

# SEQUENCE DETECTORS AS A BASIS OF GRAMMAR IN THE BRAIN

Friedemann Pulvermüller

*MRC Cognition and Brain Sciences Unit, Cambridge*

*Running head:* Sequence Detectors

*Key words:* associative learning, cell assembly, grammar, movement detection, neuron, recursion, serial order, substitution, syntax

Includes 7.074 words; 3 Figures; 1 Table

Address for correspondence:

Friedemann Pulvermüller, Ph.D.

MRC Cognition and Brain Sciences Unit

15 Chaucer Road

Cambridge CB2 2EF, England

phone: +44 1223 355294 ext. 770

fax: +44 1223 359062

e-mail: [friedemann.pulvermuller@mrc-cbu.cam.ac.uk](mailto:friedemann.pulvermuller@mrc-cbu.cam.ac.uk)

## Abstract

Grammar processing may build upon serial-order mechanisms known from non-human species. A circuit similar to that underlying direction-sensitive movement detection in arthropods and vertebrates may become selective for sequences of words, thus yielding grammatical sequence detectors in the human brain. Sensitivity to the order of neuronal events arises from unequal connection strengths between two input units and a third element, the sequence detector. This mechanism, which critically depends on the dynamics of the input units, can operate at the single neuron level and may be relevant at the level of neuronal ensembles as well. Due to the repeated occurrence of sequences, for example word strings, the sequence-sensitive elements become more firmly established and, by substitution of elements between strings, a process called auto-associative substitution learning (AASL) is triggered. AASL links the neuronal counterparts of the string elements involved in the substitution process to the sequence detector, thereby providing a brain basis of what can be described linguistically as the generalization of rules of grammar. A network of sequence detectors may constitute grammar circuits in the human cortex on which a separate set of mechanisms establishing temporary binding and recursion can operate.

To understand the biological basis of grammar, it is necessary to recur to mechanisms of serial order processing in the nervous system. Some serial order circuits are common in the nervous system of animals and it would be surprising if the human grammar system failed to exploit them. This article highlights one circuit type, discusses mechanisms that may underlie its specificity for serial order and asks whether it can contribute to an explanation of the human ability to generalize grammar rules from a limited input of sentences. The results are related to current efforts to specify the neurobiological basis of grammar in the human brain.

### *Movement detectors*

Neurons can specifically respond to sequences of events, for example to the order of activation of two sensory cells. A simple mechanism for achieving serial order sensitivity is by means of *delay lines*, so that the two sensors, called  $\alpha$  and  $\beta$  here, feed into a third neuronal element, called  $\gamma$  (Figure 1A). If the connection from  $\alpha$  to  $\gamma$  is slower than the one from  $\beta$  to  $\gamma$ , the third element,  $\gamma$ , will be more likely to fire in response to the sequence “first  $\alpha$ , then  $\beta$ ” than to the inverted one. The stimulation of  $\gamma$  will be maximal if the conduction delay difference matches the actual time difference between the activation of sensor  $\alpha$  and that of  $\beta$  (Kleene, 1956; McCulloch & Pitts, 1943).

Whereas delay-line architectures are capable of measuring exact temporal delays, an important question in linguistics is sometimes in which *order* words occurred in the input – irrespective of the exact time lag between their occurrences. If the signal provided by input unit  $\alpha$  is passed through a low pass filter, it will be delayed and stretched in time so that the sum of the two inputs to  $\gamma$  will still be largest when  $\alpha$ 's activation precedes that of  $\beta$ . However, in contrast to the delay line mechanism, there will also be flexibility with regard to

the delays. Because the low-passed-filtered signal from  $\alpha$  has now a broad maximum, the delay between the activation of  $\alpha$  and  $\beta$  can vary within certain bounds while still allowing  $\gamma$  to extract the information which of the sensors was activated first.

Serial-order mechanisms of these kinds have been found in various animals. The filter model first proposed for the visual system of insects (Reichardt & Varju, 1959; Varju & Reichardt, 1967) postulated that the filtered signal from  $\alpha$  and that from  $\beta$  are submitted to a multiplication element whose output provides information about the movement direction of perceived objects (Egelhaaf, Borst & Reichardt, 1989). Related mechanisms involving inhibitory circuits exist in the retina of mammals (Barlow & Levick, 1965; Barlow, Hill & Levick, 1964). The direction sensitivity of neurons in the visual cortex is likely due to intra-cortical interactions between excitatory and inhibitory effects on sequence-detecting complex cells (Hubel, 1995). A structure in which sequence detectors primarily depend on exact delay lines may be the cerebellum, where the output of Pukinje cells includes information about precisely timed sequential input to parallel fibers (Braitenberg, Heck & Sultan, 1997; Heck, 1993; Heck & Sultan, 2002). Delay line architectures also play a role in the generation of precisely timed sequences of neuronal firing (Abeles et al., 1993; Bienenstock, 1996; Prut et al., 1998) and may therefore be relevant for the programming of action sequences and for the detection of stimulus sequences as well.

### *Sequence detection through links with different weights*

Whereas delay lines, filter properties and non-linear interactions can enable the circuit in Figure 1A to detect the order of neuronal events, a most elementary mechanism exists that serves the same purpose. If two neuronal elements  $\alpha$  and  $\beta$  both project to a third element  $\gamma$

and the two *connections have different weights*, the third element  $\gamma$  will likely process information about the serial order of its inputs. Consider the case where the activation functions of both  $\alpha$  and  $\beta$  fall off exponentially<sup>1</sup> with time and the stronger of the two inputs to  $\gamma$  is just not sufficient to activate it. In this case, the stronger input following the weaker one is likely to make the third neuronal element exceed its firing threshold, whereas the weak input following the strong one will be without effect (Figure 1B). The critical variable is the amount of decay during the delay  $\Delta t$  with which the two input units fire. Trivially, this decay is stronger for the stronger input than for the weak one. For this very reason, the strong input placed after the weak one is more effective in firing the third element than the reverse sequence.

In a neuron network such as the neocortex, projections of two individual neurons on a third one are common. Among the many neuron triplets that are connected in this way, there is probably a minority for which both connections are identical. The more usual case is that one connection dominates the other one, thus providing the circuit for a potential sequence detector. If only pyramidal cells, the most common neuron type in the cortex (Braitenberg & Schüz, 1998), are considered and their interactions through glutamate synapses, it may be appropriate to assume that EPSPs have similar decay slopes. Ignoring their initial brief escalation, they can be approximated by exponential functions. For the model to operate, the input units must not usually fire simultaneously, but with a minimal delay  $\Delta t$  or larger. This condition is satisfied if a serial input of discrete elements is considered, for example a sequence of discrete phonemes or words in spoken language. Many cortical cells may therefore be sequence detectors due to links with different weights.

---

<sup>1</sup> The mechanism may be effective for linear decay functions as well. However, exponential decay seems to be

It is relevant to describe properties and limits of the outlined mechanism of sequence detection which, as will be argued below, may be relevant at the single cell level and at the cell assembly level as well. The input  $I$  to the third element, the sequence detector  $\gamma$ , can be described as

$$I = w e^{-A t} \quad (1)$$

where  $w$  is the weight of the connection between input unit and sequence detector,  $A$  is a constant (which, for the sake of simplicity is assumed to equal 1), and  $t$  is time ( $A, t \geq 0$ ). The two input units  $\alpha$  and  $\beta$  project to  $\gamma$  with different strengths or weights,  $w_\alpha < w_\beta$ . The input to  $\gamma$  that determines its firing can be calculated according to Eq. (2) if  $\alpha$  fires before  $\beta$ , and according to Eq. (3) if  $\alpha$  fires after  $\beta$ .

$$I_{\alpha\beta} = w_\alpha e^{-A \Delta t} + w_\beta \quad (2)$$

$$I_{\beta\alpha} = w_\beta e^{-A \Delta t} + w_\alpha \quad (3)$$

where  $\Delta t > 0$  is the delay between the firing of the two input units. Since  $w_\alpha < w_\beta$ ,  $w_\beta$  can be expressed as  $w_\alpha + c$ , where  $c$  is a constant  $> 0$ .

$$I_{\alpha\beta} = w_\alpha e^{-A \Delta t} + w_\alpha + c$$

$$I_{\beta\alpha} = (w_\alpha + c) e^{-A \Delta t} + w_\alpha = w_\alpha e^{-A \Delta t} + w_\alpha + c e^{-A \Delta t}$$

---

more realistic neurophysiologically, at the level of both single neurons and neuronal ensembles.

Since  $e^{-A \Delta t} < 1$ , it follows that  $I_{\alpha\beta} > I_{\beta\alpha}$ . Thus, unequal connections of two sensors to a third neuronal unit imply that the third element receives more input and is more likely to respond to one order of sensor activation than to the reverse sequence.

To obtain a functional sequence detector, the activation threshold  $\theta$  of  $\gamma$  must be greater than the stronger of the two inputs, thus  $\theta > w_\beta$ . Further, it should be excluded that a weak input riding on a strong one can cause  $\gamma$  to reach the threshold. This requires that

$$w_\beta e^{-A \Delta t} + w_\alpha < \theta \quad \Rightarrow \quad \Delta t > 1/A \ln \{w_\beta/(\theta-w_\alpha)\} \quad (4)$$

This provides a lower bound for the lag between the firing of the two input units, which depends on the relationship between  $\theta$ ,  $w_\beta$ , and  $w_\alpha$ . The lag cannot become too long as well. If the strong input follows the weak one, the remainder of the latter must still be strong enough so that the two together can exceed the threshold.

$$w_\beta e^{-A \Delta t} + w_\alpha > \theta \quad \Rightarrow \quad \Delta t < 1/A \ln \{w_\alpha/(\theta-w_\beta)\} \quad (5)$$

Evidently, the closer  $\theta$  and  $w_\beta$  are together, the longer the delays between the inputs can become. Figure 1C illustrates the input to  $\gamma$  within the limits defined by Eqs. (4) and (5). Within these limits,  $\gamma$  shows above-threshold activation when the input sequence  $\alpha\beta$  is present, but below-threshold activation to the reverse sequence,  $\beta\alpha$ .

*Sequence detection among neuronal ensembles*

In the nervous system, sequence detectors may operate at different levels. Apart from individual neurons, for example cortical pyramidal cells, larger neuron aggregates may operate as sequence-detecting devices. Sequence detection on the basis of different weights of connections between individual neurons is not likely to be robust. Excitatory post-synaptic potentials (EPSPs) decay within 20 to 50ms and two EPSPs arriving asynchronously can thus only momentarily exceed the firing threshold of the third neuronal element. Such brief stimulation is unlikely to produce a sustained and reliable output indicating that a sequence has been detected. However, if a large group of neurons all connected in the same way to pairs of input units are stimulated by a given sequence, a slight temporary enhancement of the firing probability in each of them may produce a reliable network response. Neuronal ensembles including up to  $10^6$  neurons have been proposed as information processing devices in the cortex (see Palm, 1993). Within such a strongly connected population, the enhancement of the firing probability of a subset of the population can lead to the reliable activation of the entire ensemble (Braitenberg, 1978; Palm, 1982; Sakurai, 1999; Sommer & Wennekers, 2001). It may therefore be that sequence detection relying on unequal weights is more robust at the level of neuronal ensembles than at that of single neurons.

Sequence detection at the ensemble level may have an additional advantage that is of special relevance for language processing. Due to reverberating activity among its neurons, ensembles may exhibit a substantially slower decay of activity than a single isolated cell, thus allowing for sequence detection on a much longer time scale. Slow decay of activity has been observed in neurophysiological studies of so-called memory cells that can be temporarily activated in short-term memory tasks. During the memory interval, some of these cells show activation and subsequent exponential decay of activity over tens of seconds (Fuster, 1995).

Surprisingly, memory cells with high stimulus-specificity were found in different cortical areas distant from each other, for example in the inferior frontal and inferior temporal lobes. It has therefore been suggested that memory cells are part of distributed neuron ensembles processing specific stimulus information that has to be kept active in memory (Fuster, 1997). If memory cells with similar stimulus-specificity are found in two areas, local cooling of one of the areas may disturb neuronal dynamics and stimulus specificity in the other area (Fuster, 1997). Further evidence that reverberation in neuronal ensembles underlie the stability and slow decay of activity in memory cells come from simulation studies using associative memories (Zipser et al., 1993).

If the neuronal units in Figure 1A are conceptualized as cell assemblies including memory cells, the corresponding decay functions will span several seconds. A sequence detector at the ensemble level can therefore process information about sequences at a time scale relevant for spoken sentence processing. For the function of the ensemble model, it is essential that the two input units do not differ with regard to the exponential coefficients of their decay functions. Simulation results (for example Zipser et al., 1993) indicate that different association strengths between neurons may lead to variable dynamics. To obtain constant decay slopes in several ensembles in one network, it is therefore necessary that dynamics be determined by factors generally effective for all excitatory neurons, for example local inhibition or fatigue effects. Furthermore, it is important that the decaying ensembles do not substantially interfere with each other, which may require that they are spatially distinct. Without any doubt, it is an important endeavor to further investigate the relevant mechanisms by performing simulation studies of sequence detection at the ensemble level.

*Sequence detectors operating on words*

The postulate to be further explored below is the following: Neuronal ensembles that correspond to words and meaningful language units (morphemes) feed into sets of neurons that respond specifically to the order of their activation. These sequence detectors are an essential part of the brain mechanisms of grammar. Here is an outline of the local and global circuits that may be relevant.

If a local cluster of neurons is activated in a particular context, its neurons will fire. In the *immediate vicinity* of the cluster, its *halo* (Braitenberg, 1978), there are neurons that are normally not included in the activation process unless they receive some additional input. Such priming may be provided by neuronal activity related to the processing of the preceding context (Milner, 1957). Local neuron clusters included in ensemble  $\beta$  may therefore be surrounded by neurons of its halo that are context-dependent in the sense that they require priming by a different input in order to become active after an ignition of  $\beta$ . If assembly  $\beta$  ignites in the context of an earlier activation of ensemble  $\alpha$ , a subset of the neurons in the halo of  $\beta$ , called  $\gamma$ , may therefore become active. Neuroanatomical data show that between adjacent cortical neurons are more likely to exist and that they are, if present at all, more effective than long-distance connections between neurons far apart. Long-distance links reach apical dendrites of pyramidal cells where EPSPs are less effective in firing the post-synaptic cell than the inputs to basal dendrites through which local cortical connections operate (Braitenberg & Schüz, 1998). Therefore, halo neurons in the vicinity  $\gamma$  of a local neuron cluster of  $\beta$  receive strong input from  $\beta$  itself and additional weaker input from a distant cluster of  $\alpha$ . Figure 2 illustrates the envisaged arrangement in the cortex, by showing one representative neuron of each relevant neuron population. The stronger weights and larger

number of connections between  $\beta$  and  $\gamma$  as compared with the  $\alpha$ -to- $\gamma$  projections are the basis of sequence detection according to the mechanism discussed earlier. Each local cluster of neurons together with its immediate vicinity may correspond to the local compartments postulated by Braitenberg (Braitenberg, 2001) and to Hubel and Wiesel's hypercolumns (Hubel, 1995). It has been argued that the cortex can be modeled as a network with a near-complete set of connections between compartments each containing the square root of the total number of cortical neurons, that is,  $\sim 10^5$  in the human cortex (Braitenberg, 2001). This wiring allows for numerous (up to  $\sim 10^5$ ) temporal sequences of neuronal elements to be potentially represented in the vicinity of each local neuron cluster. It is therefore feasible that all possible links representing pairs of ten thousands of vocabulary elements have a corresponding neuronal link in the cortex at an initial state, and that only some of these links are being established due to frequent perception of the respective word sequences.

Neuroimaging data indicate that distributed neuron populations are the cortical correlate of individual words and morphemes (Pulvermüller, 1999). Each of these neuronal ensembles, or word webs (Pulvermüller, 2001), may include several local neuron clusters distributed over the perisylvian language areas of the cortex, and each of these local clusters may, in turn, have an immediate vicinity with context-sensitive neurons. When a spoken string of words is being perceived, the distributed cell assemblies may ignite in the order determined by the string in the input. Neurons sensitive to the order of activation of pairs of word-related ensembles may be localized in the perisylvian areas as well, in the immediate vicinity of the word-related networks. Now, a word sequence AB perceived by a listener would be accompanied by the following sequence of events in the listener's brain: First, the neuronal correlate of the word A, the neuron ensemble  $\alpha$ , becomes active. Second, the brain

correlate of the second word B, assembly  $\beta$ , ignites. Third, those neurons in the vicinity of the local clusters of ensemble  $\beta$  that happen to receive input from both  $\beta$  and  $\alpha$ , fire as well (Pulvermüller, 2002).

Sequence detectors sensitive to ordered word pairs may develop in a randomly connected network of neurons on the basis of correlation learning. Since the neurons sensitive to a particular word-pair sequence fire with good correlation when the respective sequence is encountered, the partaking neurons may strengthen their connections and develop into a functionally connected neuron set, the sequence detector at the ensemble level. Connections between neuronal groups in the cortex are usually reciprocal. Local connections do usually not exhibit directional preferences and most long range connections connect areas in both directions (Braitenberg & Schüz, 1998; Young, Scannell & Burns, 1995). This suggests that also the connections between sequence detectors and word-related networks are reciprocal.

### *Substitution of language elements in sentences*

Modern linguists have always made the important point of the sparseness of the language input (the infant encounters only a restricted set of sentences) and its apparent discrepancy with the child's production of virtually endlessly many new sentences. Because the child can produce and understand new sentences it never encountered before, it is impossible that all serial order knowledge relevant for speaking and understanding a language is being acquired and laid down in the form of separate sequence detectors for individual word pairs. The child has an ability to *generalize patterns* from encountered strings and apply them to new word sequences. In this and the following section, it will be shown that a sequence detector model accounts for generalization on the basis of auto-associative learning.

In linguistics, algorithms are being used to describe the order of words in sentences (Harris, 1951; Hauser, Chomsky & Fitch, 2002). A generative algorithm may include a formula such as (6):

$$c \rightarrow a b \quad (6)$$

Eq. (6) indicates that a sequence of type  $c$  includes an element of type  $a$  followed by one of category  $b$ . Element types can be further defined, either by other rules of the form of Eq. (6), or by listing the set of elements included in each category.

$$a = \{A_1, A_2, \dots, A_i, \dots, A_m\} \quad (7)$$

$$b = \{B_1, B_2, \dots, B_j, \dots, B_n\} \quad (8)$$

According to Eqs. (7) and (8), the symbols  $a$  and  $b$  can be replaced by elements of a set of indexed capital letters, which are used as symbols of words and morphemes here. The categories  $a$  and  $b$  can be called *substitution classes*, because in a string  $A_i B_j$  each of the two elements  $A_i$  and  $B_j$  can be substituted by any other element of the same category.  $A_i$  is called a *complement* of  $B_j$  and vice versa,  $B_j$  is a complement of  $A_i$ . Further,  $a$  and  $b$  are called *complementary substitution classes*.

The strength of this formal approach to the serial order of symbols is that it does not require sequence detectors for each pair allowed for by the algorithm (6) to (8). The three rules together define a set of  $m \cdot n$  strings, and increasing the number of formulas will multiply the number of strings covered. In generative linguistics, it is commonly assumed that

algorithms such as (6) develop on the basis of genetically pre-programmed knowledge after triggering through even extremely sparse input and that all that remains for learning is to connect words and morphemes to categories, as indicated by (7) and (8). This solves the sparseness of the input problem, by introducing the new variable of inborn grammatical knowledge however. Here, it is shown that deduction of abstract knowledge and generalization, describable in terms of rules such as (6) to (8), is possible in an associative memory with appropriate structure.

What is the data set a learning mechanism for serial order in language can operate on? A typical situation is that very common word strings are frequently being encountered and that one of the words in the string is occasionally replaced by another one. If this happens repeatedly, it is possible to define classes of substituted elements, for example by using formulas such as (7) and (8) above. These formulas would summarize the knowledge that substitutions have been observed and that, at different occasions, one of the elements of set *a* has been encountered together with (followed by) one of the members of set *b*. There is no implication that each of the elements of substitution class *a* has co-occurred with each of those of *b*. Examining large corpora of spoken text, listing the classes of substituted lexical elements, and describing them by formulas such as (6) to (8) is a fruitful strategy when encountering a new language not described before (Bloomfield, 1933).

To provide an example of how words co-occur and are substituted with one another in standard language, Table 1 shows a matrix in which dots indicate co-occurrences of a few nouns (left column) and verbs (top row) in sentences of a fairy tale. It can be seen that the nouns  $A_1$ - $A_6$  are multiply substituted with one another in the context of the verbs  $B_1$ - $B_{12}$ , and that, vice versa, the verbs  $B_1$ - $B_{12}$  are exchanged with one another in the context of the nouns

$A_1$ - $A_6$ . In contrast, nouns or pronouns  $A_7$ - $A_9$  co-occur with other verbs ( $B_{13}$ - $B_{16}$ ). Because of the specific multiple exchanges between the item groups  $A_1$ - $A_6$  and  $B_1$ - $B_{12}$ , they can be defined as complementary substitution classes, although not a single item from any of these groups occurs together with all members of the complementary category. In the same way, large text corpora can be used to define groups of lexical items that are preferentially substituted with one another in particular contexts. Thus, specific substitution patterns define complementary groups of lexical items.

#### *Auto-associative substitution learning (AASL) and generalization*

Similar to a language researcher encountering a new language, an associative memory such as the cortex can exploit the information about co-occurrences of lexical elements in strings. The type of information that becomes available can be illustrated by using a most parsimonious example, a word string of two lexical elements in which each element can be replaced by another one.

The encountered sequences are the word sequence  $A_1B_1$  and strings in which each of the lexical elements is replaced by a different one, i.e.  $A_1B_2$  and  $A_2B_1$ . At the neuronal level, encountering the word string  $A_1B_1$  would lead to the strengthening of connections between the two corresponding neuronal ensembles,  $\alpha_1$  and  $\beta_1$ , through their connecting sequence detector in the halo regions of  $\beta_1$  (Figure 3). Note that there are also sequence-sensitive neurons in the halo region of  $\alpha_1$  that would respond to the reverse sequence,  $\beta_1\alpha_1$ , but, since this sequence is not present in the input, this potential link between  $\alpha_1$  and  $\beta_1$  will not be established and may therefore be pruned. Further indirect connections, between  $\alpha_2$  and  $\beta_1$  and also between  $\alpha_1$  and  $\beta_2$ , are being strengthened when the word sequences  $A_2B_1$  and  $A_1B_2$  are encountered. After

these learning steps, a connection between  $\alpha_2$  and  $\beta_2$  – the neuronal correlate of the new word sequence  $A_2B_2$  which has not occurred in the input before – is now present as well. The connection is indirect, from  $\alpha_2$  to the halo and center of  $\beta_1$ , to the center of  $\alpha_2$  and finally to the halo and center of  $\beta_2$ . Through connections strengthened by associative learning,  $\alpha_2$  can now influence the firing probability of both  $\beta_1$  and  $\beta_2$ .<sup>2</sup> This is so although the word sequence  $A_2B_2$  has never been present in the input and the corresponding sequence of neuronal activations  $\alpha_2\beta_2$  has never occurred in the example illustrated here.

It is therefore not surprising that new strings are facilitated in a cortex-like associative memory. Because the mechanism crucially depends on learning in an auto-associative memory in which a set of neuronal elements are connected reciprocally with each other in the first place, it is called *auto-associative substitution learning* or AASL.

The newly established link between  $\alpha_2$  and  $\beta_2$  is through several steps and the facilitatory influence it mediates may therefore be weak. However, a realistic scenario will not only involve one such link, but several of them instead. Considering again the substitution matrix of lexical elements in Table 1A, it becomes clear that most lexical items included in the substitution classes  $A_1$ - $A_6$  and  $B_1$ - $B_{12}$  have been members of two or more strings in which they co-occurred with complements of the respective other category. Larger corpora reveal abundant substitutions in which most lexical items participate. This implies that, even though each string may set up an indirect and therefore probably weak link, the cumulative effect of numerous such links activated at the same time could provide substantial priming of a lexical item occurring in a new context.

---

<sup>2</sup> It also stimulates the neurons in  $\alpha_1$ . Activation of  $\alpha_1$  can be ruled out by the additional condition that activation of an ensemble requires priming of both its center and at least one section of its halo.

The important point here is that the link between  $\alpha_2$  and  $\beta_2$  is strengthened at all although the respective stimulus sequence was never present, and that this strengthening is a consequence of associative learning and the substitution of string elements. Even if the influence this and similar links provide is weak, an elevation of the firing probability of neurons in a neuronal ensemble makes its activation more likely. In the model, this means that an old (already encountered) word has an enhanced probability of being processed in a new context. In production, the new string is more likely to be articulated and in perception, the new incoming string is more likely to be processed in the same way as already encountered ones. This is all that is required to account for the so-called creative use of language elements, that is, the occasional processing of a new string comprised of old elements.

Whenever AASL-driven generalization takes place, more than one pair of input units and, thus, at least two word-pair detectors previously established by associative learning become active in close temporal succession. In the example of AASL-driven processing of the string  $A_2B_2$ , the word-pair detectors of strings  $A_1B_1$ ,  $A_1B_2$ , and  $A_2B_1$  were active near-simultaneously. If neuronal elements are frequently active together, they should strengthen their mutual connections, so that the connection strength reflects the strength of their correlation. Note that words may occur in different contexts whereas the sequence detectors are specific to the respective pair sequences, and, therefore, the correlation of the activation of sequence detectors is necessarily greater than that of word representations. For this reason, AASL should primarily link sequence detectors with one another. The likely result is that higher-order sequence detectors develop that incorporate several of the sequence detectors sensitive to single word pairs. In the end, the higher-order sequence detector will become active whenever a pair of elements from two substitution classes with regular serial order

relationship is being processed. The higher-order, or *generalized sequence detector* would thus represent a sequence of complementary *sets* of lexical items. Therefore, the postulated generalized sequence detectors built up by AASL connect complementary substitution classes of lexical items as described formally by (6) – (8).

It cannot be denied that connections corresponding to a string never encountered before are strengthened by associative learning. The strengthening of these connections is a necessary consequence of (1) associative learning principles and (2) a pattern of substitution between string elements. The pattern of substitution needs to be as exemplified in the minimal example, illustrated in Table 1, and schematized by the formulas (6) – (8). A member of one class of elements needs to be followed by a member of a second class, and there needs to be ample substitution between the members of each class. Each lexical element will, so to speak, be glued to the complementary substitution class through a generalized sequence detector. The established links do not require that all possible combinations of elements from the two substitution classes have actually been encountered together in the input. It is necessary, however, that each member of a substitution class participates in ample substitutions that link it to the sequence detector connecting both substitution classes. As Table 1 shows, it is realistic to assume such multiple substitutions that provide links between a new lexical item and several of its complements from the same substitution class.

Generalization in AASL depends on specific pattern of substitutions of lexical elements between encountered strings. It is not explained by properties of the input, for example surface similarities of the lexical elements, or by a general default (Hare, Elman & Daugherty, 1995; Pinker & Ullman, 2002; Pulvermüller, 1998; Rumelhart & McClelland, 1987). Which mechanism guarantees that AASL-driven generalization is specific and does not

apply to lexical elements that do not participate in the substitution process? In the matrix presented in Table 1, the substitution pattern becomes manifest in a square patch with a high density of dots, corresponding to the frequent cross-exchange of lexical items from two groups, leading to strengthened neuronal links between the respective groups of neuronal representations via a sequence detector. There are items (at the bottom and on the right) that do not participate in the substitution process and there is no reason why these items should be linked to the sequence detector in question. Looking at large corpora, several groups of lexical items can be identified that, in given contexts, exhibit a high probability of within group substitution whereas the probability to be exchanged with non-group members in the same contexts is small. The specific patterns of substitution give rise to separate sequence detectors each connected to the members of its specific substitution classes, but not to non-members. Generalizations mediated by sequence detectors can therefore only occur for members of a substitution class in the context of a member of a complementary substitution class. This is exactly the situation described by the algorithm (6) - (8).

At this point, it may be appropriate to state that an auto-associative memory can solve some of the riddles of grammatical generalization on the basis of AASL. The representations of strings between which elements are substituted interfere with each other in a predictable manner. When auto-associative memories are being used for face or syllable recognition, interference between patterns is an undesired consequence of a too high overlap between patterns (Palm, 1980; Willshaw, Buneman & Longuet-Higgins, 1969). However, pattern interference may have a positive side as well, namely generalization (Willshaw, 1972). The consequence of interference specifically between the representations of sequences that share discrete string elements can be described by rules in the linguistic domain.

### *Recursion, pushdown storage, and dynamic binding*

Generalized sequence detectors may be the basic building blocks of the intra-cortical grammar, but they are not sufficient for grammar processing, because they do not capture what linguists characterize as recursive rule application. In the theory of abstract automata, a device equipped with one or more pushdown stack memories is used to model repeated application of the same rule in the course of the processing of a single sentence (Chomsky, 1963; Joshi, 1990). A neurobiological correlate of a pushdown memory has recently been discussed in the context of neurophysiological data (Pulvermüller, 1993; Pulvermüller, 2002). It builds upon activity dynamics of neuronal representations and resembles mechanisms proposed for verbal and nonverbal short-term memory (Milner, 1999; Page & Norris, 1998). The putative neurobiological mechanism of pushdown memory is complementary to that provided by sequence detectors. It is conceivable that local syntactic relationships between syntactic constituents are implemented by sequence detectors whereas embedding of phrases and sentences into other sentences is realized by the pushdown mechanism. The role of these putative mechanisms can be specified within the larger context of neuronal grammar and associative language theory (Crespi-Reghizzi, Pradella & San Pietro, 2001; Pulvermüller, 2003).

### *Conclusions*

This article presents a new view on the sequence detection mechanism that might be relevant for grammar processing in the human brain. Sequence detectors similar to those in the nervous systems of animals, but operating at the level of neuronal ensembles, may store

encountered sequences of word pairs. Multiple substitutions of words and morphemes in sentences leads to auto-associative substitution learning (AASL), that is, to the strengthening of multiple connections between sets of complementary lexical representations. The multiple connections between substitution classes also link the neuronal counterparts of lexical elements that have never occurred together in the input before, thereby explaining one form of grammatical generalization. Furthermore, AASL-driven generalization connects word-pair detectors to one another and yields generalized sequence detectors that connect the members of entire lexical classes. These may be but one essential part of the grammar machinery in the human brain.

*Acknowledgments:*

I am grateful to Horace Barlow, Valentino Braitenberg, William Marslen-Wilson, Peter Milner, Risto Näätänen, Yury Shtyrov, Ian Nimmo-Smith, Fritz Sommer, Thomas Wennekers and two anonymous referees for help and comments at different stages of this work. Supported by the European Science Foundation and the Medical Research Council.

## References

- Abeles, M., Bergman, H., Margalit, E. & Vaadia, E. (1993). Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *Journal of Neurophysiology* **70**, 1629-1638.
- Barlow, H. & Levick, W. R. (1965). The mechanism of directionally selective units in rabbit's retina. *Journal of Physiology* **178**, 477-504.
- Barlow, H. B., Hill, R. M. & Levick, W. R. (1964). Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *Journal of Physiology* **173**, 377-407.
- Bienenstock, E. (1996). On the dimensionality of cortical graphs. *Journal of Physiology, Paris* **90**, 251-256.
- Bloomfield, L. (1933). *Language*. Holt, New York.
- Braitenberg, V. (1978). Cell assemblies in the cerebral cortex. In *Theoretical approaches to complex systems. (Lecture notes in biomathematics, vol. 21)* (ed. R. Heim and G. Palm), pp. 171-188. Springer, Berlin.
- Braitenberg, V. (2001). Brain size and number of neurons: an exercise in synthetic neuroanatomy. *Journal of Computational Neuroscience* **10**, 71-77.
- Braitenberg, V., Heck, D. & Sultan, F. (1997). The detection and generation of sequences as a key to cerebellar function: experiments and theory. *Behavioral and Brain Sciences* **20**, 229-245.
- Braitenberg, V. & Schüz, A. (1998). *Cortex: statistics and geometry of neuronal connectivity*, 2 edition. Springer, Berlin.

- Chomsky, N. (1963). Formal properties of grammars. In *Handbook of mathematical psychology. Volume 2* (ed. R. D. Luce, R. R. Bush and E. Galanter), pp. 323-418. Wiley, New York, London.
- Crespi-Reghizzi, S., Pradella, M. & San Pietro, P. (2001). Associative definitions of programming languages. *Computer Languages* **27**, 105-123.
- Egelhaaf, M., Borst, A. & Reichardt, W. (1989). Computational structure of a biological motion-detection system as revealed by local detector analysis in the fly's nervous system. *Journal of the Optical Society of America (A)* **6**, 1070-1087.
- Fuster, J. M. (1995). *Memory in the cerebral cortex. An empirical approach to neural networks in the human and nonhuman primate*. MIT Press, Cambridge, MA.
- Fuster, J. M. (1997). Network memory. *Trends in Neurosciences* **20**, 451-459.
- Hare, M., Elman, J. L. & Daugherty, K. G. (1995). Default generalisation in connectionist networks. *Language and Cognitive Processes* **10**, 601-630.
- Harris, Z. S. (1951). *Structural linguistics*. Chicago University Press, Chicago.
- Hauser, M. D., Chomsky, N. & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569-79.
- Heck, D. (1993). Rat cerebellar cortex in vitro responds specifically to moving stimuli. *Neuroscience Letters* **157**, 95-98.
- Heck, D. & Sultan, F. (2002). Cerebellar structure and function: making sense of parallel fibers. *Hum Mov Sci* **21**, 411-21.
- Hubel, D. (1995). *Eye, brain, and vision*, 2 edition. Scientific American Library, New York.
- Joshi, A. (1990). Processing crossed and nested dependencies: an automaton perspective on the psycholinguistic results. *Language and Cognitive Processes* **5**, 1-28.

- Kleene, S. C. (1956). Representation of events in nerve nets and finite automata. In *Automata studies* (ed. C. E. Shannon and J. McCarthy), pp. 3-41. Princeton University Press, Princeton, NJ.
- McCulloch, W. S. & Pitts, W. H. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* **5**, 115-133.
- Milner, P. M. (1957). The cell assembly: Mk. II. *Psychological Review* **64**, 242-252.
- Milner, P. M. (1999). *The autonomous brain*. Laurence Erlbaum Associates, Mahwah, NJ.
- Page, M. P. & Norris, D. (1998). The primacy model: a new model of immediate serial recall. *Psychological Review* **105**, 761-781.
- Palm, G. (1980). On associative memory. *Biol Cybern* **36**, 19-31.
- Palm, G. (1982). *Neural assemblies*. Springer, Berlin.
- Palm, G. (1993). On the internal structure of cell assemblies. In *Brain theory: spatio-temporal aspects of brain function* (ed. A. Aertsen), pp. 261-270. Elsevier, Amsterdam.
- Pinker, S. & Ullman, M. (2002). Combination and structure, not gradedness, is the issue. *Trends Cogn Sci* **6**, 472-474.
- Prut, Y., Vaadia, E., Bergman, H., Haalman, I., Slovin, H. & Abeles, M. (1998). Spatiotemporal structure of cortical activity: properties and behavioral relevance. *J Neurophysiol* **79**, 2857-74.
- Pulvermüller, F. (1993). On connecting syntax and the brain. In *Brain theory - spatio-temporal aspects of brain function* (ed. A. Aertsen), pp. 131-145. Elsevier, New York.
- Pulvermüller, F. (1998). On the matter of rules. Past tense formation as a test-case for brain models of language. *Network: Computation in Neural Systems* **9**, R 1-51.

- Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences* **22**, 253-336.
- Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences* **5**, 517-524.
- Pulvermüller, F. (2002). A brain perspective on language mechanisms: from discrete neuronal ensembles to serial order. *Progress in Neurobiology* **67**, 85-111.
- Pulvermüller, F. (2003). *The neuroscience of language*. Cambridge University Press, Cambridge.
- Reichardt, W. & Varju, D. (1959). Übertragungseigenschaften im Auswertesystem für das Bewegungssehen. *Zeitschrift für Naturforschung* **14b**, 674-689.
- Rumelhart, D. E. & McClelland, J. L. (1987). Learning the past tense of English verbs: implicit rules or parallel distributed processing. In *Mechanisms of language acquisition* (ed. B. MacWhinney). Erlbaum, Hillsdale, NJ.
- Sakurai, Y. (1999). How do cell assemblies encode information in the brain? *Neurosci Biobehav Rev* **23**, 785-96.
- Sommer, F. T. & Wennekers, T. (2001). Associative memory in networks of spiking neurons. *Neural Netw* **14**, 825-34.
- Varju, D. & Reichardt, W. (1967). Übertragungseigenschaften im Auswertesystem für das Bewegungssehen II. *Zeitschrift für Naturforschung* **22b**, 1343-1351.
- Willshaw, D. J. (1972). A simple network capable of inductive generalization. *Proc R Soc Lond B Biol Sci* **182**, 233-47.
- Willshaw, D. J., Buneman, O. P. & Longuet-Higgins, H. C. (1969). Non-holographic associative memory. *Nature* **222**, 960-2.

Young, M. P., Scannell, J. W. & Burns, G. (1995). *The analysis of cortical connectivity*.

Springer, Heidelberg.

Zipser, D., Kehoe, B., Littlewort, G. & Fuster, J. (1993). A spiking network model of short-term active memory. *Journal of Neuroscience* **13**, 3406-20.



## FIGURE CAPTIONS

Figure 1: (A) A common circuit for serial order processing is illustrated. Two input units,  $\alpha$  and  $\beta$ , project to a third unit,  $\gamma$ , which is sensitive to the serial order of activation of the input units. The connection from  $\alpha$  to  $\gamma$  is qualitatively different from that linking  $\beta$  to  $\gamma$ . Differences may exist in conduction velocities, filtering of the signal, or the weights of the connections (insets). (B) If strong and weak signals that decay exponentially arrive at a subsequence neuronal unit at different times, the order of the inputs influences the strength of the response. The strong signal followed by the weak one ( $I_{\beta\alpha}$ ) yields a smaller maximum activation than the reverse sequence, the weak signal followed by the strong one ( $I_{\alpha\beta}$ ). (C) Within certain limits, the weak input followed by the strong one (first  $\alpha$ , then  $\beta$ ) can produce above threshold activation, whereas the reverse sequence (strong-weak; first  $\beta$ , then  $\alpha$ ) produces a sub-threshold response.

Figure 2: Schematic illustration of a cortical circuit that may underlie sequence detection. Two local clusters of neurons, labeled  $\alpha$  and  $\beta$ , are illustrated. One representative neuron is drawn from each population. In the immediate vicinity or halo of  $\beta$ , there are neurons that receive strong input from the  $\beta$  core and weak input from other cortical loci. A patch of neurons in the halo receives additional weak input from  $\alpha$ . These respond preferably to the sequence “first  $\alpha$ , then  $\beta$ ”.

Figure 3: Illustration of the putative basis of auto-associative substitution learning (AASL). Single letters correspond to lexical elements and to local neuron clusters included in their

neuronal representations. Letter pairs represent strings of lexical elements and the corresponding neuronal sequence detectors. For further explanation, see text.

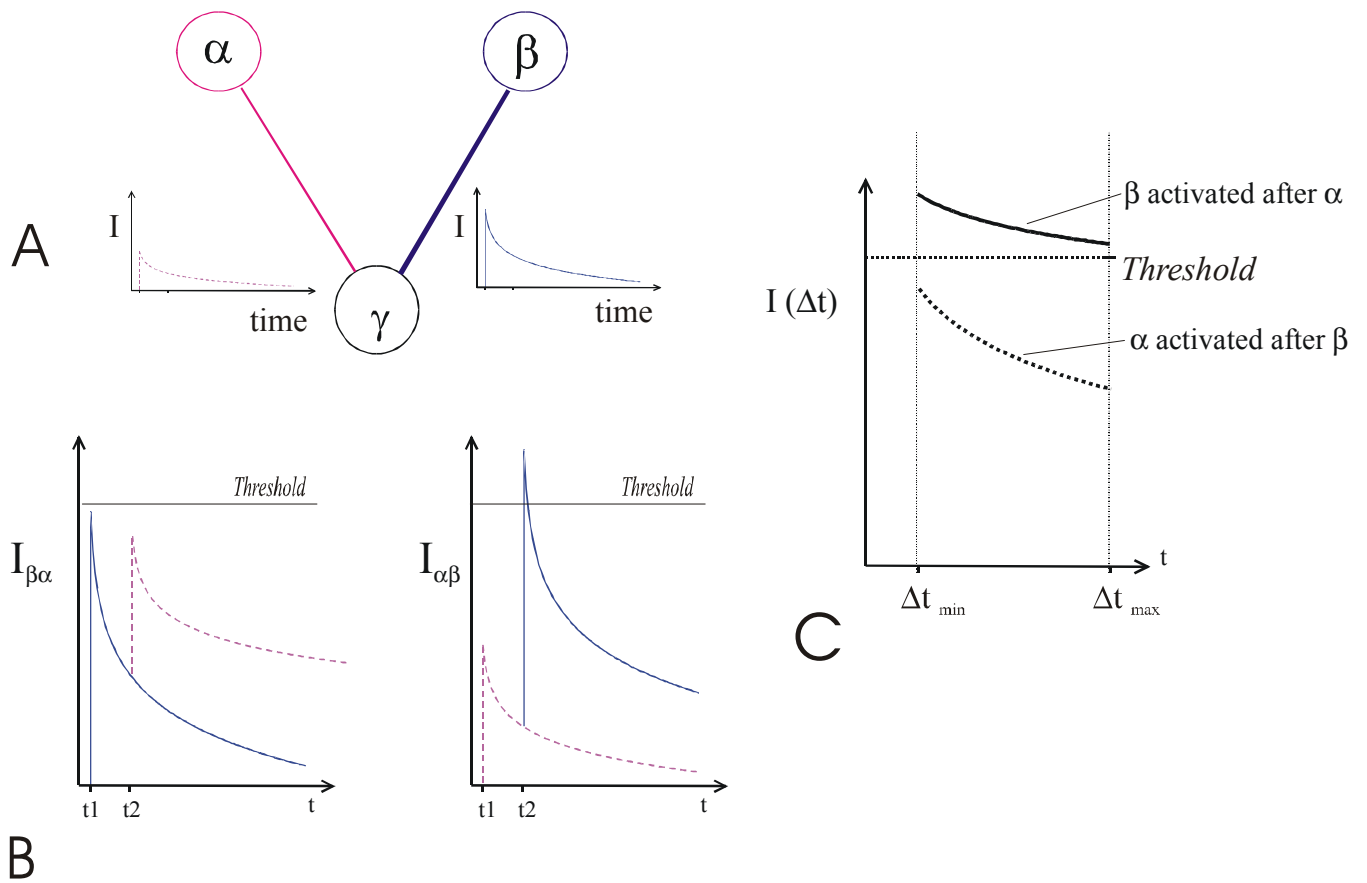


Fig. 1

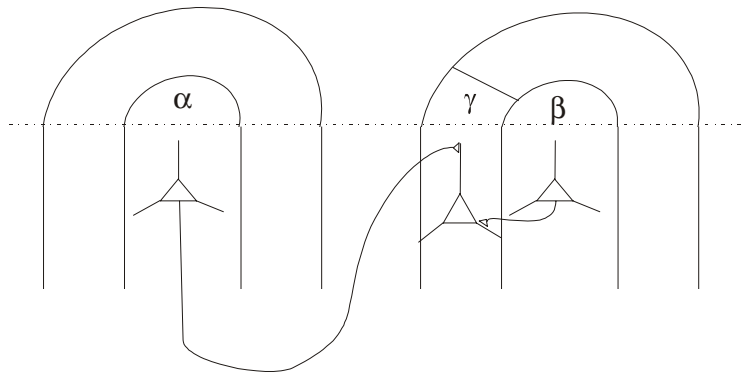


Fig. 2

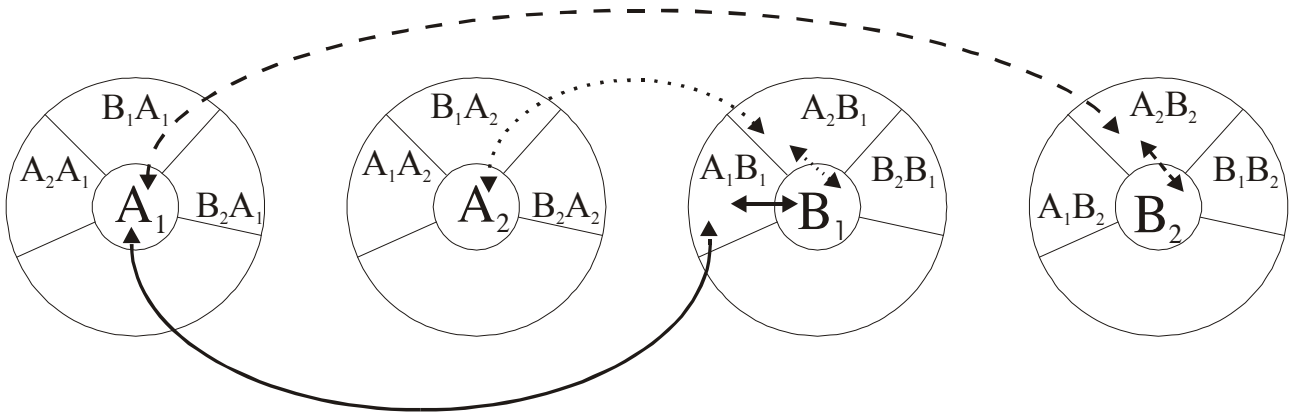


Fig. 3