

Learning to Associate Object Categories and Label Categories: A Self-Organising Model

Julien Mayor (julien.mayor@psy.ox.ac.uk)

Department of Experimental Psychology, South Parks Road
Oxford, OX1 3UD, United Kingdom

Kim Plunkett (kim.plunkett@psy.ox.ac.uk)

Department of Experimental Psychology, South Parks Road
Oxford, OX1 3UD, United Kingdom

Abstract

We present a model of early lexical acquisition. Successful word learning builds on pre-existing, self-organising categorisation capacities and through joint attentional events between the infant and the caregiver. Our model successfully accounts for the emergence of a *lexical constraint*, taxonomic responding, as well as a steep increase in the rate of acquisition of words, without any need for a new mechanism or rule. We discuss the role played by synapto-genesis in early infancy and pruning at the end of childhood in relation to word learning.

Keywords: Early word learning; lexical constraints; taxonomic responding; vocabulary spurt; self-organising model.

Introduction

A central issue in early lexical acquisition is how infants learn to understand the meaning of words. In a typical labelling situation, the caregiver points at an object (*Fido* the dog) and says “Look, this is a dog!”. The infant has then to rule out a huge number of possible meanings. For example, the word might refer to the size, to the shape, to the colour, to the individual *Fido*, etc... (Quine, 1960). However, infants reliably interpret the word “dog” as a label that can be used for this dog **and** for all dogs.

An influential solution to this conundrum is that infants make use of linguistic constraints to narrow the hypothesis space in order to assign meaning to words. These constraints provide sufficient information to guide the infant in apprehending the difficult task of word learning. One of these constraints is the taxonomic assumption (Markman & Hutchinson, 1984). In a “weak” form, it was introduced as the infants’ relative preference over a thematic extension of labels: if a dog is labelled, infants pick the taxonomic match (e.g., a cat) rather than the thematic match (e.g., a bone). In its strong form, taxonomic responding assumes that “when infants embark upon the process of lexical acquisition, they are initially biased to interpret a word applied to an object as referring to that object and to other members of its kind” (Waxman & Markow, 1995). In other words, from a single labelling event, the infant infers that every object that belongs to the same category has the same name. This capacity provides a powerful tool for identifying objects and for communication, since it enables individuals to refer to objects they have never seen before.

Nazzi & Bertoni (2003) have proposed that the period of lexical development around 18 months, known as the vocabulary spurt, reflects qualitative changes in the way words

are acquired. They define a qualitative transition between proto-words (pairing of an under-specified sound pattern to a specific object) to genuine referential use of words (pairing of a specified sound pattern and a category of objects). This qualitative change towards a referential use of words has been associated with the emergence of constraints on lexical acquisition (Golinkoff, Mervis, & Hirsh-Pasek, 1994; Markman, 1989). Nazzi and Bertoni (2003) further proposed that the vocabulary spurt corresponds to a shift from associationist to referential lexical acquisition mechanisms. Similarly, Markman suggests that word learning is divided in two stages; first, simple associative mechanisms account for early slow learning. The increase in speed of word learning is attributed to the application of lexical constraints.

The goal of our research is to understand the mechanisms that enable children to respond in a taxonomic fashion. This mechanism has to account for a qualitative change in the way infants use words – from an associative to a referential use – as well as a quantitative shift in the number of words that belong to their lexicon. We introduce a model of early word learning which exhibits a rapid increase in the rate of acquisition of words as well as taxonomic responding as an emergent property, despite the fact that the model relies on continuous changes and makes no use of a new learning mechanism. We show within a modelling framework that efficient pre-established categorisation capacities are a prerequisite to successful word learning. Joint attentional activities, involving just a single labelling event, provide the foundations for taxonomic responding, in which the label is successfully extended to all members of the category. Once a label is learnt for the whole category, the word is considered to be part of the infants’ lexicon.

To our knowledge, no existing model has attempted to model the generalisation properties of word-object associations from single labelling events. However our model shares a very similar architecture to previous models of early lexical acquisition (Miikkulainen, 1997; Li, Farkas, & MacWhinney, 2004; Li, Zhao, & MacWhinney, 2007).

The Model

Overview

We identified two main contributing factors for successful word learning: pre-lexical categorisation abilities and joint at-

tentional activities. The model combines these two contributing effects; unsupervised categorisation capacity corresponding to perceptual refinement and activity-dependant synaptic learning, implementing the role of joint attentional activities.

Uni-modal cortical maps (self-organising maps or SOMs, Kohonen, 1984) receive visual and acoustic input. The self-organising maps extract statistical regularities from the input and form categories of similar objects and of similar sound patterns without explicit supervision. They mimic the refinement of the infants’ perceptual systems in early infancy.

We mimic joint attentional activities by presenting objects and their labels simultaneously. Synapses connecting active neurones on both maps are reinforced through Hebbian learning. Hebbian synapses connecting the two SOMs are modulated by the joint presentations of objects and their labels so as to form word-object pairs, as shown in Fig. 1.

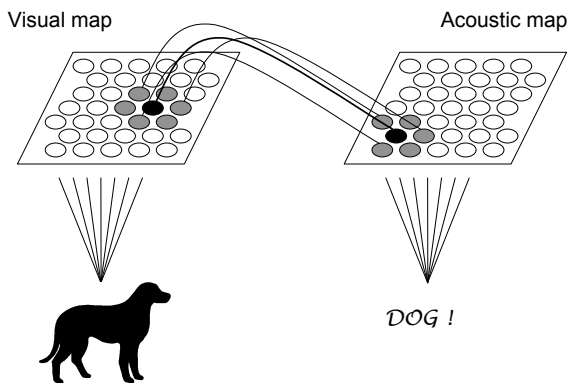


Figure 1: Sketch of joint attentional activity in the network: Hebbian synapses are modulated by the coherent activity patterns induced on both maps due to the simultaneous presentation of an object and its corresponding label.

Due to the topographical organisation of the maps that takes place during development, many neighbouring neurones on each map will be activated by the presentation of an object and a sound pattern. Hebbian learning then takes place for many neurones on each map. Therefore, the association between the paired object and its corresponding sound pattern will be *generalised* and will automatically build associations between all objects in that category to all sound patterns corresponding to the appropriate label. A single labelling event induces a taxonomic response and the label is extended to all objects of like type: the novel word is learnt. In a first experiment, joint attentional activities take place at regular intervals. The generalisation of word-object associations is compared when the system is confronted with two rates of joint attentional events, modelling the role of a high vs. low interacting caregiver on lexical acquisition. In a second experiment, taxonomic responding is assessed after having a

single labelling event per category only, taking place early or late in development. This way, we gain an insight into the learnability of a word in a developmental perspective. This procedure allows for distinguishing the relative roles of perceptual refinement (Self-Organising Map structuration) and joint attentional events.

Formal Presentation

Self-Organisation of the Uni-Modal Maps: Pre-Lexical

Categorisation The algorithm of self-organisation is the standard Kohonen algorithm (Kohonen, 1984). Each map (acoustic and visual) consists of an hexagonal grid of neurones receiving acoustic and visual inputs, respectively. Each neurone k is associated with a vector m_k . For the presentation of each input pattern x , the vectors m_k are modified according to the following procedure: We find the Best Matching Unit (BMU) i , defined by the condition $\|m_i - x\| \leq \|m_j - x\|, \forall j$. By extension, we identify the second best matching unit, the third, and so on. The standard weight update rule is applied with a learning rate that decays over time, $\alpha(t) = \frac{0.05}{1+t/2000}$ and a Gaussian neighbourhood function of the distance $\|r_i - r_k\|$ between neurones i and k on the map, $N(i, k)_t = e^{-\|r_i - r_k\|^2 / 2\sigma^2(t)}$, that shrinks linearly over time, from $\sigma(0) = 4$ to $\sigma(T_{max}) = 1$. For every epoch of training, corresponding to one presentation of each input pattern in the data-set, we compute the averaged quantisation error, as a measure of weight alignment to the input, so that the Euclidian distance between input patterns and their respective best matching units is: $E = \langle \|x - m_c(x)\| \rangle_x$ where $m_c(x)$ is the best matching unit for input pattern x . In all experiments, map sizes were fixed to a 25x25 hexagonal grid of neurones.

Coding the Inputs

Objects are represented as distorted dot patterns. Random dot patterns have been extensively studied by Posner and colleagues (1967) and have been shown to provide a good approach to the study of concept learning through the psychophysics of form similarity. We created 100 patterns (prototypes) by randomly distributing 9 dots over a 30x30 square. We then generated 24 tokens in each category, consisting of 8 tokens at each of 3 levels of distortion. Distortions are achieved by moving each dot by an amount drawn from a normal distribution with a standard deviation of 1, 2 and 3 respectively. Prototypes were not included in the data set.

There is little consensus in the field as to what acoustic information infants use when identifying words. A series of studies emphasise the fact that infants pay attention to much more than simple phonological features. In particular, it has been shown that infants are sensitive both to stress and phonetic information, make allophonic distinctions and pay attention to co-articulation (Jusczyk, 1997). All of these sensitivities to the speech signal may have an important impact on early lexical development. Therefore, we exploit the whole acoustic signature of word tokens in order to avoid discarding relevant acoustic information. In a preliminary study, we measured the variance in the sound pattern both across differ-

ent speakers (referred to as “high variance”) and within repetitions of words by the same speaker (“low variance”). We then extracted the acoustic signature from raw speech waveforms for 100 acoustic categories (words) produced by a female native speaker. We generated 24 acoustic signatures for each acoustic category, centred around the recorded speech signals. Each category consisted of 80% low variance modifications of the prototype, mimicking the high occurrence of the mother’s speech, and of 20% high variance modifications, thereby providing the model with a variance in word pronunciation introduced by different speakers. Tokens were then normalised in length and sampled at regular intervals, 4 times per syllable. After sampling, the sounds were filtered using the Mel Scale in order to approximate the human ear sensitivity. Input vectors were concatenations of four 7-dimensional mel-cepstrum vectors, derived from FFT-based log spectra (Davis & Mermelstein, 1980).

Training the Cross-Modal Associations; Joint Attentional Activities As maps become structured due to the presentations of the images and acoustic tokens in the data set, we mimic joint attentional activities between the caregiver and the infant by presenting simultaneously to both maps a randomly picked object from the data set and an acoustic token randomly picked within the matching category. We build cross-modal associations by learning Hebbian connections between both maps.

As a further simplification of the model, we use bidirectional synapses whose amplitudes are modulated by the activity of the connecting neurones. We define the neural activity of a neurone k to be $a_k = e^{-q_k/\tau}$ where q_k is the quantisation error associated with neurone k and $\tau = 0.5$ is a normalisation constant.

It is known that the number of synapses is not constant during life. There is first a process of synapto-genesis, a maturational process in which neurones make contact with other neurones. Synapses are then pruned according to an experience-dependent process (Huttenlocher, 1979) In the first stage, the process is essentially “blind”; synapses are formed at random locations. We mimic this phenomenon by linking together only a percentage of randomly picked neurones on both maps, a percentage that increases with age. The number of synapses peaks at around 2 years of age (Huttenlocher, 1979). To a first approximation, we model synapto-genesis as a linear increase from 1% connectivity to a full connectivity after 500 epochs of training. Later we address the role played by synapto-genesis in greater detail. The number of synapses then decreases by a process of elimination of “weak” synapses. This synaptic pruning mechanism is experience-driven. The process of synaptic pruning is thought to minimise energy consumption as well as optimise network properties (Chechik, 1998)). It is also known that the total number of neurones decreases over time. However, we do not address this issue in our model.

All S synapses were first randomly initialised with a normal distribution centred on $1/\sqrt{(S)}$ and with a standard deviation

of $\frac{1}{\sqrt{(1000*S)}}$. Synapse amplitudes are modulated according to a standard Hebb rule with saturation. Therefore, synaptic weights stay in a *physiological* range even for high neural activities. The synapse connecting neurone i from the visual map to the neurone j of the acoustic map is computed as follows: $w_{ij}(n+1) = w_{ij}(n) + 1 - e^{-\lambda a_i a_j}$ where n refers to the index of the word-object pairing and $\lambda = 10$ is the learning rate. The free parameter λ can be chosen on a wide range of values with qualitatively similar results. After every word-object presentation, weights are normalised so as to model limited synaptic resources: $\sum_{ij} w_{ij}^2 = 1$.

Assessment of the Model In order to assess the performance of the model, it is necessary to introduce two definitions. The first definition introduces a measure of the magnitude of taxonomic responding. It may be interpreted as the learnability of a word. Real word learning situations involve more than a single labelling event per category. Hence, the second definition defines a criterion for deciding when a word is learnt. All results refer to these measures.

Definition 1 We refer to the *taxonomic factor* as being the percentage of objects that are associated with their correct label, following a *single labelling event* per category.

Definition 2 We consider a word to be part of the lexicon if more than 80% of the objects of that category are associated with the correct label.

After training on cross-modal pairings, we assess the capacity of the network to extend the association of a presented word-object pair to non-paired items that belong to the same category. Following one simultaneous presentation of a word-object pair per category, weights are fixed. All images in the dataset are classified according to whether the induced activity on the acoustic map corresponds to the activation of the appropriate label (corresponding to the minimal Euclidian distance). Similarly, acoustic tokens are classified according to the activity induced onto the visual map. The average of both classification measures is the taxonomic factor. All results reported are averaged over 35 independent simulations.

Results

We present three experiments highlighting contributing factors to successful word learning. First, we present the time course of vocabulary acquisition under two conditions; a high versus a low level of regular joint attentional activities. We demonstrate that the amount of joint attentional activity correlates with the total vocabulary size. The rate of lexical acquisition undergoes a rapid increase in both conditions, despite the fact there is no change in the underlying neural mechanisms. In a second experiment, we show that the taxonomic factor increases abruptly at the same time as the network undergoes accelerated lexical growth. The increased rate of word learning is controlled by the taxonomic factor. However, it should be noted that taxonomic responding is an emergent property of the architecture when confronted with a structured environment. The taxonomic factor is directly determined by the pre-established categorisation skills, in the

form of well-structured maps. A further experiment highlights the role played by synapto-genesis and later pruning for taxonomic responding, emphasising the fact that early synapto-genesis supports an increase in taxonomic responding. Once the infant, and the model, have gained sufficient experience from the environment, synapses can be pruned without affecting the word-learning capacity, thereby optimising the system from a metabolic point of view.

Time Course of Lexical Acquisition: Vocabulary Spurt

We start by monitoring the evolution of vocabulary size with time when the model is trained in the word-object environment. Vocabulary size (according to Definition 2) is measured at different points of development. After n epochs, each map has been trained n times on all images and words in the data set. Simultaneously, the number of synapses between the maps increases linearly from a 1% random connectivity at the start of learning to full connectivity at epoch 500, mimicking qualitatively the synapto-genesis occurring in early childhood. Labelling events take place at regular intervals; we present the network with two environments, one with a small number of joint-attentional activities (0.2 labelling events per category and per epoch) and one with a high number of joint-attentional activities (0.5 labelling events per category and per epoch).

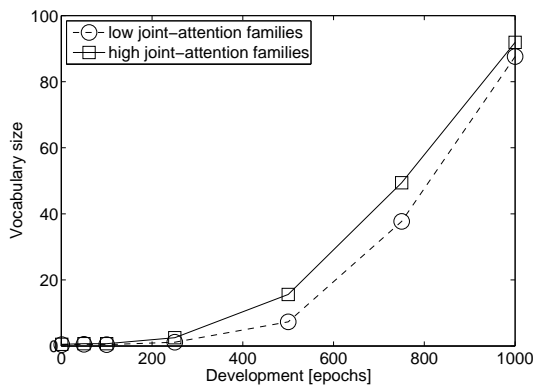


Figure 2: Time course of vocabulary growth when immersed in an environment with a high number of joint attentional activities (\square) and with a low number of joint attentional activities (\circ). Our model confirms experimental findings that joint attentional activities is a predictor of language development. In both conditions, the rate of lexical acquisition undergoes a rapid increase, similar to a vocabulary spurt.

Fig. 2 depicts the evolution of total vocabulary size during development. First, note the nonlinear increase in the size of the lexicon at around 500 epochs, similar to a vocabulary spurt. Second, when the network is confronted with an environment with joint attentional activities of low intensity, vocabulary growth is slower and later vocabulary size is reduced when compared to an environment rich with joint atten-

tional events. This finding provides a theoretical account of the corresponding experimental findings (Tomasello & Farrar, 1986).

Role of Pre-Lexical Categorisation Capacity

We now identify the contributing effect of pre-lexical categorisation capacity to taxonomic responding. Similarly to the previous experiment, we train the maps with the (non-simultaneous) presentation of objects and labels. However this time, we control joint attentional activities by having only one labelling event per category taking place after the maps have been trained for n epochs. For each point on Fig. 3, the network has seen one word-object pair per category only. We assess the taxonomic factor, according to Definition 1. This procedure reveals the time course of the taxonomic factor. Equivalently, it defines the likelihood, at different points in time, that a word is learnt following one labelling event. As the maps gain experience with the structured environment, they organise topologically so that objects of like kind activate neighbouring neurones on the maps. Generalisation of word-object associations is thereby facilitated. Taxonomic responding correlates with the quality of the map structure. The time course of the taxonomic factor shows a steep increase at around 500 epochs, as shown in Fig. 3. Taxonomic responding is an emergent property of the network. It is not

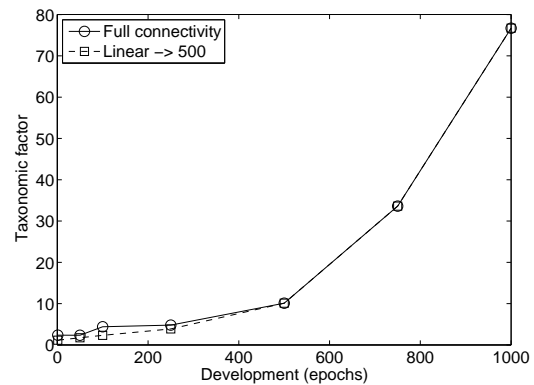


Figure 3: Role of pre-lexical experience. As the maps gain experience with the visual and acoustic environment, they undergo an organisation in their structure. The improvement in structural organisation impacts directly the system's ability to generalise associations from a single labelling event. The taxonomic factor displays a steep increase from about 500 epochs of training. The exact time course of synapto-genesis has no qualitative impact on the emergence of taxonomic responding.

surprising that the nonlinear increase in the taxonomic factor seen in Fig. 3 coincides with the onset of the vocabulary spurt observed in Fig. 2. It is, however, noteworthy that the time course of the taxonomic factor is not qualitatively different when the maps are fully-connected throughout development (Fig. 3 solid line) and when there is a linear increase

in connectivity (Fig. 3 dashed line). This suggests that the fact that the onset of the vocabulary spurt coincides with the peak of synaptic density is accidental, since a constant full connectivity still exhibits the nonlinearity at the same time in development. We explore the role of synaptic connectivity in more detail in the next experiment.

The Role of Synapto-Genesis and Pruning

In early childhood, the number of synapses increases during the first 2 years of life (Huttenlocher, 1979). The process is determined genetically and is essentially independent of the environment. The number of synapses remains constant up to about 10–15 years, when the number of synapses decreases, according to a pruning mechanism. Weak synapses are eliminated so as to keep the most relevant synapses. We model both synapto-genesis and synaptic pruning in order to identify their possible impact on word learning capacity. Synapto-genesis is modelled by having an increase in the number of *random* connections between the maps. Once maximal connectivity is reached (full connectivity in our model) weak synapses are eliminated. At different points in the process, and with well-structured maps, we assess the capacity to generalised a word-object association; the taxonomic factor.

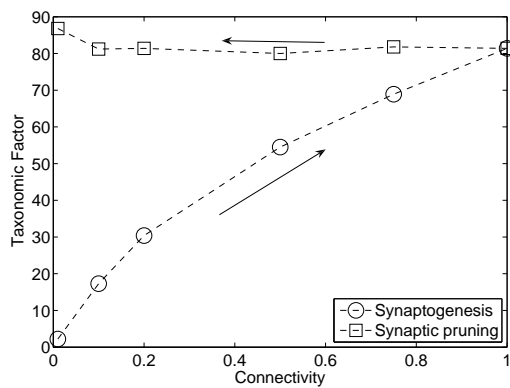


Figure 4: Role of synapto-genesis and pruning on word learning capacity. In early infancy the number of synapses increases, independently of the statistical regularities found in the environment. Generalisation of word-object associations improve with the number of randomly generated synapses. Later in life, when synapses are pruned according to their strength, we show that the ability to learn a word stays constant and even improves with a smaller number of synapses. The full process of synapto-genesis and synaptic pruning satisfies both metabolic constraints and word learning capacity.

Fig. 4 shows the taxonomic factor as a function of the synaptic connectivity between the maps, both with random connections (synapto-genesis) and with the strongest connections (synaptic pruning). First, there is a monotonic increase in taxonomic responding as the number of synapses grow. Taxonomic responding increases during the synapto-genesis period. Once the system has established the relative strengths

of synapses, the pruning process can take place. We see in Fig. 4 that pruning can be done without affecting word learning capacity. The taxonomic factor even peaks for a connectivity of about 0.01, when the normalisation procedure is not penalised by scaling irrelevant weights. It is interesting to note that synaptic pruning does not take place before the person has experienced most word-label associations, at around 15 years of life. The choice of which synapses to prune is then stable and can be done without affecting word learning performance.

Discussion

We have proposed a model of early word learning which is based on core aspects of the infants development. First, infants explore their environment and their perceptual systems become structured to reflect the regularities they experience in the environment, in the absence of constant supervision. Structure emerges in their visual system so that they are able to group together objects that look alike. Similarly, in the auditory system, the infant learns to identify word forms. The perceptual systems unfold in service of efficient word learning and through joint interactions (labelling events) with the caregiver, the infant learns to associate a label with its corresponding object. Based on a single labelling event (or very few of them), the infant excels at generalising this pairing to associate the word category to the object category.

Our model learns to categories words and objects in an unsupervised way, using self-organising maps. Once structure emerges, joint presentations of words and objects are generalised efficiently to associate all members of the word category to all members of the object category. Successful generalisation of word-object associations, following a single labelling event, build on “pre-lexical” categories formed in an unsupervised way by the maps. It is important to note that we provide a general mechanism for generalising associations between two categories, based on the association of one exemplar in each category. This process is independent of the mechanism of formation of these categories. Categories do not need to be purely perceptually based. The only assumption we make is that most members of a category are closer to each other than to members of another category. We showed that taxonomic responding, defined equivalently as the facility of learning a word from a single labelling event, correlates with the emergence of map structure. We thereby provide a mechanistic account of experimental findings showing a positive correlation of the infants’ perceptual systems with word learning capacity (Tsao, Liu, & Kuhl, 2004).

In addition, we have shown that an increase in the number of labelling events, or an increase in the time spent in joint attentional activities impacts favourably the size of the lexicon. Again, this provides a theoretical account of the positive correlation between joint attention and language acquisition (Tomasello & Farrar, 1986) and of the dependence of the vocabulary size on the social environment (high-talking versus low talking families) (Hart & Risley, 1995).

We also address the role of synapto-genesis and synaptic pruning with regards to word learning. In order to provide the best substrate for forming new associations between the visual map and the acoustic map, a high number of synapses is preferable, so that the system can rely on the presence of a connection between any neurones on each map. From this perspective the increase in the number of synapses observed in early childhood can be understood in terms of offering the system with a high number of possible connections to choose from, in order to form cross-modal connections. This process is considered to be hard-wired, independently of the structure of the environment. At the transition into adulthood, the number of synapses decreases according to the relative roles they play after being confronted with the environment. Weak synapses are pruned and only well-tuned synapses are kept. A reduction of the number of synapses lowers the metabolic demands of the brain. We showed that keeping only the strongest synapses can be achieved without affecting successful word-object associations. However, this selection process cannot take place before the learner has experienced most of the environmental conditions. If synaptic pruning takes place too early in development, the relative strengths of synapses are not optimally assigned and potentially useful synapses would be wrongly suppressed, inducing sub-optimal word learning capacity. The full developmental trajectory of “blind” synapto-genesis and later environmentally-driven synaptic pruning provides a biologically plausible access to the identification of the synapses required to be maintained. It is reasonable to suppose that such a developmental trajectory would be beneficial in perfecting the execution of a wide range of cognitive processes beyond language.

Conclusion

We have presented a model of word learning that reproduced some crucial aspects of the developmental trajectory of early lexical acquisition. We propose that successful word learning is facilitated by pre-lexical categorisation mechanisms. Through joint attentional activities, following a single labelling event, the association built through the simultaneous presentation of an object and its corresponding label is generalised to all objects of like kind and to all acoustic tokens corresponding to the same label. This generalisation of associations, commonly referred to as taxonomic responding, coincides with an increase in the pace of word acquisition. It provides a mechanistic account of the emergence of a “lexical constraint”, taxonomic responding, and highlights its role in word learning. This is achieved without postulating qualitative shifts in the learning mechanisms underlying the transition from early to later lexical development. We identified a developmental trajectory for selecting the best synapses, satisfying both metabolic optimisation and an excellent word learning capacity.

References

Chechik, G. (1998). Synaptic Pruning In Development:

- A Computational Account. *Neural Computation*, 10(7), 1759–1777.
- Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4), 357–366.
- Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*, 21, 125–155.
- Hart, B., & Risley, T. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Brookes Publishing Company, Inc., PO Box 10614, Baltimore, MD 21285-0624 (\$22).
- Huttenlocher, P. (1979). Synaptic density in human frontal cortex-developmental changes and effects of aging. *Brain Res*, 163(2), 195–205.
- Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.
- Kohonen, T. (1984). *Self-organization and associative memory*. Berlin: Springer.
- Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks*, 17, 1345–1362.
- Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic Self-Organization and Early Lexical Development in Children. *Cognitive Science*, 31(4), 581–61232.
- Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge: MIT Press.
- Markman, E. M., & Hutchinson, J. (1984). Children’s sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, 16, 1–27.
- Miikkulainen, R. (1997). Dyslexic and category-specific aphasic impairments in a self-organizing feature map model of the lexicon. *Brain and Language*, 59(2), 334–366.
- Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: two modes of word acquisition? *Developmental Science*, 6(2), 136–142.
- Posner, M., Goldsmith, R., & Welton Jr, K. (1967). Perceived distance and the classification of distorted patterns. *J Exp Psychol*, 73(1), 28–38.
- Quine, W. V. O. (1960). *Word and object*. Cambridge, Mass.: MIT Press.
- Tomasello, M., & Farrar, M. (1986). Joint Attention and Early Language. *Child Development*, 57(6), 1454–1463.
- Tsao, F., Liu, H., & Kuhl, P. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Development*, 75, 1067–1084.
- Waxman, S., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, 29, 257–302.