Can Neural Adaptation Explain Word Choice?

Jörn Klinger (jklinger@mail.utexas.edu)
Department of Linguistics, University of Texas at Austin,
Austin, TX 78712 USA

Julien Mayor (j.mayor@bcbl.eu)
Basque Center on Cognition, Brain and Language, Donostia - San Sebastian, Spain.

Abstract

Speakers refer to objects using terms on various levels of description. Labeling a poodle, they use the subordinate term "poodle" or the basic term "dog". Our model attributes these effects to visual context, relying on the domain general mechanism of neural adaptation. Two SOMs represent visual and auditory categories. Word learning is modeled via simultaneous presentation of item and word form and the creation of Hebbian synapses. Neural adaptation causes a decrease of activation in repeatedly activated nodes. We predicted that as a result of this, when presented alone or alongside a distractor from another basic category, an item will be referred to by its basic term, while the presence of a distractor from the same basic category will induce a shift to the target's subordinate label. Three simulations taking into account the relative frequency of an item's basic and subordinate level labels supported this hypothesis.

Keywords: Self-Organizing Maps; Neural Adaptation; Language Production.

Introduction

The language acquisition literature contains much evidence that children are reluctant to learn multiple labels for objects (Markman, 1990). They nonetheless grow up to become adult speakers who competently use multiple terms of reference for a given item. One group of words that requires mastering the use of multiple labels for an item are hierarchical category level terms ("dog" and "poodle"). They can all describe the same referent (e.g. a poodle), yet, in most cases, they are not mutually exchangeable – see Rosch & Mervis' seminal paper (1975). Research in language acquisition suggests that context can help children learn second labels for familiar items (Grassmann & Tomasello, 2010) and that children can tailor their utterances to meet the requirements of functional context (Deutsch & Pechmann, 1982; Matthews, Lieven, Theakstone & Tomasello, 2006; Matthews, Lieven & Tomasello, 2007; Bannard, Klinger & Tomasello, under revision). The present research is also concerned with adjusting one's utterance to the requirements of the context, but focuses on adult speakers. Our hypothesis is that adult speakers' use of basic and subordinate terms of reference (i.e. when to say "dog" or "poodle") is modulated by visual context. If a target item is seen alone or among unrelated items production of a basic level term of reference is sufficient to unambiguously identify the target. If the target is seen amidst very similar objects a more precise description is needed to unambiguously identify the target, thus a switch to the subordinate level of reference is expected. This can be modeled via the simple domain-general mechanism of neural adaptation.

This hypothesis is made explicit through a neuro-computational model, which simulates a speaker facing the previously described situation and implements neural adaptation.

Figure 1. Illustration of a visual context in which the basic level term is sufficient to identify a target object (left) and one in which a shift to a subordinate level term is required (right).

Modelling background

The model presented here builds on a tradition of supervised connectionist models of word learning and production, but is mostly based on Mayor & Plunkett's unsupervised account (2010). While word learning itself is considered a supervised activity, the acquisition of perceptual categories is not, i.e. does not require a label or a teacher. Hence Mayor and Plunkett adopted an architecture using SOMs (self-organizing maps; Kohonen, 1984, also interpreted in biological terms; Kohonen, 1993), which can account for taxonomic responding and fast mapping, while having unsupervised category formation.

Self-organizing maps are topological maps that extract statistical regularities from input and thereby effectively cluster objects, which have common properties. After self-organization is complete, similar objects activate neighbouring units on the map. Connections between two SOMs can be modulated by the activity of individual neurons on each map via Hebbian synapses (Hebb, 1949).

Mayor and Plunkett assumed that pre-lexical categorization and joint-attention events are crucial factors in word learning, hence their model consists of two SOMs (visual and auditory), which are organized by their respective input before any labelling events takes place.

After self organization, joint attention is mimicked through the simultaneous presentation of an object to the visual map and a word form to the auditory map. Connections between the activated neurons on each map are strengthened via Hebbian learning. Due to the topological structure of the SOMs, many neighbouring units are activated on each map and Hebbian learning occurs on the synapses connecting these units as well. This allowed for a single labelling event to be sufficient to induce taxonomic responding. In addition, the model also mimicked a typical learning behaviour, e.g. a vocabulary spurt, a reduction in over-extensions with increasing vocabulary, etc.

In spite of its accurate account of the above-mentioned phenomena, Mayor & Plunkett (2010) do not account for context-dependent use of hierarchical category terms. The aim of the present model in now to mimic the modulation of word choice in terms of low-level neural mechanisms; using the simple domain general mechanism of neural adaptation. Neural or sensory adaptation (see e.g. Kaplan, Sontagand & Chown; Grill-Spector, Henson & Martin, 2006) is a decrease over time in the responsiveness of the sensory system to a constant stimulus. On a behavioural level it surfaces as habituation, causing participants to be less and less sensitive to a repetitive stimulus. In terms of modelling, it means that a node that receives activation multiple times within a short period of time will experience a decrease in activation. Thus, the mechanism would explain hierarchical word choice in context as follows: When two items from the same basic category (e.g. a Poodle and a Labrador) are presented to the model at the same time, they both activate their basic label (e.g. "dog" is activated twice) and their respective subordinate labels (e.g. "Poodle" and "Labrador" both are activated once). Neural adaptation would then cause the activation of the twice-activated basic label to decrease, resulting in a subordinate term winning the competition and being produced.

Method

Similarly to Mayor & Plunkett (2010), the model consisted of two separate SOMs, one visual and one auditory, which received visual and auditory input respectively. We also assumed that when it comes to word learning, infants have already acquired the ability to segment objects out of complex visual scenes (Kellman, Spelke & Short, 1986; Kaufmann-Hayoz, Kaufmann & Stucki, 1986) as well as labels from a flow of speech (Jusczyk & Aslin, 1995; Saffran, Aslin & Newport, 1996). Each SOM was formed via presentation of the respective set of items. These visual stimuli were 20-dimensional real-valued vectors, which were clustered in 2 basic level categories (e.g. "dog" and "car"), each consisting of 10 subordinate categories (e.g. "poodle", "limousine"), which were, in turn, made up out of 20 exemplars each (i.e., for instance, individual poodles or limousines).

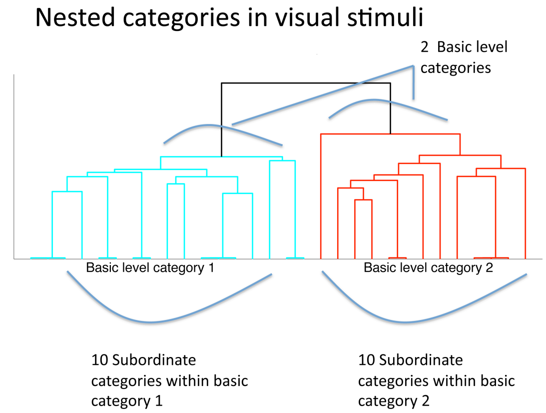

Figure 2. Graphical representation of the nested categories in the visual data. The large turquoise and red stripes represent the 2 basic categories, while the little feet on the bottom of the diagram represent the subordinate categories within the two basic categories.

In creating the stimuli we used a simple randomization algorithm that ensured that subordinate categories were closest to their respective basic category's prototype and that, in turn, exemplars of each subordinate category were closest to their respective subordinate prototype and accordingly closer to their respective basic category prototype than to the other basic category's prototype.

The auditory stimuli, in contrast, were not organized in such a nested fashion. Whereas visual categories display a hierarchical structure, such that "Labrador" shares many features with other subordinates of the broader category "dog", this is not necessarily the case for auditory categories. Similar sounding words are clustered together, but there is usually no sound symbolism such that "car" is a member of a hypothetical broader category "carpet". It has to be noted though that certain terms referring to an item on a subordinate level such as racing-car share parts with their respective basic level label (in this case "car"). In its current state the model assumed that such labels have separate entries on the auditory map. Whether this way of storage reflects reality is a matter of experimental investigation, but not central to the current model.

All visual and auditory stimuli were represented as 20-dimensional vectors. Coding the input in this abstract way, allowed us to remain agnostic to the nature of attributes involved in category formation. The total number of 400 visual and 440 auditory items, after excluding prototypes, was then presented to train the respective SOMs, mimicking exposure to objects to words. These uni-modal self-organising maps used the standard Kohonen learning algorithm (Kohonen, 1984) - Each map consisted of a hexagonal grid of 64 (8x8) units. Each unit k was associated with a vector m_k. Upon presentation of each item x the vectors m_k (like in Mayor & Plunkett, 2010) were modified by finding the Best Matching Unit (BMU) c, defined by the following condition:

$$||m_c(x) - x|| \leq ||m_j - x|| \qquad \forall j \tag{1}$$

with $\| \bullet \|$ measuring the standard Euclidean distance. Similarly, the second and third BMU could be identified. Then the standard weight update rule was applied with a learning rate that decayed over time $a(t) = 0.05/1 + t/200$ and a Gaussian neighbourhood function of the distance d_{ik} between units i and k on the map (Equation 2), that shrinks linearly over time from $\sigma(0) = 4$ to $\sigma(T_{\max}) = 1$.

$$N(i,k)_t = e^{-d_{ik}^2/2\sigma^2(t)} \tag{2}$$

An average quantization error was defined, such that the Euclidean distance between input patterns and their respective BMU was:

$$\left\langle \| x - m_c(x) \| \right\rangle_x \tag{3}$$

where $m_c(x)$ is the best matching unit for input pattern x and $< \bullet >$ indicates an averaging over all input patterns. This quantization error E is not a traditional error teacher signal, but a global measure of weight alignment to the input in the map. In forming the SOMs, we used a batch version of Kohonen's algorithm (1984).

Then associations across SOMs were trained. This only happened after the maps were entirely formed, emulating that the speaker had fully formed visual and auditory categories. Such a simplification was possible, since the current research did not focus on developmental aspects. We mimicked joint attention activities (i.e. labelling events) between caregiver and infant by simultaneously presenting an item from the visual set to the visual map and a random exemplar from the respective auditory category to the auditory map. Each visual stimulus was in this way associated with both, a subordinate and a basic level label. Thus, each instance of the visual stimulus "poodle" was simultaneously presented and therefore associated with a number of instances of the word form "dog", as well as a number of instances of the word form "poodle". This way, the model was able to learn that two labels can refer to the respective item.

We built cross-modal connections by learning Hebbian connections between both maps. The amplitude of these bidirectional connections is modulated by the activity of the connection units. We defined the neural activity of a unit k to be $a_k = e^{-q_k/r}$, where q_k is the quantization error associated with unit k and $r = 0.5$ was normalization constant. The amplitudes of those connections were modulated according to the standard Hebb rule with saturation, which allows for keeping weights within physiological range even for high neural activities (Mayor & Plunkett, 2010). The connections between unit i on the visual map and unit j on the auditory map was computed as follows:

$$w_{ij}(n+1) = w_{ij}(n) + 1 - e^{-\lambda a_i a_j} \tag{4}$$

where n refers to the index of the item-word pairing and $\lambda = 0.3$ is the learning rate. It was set to that value, since it offered a good compromise between quick learning and establishing many meaningful connections.

Since it was the model's objective to emulate experienced speakers who knew all the items and their corresponding hierarchical word forms, the model's performance was only tested after cross-modal associations between items and word forms had been fully established. The frequency ratio between basic and subordinate labelling events differs from object to object. We mimicked this change in frequency by training cross-modal associations with different ratios of basic and subordinate word forms in each simulation.

Neural adaptation

Neural adaptation was implemented into the auditory map by making the activation of a unit a function of newly received and previous activation, such that multiple activation of a unit within a short period of time leads to a drastic decrease in activation:

$$a_n^{NA}(s) = a_n(s) * e^{(-a_n^{NA}(s-1)/tau)} \tag{5}$$

where $a_n(s)$ is the activation of a unit n at presentation number s, $a_n(s)$ is the activation of a unit before adaptation is applied and tau is a constant determining the strength of adaptation with a low tau value standing for drastic and a high tau value standing for mild adaptation. Thus, if a stimulus has never been presented before, the activity of the corresponding nodes at time s-1 is virtually 0; adaptation has no effect on word choice. However, after a number of presentations, adaptation reduces the activation level for the nodes subjected to repeated stimulation. Note that neural adaptation was implemented into the model at the level of the cross-modal activation flow transferred via Hebbian connections, and not to activation arising from direct presentation of stimuli to the maps. While this facilitated making the model work, it also appears to be plausible, since quickly-occurring drastic adaptation of visual nodes that receive activation through the presentation of visual stimuli seems counter intuitive - when looking at more than one dog, one certainly still sees the dogs.

Simulation 1

This simulation investigated possible effects of visual context via the implementation of neural adaptation. The SOMs were trained according to the previous section; thereafter associations between SOMs were established. In this particular simulation, cross-associations were formed using a ratio of 85% basic level and 15% subordinate level labelling events for each visual stimulus, thereby representing items for which caregivers predominantly use basic level terms in labelling situations. For example, imagine an object "poodle", which would typically be named "dog" in 85% and "poodle" in 15% of the labelling events in acquisition.

This simulation attempted to mimic the situation of a speaker seeing two objects from the same basic, but from different subordinate categories (e.g. a poodle and a Labrador, see Figure 1, right image) who named one of those objects using either a basic or subordinate level term. Thus, to evaluate the model's performance, 400 pairs of visual stimuli were presented to the visual map. Each pair consisted of a distracter (presented first) and the target (presented one time step after). The distracter was always from the same basic, but never from the same subordinate level category, as the target. It was then noted, whether the model produced a correct target label upon presentation of a pair of visual stimuli and whether the word produced was a basic or subordinate level term. In a control condition, distracter and target belonged to different basic level categories (see Figure 1, left image), simulating a speaker wanting to identify a referent presented together with an unrelated object.

Several simulation runs were undertaken, covering adaptation rates for tau values from 0.15 (weak adaptation) to 0.01 (very strong adaptation).

Simulation 1 – Results & Discussion

The results for items with a high frequency of basic level labelling events (85%) in acquisition suggested that with only weak adaptation (*tau* < 0.08) the basic level label is preferred when a target is shown together with a distracter from the same basic level category (see Figure 3).

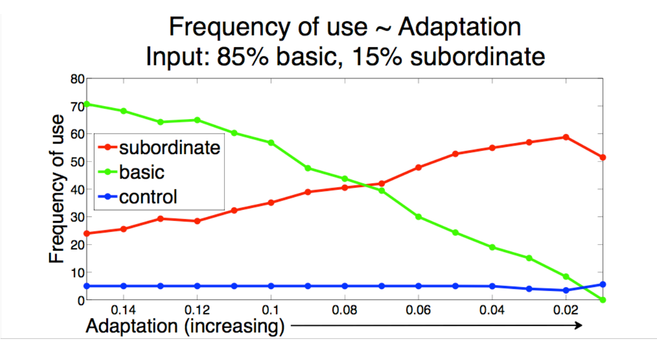

Figure 3. Mean frequency of use of an item's basic and subordinate level labels for different rates of adaptation across 100 simulation runs (10 per tau value). The red line represents the percentage of subordinate, the green line the percentage of basic terms produced. The blue line indicates the number of subordinate terms produced in the control condition.

For tau = 0.15, the model produced 70.07% basic and 23.95% subordinate level terms. As adaptation rate increased this ratio shifted towards production of subordinate level terms. For tau = 0.07 the model produced an almost equal amount of basic (43.75%) and subordinate level terms (40.5%). The remaining 20% were mapping errors. As adaptation became even more powerful, subordinates were produced more frequently than basic level terms, e.g. 8.4% basic and 58.75% subordinate terms for tau = 0.02. While not taking away from the model's overall predictions, it still has to be noted that error rate increased as neural adaptation became stronger, indicating that adaptation added some noise to the system.

In the control condition, where a target was shown alongside a distractor from a different basic level category, almost no subordinate terms were produced, regardless of power of adaption. The simulation predicted that, given a sufficiently high rate of adaptation, visual context would prompt a shift from basic to subordinate level terms used when labelling an item.

Such a behaviour was produced through neural adaptation: When a target and a distracter from the same basic level category were presented to the model's visual map, they – via Hebbian connections - both activated the same basic level term on the auditory map and each their respective subordinate level term. Neural adaptation then caused that the units representing the basic level term ("dog"), due to being activated multiple times, to exhibit a strong decrease in activation, such that the subordinate level term ("poodle") won the competition and was produced. In the control condition, target and distracter belonged to different basic level categories, such that no neurons were activated twice and no adaptation took place, such that the basic level terms won the competition.

The findings of this simulation hence suggest that a single domain-general mechanism, like neural adaptation, would be able to account for adult speakers' switch from basic to subordinate level terms of reference when unambiguously identifying a referent alongside a distracter from the same basic level category and the absence of such an effect when both items belong to different basic categories.

Simulation 2

Unlike the first simulation, this simulation dealt with items that were mostly labelled with their subordinate level term during acquisition (e.g. the object "eagle", which would be referred to with "bird" less often than with its subordinate term "eagle". The SOMs were formed in the same way as in simulation 1, but trained cross-modal associations using a ratio of 20% basic and 80% subordinate level labelling events for each. The models performance was also evaluated in the same way as in Simulation 1.

Simulation 2 – Results & Discussion

For items that are predominantly associated with subordinate level terms during labelling events, we observed a higher amount of subordinate labels (above 60%) in production, while the frequency of basic level labels was overall lower, as can be seen in Figure 4.

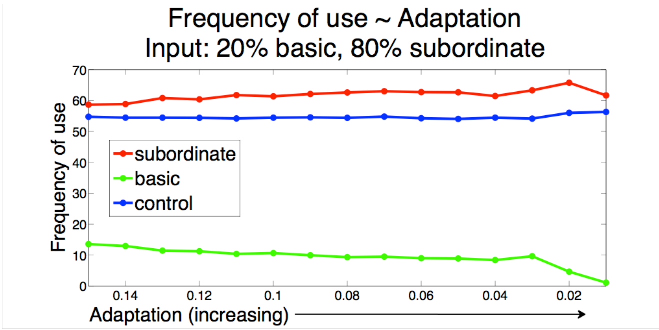

Figure 4. Mean frequency of use of an item's basic and subordinate level labels for different rates of adaptation. The red line represents the percentage of subordinate terms produced, the green line the percentage of basic terms produced. The blue line indicates the number of subordinate items produced in the control condition.

Higher rates of adaptation further increased the use of subordinate level terms, while the use of basic level terms decreased to a minimum. A slightly higher number of false labels was produced.

The high percentage of subordinate level terms produced, even for lower rates of adaptation, suggested that contextual effects were less pronounced for items that are referred to mostly by their subordinate term in labelling events. The lesser increase in use of subordinate terms with increasing adaptation rate, can be explained in terms of functional demands of the context: The subordinate level term is already sufficient to distinguish the item in question from other members of the same basic level category, such that context does not require the model or the speaker to shift to an even more specific term. It is worth noting that in the control condition, around 55% of subordinate terms were produced, regardless of adaptation. This suggested that an item's frequency of basic and subordinate labels in production was – independent of context and neural adaption – a function of the frequency of basic and subordinate terms used in labelling events for that item in acquisition. This prediction was further explored in a third simulation.

Simulation 3

The third simulation set out to investigate the effect of frequency of basic and subordinate level terms for an item in acquisition on the frequency of use of the respective terms in production. After training the SOMs, cross-modal associations were formed via Hebbian learning upon simultaneous presentation of visual and auditory stimuli to the respective SOM. Each item was associated with both a basic and a subordinate category level term. Thereafter, we had the model form cross modal associations for a variety of ratios of basic and subordinate labels for an item. These ranged from 90% basic and 10% subordinate level labelling events to the inverse where 10% of the labelling events associated basic and 90% associated subordinate level terms. For each ratio, the model was then tested by presenting each visual stimulus separately and checking whether the basic or subordinate label was produced. No adaptation was implemented.

Simulation 3 – Results & Discussion

The findings of simulation 3, visualized in Figure 5, indicated that the ratio of use of basic and subordinate labels for an item was a function of their frequency ratio in acquisition. Thus, the simulation predicts that when disregarding possible effects of context and not implementing neural adaptation, the model will use hierarchical category level terms based on their frequency in acquisition when presented with a single visual stimulus.

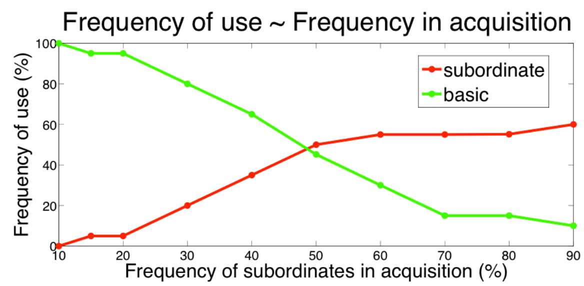

Figure 5. Mean frequency in acquisition plotted against frequency of use in production across 90 simulation runs. The green line indicates percentage of use of basic, the red line percentage of use of subordinate level labels in production.

As seen in Figure 5, a very low frequency of subordinate labels (less than 10%) for an item in acquisition lead to the model not producing the subordinate form in a neutral context at all. In the cases where the frequency of subordinates in acquisition is between 15% and 45%, production frequency of subordinates gradually increased. A 50/50 frequency ratio in production was reached when the frequency ratio in acquisition was also at around 50/50. Further shifting this ratio in acquisition towards subordinates (50% to 90% subordinate labelling events) reduced the production frequency of basic level terms, while slightly increasing the frequency of subordinate terms. As the frequency of subordinates in acquisition rose well over 50%, the model attempted to produce more subordinate terms, but also started to produce an increasing amount of errors - with a frequency ratio of 10% basic and 90% subordinate level terms in acquisition, the model produced 10% basic and 60% subordinate level terms, plus 30% incorrect subordinates.

The main findings of this third simulation are plausible both with regards to the model's architecture as well as in terms of implications for speakers' behaviour. Since the frequency with which an item is associated with a word during formation of the cross-modal associations has a direct impact on the strength of the respective connection, we expected more frequent labelling events associating either a basic or subordinate level label with an item to result in a propensity for referring to that item with the respective hierarchical term in production. The prediction is that speakers would display a similar behaviour.

General Discussion & Conclusion

The simulations reported above showed the following:
1) When labelling an item presented together with a distracter from the same basic level category, the model shifts from basic to subordinate level terms.
2) A simple domain general mechanism, like neural adaptation, can account for this shift from basic to subordinate level labels prompted by visual context.
3) The amplitude of the effect of adaptation is stronger for items with a higher frequency of basic level labelling events in acquisition than for those with a lower frequency of basic level labelling events in acquisition.
4) Frequency ratio between basic and subordinate terms in production is a function of the frequency of basic and subordinate labelling events for an item in acquisition.

These predictions can be tested in further experimental studies. The simplest premise, the switch to subordinate level terms in the presence of a distracter from the same basic category, requires a simple picture-naming task in which participants are required to unambiguously identify a referent alongside either an item of the same or a different basic category. To see whether neural adaptation is a plausible explanation for effects of visual context, one could contact a very similar study in which participants always name an item by its basic label. If adaptation occurrs, a difference in responses should arise if the target is presented alongside a distracter from the same basic level category and thus forcing a subordinate label.

In terms of limitations, when presented with two subordinate items (e.g. two poodles) the current model would fail to revert to a level of description more specific than the subordinate. In that situation, humans might use adjectives to reach a finer level of description, the model would not. However, this could in principle be accommodated by training the model with labels at multiple levels in the hierarchy; adaptation would then cascade down and naming would become more and more specific as the context becomes stronger.

In conclusion, we have presented a neuro-computational model that predicts a shift from basic to subordinate level terms of reference for items driven by visual context by relying on the single domain-general mechanism of neural adaptation. The present research is a low-level associative account of a phenomenon so far described on a high socio-pragmatic level, but makes no claims about the nature of the relationship between such low-level mechanisms and richer processes, nor is it concerned with high-level processes such as intention reading or inferences about goals and mental states of speakers. It rather attempts to show that high- and low-level accounts need not be antagonistic, but can complement each other.

References

Bannard, C., Klinger, J. & Tomasello, M. 3-year-olds' copying of novel linguistic material is sensitive to semantic and pragmatic factors but is not always fully insightful Under revision.

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition*, 11, 159 – 184.

Grassmann, S. & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science*, 13, 252-63.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, 10, 14-23.

Hebb, D. (1949). *The organization of behavior: A neuropsychological theory*. New York, NY: John Wiley & Sons.

Jusczyk, P., & Aslin, R. N. (1995). Infant's detection of sound patterns of words in fluent speech. *Cognitive Psychology*, 29, 1–23.

Kaplan S., Sontag, M. and Chrown, E. (1991). Tracing recurrent activity in cognitiveelements (TRACE): A model of temporal dynamics in a cell assembly. *Connection Science*, 3, 179-206.

Kellman, P., Spelke, E., & Short, K. (1986). Infant perception of object unity from translatory motion in depth and vertical translation. *Child Development*, 57, 72-86.

Kohonen, T. (1984). *Self-organization and associative memory*. Berlin: Springer.

Kohonen, T. (1993). Physiological interpretation of the self-organizing map algorithm. *Neural Networks*, 6, 895 – 905.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, 14, 57–77.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics* , 27, 403-422.

Matthews, D., Lieven, E., & Tomasello, M. (2007). How toddlers and preschoolers learn to uniquely identify referents for others: A training study. *Child Development*, 76, 1744-1759.

Mayor, J. & Plunkett, K. (2010). A neuro-computational account of taxonomic responding and fast mapping in early word learning. *Psychological Review*, 117, 1-31.

Rosch, E. & Mervis, C. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, 7, 573-605.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.