

THE EXPRESSION OF COGNITIVE CATEGORIES



THE EXPRESSION OF TIME

EDITED BY WOLFGANG KLEIN AND PING LI

MOUTON DE GRUYTER

Computational modeling of the expression of time

Ping Li and Xiaowei Zhao

1. Introduction

To describe and exchange information about time is an important part of human activities. The expression of time is therefore one of the central conceptual domains of our language use (see chapter 1 of this book and Bates, Elman & Li 1994). When we are talking, we describe situations as being in the past, present or in the future, and we talk about events as ongoing or completed. There are two key linguistic categories for expressing temporal concepts in the world's languages: tense and aspect. Tense is often concerned with the chronological ordering of situations that happen at different time points, and is often used to locate the relationship between time of event and time of speech. In contrast, aspect typically characterizes how a speaker views the temporal contour of a situation described, for example, the beginning, continuation, or completion of a situation.

As temporal contours and relationships of events figure prominently in people's speech activities, it is important for any given human language to have a capable system for expressing these events and relationships, and for speakers/listeners to learn and process this system. Empirical evidence appears to support the idea that tense and aspect are among the earliest linguistic devices acquired by children, and as such the scientific study of the expression of time provides significant insights into not only how young children acquire temporal notions, but also what psycholinguistic mechanisms underlie the general acquisition processes.

In this chapter, we review computational models of the expression and acquisition of temporal concepts in language. With the advancement of modern computers in the last decades, computational modeling has become a very powerful methodology in many disciplines, including cognitive science and psycholinguistics. With respect to our focus here, computational models can help us introduce explicit, controllable and testable mechanisms to understand the linguistic phenomena related to temporal expressions. In addition, computational models often include certain levels of simplification in terms of language details, which makes them easier to study than traditional empirical studies that are often costly. The models and simplified

datasets both make it possible for us to examine the underlying mechanisms of temporal concepts without getting entangled by specific details or certain noise in the linguistic input.

The computational modeling of tense has attracted much attention in psycholinguistics and cognitive science since Rumelhart and McClelland (1986) introduced a simple feed-forward neural network model¹ (the R&M model) to account for children's acquisition of English past tense. Given the prominence of their model and the subsequent debates in the field, we will only provide a very brief review on the acquisition of the English past tense. The primary focus in our discussion will be on various computational models that account for the expression and acquisition of aspect. Of course one needs to realize that the expressions of tense and aspect are often closely correlated in many languages; for example, the English past tense marker *-ed* marks both the past tense and the perfective aspect (Comire 1976). Thus, in our discussion we often need to speak of the acquisition of tense and aspect together.

¹ A neural network model is a computational model made of information processing units (neurons) that are connected in a network. The construction and learning of neural network models are often based on considerations of neural information processing. Different from traditional digital computers, the computation in a neural network is based on the connection change among the parallel working units, which has made it a great success in many scientific disciplines during last two decades, such as cognitive science, linguistics, psychology, to name a few. Neural network modeling is also called connectionism or PDP (parallel distributed processing). It views knowledge representation and acquisition as distributed, parallel and interactive in nature. First, a given concept is represented not by a single unit or node but by multiple units or nodes in concert, the result of which is a pattern of activation of relevant micro-features that distribute across multiple units. Second, in terms of knowledge acquisition, connectionism argues for learning through the adaptation of weights, the strengths of connections that hold between multiple and parallel working units, which can also serve as a simplification of the synaptic connections among real neurons. There have been various algorithms developed for adjusting the weights to an optimal set of configurations, which may lead to the appropriate activation patterns of units that represent new knowledge. Third, the interactive activities among multiple units and the learning environment play very important roles in the information processing. For example, PDP argues that linguistic representations can be best understood as the properties that emerge out of learning (i.e., 'emergent properties') rather than built in *a priori*, owing to the interaction of the learning system with the linguistic environment.

1.1. Computational models of tense

The English past tense includes both regular (e.g. *work-worked*) and irregular forms (e.g. *go-went*). When children learn these forms, they sometimes make “over-generalized” error such as producing *goed* as the past tense of *go*, and *breaked* as the past tense of *break*. In addition, studies have shown a “U-shape” trajectory in children’s learning of irregular past tenses. At first, they seem to have mastered only a few but correct inflectional forms of verbs; then they forget the correct forms and make many “overgeneralized” errors; finally, children grasp the usage of irregular past tense forms as well as thousands of other regular verbs. (Berko 1958; Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett 1996; Marcus, Ullman, Pinker, Hollander, Rosen, & Xu 1992). These phenomena have been traditionally interpreted by some investigators as indicating two totally separate mechanisms for children to learn verb past tenses: one dictionary-like rote mapping for irregular verbs, and the other the explicit representation of a rule of adding suffix *-ed* to regular verbs (Pinker 1994). According to this “dual mechanisms” theory, at first, children learn past tenses only by rote learning, and then almost suddenly, they discover the existence of the rule controlling the formation pattern of regular past tenses; they then apply this rule to any new verb, including irregular verbs, thus causing the so-called “overgeneralized” errors. In the end, children realize that there are exceptions in English past tenses, and then correct their errors and produce the right irregular forms. This type of “dual mechanisms” theory fits well with the general assumption about language as a kind of general symbolic machinery that Chomsky and his followers have advocated.

The symbolic theory of language dominated the psycholinguistic view of morphology and its acquisition for a long time. In 1986, Rumelhart and McClelland (R&M) introduced a simple feed-forward neural network that clearly shows that overgeneralizations and “U-shaped” learning of English past tense may be due to a single mechanism based on neurobiologically plausible features, without the need of two explicit and distinct mechanisms. Basically, the R&M model is a model of pattern associator that can make the strong connection between a verb stem and its phonological form of past tense. There has been much computational modeling work inspired by R&M’s model as well as heated debate on what drives the learning of the English past tense.

The R&M model was criticized by Pinker and his colleagues (Pinker & Prince 1988; Marcus et al. 1992) on grounds that the model used unrealistic input and training schedules and the model was unable to capture subtle

error patterns in realistic speech. In response to these criticisms, other neural network models equipped with “hidden units” and the “back propagation” learning algorithm have been successfully applied to simulate the acquisition of the English past tense and overcome the shortcomings of the R&M model, such as the model introduced by Plunkett and Marchman (1991) and that by MacWhinney and Leinbach (1991). On the other hand, Ling and Marinov (1993) provided a symbolic pattern associator (SPA) in support of the “dual mechanisms” assumption. The authors claimed that their model outperformed both the R&M model and the MacWhinney & Leinbach (M&L) model when compared with the real data extracted from human subjects. A further detailed comparison of SPA and M&L models (MacWhinney 1993), however, showed that actually the M&L model performed as well as the SPA model in all the past-tense learning tasks; in addition, there were two artificial parameters, which did not have much empirical evidence, that played extremely important roles in the emergence of the U-shaped curve in the SPA model.

In short, in the last two decades, there was a great deal of interest in the computational modeling of the acquisition of tense, focusing on the capacity of neural network models of the English past tense acquisition. The center of the debate was whether the acquisition of grammar can be viewed as the acquisition of symbolic rule systems (in the views of the “dual mechanisms” theory) or whether it can be treated as a statistical learning process (in the views of connectionist theory). The debate is far from being resolved, but the reader is encouraged to consult Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett (1996) for integrative discussions.

1.2. Computational Models of Aspect

For the remainder of this chapter, we will focus on computational models of the expression and acquisition of aspect, another important temporal concept in languages. Although in the computational linguistics literature, a few computational models have been applied to study aspect categories (as well as tense categories) and analyze the temporal relationships between clauses in terms of event time, speech time, and reference time (Passonneau 1988), computational models of the expression and acquisition of aspect fell far shorter compared with those of the acquisition of tense. Given that previous studies often focused on how to classify verbs into appropriate aspect classes according to the relevant linguistic features and contexts in order to reason about time, we will first discuss here different aspect categorization theories.

1.2.1. Two Kinds of Aspect

Linguists generally distinguish between two kinds of aspect, grammatical aspect and lexical aspect (under various names; see chapter 2, sections 3 and 4, and Li & Shirai 2000, for reviews). Grammatical aspect is related to aspectual distinctions which are often marked explicitly by linguistic devices, such as the inflectional suffixes and auxiliaries in English. It is also known as the viewpoint aspect (Smith 1997) which refers to a particular viewpoint toward the situation being talked about. According to Comrie (1976), there are two major categories of grammatical aspect: imperfective and perfective. Imperfective aspect presents a situation with an internal point of view, often as ongoing (progressive) or enduring (continuous), whereas perfective aspect presents a situation with an external perspective, often as completed. In English, the imperfective-perfective contrast is realized in the difference between the progressive *be V-ing* and the past-perfective *-ed*.

Lexical aspect, on the other hand, refers to the characteristics inherent in the temporal meanings of a verb, for example, whether the verb encodes an inherent end point of a situation, or whether the verb is inherently stative (i.e., continuous and homogeneous) or punctual (i.e., momentary and instantaneous). Most researchers adopt Vendler's (1957) classification as the standard treatment of inherent semantics of verbs, which involves four categories: activities, accomplishments, achievements, and states. A new category has been lately added, which is the so-called "point activities" (Moens & Steedman 1988) or "semelfactives" (Smith 1997). Activity verbs like *walk*, *run* and *swim* encode situations as consisting of successive phases over time with no inherent end point. Accomplishment verbs like *build a house* also characterize situations as having successive phases, but unlike activities they encode an inherent endpoint (e.g., house-building has a terminal point and a result). Like accomplishments, achievement verbs also encode a natural endpoint, but unlike accomplishments and activities they encode events as punctual and instantaneous, that is, as having no duration, such as in *fall*, *recognize a friend* and *cross the border*. State verbs encode situations as homogeneous, with no successive phase or endpoints, involving no dynamicity, such as *know*, *want* and *love*. Finally, the semelfactive verbs involve dynamicity, and encode instantaneous events, but these verbs do not have an inherent end point, like *cough* or *hiccup* in English. In addition, on the basis of whether the verb encodes endpoints, linguists also call activity, state, and semelfactive verbs "atelic" (no endpoint), and accomplishment and achievement verbs "telic" (with endpoint).

In English, grammatical aspect and lexical aspect often interact with each other in complex fashions. Uses of the inflectional suffixes, *-ing*, *-ed* and *-s* are in many cases constrained. For example, progressive aspect *-ing* does not occur often with state verbs; thus while “John knows the boy” is good, “John is knowing the boy” sounds odd (Smith 1983). There are also combinatorial constraints between *-ing* and event verbs; for example, “The book is falling off the shelf” is odd when used to refer to the actual falling down, but is good when used to mean a preliminary stage (i.e., prior to actual falling; Smith 1997). These kinds of constraints may reflect the intricate relationships between language use and characteristics of the described event. For example, as pointed out by Brown (1973), many events with an end result last for such a short period of time that any description of them is unlikely to occur during the period, such as the actions of *fall*, *drop*, and *break*. Thus it is rare for speakers to describe the “ongoing-ness” of such events with *-ing* but more natural for them to describe the “completeness” using past-perfective forms.

1.2.2. Computational models of aspectual classification

As we mentioned before, previous computational models about aspect often focused on aspectual classification. For example, in a study conducted by Bennett, Herlick, Hoyt, Liro and Santisteban (1989), using three aspectual features the authors introduced a five-way aspectual classification system to distinguish verbs into the five lexical aspects as we described in section 1.2.1. The three features are \pm dynamic, \pm atomic and \pm telic. Their feature-based descriptions of the five aspectual types are shown as follows (adapted from Bennett et al. 1989):

Accomplishment:	[+d +t -a]
Achievement:	[+d +t +a]
(Extended) Activity:	[+d -t -a]
(Point) Activity/Semelfactive:	[+d -t +a]
State:	[-d -t -a]

This classification method is consistent with Smith’s (1991) aspectual analysis except that Bennet et al. used the term “atomic” here to represent the “punctual” feature in Smith’s analysis. In addition, the authors further argued that not only the verb features but also some other sentential features (e.g. certain tenses, temporal adverbials) can affect the aspectual situation

of verbs in certain sentences. The authors discussed about eleven such sentential feature which can also operate on the [telic] and [atomic] features. For example, in the sentence *John ran*, the verb *ran* keeps its original aspect type of activity. But in some other sentences like *John ran a mile*, *John ran to the park*, and *John ran until 8 o'clock*, the noun phrase, the preposition phrase and the durative adverbial after the verb *ran* imply the endpoint of this action (+telic), and thus the lexical aspect of this verb changes to accomplishment in these sentences.

Bennett et al.'s work showed that aspectual classification is a complex process that depends on both verbs' intrinsic temporal properties and the syntactical features. However, these authors did not show how to use specific methods to realize these classifications computationally. A recent study conducted by Siegel and McKeown (2000) attempted to fill the gap. They declared that the co-occurrence frequencies between the verb and certain linguistic modifiers can reliably predict the verb's aspectual class. The authors generalized 14 so-called "linguistic indicators" (the co-occurrence frequency measures) as the basis to classify verbs into state vs. event verbs, and further into culminated (telic) vs. nonculminated (atelic) verbs in the event category. Specifically, they used three supervised machine/statistical learning methods to combine the 14 linguistic indicators for aspectual classification. The three methods are: logistic regression, decision trees, and genetic algorithm (GA).

Logistic regression is a multivariate statistical method that can derive an overall variate based on the weighted nonlinear combination of the 14 variables or the linguistic indicators. The overall variate can increase the classification performance of the model. The decision tree is a traditional data mining method based on many choice points. At each point, according to the value of a specific linguistic indicator, the system makes an *if-then-else* choice to decide which one of the two possible classes a verb should belong to. When facing a classification task for a verb, the method will start from the root of the decision tree, undergo a series of tests on choice points, and then end at a leaf (thus labelled by an aspectual type). The decision tree enables the complex interaction of different indicators in the system. Finally, genetic algorithm (GA) is a novel method in computer science based on the concept of Darwinian natural selection and survival of the fittest (Holland 1975). This method enables the generation and evolution of arbitrary mathematical combinations of the 14 linguistic indicators to classify the verb aspects. All the three methods have shown good performance on the two-way classification of lexical aspect (state vs. event; culminated vs. non-culminated).

Neural network models can also be used in aspectual classifications. Based on the idea that people should be able to extract aspectual features and meanings from syntactic representations, Scheler (1997) introduced a model that can extract the grammatical categories of English aspects (progressive vs. simple) and Russian aspects (imperfective vs. perfective). The model includes four main modules:

- (1) An automatic tagger that can provide syntactic tags to the words in the input text in the system, thus transferring the text to specialized syntactic representations;
- (2) A process that transforms the syntactic representation into the semantic representation, a syntactic-to-semantic pattern association task;
- (3) A set of semantic features describing aspectual meanings of verbs, such as event type, action status, and habituality;
- (4) A process that maps the individual aspectual meanings to the grammatical categories for each language, which is the final pattern classification task.

The modules (2) and (4) are the core parts of the model, and Scheler used two standard back-propagation neural networks with hidden layers to simulate the processes of the two modules. For module (2), the author constructed a network with an input layer of 25 binary neurons (which represent the 6 slots syntactic features), two hidden layers with 15 and 12 neurons respectively, and an output layer with 34 binary neurons representing the 15 semantic features. The network was trained to associate the syntactic patterns of verbs to their semantic representations. The network successfully learned 87 percents of the total syntactic-semantic pattern pairs, although the generalization performance was not as good as the learning. Based on the simulating results, the author argued that most of the information needed to extract semantic features for aspects is based on local syntactic features. For module (4), the author used a $34 \times 5 \times 2$ network to classify verbs into different grammatical aspectual categories according to their semantic features. The network has an input layer with 34 binary neurons, a hidden layer with 5 neurons, and an output layer with 2 neurons to represent the grammatical aspectual types. The network performed well in both learning and generalization of the patterns. This model was a first full-scale back-propagation connectionist model for aspectual classification. A problem with the model is that the two neural networks for modules (2) and (4) in the model were isolated and did not communicate with each other. In addition, the generalization ability of the syntactic-semantic association network

was not good. In section 2, we will introduce a self-organizing neural network model that can correct these problems.

The aspectual models discussed above are all about the expression or classification of aspectual categories. In contrast to this line of research, scholars have also been interested in how children acquire the classification of lexical aspect and how this acquisition influences their use of grammatical aspect, which brings us to the next section.

1.2.3. Aspect acquisition in child language

A core issue in the study of aspect acquisition is how children acquire the two kinds of aspect (grammatical aspect and lexical aspect, see section 1.2.1) and their interactions in different languages. It has been now well established that there is a strong association between lexical aspect and grammatical aspect in child language: children initially tend to restrict tense-aspect morphology to specific categories of lexical aspect. For example, English-speaking children initially tend to use progressive marker *-ing* only with atelic, activity verbs, whereas past-perfective marker *-ed* only with telic verbs (accomplishment and achievements) at an early stage of development (McShane & Whittaker 1988; Shirai & Andersen 1995). This restricted or “undergeneralized” pattern of use has led to intense debate with respect to various theoretical frameworks (see Li & Shirai 2000 for review). An early suggestion from Bickerton (1984) was that children have innate semantic categories that roughly correspond to the lexical aspect distinctions of verbs (e.g., punctual-nonpunctual, state-process distinctions), and these categories are biologically programmed as part of a Language Bioprogram. Bickerton relied on both data from creole languages and child language acquisition to support his proposal that children’s acquisition of tense-aspect morphology has a biological basis. Subsequent crosslinguistic studies, however, have provided counter evidence to this hypothesis (e.g., Li & Bowerman 1998; Shirai & Andersen 1995), and led researchers to propose a variety of input-driven hypotheses about how children acquire tense-aspect morphology and lexical semantics of verbs (see Li & Shirai 2000 for a review).

The goal of our computational modelling is to provide mechanistic accounts of how empirically observed patterns could emerge out of simple computational principles. In previous empirical studies (Li & Bowerman 1998; Li & Shirai 2000), we proposed that the initial lexical-morphological associations could arise as a result of the learner’s analyses of the verb-morphology co-occurrence probabilities in the input environment of the

language learner. In parental speech, there are probabilistic associations between progressive markers and atelic verbs, and between perfective markers and telic verbs (Shirai & Andersen 1995). Children's initial under-generalizations (restricted uses of morphology) might reflect their analyses of these probabilities. In the next section, we will discuss a neural network model which can analyze these probabilities and arrive at patterns that resemble children's patterns of acquisition.

2. A self-organizing neural network model of the acquisition of aspect

In the last few years we have explored self-organizing neural networks as candidates of cognitively and neurally plausible models of language acquisition (Li 2003, 2006; Li, Farkas & MacWhinney 2004; Li, Zhao & MacWhinney 2007; Zhao & Li 2005, 2007, 2008, in press). Compared to other developmental neural network models, most of which rely on supervised learning algorithms (e.g., back-propagation, see models reviewed in Elman et al. 1996), self-organizing neural networks, especially the so-called self-organizing maps (SOM), have several important properties that make them particularly well suited to the study of lexical and morphological acquisition (see Li 2003, 2006 for discussion).

First, these models belong to the class of unsupervised learning networks that require no explicit teacher; learning is achieved by the system's organization in response to the input. Such networks provide computationally more relevant models for language acquisition, given that in real language learning children do not receive constant feedback about what is incorrect in their speech (see Li 2003; MacWhinney 1998, 2001; Shultz 2003 for discussion). Second, self-organization in these networks allow for the gradual formation of structures on 2-D maps, as a result of extracting an efficient representation of the complex statistical regularities inherent in the high-dimension input space (Kohonen 2001). Third, the self-organizing map forms topography-preserving structures, which means nearby areas in the map respond to inputs with similar features. This property allows us to model the emergence of semantic categories as a gradual process of lexical learning. Finally, several self-organizing maps can be connected via Hebbian learning, a well-established biologically plausible learning principle, according to which the association strength between two neurons is increased if the neurons are both active at the same time (Hebb 1949). Although Hebbian learning itself is not an inherent property of the self-organizing algorithm, when incorporated, the SOM model would have strong implications for

language acquisition: it can account for the process of how the learner establishes relationships between word forms, lexical semantics, and grammatical morphology, on the basis of how often they co-occur and how strongly they are co-activated in the representation.

A number of recent models have taken advantage of the properties discussed above to examine language processing and language acquisition. These include DISLEX (Miikkulainen 1997), DevLex (Li, Farkas, & MacWhinney 2004), DevLex-II (Li, Zhao & MacWhinney 2007), and SEMANT (Silberman, Bentin, & Miikkulainen 2007). In particular, we have applied the DISLEX and DevLex-II models to the study of the acquisition of grammatical aspect (*-ing*, *-s* and *-ed*) in connection with the acquisition of semantic categories of lexical aspect (Li 2000; Li & Shirai 2000; Zhao & Li, in press). In what follows, we will provide a review of the findings from our DevLex-II model that has been used to simulate aspect acquisition² (Zhao & Li, in press).

2.1. A sketch of DevLex-II

DevLex-II is a multi-layer self-organizing neural network for modeling early lexical acquisition. It is based on and adapted from the DevLex model (Li, Farkas, & MacWhinney 2004). It has been developed to account for empirical phenomena in early lexical acquisition (e.g., ‘vocabulary spurt’) and bilingual lexical development (Li, Zhao, & MacWhinney 2007; Zhao & Li 2005, 2007, 2008). The basic structure of DevLex-II is shown in Figure 1.

² Here we only provide a review of simulation results based on DevLex-II. For detailed results about aspect acquisition with the DISLEX model, see Li (2000) and Li and Shirai (2000: ch. 7).

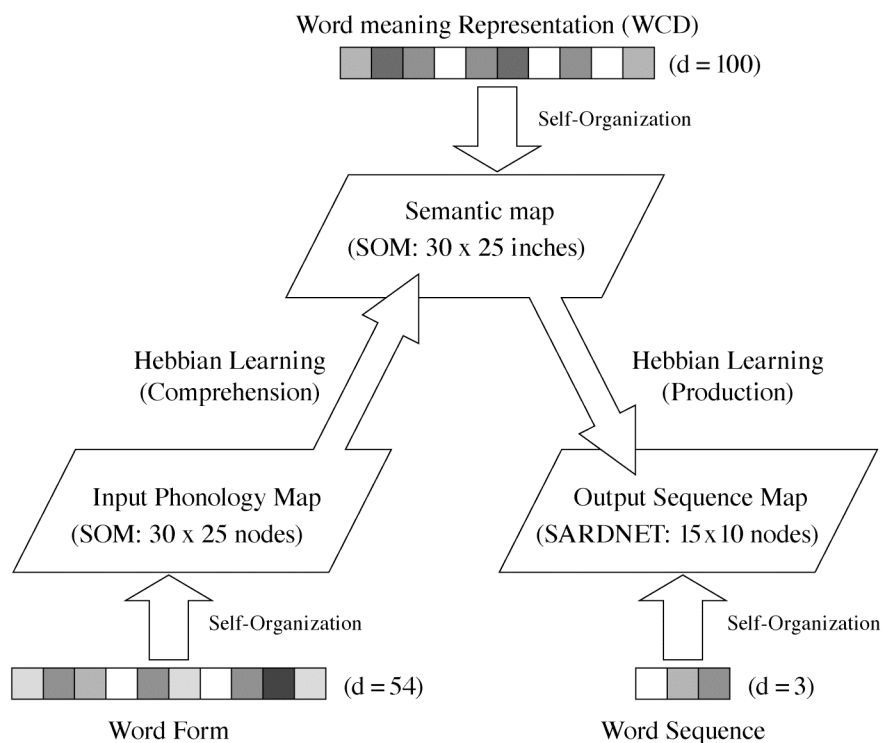


Figure 1. The architecture of the DevLex-II model. Each of the three self-organizing maps (SOM) takes input from the lexicon and organizes phonology, semantics, and phonemic sequence information of the vocabulary, respectively. The number of nodes in each map is indicated in parentheses. The dimension of the input vector for each map is indicated by 'd = ' in parentheses next to the input representation symbols. The maps are connected via associative links updated by Hebbian learning.

DevLex-II uses three layers of SOMs to process three basic levels of linguistic information: phonological content, semantic content, and output phonemic sequence. The phonological layer and semantic layer operate according to the standard SOM algorithm (see Kohonen 2001, for details). The standard SOM constructs a two-dimensional topographic map for the organization of input representations, where each node (or "neuron") is a location on the map that has input connections to receive external stimulus patterns. At each training step of SOM, an external input pattern (e.g., the phonological or semantic information of a word in our study) is randomly chosen

and presented to all the nodes on the map; this activates many nodes on the map, according to how similar by chance the input pattern is to the weight vectors of the nodes, and the node that has the highest activation is declared the winner (the Best Matching Unit or BMU). Once a node becomes active in response to a given input, the weight vectors of that node and its neighboring nodes are adjusted, so that they become more similar to the input and the nodes will respond to the same or similar inputs more strongly the next time. In this way, every time an input is presented, an area of nodes will become activated on the map (the “activity bubbles”) and the maximally active nodes are taken to represent the input. Initially activation occurs in large areas of the map, that is, large neighborhoods, but gradually learning becomes focused and the size of the neighborhoods reduces. This process continues until all the inputs have found some maximally responding nodes as their BMUs. As a result of this self-organizing process, the statistical structures implicit in the input are represented as topographical structures on the 2-D space. In this new representation, similar inputs will end up activating nodes in nearby regions, yielding meaningful activity bubbles that can be visualized on the map.

The addition of the phonemic sequence layer represents a step forward from the original DevLex model, and is inspired by models of word learning based on temporal sequence acquisition (e.g., Gupta & MacWhinney 1997). It is designed to simulate the challenge to young children when they need to develop better articulatory control of the phonemic sequences of words. Just as the learning of auditory sequences requires the mediation of memory systems, the learning of articulatory sequences requires support from the rehearsal in phonological working memory (Gathercole & Baddeley 1993; Gupta & MacWhinney 1997). In our implementation of this idea, the activation pattern corresponding to the phonemic sequence information of a word is formed according to the algorithm of SARDNET (James & Miikkulainen 1995), which works slightly differently from the standard SOM algorithm. At each training step, phonemes are input into the sequence map one by one, according their order of occurrence in the word. The winning unit of a phoneme is found and the responses of nodes in its neighborhood are adjusted. Once a unit is designated as the winner, it becomes ineligible to respond to the subsequent inputs in the sequence. In this way, same phonemes in different locations of a word will be mapped to different (but adjacent) nodes on the map as a result of the network’s topography-preserving ability. When the output status of the current winner and its neighbors is adjusted, the activation levels of the winners responding to phonemes before the current phoneme will be adjusted by a number γ^d , where γ is a constant and d

is the distance between the locations of the current phoneme and the previous phoneme that occurred in the word. This adjustment is intended to model the effect of phonological short-term memory during the learning of articulatory sequences; the activation of the current phoneme could be accompanied by some rehearsal of previous phonemes due to phonological memory, which deepens the network's or the learner's impression of previous phonemes. The γ here is chosen to be less than 1 (0.8 in our case), in order to model the fact that phonological memory tends to decay with time. For further details of the DevLex-II model, see Li, Zhao, and MacWhinney (2007).

The associative links between any two layers of maps in DevLex-II are trained by Hebbian learning, such that the activation of a word on the form map can evoke the activation of a word on the meaning map via form-to-meaning links, thereby modeling word comprehension, and the activation of word meaning can cause the formation of word sequence via meaning-to-sequence links, thereby modeling word production. In DevLex-II, we say that a word has been learned in comprehension when a node in the destination map (word meaning map) becomes consistently activated as the 'winner' for a given input from the source map (word form map). We say that a word has been learned in production when several nodes in the word sequence map become activated sequentially as winners that represent the word's phonemes.

2.2. Input representations for DevLex-II

As with DevLex, we used the PatPho system to construct the phonological patterns for word forms. PatPho is a generic phonological pattern generator for neural networks, which fits every word (up to tri-syllables) onto a template according to its vowel-consonant structure (Li & MacWhinney 2002). PatPho uses the same phonological method as in MacWhinney and Leinbach (1991), but relies on articulatory features of phonemes (Ladefoged 1982) to represent the phonemes, *C*s and *V*s, and a phoneme-to-feature conversion process to produce real-value or binary feature vectors for any word up to three syllables. In short, PatPho can code each input word in our simulation by the template *CCCVVCCCVVCCCVVCCC*, and then replace each phoneme with its appropriate representation using real-value or binary numbers.³ For example, the verb *pick* with its progressive marker *-ing* would be

³ In this simulation we used the real-value vectors.

encoded as pCCɪVkCCɪVɨCCVVCCC, and is represented as the following vector:

/pɪkɪŋ/: 1-0.45-0.733 0-0-0 0-0-0 0.1-0.1-0.185 0-0-0 1-0.921-0.733 0-0-0
 0-0-0 0.1-0.1-0.185 0-0-0 0.75-0.921-0.644 0-0-0 0-0-0 0-0-0 0-0-0
 0-0-0 0-0-0 0-0-0

In this representation, the first three units [1-0.45-0.733] indicate the phonetic features of phoneme /p/, the second and third sets of three units indicate that no more consonants follow /p/ in this word (hence zeros). The representation is left-justified, which means that in a given syllable, the representation of the phoneme is pushed toward the left side of the template (rather than the right side).

On the output sequence map, the phonemes of a word are processed one by one, so we need representations for each of the 38 English phonemes. Using the method of PatPho, we can represent these phonemes by three-dimensional real-value vectors. In particular, in the vector, the first dimension indicates whether the phoneme is a vowel or a consonant, and in the case of a consonant, whether it is voiced or voiceless. The second dimension indicates the position for vowels and manner of articulation for consonants and the third dimension indicates the sonority for vowels and place of articulation for consonants (see Li & MacWhinney 2002).

With respect to the semantic representation of the input, we used a special recurrent network called WCD (word co-occurrence detector) to generate vectors. WCD allows us to generate vectors that dynamically change with the learning history: lexical representations enrich over time as a function of learning the number of co-occurring words in the input sentences. Metaphorically, this learning scenario can be compared to filling the holes in a Swiss cheese: initially there may be more holes than cheese (*shallow* representations) but the holes get filled up quickly as the co-occurrence context expands with more words being acquired (*rich* representations).

Briefly, WCD works as follows (see Farkas & Li 2001, 2002; Li, Farkas & MacWhinney 2004, for details). It reads through a stream of input sentences one word at a time, and learns the transitional probabilities between words which it represents as a matrix of weights. Given a total lexicon sized N , all word co-occurrences can be represented by an $N \times N$ contingency table, where the representation for the i_{th} word is formed by concatenation of i_{th} column vector and i_{th} row vector in the table. Hence, the two vectors correspond to the left and the right context, respectively; WCD transforms these probabilities into normalized vector representations for word meanings,

which in turn are read by self-organizing maps (after Random Mapping, a procedure to achieve reduced uniform dimension of vectors; see Ritter & Kohonen 1989). Here, different inflectional forms of the same verb are considered as different items, and therefore WCD will derive different (but also similar) representations for them. For example, *playing* and *played* will be represented distinctly, but since the co-occurrence contexts for these words will overlap significantly (e.g., the co-occurring words tend to be *ball*, *toys*, etc.), the representations for them will also tend to be similar. An example of a semantic representation generated by WCD is shown below (only part of the vector is shown, the full vector contains 2002 units). Here, every two units of the vector represent the normalized co-occurrence possibility between the verb *picking* and another word in the lexicon. The odd unit represents the possibility that *picking* happens before a given word, and the even unit represents the possibility that *picking* follows a given word in the context.

Picking: 0.000000 0.000000 0.006004 0.007211 0.003548 0.017577 0.000000
 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.001202
 0.000000 0.000300 0.000000 0.000000 0.000000 0.000000 ...

DevLex-II also uses as its input data the parental or caregivers' speech in the CHILDES database (MacWhinney 2000). Here we extracted all parental or caregivers' utterances from the complete English database (as of 2002). A verb type was chosen as input if it occurred in the parental speech for fifty or more times in a given age period. The verbs were divided into four stages to be presented to the network, according to the age groups (Age 1;6, 2;0, 2;6, 3;0) at which they occurred. To increase the accuracy of WCD representations, we also analyzed the selected verbs along with the nouns, adjectives, and closed-class words from the MacArthur-Bates Communicative Development Inventories (the CDI, Toddler's List; Dale & Fenson 1996; homographs and homophones, word phrases, and onomatopoeias were excluded). These CDI words along with the verbs that fit our selection criterion (a total of 1001 words) served as the input contexts of WCD. We computed the semantic representations of the vocabulary at each of the four growth stages, resulting in 4 different data sets with increasing complexity in semantic representation. The four growth stages had the following vocabulary composition:

- (1) Input Age 1;6 (13–18 months): a total of 62 verb types fit our selection criteria for the period before age 1;6; 35 of these verbs occurred with *-ing*, 13 with *-ed*, 14 with *-s*.

- (2) Input Age 2;0 (19–24 months): 100 verb types were selected, which included the new words as well as words from the previous stage; 58 occurred with *-ing*, 19 with *-ed*, 23 with *-s*.
- (3) Input Age 2;6 (25–30 months): 154 verb types were selected, among which 86 occurred with *-ing*, 32 with *-ed*, 36 with *-s*.
- (4) Input Age 3 (31–36 months): A total of 184 verb types were selected, out of which 97 verbs occurred with *-ing*, 41 with *-ed*, and 46 with *-s*. This stage included all verbs that occurred in previous stages plus new ones.⁴

As shown in Figure 1, the size of the network was 30 x 25 nodes for the phonological map and the semantic map, and 15 x 10 nodes for the phonemic map. These numbers were chosen to be large enough to discriminate among the words and phonemes in the lexicon, while keeping the computation of the network tractable. For each training stage, the network was trained for 50 epochs, which means that each verb in a given stage was presented to each map 50 times.

2.3. Association of lexical and grammatical aspect in DevLex-II

An important rationale behind our simulations is for us to understand the role of linguistic input in guiding children's acquisition of lexical and grammatical aspect. Here, we wanted to test whether DevLex-II, endowed with self-organization and Hebbian learning principles, is able to display learning patterns as the child does. Our networks receive phonological and semantic representations of input words based on actual adult speech along with phonemic sequence (morphology) information of these words. If the network is able to produce patterns like those we found in children's speech on the basis of learning of the input, we can then conclude that self-organization and Hebbian learning provide the necessary kinds of mechanisms that drive the formation of patterns in children's acquisition. In this way, our modeling enterprise sheds light on the mechanisms that underlie the learning process.

⁴ Bare verb forms were excluded from our simulations, as well as irregular past tense forms and non-verbs. Exclusion of these forms simplifies the simulation task and makes the analysis more tractable. Our major goal here is to demonstrate whether the use of verbal suffixes is correlated with the lexical aspect of verbs, and as such our simulations focused on suffixed verbs.

Table 1 (adapted from Table 2 of Zhao & Li, in press) provides a summary of the major patterns from the DevLex-II models, according to the tense-aspect suffixes the model produced at different learning stages. The table presents the results of the networks' production of three suffixes, *-ing*, *-ed*, and *-s*, with three types of verbs, activity, telic and stative.⁵ The results were based on the analysis of the networks' production ability; that is, how semantic representations induce activations on corresponding feature map (phonemic sequence map in DevLex-II) through associative pathways. The analysis was done by inspecting the nodes that each verb on the semantic map activated, after the network had been trained for a specified number of epochs at each stage (50 epochs for DevLex-II).

The testing of DevLex-II's word production ability is as follows. At the end of each training stage, verb types in the lexicon of the current stage are presented to the semantic map one by one. For a verb type, its best matching unit or BMU on the semantic map is found, and in turn this node propagates its activation to the output sequence map through the associative links. Several nodes in the sequence map become activated sequentially as winners that represent the word's phonemes. Then the network checks to see if every node is the BMU of a unique phoneme, according to the Euclidean distance between its input weight vector and the feature representation of every phoneme. If it is, the phoneme closest in Euclidean distance to the current winner becomes its retrieved phoneme; if it is not, the pronunciation of this phoneme has failed. Finally, the pattern of the retrieved phoneme sequence is treated as the output of word production. When the retrieved phonemic sequence matches up with the actual word's phonemic sequence, we say that the word has been correctly produced. For example, if the word *kicking* is shown to the semantic map, correct production occurs only when the consecutively activated nodes on the output phonemic map are the BMUs for /k/ /ɪ/ /k/ /ɪ/ /ŋ/ in this particular sequence.

⁵ Our analyses below deviate slightly from a strict five-way classification, because accomplishment and achievement verbs are often difficult to separate without an extensive analysis of the sentence and speech context. Thus in what follows telic verbs include both accomplishments and achievements.

Table 1. Percentage of use of tense-aspect suffixes with different verb types across input age groups in DevLex-II's production and in parental input data (including verbs with multiple suffixes)

VERBS	TENSE-ASPECT SUFFIXES											
	Age 1;6			Age 2;0			Age 2;6			Age 3;0		
	<i>-ing</i>	<i>-ed</i>	<i>-s</i>	<i>-ing</i>	<i>-ed</i>	<i>-s</i>	<i>-ing</i>	<i>-ed</i>	<i>-s</i>	<i>-ing</i>	<i>-ed</i>	<i>-s</i>
Network Production												
Activity	73	0	29	69	27	33	61	24	35	62	30	37
Telic	27	75	14	21	53	28	32	62	27	31	62	26
Stative	0	25	57	10	20	39	7	14	38	7	8	37
Parental Input Data												
Activity	63	23	29	62	26	26	63	22	33	60	29	35
Telic	31	62	29	31	58	26	29	66	25	32	59	24
Stative	6	15	43	7	16	48	8	12	42	8	12	41

The results of Table 1 are highly consistent with empirical patterns observed in early child language: the use of imperfective aspect is closely associated with activity verbs that indicate ongoing processes, while the use of perfective aspect is closely associated with telic verbs that indicate actions with endpoints or end results. In particular, in early child English, the progressive marker *-ing* is highly restricted to activity verbs, the perfective/past marker *-ed* restricted to telic verbs, and the third person singular *-s* restricted to stative verbs (Bloom, Lifter and Hafitz 1980; Brown 1973; Clark 1996; Shirai 1991). Our network, having taken in input patterns based on realistic parental speech, behaved in the same way as children do. For example, at Input Age 1;6, the networks produced *-ing* predominantly with activity verbs (73%), *-ed* overwhelmingly with telic verbs (75%), and *-s* with stative verbs (57%). Such associations were strong at all four stages (especially for *-ing* and *-ed*), but they tended to become weaker over time.

Interestingly, when we analyzed the actual input to our networks (based on parental speech), we found similar patterns. Table 1 also shows the percentages of the use of suffixes with different verb types in the input data for DevLex-II. An analysis of the table indicates that in the input data there are also clear associations between *-ing* and activity verbs, *-ed* and telic verbs, and that these associations are strong throughout the four stages, as also found previously by Shirai (1991) and Olsen, Weinberg, Lilly and Drury (1998). The degree to which the networks' production matches up with the input patterns indicates that DevLex-II was able to learn on the basis of the information of the co-occurrences between lexical aspect (verb types) and

grammatical aspect (verb morphology). This learning ability was due to the networks' use of Hebbian associative learning in computing if the semantic, phonological, and phonemic properties of a verb co-occur and how often they do so.

To see the data more clearly, we illustrate the patterns with Figure 2⁶ to show the percentages of the use of suffixes with different verb categories in both DevLex-II's productions (Figure 2a) and in parental input data (Figure 2b) at Input Age 1;6. Comparing Figure 2a and 2b, we can see that the network's production patterns are consistent with patterns in the parental input data, but the network showed more restricted use of the suffixes rather than a verbatim replication of the input association patterns.

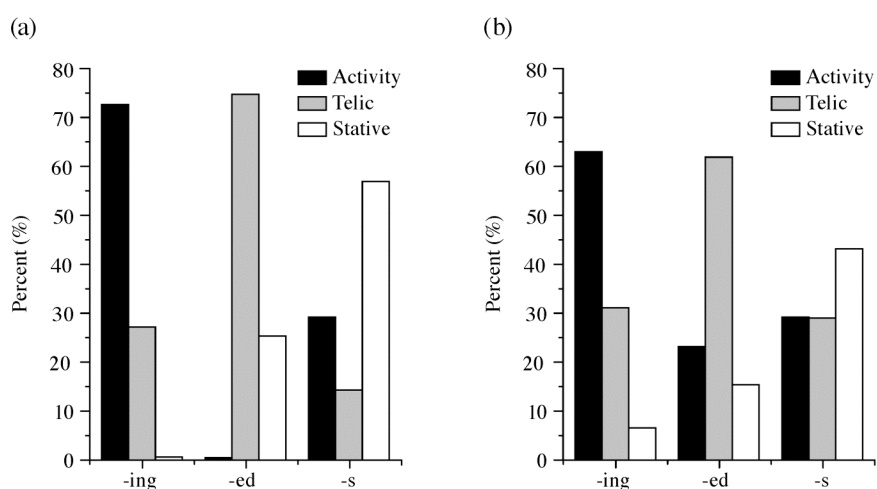


Figure 2. Percentages of the use of suffixes with different verb types at Input age 1;6 in (a) network productions in DevLex-II, and (b) parental input data. Data are based on Table 1.

2.4. Structured semantic representations of aspect in DevLex-II

Elsewhere we have proposed an account of semantic development as an emergent process in which semantic features are connected in a system to support lexical categories, like in the formation of semantic cryptotypes (Li & MacWhinney 1996; Li 2003; Li, Farkas & MacWhinney 2004; Hernandez,

⁶ Figures 2–4 in this chapter were adapted from Zhao & Li (in press).

Li, & MacWhinney 2005; see also Rogers & McClelland, 2004, for similar discussions). The basic idea is that a given verb may be represented with multiple linguistic features, and the features themselves often co-occur and overlap in different verbs. For example, the verb *screw* may be viewed as having both a meaning of circular movement and a meaning of binding or locking, and the verb *zip* may be viewed as sharing both the “binding/locking” meaning and the “covering” meaning. Moreover, both *screw* and *zip* involve hand movements. Features may also vary in the strength with which they are represented in different verbs. For example, the verb *wrap* may be viewed as having the covering meaning. However, in some cases, the action of wrapping may also involve circular movements. Children may acquire such complex feature-to-verb relationships through statistical analyses of the co-occurrences of verbs with situational contexts, with other words, and co-occurrences of particular grammatical morphemes with semantic features (see discussion in Siegel and McKeown’s 2000 work reviewed in section 1.2.2), leading to feature-based organization of verb categories. In the simulations here, we provided our networks with verbs that are represented with multiple semantic/syntactic features (lexical co-occurrence constraints, extracted by WCD), and we wanted to see how categories of lexical aspect could emerge from the self-organizing learning process.

A particularly useful property of self-organizing feature maps is that the statistical structures in the representations can be clearly visualized as activity bubbles or patterns of activity on a two dimensional map in a topography-preserving structure. Given that DevLex-II represented semantic information from the high-dimensional space of verb usage in parental input, we hypothesize that verbs with similar aspectual properties should cluster together on the feature map. Figure 3 present a snapshot of DevLex-II’s self-organization of the semantic representations of verbs (with suffixes) at the end of the learning process (i.e., Stage 4, Input Age 3;0).

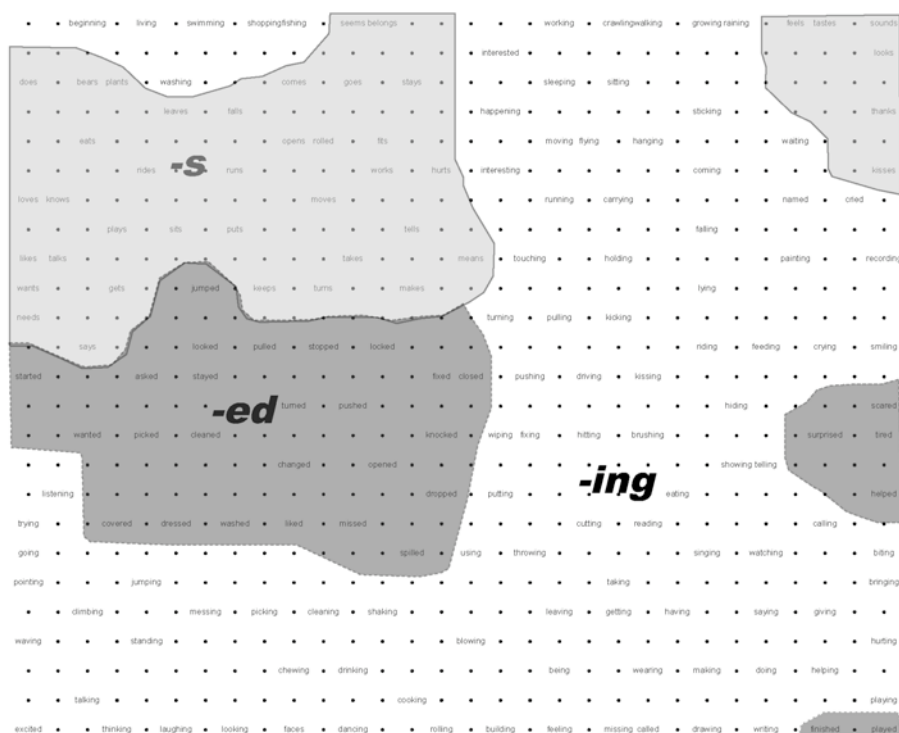


Figure 3. Emergent semantic representations in DevLex-II after Input age 3;0. Differently shaded regions indicate different aspect categories corresponding to different suffixes *-ed*, *-ing*, and *-s*. Within each category, verbs with the same lexical aspect are often grouped together; see text for discussion.

An examination of this map shows that the network has clearly developed structured semantic representations that correspond to different lexical aspect categories. It formed clear clusters of verbs by mapping verbs with similar combination of semantic features onto nearby regions of the map. We can make several interesting observations on the basis of these results:

- (1) The most obvious structure of the map is that the words can be roughly divided into three main clusters according to the suffix that a verb stem takes, *-ing*, *-ed*, or *-s* (see Figure 3);
- (2) Within each cluster, there are also groups that correspond to categories of lexical aspect such as telic verbs, activity verbs, and stative verbs. For example, towards the lower left-hand corner of the larger part of the *-s* cluster (the light gray area), stative verbs, like *loves*, *knows*, *likes*,

wants, and needs are mapped to the same region. Another example can be found in the *-ing* cluster (the area without shading): although most verbs clustered in this area are activity verbs such as *working, sitting, crawling, walking, sleeping, etc.*, there is also a cluster of telic verbs (at the middle-to-lower portion of the map) such as *wiping, fixing, hitting, putting, cutting, throwing, making, and getting*;

- (3) The distribution of lexical aspect is closely related to the distribution of grammatical aspect. Not only it is the case that the *-ed* cluster contained mostly telic verbs and the *-ing* cluster mostly activity verbs, but also telic verbs that take *-ing* were closer to the *-ed* cluster (e.g., *going, jumping, messing, picking* and *cleaning*, all bordering the *-ed* cluster);
- (4) Verbs with the same stem but different suffixes are also often mapped to regions not far away from one another, for example, *fixing* and *fixed*, *pushing* and *pushed*, *turns* and *turned* at the middle area of the map, and *playing* and *played* at the lower right corner of the map.

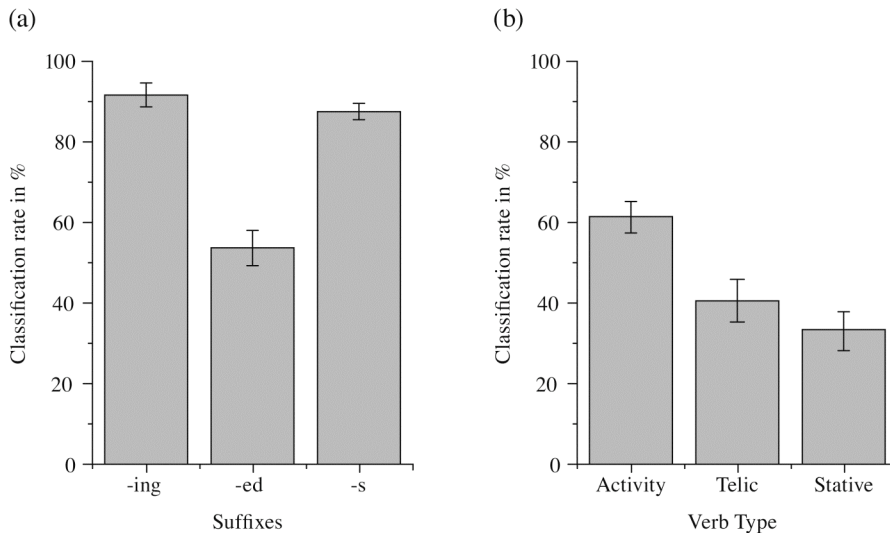


Figure 4. Classification rates calculated by a 5-NN classifier according to lexical and grammatical representations of the verbs in DevLex-II's semantic map. Classifications are based on: (a) the suffix that a verb stem takes: *-ing*, *-ed*, or *-s*; (b) the lexical aspect of the verb: activity, telic, or stative. The error bars indicate the standard deviations based on 5 trials.

The emergence of structured semantic representations in our model can also be verified by a simple method called *k*-nearest neighbor (*k*-NN) algorithm (Duda, Hart & Stork 2000). As a classical method in the field of pattern recognition for classifying objects into different classes, the basic idea of *k*-NN is to predict the class of a point in a dataset according to the most frequent class label of its *k* nearest neighbors. Implementing this method in our semantic map (see also Li, Farkas & MacWhinney 2004), we can evaluate if a verb in our lexicon was mapped to a node close in Euclidean distance to other verbs belonging to the same class. This allows us to have a rough idea of the overall compactness of different lexical classes. Here, we conducted a 5-NN analysis of verb representations on the semantic map according to the suffix a verb stem takes, *-ing*, *-ed*, or *-s*. As shown in Figure 4a, the semantic map has developed clear clusters for different suffixes: for the category of *-ing*, the classification rate is about 92%, which means that 92 percents of verbs suffixed with *-ing* are located within a nearest neighborhood according to *k*-NN; for the categories of *-s* and *-ed*, the classification rates are 88% and 60%, respectively. We also conducted a 5-NN analysis of the verbs according to their lexical aspect properties: as shown in Figure 4b, the classification rates for activity verbs, telic verbs, and stative verbs are 61%, 41%, and 34%, respectively. The relatively low classification rates of the verb categories, compared with those of suffix categories, indicate that the organization of verb meanings according to lexical aspect is subordinate to the organization of verb suffixes on the map. In general, these quantitative analyses are consistent with our visual analyses of the semantic maps.

The above observations lead us to conclude that the map has formed structured representations for grammatical aspect markers, such as *-ed*, *-ing*, *-s*, and that the interaction between grammatical aspect and lexical aspect is reflected in the correlation between grammatical morphology and verb types, and in the categories of lexical aspect such as activity, telic, and stative. The results from our modeling offer a new way of thinking about the representation of lexical aspect and its interaction with grammatical aspect. Verbs in a lexical aspect category form complex relationships, in that they vary in (1) how many linguistic features are relevant to the category, (2) how strongly each feature is activated in the representation of the category, and (3) how features overlap with each other across category members. For example, *spill* may be viewed as indicating both a punctual and a resultative meaning; *close* may involve both a change of state and a completive meaning; and the feature "punctual" may be represented more strongly in *jump* than in *fall*: in a natural setting a single jump occurs instantaneously, whereas falling need not (e.g., we could still say that a leaf fell

from a tree even if it drifted down slowly).⁷ With varying degrees of connections from semantic features to verb forms, verbs can form clusters or categories that differ overall in lexical aspect. Traditional analytical methods from linguistics and psycholinguistics are much less effective in dealing with these complex semantic relationships (but see Siegel and McKeown's recent attempt to combine different linguistic indicators, reviewed earlier). By contrast, neural network models that rely on distributed feature representations and nonlinear learning are ideally suited to accounting for the properties of feature overlapping and weighted feature composition. DevLex-II provides a clear example for how we may solve complex semantic problems via weighted feature composition (see also Li & MacWhinney 1996; Li 2003 2006; Li, Farkas & MacWhinney 2004).

3. Conclusion

In this chapter, we presented an overview of the computational models of the expression and acquisition of temporality in languages. For the issue of tense, we provided a brief introduction to computational modeling of children's acquisition of the English past tense, and the heated debate on single versus dual mechanisms revolving around this issue. For the issue of aspect, we reviewed a few computational models of aspectual classification, and introduced in detail our DevLex-II model for simulating aspect acquisition in languages. Our review clearly shows that the expression and acquisition of aspect are very complex processes that depend both on the verb's internal semantic meanings about temporal concept and on the syntactic features of the sentence where the verb occurs. Our DevLex-II model successfully simulates the acquisition of lexical and grammatical aspect, and provides insights into issues regarding the role of linguistic input, the emergence of lexical categories of verbs, and the development of prototypical to non-prototypical associations.

Self-organization and Hebbian learning in our model are two important computational principles that can account for the psycholinguistic processes in the acquisition of lexical and grammatical aspect. Our simulations dem-

⁷ To see this, in Figure 3, we can find that the word *jump*, with its progressive marker *-ing*, is much closer than *falling* to the *-ed* cluster. Note that *jump* can also be construed iteratively, so that *jumping* refers to a series of jumps, which is why we may see that children use *jumping* more frequently than *jumped* in natural speech.

onstrate that the network is able to display patterns of association as observed in empirical acquisition studies, on the basis of its analyses of input characteristics. In particular, self-organization of the semantic structure of verbs leads to the formation of lexical aspect categories and grammatical aspect categories, on the basis of the network's analysis of the complex feature-to-verb and verb-to-morphology relationships in language use. In addition, our model clearly shows that simple but biologically plausible computational principles in self-organizing neural networks can account for empirically observed patterns in children's acquisition of lexical aspect and grammatical morphology, without a priori stipulations about the structure of meaning or concept.

Contributing to the debate on single versus dual mechanisms for learning, our model also specifically suggests that the learning of grammatical suffixes is not simply the learning of a rule (such as adding *-ing* or *-ed* to a verb to mark the progressive aspect or the perfective aspect), but the accumulation of associative strengths that hold between a particular suffix and a complex set of semantic features distributed across verb forms (which support the emergence of a lexical aspect category). This learning process can be best described as a statistical, probabilistic process in which the learner implicitly tallies and registers the frequency of co-occurrences (strengthening what goes with what) and co-occurrence constraints (inhibiting what does not go with what) among the semantic features, lexical forms, and tense-aspect suffixes. The co-occurrence-and-constraint process is clearly modeled in our network by Hebbian learning of the associative connections between forms and meanings (see also a detailed description in Zhao & Li, in press).

To conclude, the models we reviewed here clearly serve to demonstrate the utility of computational modeling (especially connectionist modeling) for unraveling mechanisms underlying the expression and acquisition of tense and aspect in languages. With the rapid development of computing techniques and the advancement of computational modeling, we are hopeful that the detailed cognitive and psycholinguistic mechanisms can be clearly revealed in the fields of language acquisition, language representation, and language processing.

References

- Bates, Elizabeth, Jeffrey Elman, and Ping Li
 1994 Language in, on, and about time. In *The Development of Future-Oriented Processes*, M. Haith, J. Benson, R. Roberts and B. Pennington (eds.), 293–321. Chicago, IL: University of Chicago Press.
- Bennett, Winfield, Tanya Herlick, Katherine Hoyt, Joseph Liro and Ana Santisteban
 1989 Toward a computational model of aspect and verb semantics. *Machine Translation* 4: 247–280.
- Berko, Jean
 1958 The child's learning of English morphology. *Word* 14: 150–177.
- Bickerton, Derek
 1984 The language bioprogram hypothesis. *Behavioral and Brain Sciences* 7: 173–188.
- Bloom, Lois, Karin Lifter, and Jeremie Hafitz
 1980 Semantics of verbs and the development of verb inflection in child language. *Language* 56: 386–412.
- Brown, Roger
 1973 *A First Language*. Cambridge, MA: Harvard University Press.
- Clark, Eve V.
 1996 Early verbs, event-types, and inflections. In *Children's Language*, Vol. 9, C. E. Johnson & J. H. V. Gilbert (eds.). Mahwah, NJ: Erlbaum.
- Comrie, Bernard
 1976 *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge: Cambridge University Press.
- Dale, Philip S. and Larry Fenson
 1996 Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers* 28: 125–127.
- Duda, Richard O., Peter E. Hart, and David G. Stork
 2000 *Pattern Classification*, 2nd Ed.. New York, NY: John Wiley & Sons.
- Elman, Jeffrey L., Elizabeth A. Bates, Mark H. Johnson, Aannette Karmiloff-Smith, Domenico Parisi, and Kim Plunkett
 1996 *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.
- Farkas, Igor and Ping Li
 2001 A self-organizing neural network model of the acquisition of word meaning. In *Proceedings of the Fourth International Conference on Cognitive Modeling*, E. M. Altmann, A. Cleeremans, C. D. Schunn, & W. D. Gray (eds.), 67–72. Mahwah, NJ: Lawrence Erlbaum.
- 2002 Modeling the development of the lexicon with a growing self-organizing map. In *Proceedings of the Sixth Joint Conference on Information Science*, H. J. Caulfield et al. (eds.), 553–556. Association for Intelligent Machinery, Inc.

- Gathercole, Susan E. and Alan D. Baddeley
1993 *Working Memory and Language*. Hillsdale, NJ: Erlbaum.
- Gupta, Prahlad and Brian MacWhinney
1997 Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language* 59: 267–333.
- Hebb, Donald O.
1949 *The Organization of Behavior: A Neuropsychological Theory*. New York, NY: Wiley.
- Hernandez, Arturo, Ping Li, and Brian MacWhinney
2005 The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences*, 9, 220–225.
- Holland, John H.
1975 *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: The University of Michigan Press.
- James, Daniel L. and Risto Miikkulainen
1995 SARDNET: A self-organizing feature map for sequences. In *Advances in Neural Information Processing Systems 7*, (NIPS'94) G. Tesauro, D. S. Touretzky, and T. K. Leen (eds.), 577–584. Cambridge, MA: MIT Press.
- Kohonen, Teuvo
2001 *The Self-Organizing Maps*. 3rd Ed. Berlin: Springer.
- Ladefoged, Peter
1982 *A Course in Phonetics*. 2nd Ed. San Diego, CA: Harcourt Brace.
- Li, Ping
2000 The acquisition of tense-aspect morphology in a self-organizing feature map model. In *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society*, Lila R. Gleitman & Aravind K. Joshi (eds.), 304–309. Mahwah, NJ: Lawrence Erlbaum.
- 2003 Language acquisition in a self-organizing neural network model. In *Connectionist Models of Development: Developmental Processes in Real and Artificial Neural Networks*, P. Quinlan (ed.), 115–149. Hove & Briton: Psychology Press.
- 2006 In search of meaning: The acquisition of semantic structure and morphological systems. In *Cognitive Linguistics Investigations: Across Languages, Fields and Philosophical Boundaries* J. Luchjenbroers (ed.), 109–137. Amsterdam: John Benjamins.
- Li, Ping and Melissa Bowerman
1998 The acquisition of lexical and grammatical aspect in Chinese. *First Language* 18: 311–350.
- Li, Ping and Brian MacWhinney
1996 Cryptotype, overgeneralization, and competition: A connectionist model of the learning of English reversive prefixes. *Connection Science* 8: 3–30.

- 2002 PatPho: A phonological pattern generator for neural networks. *Behavior Research Methods, Instruments, and Computers* 34: 408–415.
- Li, Ping and Yasuhiro Shirai
2000 *The Acquisition of Lexical and Grammatical Aspect*. Berlin/New York: Mouton de Gruyter.
- Li, Ping, Curt Burgess, and Kevin Lund
2000 The acquisition of word meaning through global lexical co-occurrences. In *Proceedings of the Thirtieth Stanford Child Language Research Forum*, E. V. Clark (ed.), 167–178. Stanford, CA: Center for the Study of Language and Information.
- Li, Ping, Igor Farkas, and Brian MacWhinney
2004 Early lexical development in a self-organizing neural networks. *Neural Networks* 17: 1345–1362.
- Li, Ping, Xiaowei Zhao, and Brian MacWhinney
2007 Dynamic self-organization and early lexical development in children. *Cognitive Science* 31: 581–612.
- Ling, Charles X. and Marin Marinov
1993 Answering the connectionist challenge. *Cognition* 49: 267–290.
- MacWhinney, Brian
1993 Connections and symbols: Closing the gap. *Cognition* 49: 291–296.
1998 Models of the emergence of language. *Annual Review of Psychology* 49: 199–227.
2000 *The CHILDES project: Tools for Analyzing Talk*. Hillsdale, NJ: Lawrence Erlbaum.
2001 Lexicalist connectionism. In *Models of Language Acquisition: Inductive and Deductive Approaches*, P. Broeder & J. M. Murre (eds.), 9–32. Oxford, UK: Oxford University Press.
- MacWhinney, Brian and Jared Leinbach
1991 Implementations are not conceptualizations: Revisiting the verb learning model. *Cognition* 40: 121–157.
- Marcus, Gary F., Michael Ullman, Steven Pinker, Michelle Hollander, T. John Rosen, and Fei Xu
1992 Overregularization in language acquisition. *Monographs of the Society for Research in Child Development* 57(4): 1–182.
- McShane, John and Stephen Whittaker
1988 The encoding of tense and aspect by three- to five-year-old children. *Journal of Experimental Child Psychology* 45: 52–70.
- Moens, Marc and Mark Steedman
1988 *Computational Linguistics* 14: 15–28.
- Miikkulainen, Risto
1997 Dyslexic and category-specific aphasic impairments in a self-organizing feature map model of the lexicon. *Brain and Language* 59: 334–366.

- Olsen, Mari Broman, Amy Weinberg, Jeffrey P. Lilly and John E. Drury
1998 Mapping innate lexical features to grammatical categories: acquisition of English *-ing* and *-ed*. In *Proceedings of the 20th Annual Conference of the Cognitive Science Society*, M. Gernsbacher and S. Derry (eds.), 794–799. Mahwah, NJ: Lawrence Erlbaum.
- Passonneau, Rebecca J.
1988 A computational model of the semantics of tense and aspect. *Computational Linguistics* 14: 44–60.
- Pinker, Steven
1994 *The Language Instinct: How the Mind Creates Language*. New York, NY: Harper Collins Publishers Inc.
- Pinker, Steven and Alan Prince
1988 On language and connectionism: Analysis of a Parallel Distributed Processing Model of language acquisition. *Cognition* 29: 73–193.
- Plunkett, Kim and Virginia A. Marchman
1991 U-shaped learning and frequency effects in a multilayered perceptron: Implications for child language acquisition. *Cognition* 38: 43–102.
- Rogers, Timothy T. and James L. McClelland
2004 *Semantic Cognition: A Parallel Distributed Processing Approach*. Cambridge, MA: MIT Press.
- Ritter, Helge and Teuvo Kohonen
1989 Self-organizing semantic maps. *Biological Cybernetics* 61: 241–254.
- Rumelhart, David E. and James L. McClelland
1986 On learning the past tense of English verbs. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, J. L. McClelland and D. E. Rumelhart (eds.), 216–271. Cambridge: MIT Press.
- Scheler, Gabriele
1996 Learning the semantics of aspect. In *New Methods in Language Processing*, H. Somers and D. Jones (eds.), 83–95. London: Routledge.
- Shirai, Yasuhiro
1991 Primacy of aspect in language acquisition: Simplified input and prototype. Ph.D. dissertation, Applied Linguistics, University of California, Los Angeles.
- Shirai, Yasuhiro and Roger W. Andersen
1995 The acquisition of tense-aspect morphology: A prototype account. *Language* 71: 743–762.
- Shultz, Thomas R.
2003 *Computational Developmental Psychology*. Cambridge, MA: MIT Press.
- Silberman, Yaron, Shlomo Bentin, and Risto Miikkulainen
2007 Semantic boost on episodic associations: An empirically-based computational model. *Cognitive Science* 31: 645–671.

- Siegel, Eric V. and Kathleen R. Mckeown
2000 Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights. *Computational Linguistics* 26: 595–628.
- Smith, Carlota S.
1983 A theory of aspectual choice. *Language* 59: 479–501.
1997 *The Parameter of Aspect*. Dordrecht: Kluwer.
- Vendler, Zeno
1957 Verbs and times. *Philosophical Review* 66: 143–160.
- Zhao, Xiaowei and Ping Li
2005 A self-organizing connectionist model of early word production. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, B. G. Bara, L. Barsalou, & M. Bucciarelli (eds.), 2434–2439. Mahwah, NJ: Lawrence Erlbaum.
2007 Bilingual lexical representation in a self-organizing neural network. In *Proceedings of the 29th Annual Cognitive Science Society*, D. S. McNamara & J. G. Trafton (eds.), 755–760. Austin, TX: Cognitive Science Society.
2008 Vocabulary development in English and Chinese: A comparative study with self-organizing neural networks. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, B. C. Love, K. McRae, & V. M. Sloutsky (eds.), 1900–1905. Austin, TX: Cognitive Science Society.
in press The acquisition of lexical and grammatical aspect in a developmental lexicon model. *Linguistics: An Interdisciplinary Journal of the Language Sciences*.