stimulus (say an object) and the \mathbf{PS} or the \mathbf{RE} vector, a measure of expectation of that stimulus is computed. This is the driving force behind the following simulations.

Simulation Details

Three different results were simulated, with one using an artificial language and two using natural language. These simulations demonstrate that the model is capable of accounting the following findings: (a) resolving ambiguity with visual information, (b)

generating verb frames through exposure to the visual environment, and (c) generating expectancies with verbs. For the studies using natural language, the same 300,000 sentences from the TASA corpus as described in Part 2 were used. Twenty-five percent of these sentences contained referential traces concatenated to them in memory. In addition, function words were not used to generate expectancy values for referents.

Resolving Ambiguity With Visual Information

The main experimental paradigm that has been used to study how linguistic structure and perceptual information interact is the visual world paradigm (for a review, see [Huettig, Rommers, & Meyer, 2011](#)). A typical experiment in this paradigm involves displaying a number of objects onscreen and having participants listen to a sentence. Monitoring of where participants look on the screen while listening to the sentence allows for a real-time look into the processes underlying sentence processing. One of the first studies to use this paradigm was conducted by [Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy \(1995\)](#). In this study, participants were given sentences such as “Put the apple on the towel on the table,” and eye movements to four different pictures were monitored: (a) an apple on a towel, (b) a towel, (c) a table, and (d) an irrelevant stimulus (e.g., a pencil). [Tanenhaus et al.](#) found that participants initially fixated on the apple, but when they heard “towel,” they fixated on the lone towel, because the phrase “put the apple on the towel” is ambiguous. However, when “on the towel” was heard, fixations tended to move back to the apple and then to the table. A second manipulation was examined in which [Tanenhaus et al.](#) replaced the irrelevant picture with a second referent (e.g., an apple on a napkin), making the phrase “put the apple” ambiguous as to what apple was being referred to, which was not resolved until “on the towel” was heard. This linguistic pattern was borne out in the pattern of participants’ eye movements. Unlike in the first condition, the towel by itself was not fixated, because “on the towel” was being interpreted as specifying the correct referent and not a destination. This was a significant finding as it demonstrated that the meaning of an utterance was being interpreted in an online fashion, unlike the claims of two-stage processing models.

The exemplar-based model we propose here is unable to provide a complete account of this result, mainly because of a lack of appropriate verb representations (e.g., it does not know what “put” is) or adverbs (e.g., “on”). Instead, we tested whether the representations retrieved were at least mildly consistent with the pattern that [Tanenhaus et al. \(1995\)](#) found. To test this pattern, a perceptual simulation vector was constructed in the same manner as described in Equation 7. An activation value was then calculated for all the different pictures in a scene. Composite pictures (e.g., an apple on a towel) were represented by averaging the GPR representation for, in this example, apple and towel. This activation value was a mixture of the similarity of the picture to the retrieved scene representation (the **PS** vector) and the retrieved referent of the word currently being processed (the **RE** vector), with the importance of the different information sources being controlled with the parameter π . Similarity was again computed as a vector cosine. The activation value calculated for a picture **X** is

$$Act(\mathbf{X}) = (\text{Cos}(\mathbf{X}, \mathbf{PS}) \times \pi) + (\text{Cos}(\mathbf{X}, \mathbf{RE}) \times (1 - \pi)) \quad (8)$$

where π is a value between 0 and 1, which for this simulation was set at .75. The π parameter simply modulates the contributions of already formed predictions (the perceptual simulation vector) and the current prediction information retrieved from the word being actively processed. A value of .75 indicates that most weight is being given to past predictions but that the current expectations can still have a considerable impact on the activation of different perceptual referents. The activation value is calculated for each word in the sentence (with the exclusion of some function words [e.g., “is,” “the,” “put”]). Five sentences from [Tanenhaus et al.](#) were used to ensure that the patterns of activation values were not simply attributable to one set of words.

The simulation results of both the one-referent and two-referent conditions are displayed in [Figure 4](#). This figure demonstrates that the model gave quite a good approximation of the findings of [Tanenhaus et al. \(1995\)](#), even with all the limitations in terms of the referent representations being used. For the one-referent condition, when the word “apple” was processed, this was the picture that has the highest activation level, until “towel” was heard, at which point the activation was split between the two corresponding pictures. This is not exactly what the study data demonstrated as participants tended to focus more on the towel, but it does demonstrate that both words’ referents were being successfully retrieved. When “table” was seen, that picture became the highest activated referent. The pattern of activation values for the two-referent condition was similar to this, with the exception that when the word “apple” was processed, both of the referents had the same activation value because of ambiguity about which one was being referred to. However, this ambiguity was resolved once the word “towel” was processed. The towel picture’s activation level tended to be too high, but the apple + towel picture still had the highest activation value at that point, similar to the study data. This demonstrates the power of integrating linguistic and referential traces and using these to generate simulations of the perceptual environment from exemplar memory, because it shows how referential information can be retrieved on the basis of the storage and retrieval of interconnected linguistic–perceptual traces.

Generating Verb Templates

[Wonnacott, Newport, and Tanenhaus \(2008\)](#) conducted a set of artificial language–learning experiments in which they examined how individuals are able to acquire verb–argument structure in the visual world paradigm. These experiments have been modeled sufficiently elsewhere (see [Perfors, Tenenbaum, & Wonnacott, 2010](#)), so we did not wish to be overly redundant with other work. Instead, we used the paradigm to test the ability of the present model to accomplish one simple task: to generate expectations about the visual environment from verbs. Multiple simulations in Part 2 demonstrated that verbs are able to retrieve upcoming linguistic structure, suggesting that they should also be able to accomplish the task of retrieving structure for the visual world. The main manipulation of [Wonnacott et al.](#) that allowed for this test was their combination of the visual world paradigm with an artificial language–learning paradigm. Specifically, the language contained sentences that had the verb in the first position, followed

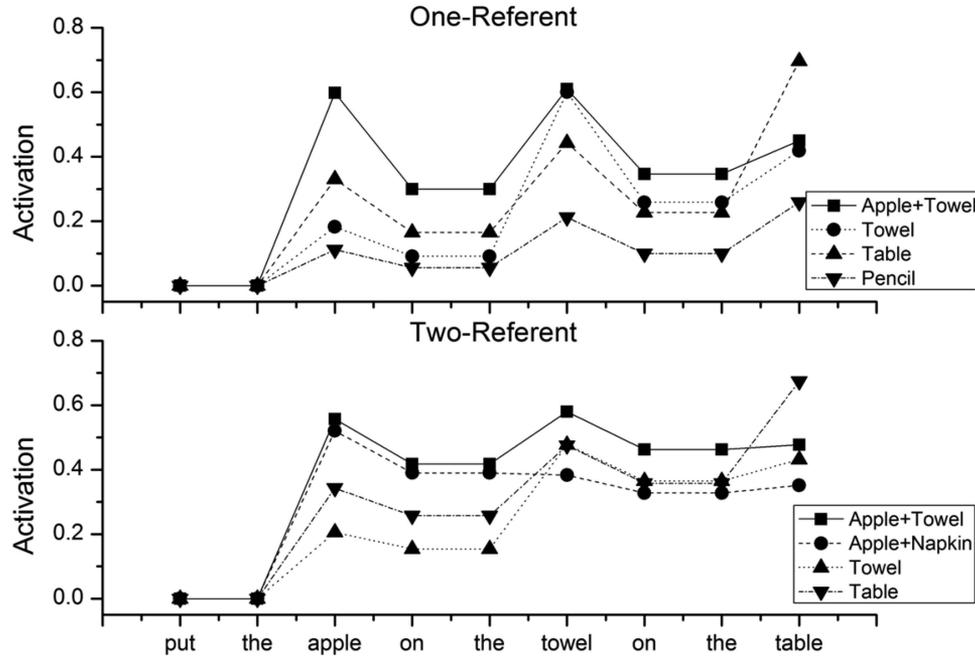


Figure 4. Simulation of Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995). This simulation demonstrates that the model is retrieving perceptual information that is an accurate reflection of the linguistic stream, which can be used to resolve ambiguity in comprehension.

by two nouns. One group of verbs then had a particle at the end of each sentence, whereas another group did not. A third group occurred equally often with either construction. Critically, the two constructions differed in which noun corresponded to the patient and which corresponded to the agent. In no-particle sentences, the first noun (the second word in the sentence) corresponded to the agent in a visual scene, and the second corresponded to the patient, whereas the opposite was true for particle sentences. We sought to determine whether the present model could generate expectations about the construction of different verbs.

To model this paradigm, the 12 verbs used in Wonnacott et al. (2008) were constructed with the GPR model (split into three groups of four), and a random environmental vector was generated to represent the perceptual referent of a word. This was also done for the five nouns used. Sentences were constructed in the following fashion: $RP^1 verb + RP^2 noun1 + RP^3 noun2 + (RP^4 particle)$, where the particle is optional depending on which sentence type is being generated. Half of the verbs in a group were low frequency (occurred in six sentences), and half were high frequency (occurred in 18 sentences). However, the referential traces were constructed in a different way than was previously done, because some mechanism for marking agents and patients was required. To accomplish this, instead of simply summing all of the perceptual traces, the representation of the agent was given a unique permutation, and the same was done for patients. For the no-particle sentences, the referential traces were constructed in the following manner: $verb + RP^a noun1 + RP^p noun2$, whereas for the particle sentences, the traces were constructed with this: $verb + RP^p noun2 + RP^a noun1$. To determine whether the verb could retrieve the correct perceptual structure, the verb was used to retrieve an RE vector by probing with the verb at Location 1. A

grammatical rating was then calculated by taking the cosine between the RE vector and a test scene composed with the structure of a particle and a no-particle scene. A criterion was used to determine whether the expectation generated was similar enough to the scene vector, and this criterion was fit to the grammaticality judgments of Wonnacott et al.'s Experiment 1. This was averaged across 25 resamples.

The results of this simulation are displayed in Table 1 along with the corresponding data from Wonnacott et al. (2008). This table shows a strong fit to the study data ($R^2 = .98, p < .001$) and demonstrates that the verbs were able to retrieve the correct expected structure about the environment. It also demonstrates that the present model is capable of generating grammatical markings through the recording of referent traces, similar to the claims of Zwaan and Madden (2005).

Table 1
Simulation of the Results of Wonnacott Newport, and Tanenhaus (2008)

Verb type	Model	Wonnacott et al. (2008)
No particle, low frequency	.44	.30
No particle, high frequency	.60	.61
Particle, low frequency	-.45	-.43
Particle, high frequency	-.59	-.54
Both	-.03	-.06

Note. Values are the difference between accepting a particle scene and nonparticle scenes.

Generating Expectations With Verbs

A landmark result in the visual world paradigm was reported in a study by Altmann and Kamide (1999), which tested the role of prediction in verb comprehension. In this study, participants were presented with sentences such as “The boy will eat the cake” and “The boy will move the cake” while four different pictures were being displayed. Of the four pictures, only one displayed an object that was edible, and it was found that in the postverbal region, people tended to focus more on this picture when the verb “eat” was heard compared with when the verb “move” was heard. This demonstrated that there was rapid online prediction occurring during sentence comprehension and that a prediction made from linguistic information also mapped onto the visual world.

To model this result, 16 sentences were taken from Altmann and Kamide (1999) along with the relevant pictures that were shown with the sentences (represented with GPR vectors of the respective items in the pictures). Activation values were computed with Equation 8, and a π value of .75 was again used. The activation values were then assessed across four different verb–picture relationships: (a) “eat”–cake, (b) “move”–cake, (c) “eat”–other (e.g., TV), and (d) “move”–other. These were averaged across 25 resamples of the environmental vectors.

The results of this simulation are displayed in Figure 5. This figure shows that when the verb was processed, it retrieved information about what that verb was likely to act on. This information manifested in greater activation values for the related noun picture when the verb was processed compared with a nonpredictive verb. However, this activation difference was reduced after the noun was processed, similar to the results of Altmann and Kamide (1999). That is, when “eat” was processed, it generated a simulation about what that verb was likely to act on (e.g., a cake). This was accomplished even across a large exemplar memory of natural language, demonstrating the power of memory retrieval and the integration of linguistic and referential traces, as Zwaan and Madden (2005) suggested.

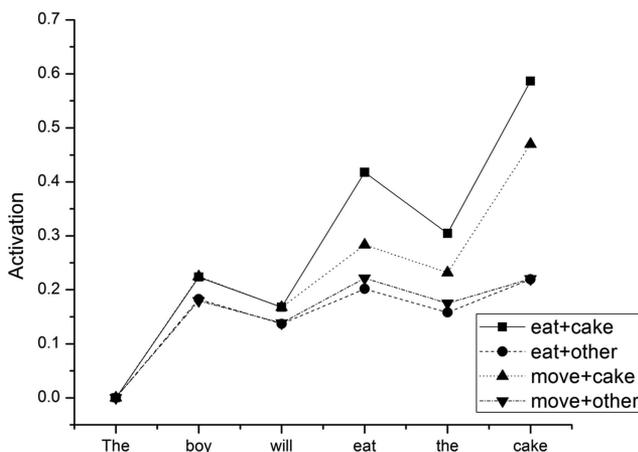


Figure 5. Simulation of Altmann and Kamide (1999), demonstrating that verb information was being used to generate predictions about the noun that was being acted on.

Discussion

The goal of this part was to use the exemplar-based framework described in Part 1 to formalize aspects of grounded theories of language. The results were promising, but clearly there is much work to be done to determine how successful this approach can be. For instance, the memory traces used will clearly have to contain more multimodal information (e.g., auditory, motor) and should also include higher level information—such as goal-based, motivational, and emotional knowledge—as usage-based theories of language propose (Tomasello, 2003). Whether these different information sources could be contained within a single trace or multiple traces concatenated together is unclear.

However, the theory outlined here and tested in Parts 2 and 3 still has not demonstrated the ability to explain one important aspect of language: how structure in language can be constructed. That is, we have only shown that an exemplar memory system can be used to learn the structure of language, not to generate it. However, if this is to be a serious model of language, it should also be able to also take an initially unstructured language and create structure through the communication of agents within an environment. This assertion is based on the proposals of Tomasello (2003) and the distributed approach to language (Dale, 2012), both of which suggest that language is constructed by social/cultural processes. We next describe how this is possible within the current framework.

Joint Memory and the Cultural Evolution of Language

The need for an innate language module has been one of the largest criticisms leveled against generative linguistics. There are many in-depth discussions of the issue (see Christiansen & Chater, 2008; Evans & Levinson, 2009), and addressing them all is beyond the scope of this article. Rather, we aim to simply demonstrate how memory may play an important role in the development of structure in language through cultural evolution. The proposal that we focus on is one offered by Christiansen and Chater (2008), who theorized that language has overtaken already evolved mechanisms in the brain to build a system of communication. That is, language is shaped by domain-independent learning mechanisms, such as memory and perceptual–motor factors, among other cognitive processes.

In this part, we explore the role that exemplar memory could have played in the cultural evolution of language. As noted by Jamieson and colleagues (e.g., Jamieson, Crump, & Hannah, 2012; Jamieson, Hannah, & Crump, 2010), there is considerable evidence that other animals also have exemplar memory systems and, hence, would be a strong candidate to be overtaken in the evolution of language under Christiansen and Chater’s (2008) framework. For example, Fagot and Cook (2006) tested the long-term memory capacity of pigeons and baboons. They found that pigeons could successfully memorize between 800 and 1,200 items before reaching the limit of performance, whereas baboons memorized between 3,500 and 5,000 items over 3 years without ever reaching a limit. Similarly, recent results have found episodic memory across experiments with rats (Kart-Teke, De Souza Silva, Huston, & Dere, 2006). These results suggest that exemplar memory may

have evolutionary origins across a number of species and is unlikely to be a human-specific learning mechanism.

On the basis of this evidence and the proposals of [Christiansen and Chater \(2008\)](#), it seems plausible that an exemplar memory system could have been overtaken for the use of language. To test this hypothesis, we conducted an iterative multiagent simulation. The iterated learning model of language evolution ([T. L. Griffiths & Kalish, 2007](#); [Kirby, 2002](#)) is a computational framework for studying the emergence of language through cultural interaction rather than through a genetic basis. Within this framework, there is a set of agents who perform two basic functions: (a) receive a set of meanings from the environment and (b) signal those meanings to other agents when prompted to do so. Given that the agents are given sufficiently capable learning mechanisms, it has been demonstrated that this type of process is capable of taking initially random input and creating structured output that contains aspects of human language ([Smith, Kirby, & Brighton, 2003](#)).

Language

This simulation used a simple language composed of sentences over different semantic categories but with no structure in terms of how the sentences were expressed. As in Part 3, words were physical referents—hence, all words corresponded to something in the environment (be it an action or an object). Each sentence consisted of three words. To simplify the simulation (because we are interested in evolution of word order), it was assumed that agents already had a mapping for words to objects. Each word and each object was represented with random vectors. The task of the agents was to generate a common set of ordering to the same environmental information. The only processing operations were a memory-retrieval process and the joint communication of agents within the community (in coherence with the proposals of usage-based theories). More details of the simulation are described in Section 4.3.

Model

The first step in this simulation was for an agent to construct an utterance on the basis of the information currently in the environment. As in the previous iterations of the exemplar model, this required a cue to generate structure from memory. The cue to generate an utterance was a scene vector (\mathbf{S}), which was constructed as described in Equation 5 by summing all of the items in the environment into a single vector. The \mathbf{S} vector was assumed to be the surrounding perceptual environment, and it was used as a retrieval cue by an agent to generate a linguistic representation of the scene on the basis of the storage of attached perceptual and linguistic information. The constructed linguistic vector is referred to as the *retrieved linguistic* (\mathbf{RL}) trace. The retrieval process returned the most likely linguistic structure to describe a scene and was used by an agent to generate an utterance. The \mathbf{RL} vector was constructed as follows:

$$\mathbf{RL}(\mathbf{S}) = \sum_{i=1}^n (\text{Sim}(\mathbf{S}, \mathbf{R}_i) > 0) \text{Sim}(\mathbf{S}, \mathbf{R}_i)^\alpha \times \mathbf{L}_i, \quad (9)$$

where \mathbf{S} refers to the scene vector, \mathbf{R} is a referent exemplar stored in memory, \mathbf{L} is the attached linguistic trace, and n is the number

of traces in the agent’s memory. This equation sums all linguistic traces in memory proportional to the similarity between the stored reference traces and the current reference vector. This process resulted in the likely linguistic representation being retrieved on the basis of the surrounding perceptual representation. In some ways, this was the inverse of the operation of the model described in Part 3: Instead of perceptual information being retrieved from linguistic cues, here linguistic information was retrieved from perceptual cues.

The \mathbf{RL} vector was then be used to generate the word ordering of the scene’s referents, accomplished by testing all possibilities for the ordering of a set of words and calculating a similarity value for each different permutation. The ordering that generated the highest activation value was the one that was generated by the agent. Similarity was calculated by taking the cosine between a constructed sentence representation (one ordering) and the \mathbf{RL} vector. A candidate sentence was constructed with the following equation:

$$\text{Sent}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, n) = RP^{n(1)}\mathbf{X} + RP^{n(2)}\mathbf{Y} + RP^{n(3)}\mathbf{Z} \quad (10)$$

\mathbf{X} , \mathbf{Y} , and \mathbf{Z} are the words under consideration, whereas n is a particular ordering of those words (any permutation of 1, 2, and 3). The sentence with the highest activation level was selected with the following:

$$\text{Utterance} = \max(\cos(\text{Sent}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, n), \mathbf{RL})_{n=1, \dots, n!}) \quad (11)$$

where n iterates through each possible ordering of the three words (there are $n!$ possible permutations of n items) and selects the ordering that is most similar to previous experience (by taking the cosine between each sentence and the retrieved linguistic vector). Once an utterance was generated, the other agents in the context formed a memory trace by concatenating the utterance and the reference vector and storing this in memory.

Thus, the construction of linguistic structure was entirely social: Agents were likely to generate utterances that were similar to what other members of the community had previously generated, given the structure of the current environment. This was an attempt to formalize aspects of the claims made by usage-based theories ([Tomasello, 2003](#)) and the distributed approach to language ([Dale, 2012](#)) by basing the communication method entirely on the social environment.

Simulation Details

The language consisted of 15 different semantic categories, which could correspond to different classes of objects, actions, and so forth. There were three words in each category, yielding a total lexicon size of 45 words. Each word also had a single perceptual referent. Both words and referents were represented with a large, sparse vector (equivalent to the environmental vectors used in the previous simulations). A vector dimensionality of 5,000 was used, with five nonzero values being contained in each vector. From these different categories, 135 different “scenes” were created by randomly selecting an item from three randomly selected categories. Thus, each word occurred on average in three different sentences, introducing ambiguity into the language. The α parameter was set at the default level of 3 for this simulation.

Communication was accomplished by selecting three random agents: One agent was designated as the communicator and the other two as listeners. A random scene was selected for the communicator to express. The task of the communicator was to generate an utterance describing that scene (which was composed of three items selected from three different categories), accomplished by cuing the memory store with the **R** vector (the sum of the items in the scene) to retrieve an utterance. All agents then stored the produced utterance (with its word ordering encoded with random permutations) along with the perceptual scene that it co-occurred with. It has been shown that the sharing of a visual experience does have a significant impact on the knowledge that is gained (e.g., Richardson & Dale, 2005; Richardson, Dale, & Tomlinson, 2009), so this type of learning mechanism is not without evidence.

Because this was an iterative-learning paradigm, the process occurred across generations. Each generation contained 30 agents, with 2,000 communications per generation. A new generation was introduced after the previous generation, which received 2,000 utterances from the previous generation. Members of the new generation then communicated amongst themselves (with 2,000 utterances), and the whole process iterated. Members of the first generation of agents were each primed with 100 sentence exemplars (randomly ordered) so that they were initially constructing random utterances. How well agents were able to converge on a common communication pattern was assessed across multiple tests.

Simulation Results

Our first analysis was conducted to simply determine the probability that agents would generate the same ordering given the same perceptual input. This phenomenon was tested by randomly selecting one scene and two different agents. The probability that the agents would generate the same ordering was assessed across 1,000 randomly selected agent pairs and scenes. The result of this simulation are displayed in the top panel in Figure 6, which shows the increase in the probability of agents generating the same word ordering across 10 generations. This figure shows a substantial increase in the structure of the utterances that the agents created. Before interaction between agents took place (Generation 0), the agents generated the same ordering only 18% of the time—approximately at chance. However, by Generation 10, the same ordering was generated 93% of the time. Thus, a substantial improvement across time was observed.

A second test was performed to determine what effect the increased structure of the language had on the predictions that the agents formed. As the language became more structured, it would necessarily require the predictions that the agents formed to become more accurate. However, determination of the increase in prediction accuracy across generations allows for an examination of the dynamics between the structure of the usage of words by a communicator and the increase in comprehension efficiency by a listener. To accomplish this, the same method of sentence prediction that was used in Part 2 was used, with the average EV for the words in an utterance measured by having one agent generate a word ordering and then calculating the resulting EVs from a second agent. The middle panel of Figure 6 displays the increase

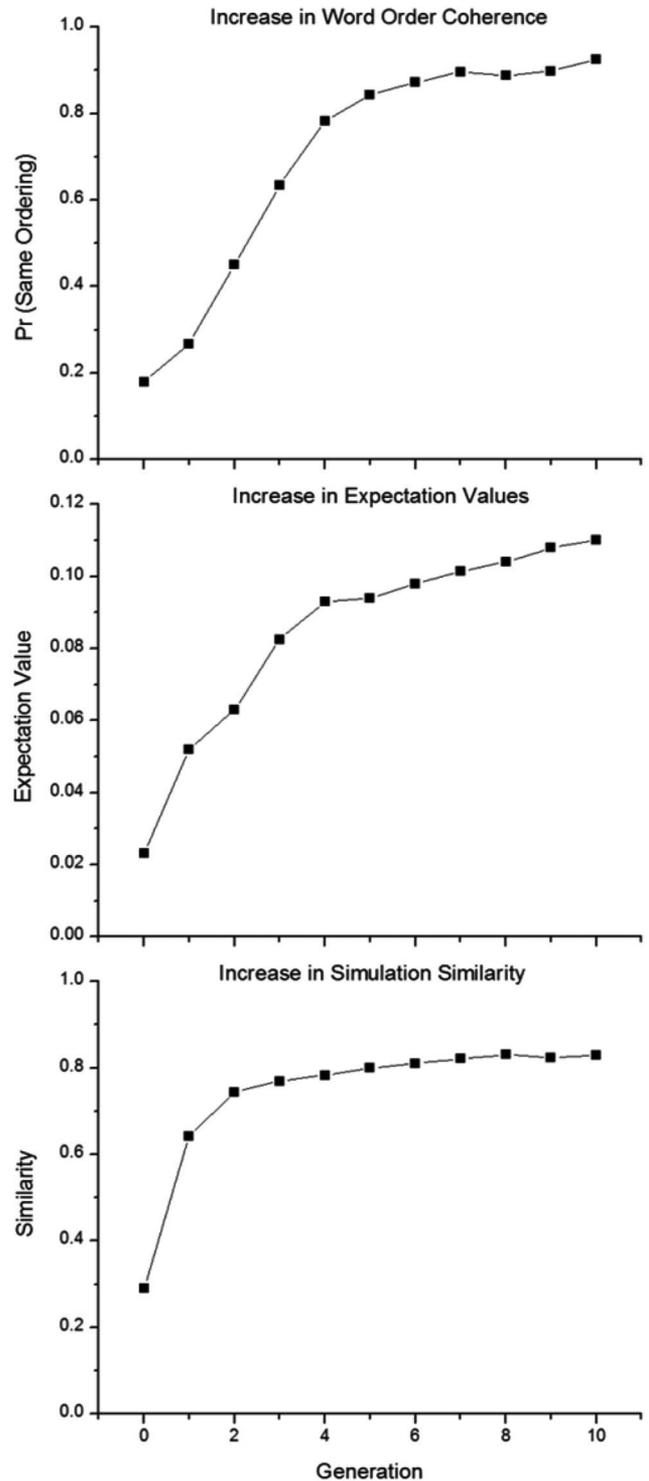


Figure 6. Increase in language structure across generations. The top panel displays the probability (Pr) of two agents generating the same ordering for a sentence. The middle panel shows the increase in expectation values that agents formed. The bottom panel displays the increase in the similarity of the perceptual simulations that agents constructed.

in average EVs across generations. A large increase in EVs was found. From Generation 0 to Generation 10, an approximate increase of 470% was seen in average EVs. This illustrates that as the structure of the language increased, a corresponding increase in the processing efficiency of the language was found.

A third test was conducted to determine how well the language was able to generate the same meaning in response to an utterance (in the form of a perceptual simulation) across different agents. To examine this pattern, three different agents were selected: The first agent was cued with a scene, and the utterance was used to generate a perceptual simulation in the other two agents. By assessing the similarity between the two resulting vectors, it was possible to demonstrate how capable the utterance was of conveying a semantic message. The increase in average similarity of perceptual simulations across generations is displayed in the bottom panel of Figure 6. This demonstrates that as the language became more structured, a corresponding increase in the ability of the language to convey the correct message (in terms of an agent's ability to construct a correct perceptual simulation) was also observed. There was a large increase in simulation similarity from Generation 0 to Generation 1, demonstrating that with even a slight increase in the structure of the language, a large increase in the communication efficiency was seen.

Another way of visualizing the increasing structure in the language is to plot the average word ordering selected across all agents for each sentence across 10 generations. To ease visualization, this was done with the first 100 sentences. The results are displayed in Figure 7 and show that across time, the word ordering for the different sentences emerged on the basis of the joint linguistic and perceptual experiences that agents received. It was not perfect, however; some of the sentences did not converge on a common ordering, but the majority of them did.

Discussion

This simulation demonstrated a simple point: that structure in language can be culturally constructed by storing the utterances

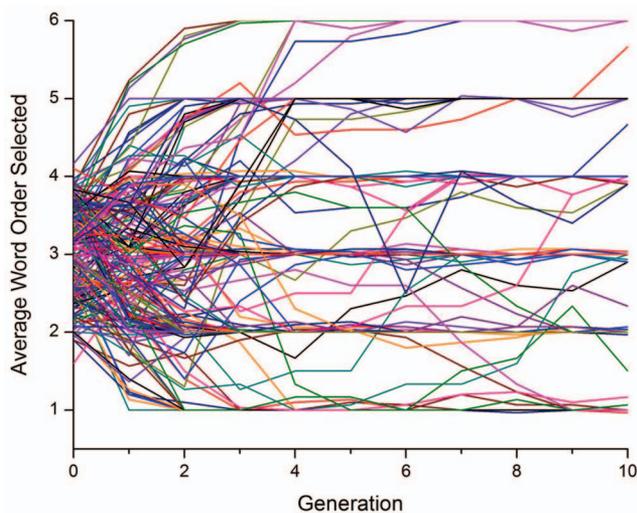


Figure 7. Convergence of word order for 100 sentences across generations for all agents in the simulation. See the online article for the color version of this figure.

that other members of one's community generate. When an agent needed to construct an utterance (given an environmental context), the order in which the words were strung together was dependent on the structure of the utterances and environments that the agent had previously experienced. Across generations, it was shown that a set of agents was able to converge on a fairly stable set of word orderings across a variety of different sentences. As in any simulation of a very complex topic such as the cultural evolution of language, a certain level of simplification was necessary, but the simulation reported here constitutes an existence proof that the joint memory of agents is a plausible mechanism that could have been used in the construction of a communication system.

General Discussion

This study outlined and tested a simple exemplar-based model of language. The model is based on the storage of exemplars of real language in memory (heavily inspired by models of semantic memory) and usage of this memory store to retrieve the expected future structure of a sentence through the use of an instance model. Specifically, we examined the possible role of exemplar memory in language processing. Part 1 described the basis of an expectation-generation mechanism in sentence processing. In Part 2, it was shown that this model could account for results in natural language sentence processing. Part 3 demonstrated how perceptual information could be built into the model using the framework provided by the interconnected trace hypothesis (Zwaan & Madden, 2005). Finally, in Part 4, a multiagent iterative-learning simulation was conducted, revealing that the same model could be used to impose structure on an initially random language, offering an existence proof that memory may have played a role in the cultural evolution of language.

Unlike the majority of language theories, this approach is not concerned with learning the rules of a language. Instead, the predicted structure of the current language environment is generated on the basis of the previous experiences one has had with language, which allows for a prediction of the structure of a current sentence to be formed. This entails that structure in language is not just based on rules and abstractions of the language input but different communication patterns used to express different types of thoughts. By storing these patterns in memory, one can create surprisingly sophisticated expectations about the forthcoming patterns in language.

An influential general theory of memory and language is specified by the complementary learning systems approach (McClelland, McNaughton, & O'Reilly, 1995; O'Reilly, Bhattacharyya, Howard, & Katz, 2013), which specifies that the brain uses two pathways to learning: one to store individualized episodic traces and one to learn latent structure across episodes. The majority of computational models of semantics, such as latent semantic analysis (Landauer & Dumais, 1997) and BEAGLE (Jones & Mewhort, 2007), also make the assumption that learning of language information requires the extraction of latent information across experience into a single representation of a word's meaning. However, the approach taken here (and earlier by Kwantes, 2005) specifies that this dichotomy between individual episodes and latent information is unnecessary. Instead, an exemplar approach proposes that abstracted information can be extracted dynamically on the basis of cues provided by the current environment. That is,

the model is capable of accounting for both aspects of the complementary learning systems approach within a single system: individual episodes are stored, and latent information is extracted from these experiences at retrieval. More research is needed to determine exactly how powerful this framework is at providing an integrated account of memory and language, but the preliminary results are promising.

Similarly, most current theories of language assume that linguistic processing is based on abstraction of linguistic input, designed to create a higher level representation of the workings of a language. However, the model described here casts doubt on whether abstraction is a general requirement in language learning, adding to an already growing literature, such as findings on the importance of item-level information in the development of language (Bannard et al., 2009; Tomasello, 2000, 2003) and recent developments in situational, grounded, and embodied theories of cognition (Barsalou, 1999, 2008; Zwaan & Madden, 2005; Zwaan & Radvansky, 1998). However, we do not want to make the claim that language does not make any abstractions. Instead, it is simply proposed that our raw experience with language, and the surrounding perceptual and social environment, provides a powerful basis for more advanced language mechanisms to operate. Undoubtedly, a complete model of language will require more than the pure instances of a language, such as the inclusion of higher level information.

A possible mechanism that could be used to add higher level information was given by Dennis (2004, 2005), who demonstrated how propositional information can be extracted from text using instance memory of sentences. This method requires much more online computation of different instances (by aligning different instances stored in memory with the current sentence through string-edit theory) across a smaller number of exemplars, whereas the model described here uses passive activation of a large number of exemplars. This difference could be bridged by only extracting proposition-like information from the current sentence, storing that within the exemplar, and then probing memory with both the linear-order and propositional information. This would likely retrieve a richer representation of the sentence from memory.

As shown in Part 3, one major advantage of the exemplar memory approach is the ease with which perceptual information can be integrated with the processing of linguistic structure. The model is an attempt to formalize processes that are similar in nature to the interconnected trace hypothesis of Zwaan and Madden (2005) and the PSS approach (Barsalou, 1999). In this approach, traces of linguistic experience are attached to their perceptual correlates in memory and stored in a referential trace. By retrieving similar linguistic traces from memory (accomplished by cuing with the curing linguistic environment), a perceptual simulation can be generated by activating the corresponding referent traces. Clearly, the simple perceptual representations that we used in the simulations in Part 3 were insufficient to create a complete formal explanation of this rather complex cognitive operation, but we achieved a first step toward this goal. The biggest challenge may be in creating representations for verbs, which will likely require some functional equivalent of embodied action. This is a concern for all models of sentence processing; because it has been fairly conclusively demonstrated that embodied action plays an important role in comprehension (Pulvermüller,

2005), verb integration is a definite challenge for computational approaches to language processing.

Here, we have described a new approach to language processing based on exemplar memory theory. The model is based on the premise that the structure of our experiences can be used to generate the expected structure of some linguistic input. We have shown that this type of process can account for a variety of different behavioral results and could also have played a role in the cultural evolution of language. This approach is promising because of both its use of real language and its simplicity, and it also displays the usefulness of examining the value of the storage and retrieval of individual experiences with language.

Résumé

Les théories du langage courantes présument généralement qu'une certaine abstraction des entrées linguistiques est nécessaire pour créer des représentations plus poussées des structures linguistiques (par ex., la grammaire). Toutefois, l'importance des expériences individuelles à l'égard du langage a récemment été mise en relief, à la fois par les théories fondées sur l'usage (Tomasello, 2003) et les théories basées sur des données empiriques et situationnelles (par ex., Zwaan & Madden, 2005). Après l'approche basée sur l'expérience, nous présentons un modèle qui réunit des exemples de phrases dans un corpus de langage naturel, en appliquant de récentes avancées tirées de modèles de mémoire sémantique. Dans ce modèle, la mémoire d'exemplaires est utilisée pour générer des attentes au sujet de la structure éventuelle de phrases, en utilisant un mécanisme pour prédire le traitement du langage (Altmann & Mirkovic, 2009). Le modèle réussit à cerner une vaste gamme d'effets sur le comportement — le traitement réduit de la subordonnée relative (Reali & Christiansen, 2007), le rôle de la contrainte contextuelle (Rayner & Well, 1996) et l'activation de la connaissance d'événements (Ferretti, Kutas & McRae, 2007), entre autres. Ensuite, nous démontrons de quelle façon la connaissance perceptive peut être intégrée dans ce cadre fondé sur l'exemple, avec le but d'ancrer le traitement du langage dans la perception. Enfin, nous illustrons en quoi un système mnésique d'exemplaires aurait pu être utilisé dans l'évolution culturelle de la langue. Ce modèle fournit la preuve qu'une grande portion du traitement du langage pourrait être ascendant, et reposer sur le stockage et la récupération d'expériences linguistiques individuelles.

Mots-clés : traitement du langage, mémoire d'exemplaires, traitement des phrases, mémoire sémantique, cognition ancrée.

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. <http://dx.doi.org/10.1515/TLR.2006.011>
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264. [http://dx.doi.org/10.1016/S0010-0277\(99\)00059-1](http://dx.doi.org/10.1016/S0010-0277(99)00059-1)
- Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements in linguistic processing. *Journal of Memory and Language*, 57, 502–518. <http://dx.doi.org/10.1016/j.jml.2006.12.004>

- Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*, 583–609. <http://dx.doi.org/10.1111/j.1551-6709.2009.01022.x>
- Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*, 463–498. <http://dx.doi.org/10.1037/a0016261>
- Bannard, C., Lieven, E., & Tomasello, M. (2009). Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 17284–17289. <http://dx.doi.org/10.1073/pnas.0905638106>
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–609.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. <http://dx.doi.org/10.1146/annurev.psych.59.103006.093639>
- Barsalou, L. W., Simmons, W. K., Barbey, A., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*, 84–91.
- 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. <http://dx.doi.org/10.1016/j.jml.2010.08.004>
- Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*, 489–508. <http://dx.doi.org/10.1017/S0140525X08004998>
- Connell, L., & Lynott, D. (2011). Modality switching costs emerge in concept creation as well as retrieval. *Cognitive Science*, *35*, 763–778. <http://dx.doi.org/10.1111/j.1551-6709.2010.01168.x>
- Dale, R. (2012). Integrating and extending the distributed approach in cognitive science. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, *13*, 125–138. <http://dx.doi.org/10.1075/is.13.1.08dal>
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1121. <http://dx.doi.org/10.1038/nn1504>
- Dennis, S. (2004). An unsupervised method for the extraction of propositional information from text. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(Suppl. 1), 5206–5213. <http://dx.doi.org/10.1073/pnas.0307758101>
- Dennis, S. (2005). A memory-based theory of verbal cognition. *Cognitive Science*, *29*, 145–193. http://dx.doi.org/10.1207/s15516709cog0000_9
- Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning & Verbal Behavior*, *18*, 1–20.
- Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33*, 547–582. <http://dx.doi.org/10.1111/j.1551-6709.2009.01023.x>
- Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*, 429–448. <http://dx.doi.org/10.1017/S0140525X0999094X>
- Fagot, J., & Cook, R. G. (2006). Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 17564–17567. <http://dx.doi.org/10.1073/pnas.0605184103>
- 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. <http://dx.doi.org/10.1037/0278-7393.33.1.182>
- Garcea, F. E., Dombrov, M., & Mahon, B. Z. (2013). Preserved tool knowledge in the context of impaired action knowledge: Implications for models of semantic memory. *Frontiers in Human Neuroscience*, *7*, 120. <http://dx.doi.org/10.3389/fnhum.2013.00120>
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279. <http://dx.doi.org/10.1037/0033-295X.105.2.251>
- Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, *31*, 441–480. <http://dx.doi.org/10.1080/15326900701326576>
- 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. <http://dx.doi.org/10.1111/j.1551-6709.2009.01027.x>
- Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition*, *111*, 151–167. <http://dx.doi.org/10.1016/j.cognition.2009.01.009>
- Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, *93*, 411–428. <http://dx.doi.org/10.1037/0033-295X.93.4.411>
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528–551. <http://dx.doi.org/10.1037/0033-295X.95.4.528>
- Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative clause sentences. *Journal of Verbal Learning & Verbal Behavior*, *20*, 417–430. [http://dx.doi.org/10.1016/S0022-5371\(81\)90533-8](http://dx.doi.org/10.1016/S0022-5371(81)90533-8)
- Huetig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*, 151–171. <http://dx.doi.org/10.1016/j.actpsy.2010.11.003>
- Jamieson, R. K., Crump, M. J. C., & Hannah, S. D. (2012). An instance theory of associative learning. *Learning & Behavior*, *40*, 61–82. <http://dx.doi.org/10.3758/s13420-011-0046-2>
- Jamieson, R. K., Hannah, S. D., & Crump, M. J. C. (2010). A memory-based account of retrospective reevaluation. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *64*, 153–164. <http://dx.doi.org/10.1037/a0020886>
- 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. <http://dx.doi.org/10.1080/17470210802055749>
- 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. <http://dx.doi.org/10.1080/17470210802557637>
- Jamieson, R. K., & Mewhort, D. J. K. (2010). Applying an exemplar model to the artificial-grammar task: String completion and performance on individual items. *Quarterly Journal of Experimental Psychology*, *63*, 1014–1039. <http://dx.doi.org/10.1080/17470210903267417>
- Jamieson, R. K., & Mewhort, D. J. K. (2011). Grammaticality is inferred from global similarity: A reply to Kinder (2010). *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *64*, 209–216. <http://dx.doi.org/10.1080/17470218.2010.537932>
- Johns, B. T., Gruenenfelder, T. M., Pisoni, D. B., & Jones, M. N. (2012). Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *Journal of the Acoustical Society of America*, *132*, EL74–EL80. <http://dx.doi.org/10.1121/1.4731641>
- Johns, B. T., & Jones, M. N. (2012). Perceptual inference through global lexical similarity. *Topics in Cognitive Science*, *4*, 103–120. <http://dx.doi.org/10.1111/j.1756-8765.2011.01176.x>
- Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *66*, 115–124. <http://dx.doi.org/10.1037/a0026727>

- Jones, M. N., Kintsch, W., & Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, 55, 534–552. <http://dx.doi.org/10.1016/j.jml.2006.07.003>
- Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114, 1–37. <http://dx.doi.org/10.1037/0033-295X.114.1.1>
- Kamide, Y., Altmann, G. T. M., & Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133–156. [http://dx.doi.org/10.1016/S0749-596X\(03\)00023-8](http://dx.doi.org/10.1016/S0749-596X(03)00023-8)
- Kanerva, P. (1988). *Sparse distributed memory*. Cambridge, MA: MIT Press.
- Kanerva, P. (2009). Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors. *Cognitive Computation*, 1, 139–159. <http://dx.doi.org/10.1007/s12559-009-9009-8>
- Kart-Teke, E., De Souza Silva, M. A., Huston, J. P., & Dere, E. (2006). Wistar rats show episodic-like memory for unique experiences. *Neurobiology of Learning and Memory*, 85, 173–182. <http://dx.doi.org/10.1016/j.nlm.2005.10.002>
- Kirby, S. (2002). Natural language from artificial life. *Artificial Life*, 8, 185–215. <http://dx.doi.org/10.1162/106454602320184248>
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44. <http://dx.doi.org/10.1037/0033-295X.99.1.22>
- Kwantes, P. J. (2005). Using context to build semantics. *Psychonomic Bulletin & Review*, 12, 703–710. <http://dx.doi.org/10.3758/BF03196761>
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–240. <http://dx.doi.org/10.1037/0033-295X.104.2.211>
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126–1177. <http://dx.doi.org/10.1016/j.cognition.2007.05.006>
- Lieven, E. V. M., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24, 187–219. <http://dx.doi.org/10.1017/S0305000996002930>
- Logan, G. D. (1988). Towards an instance theory of task automatization. *Psychological Review*, 95, 492–527. <http://dx.doi.org/10.1037/0033-295X.95.4.492>
- Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, 109, 376–400. <http://dx.doi.org/10.1037/0033-295X.109.2.376>
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419–457.
- McKoon, G., & Macfarland, T. (2002). Event templates in the lexical representations of verbs. *Cognitive Psychology*, 45, 1–44. [http://dx.doi.org/10.1016/S0010-0285\(02\)00004-X](http://dx.doi.org/10.1016/S0010-0285(02)00004-X)
- McKoon, G., & Ratcliff, R. (2003). Meaning through syntax: Language comprehension and the reduced relative clause construction. *Psychological Review*, 110, 490–525. <http://dx.doi.org/10.1037/0033-295X.110.3.490>
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37, 547–559. <http://dx.doi.org/10.3758/BF03192726>
- McRae, K., & Jones, M. (2013). Semantic memory. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 206–219). New York: Oxford University Press.
- McRae, K., & Matsuki, K. (2013). Constraint-based models of sentence processing. In R. Van Gompel (Ed.), *Sentence processing* (pp. 51–77). New York, NY: Psychology Press.
- Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609–626. <http://dx.doi.org/10.1037/0033-295X.89.6.609>
- Murdock, B. B. (1995). Developing TODAM: Three models for serial-order information. *Memory & Cognition*, 23, 631–645. <http://dx.doi.org/10.3758/BF03197264>
- Myung, J. Y., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: Manipulation knowledge of objects. *Cognition*, 98, 223–243. <http://dx.doi.org/10.1016/j.cognition.2004.11.010>
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, 115, 39–57. <http://dx.doi.org/10.1037/0096-3445.115.1.39>
- O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2014). Complementary learning systems. *Cognitive Science*, 38, 1229–1248.
- Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, 37, 607–642. <http://dx.doi.org/10.1017/S0305000910000012>
- Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review*, 15, 161–167. <http://dx.doi.org/10.3758/PBR.15.1.161>
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6, 576–582. <http://dx.doi.org/10.1038/nrn1706>
- Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3, 504–509. <http://dx.doi.org/10.3758/BF03214555>
- 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. <http://dx.doi.org/10.1016/j.jml.2006.08.014>
- Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 88–94.
- Recchia, G. L., Jones, M. N., Sahlgren, M., & Kanerva, P. (2010). Encoding sequential information in vector space models of semantics: Comparing holographic reduced representation and random permutation. In S. Ohissou & R. Catrambone (Eds.), *Cognition in flux: Proceedings of the 32nd Annual Cognitive Science Society* (pp. 865–870). Austin, TX: Cognitive Science Society.
- Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, 29, 1045–1060. http://dx.doi.org/10.1207/s15516709cog0000_29
- Richardson, D. C., Dale, R., & Tomlinson, J. M. (2009). Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science*, 33, 1468–1482. <http://dx.doi.org/10.1111/j.1551-6709.2009.01057.x>
- Sahlgren, M., Holst, A., & Kanerva, M. (2008). *Permutations as a means to encode word order in word space*. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1300–1305). Austin, TX: Cognitive Science Society.
- Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience*, 8, 657–661. <http://dx.doi.org/10.1038/nrn2213>
- Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology*, 49, 289–318. <http://dx.doi.org/10.1146/annurev.psych.49.1.289>
- Schwaneflugel, P. J. (1986). Completion norms for final words of sentences using a multiple production measure. *Behavior Research Methods, Instruments, & Computers*, 18, 363–371. <http://dx.doi.org/10.3758/BF03204419>

- Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.
- Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, 9, 371–386. <http://dx.doi.org/10.1162/106454603322694825>
- Stassenko, A., Garcea, F. E., & Mahon, B. Z. (2013). What happens to the motor theory of perception when the motor system is damaged? *Language and Cognition*, 5, 225–238. <http://dx.doi.org/10.1515/langcog-2013-0016>
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995, June 16). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634. <http://dx.doi.org/10.1126/science.7777863>
- Thiessen, E. D., & Pavlik, P. I., Jr. (2013). iMinerva: A mathematical model of distributional statistical learning. *Cognitive Science*, 37, 310–343.
- Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences*, 4, 156–163. [http://dx.doi.org/10.1016/S1364-6613\(00\)01462-5](http://dx.doi.org/10.1016/S1364-6613(00)01462-5)
- 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. <http://dx.doi.org/10.1006/jmla.2001.2836>
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 443–467. <http://dx.doi.org/10.1037/0278-7393.31.3.443>
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Wells, J. B., 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. <http://dx.doi.org/10.1016/j.cogpsych.2008.08.002>
- Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). New York: Macmillan.
- Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, 56, 165–209. <http://dx.doi.org/10.1016/j.cogpsych.2007.04.002>
- Zwaan, R. A. (2004). The immersed experienter: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 33, pp. 35–62). New York: Academic Press.
- Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In D. Pecher & R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 224–245). New York: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511499968.010>
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162–185. <http://dx.doi.org/10.1037/0033-2909.123.2.162>

Received November 25, 2014

Accepted March 13, 2015 ■

E-Mail Notification of Your Latest CPA Issue Online!

Would you like to know when the next issue of your favorite Canadian Psychological Association journal will be available online? This service is now available. Sign up at <http://notify.apa.org/> and you will be notified by e-mail when issues of interest to you become available!

Avis par courriel de la disponibilité des revues de la SCP en ligne!

Vous voulez savoir quand sera accessible en ligne le prochain numéro de votre revue de la Société canadienne de psychologie préférée? Il est désormais possible de le faire. Inscrivez-vous à <http://notify.apa.org/> et vous serez avisé par courriel de la date de parution en ligne des numéros qui vous intéressent!