Towards Sentence Understanding: Phase Arbitration in Temporal-Coding Memory Mechanism

MAKINO Takaki
Department of Information Science
Graduate School of Science

AIHARA Kazuyuki
Department of Complexity Science and Engineering
Graduate School of Frontier Science

TSUJII Jun-ichi
Department of Computer Science
Graduate School of Information Science and Technology

Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033 Japan
University of Tokyo

mak@is.s.u-tokyo.ac.jp aihara@sat.t.u-tokyo.ac.jp tsujii@is.s.u-tokyo.ac.jp

Abstract

This paper explores a mechanism of memory in human brain from a viewpoint of sentence understanding. We pointed out the following: (1) Some complexity must be incorporated into memory coding in order to be capable of representing binding in a meaning of a sentence. (2) When temporal coding is used to achieve the complexity, some mechanism is required to arbitrate phases (temporal slots) among memorized items. (3) Considering its implementation, the mechanism is likely to be global, which resembles a sort of structured memory, such as push-down stack. (4) Episodic memory, which is thought to be formed through mammal hippocampus, can be regarded as a phase arbitration mechanism and is possibly related in depth to sentence understanding.

1 Introduction

Memory is intrinsically used in human language processing. A person cannot process a long sentence at once; one divides the sentence into words (or some other units) and processes word by word, while storing partially processed results in one's memory. Moreover, the semantic information of the sentence is also supposed to be stored in the memory along the partial results. Thus, a model of sentence understanding cannot be described without a model of memory. In other words, we are able to use a sentence-understanding task to justify a model of human memory mechanism.

However, past studies of ANN-based NLP are not enough to explain the memory mechanism of sentence understanding. For example, a memory model in a simple recurrent network (Elman, 1990) suffers from *superposition catastrophe* restricting capacity of semantic representation. Although temporal coding as in Henderson's connectionist parser (Henderson, 1994) seems promising, it is still an important open problem to pursue a better model of human memory along this line.

In this paper, we explore a model of memory mechanism in the human brain from a viewpoint of sentence understanding. Especially, we point out the necessity of *phase arbitration*, a mechanism that allocates an unused pulse phase to a newly memorized item, and discuss the implication of phase arbitration.

We first identify the task of sentence understanding as a target of a neural memory mechanism and show that the superposition catastrophe prevents traditional neural networks from achieving the task. We then investigate possible sources of complexity to be added to the neural network, and show an advantage of *temporal complexity*. Further, we show the necessity of phase arbitration in a network involving temporal complexity. We empirically show that a network cannot handle temporal coding without phase arbitration. We also discuss possible models of phase arbitration, and global phase arbitration bears resemblance to storing operation on structural memory mechanism, such as queue or push-down stack.

In Section 2, the sentence-understanding task is outlined. Then, in Section 3, we explain superposition catastrophe and possible solutions, and point out the necessity of the phase-arbitration mechanism. Section 4 empirically shows the necessity of phase-arbitration mechanism through our simulation model. Finally, we

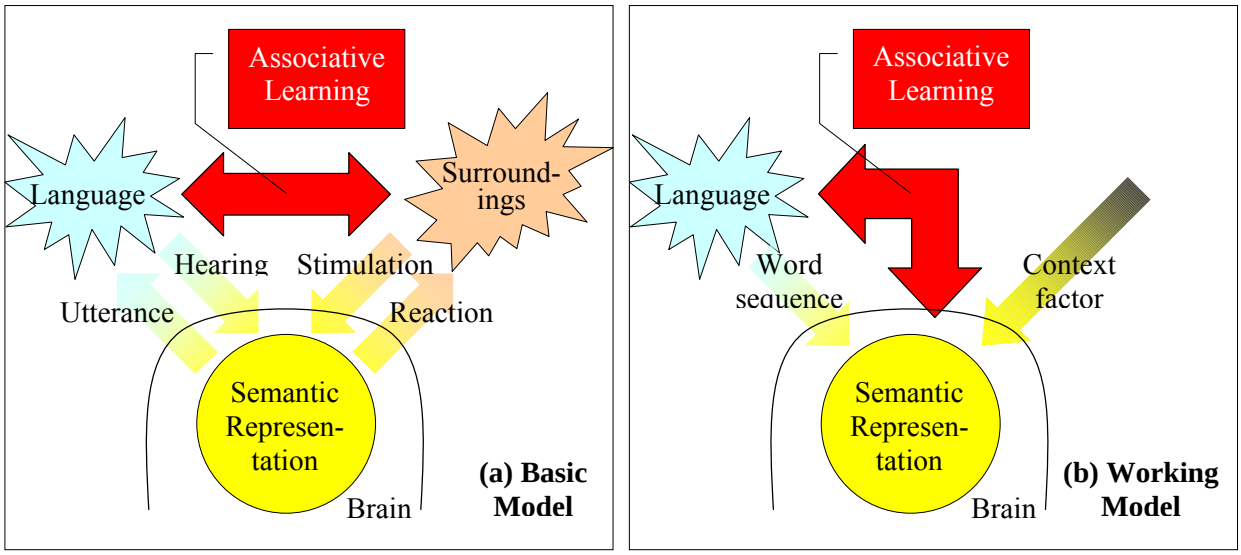


Figure 1: A Basic and Working Models of Simulation in Learning.

discuss possible implementations of the mechanism and its implications in Section 5.

2 Sentence-Understanding Task

Although it is almost impossible to define what is sentence understanding, we can say that some conditions should be satisfied in sentence understanding. By enumerating the conditions, we outline the task of sentence understanding.

When a person understands a sentence, his/her brain is activated in some pattern of activity. For example, if the sentence is describing some feeling, a part of the pattern will match to the activity caused by experiencing the feeling. In other words, the matching part of pattern provides meaning of the sentence in a form of association to the feeling. We call this part of pattern *semantic representation*.

In most cases, a meaning contains information of relations between entities or concepts. For example, a meaning of a phrase 'a white hat' contains a relation between a concept 'white' and an entity 'a hat'. We call this relation *binding*, borrowing the term from logic programming.

Although semantic representation may be unique for each person, it should have some common feature in order to be regarded as a part of sentence-understanding process, including the following:

- Semantic representation is **dynamic**, that is, available immediately after understanding. Although static memory mechanism (such as change of wiring) may concern background knowledge of semantics, it is too slow to be used in the following processes. Semantic representation should be on more dynamic and flexible medium, such as change of electric potential and functional connectivity.
- Semantic representation is **memorable** in the brain. Namely, the brain does not understand a sentence without keeping the semantic representation for a certain period.
- Semantic representation is **simple** enough, in a sense that mapping from a sentence to semantic representation can be learned by the brain. Although this paper does not concern with learning, we have to consider the learnability condition in order to obey simplicity. Infants learn the mapping from associated interaction of language and surroundings; we can prove a semantic representation satisfies the simplicity condition by showing similar learning is possible on simulation model. Our basic model of simulation to show learnability is that, as in Figure 1a, the brain learns surroundings from verbal and non-verbal interaction, and constructs appropriate semantic representation through learning. However, we first try on a working model shown in Figure 1b, in which a sequence of words and given semantic representation is posed to the brain so that the brain learns the association of sentence and semantic representation.
- Representation of a binding of an attribute and a value is **additive**, i.e., representation of a binding is overlaid activities of attribute

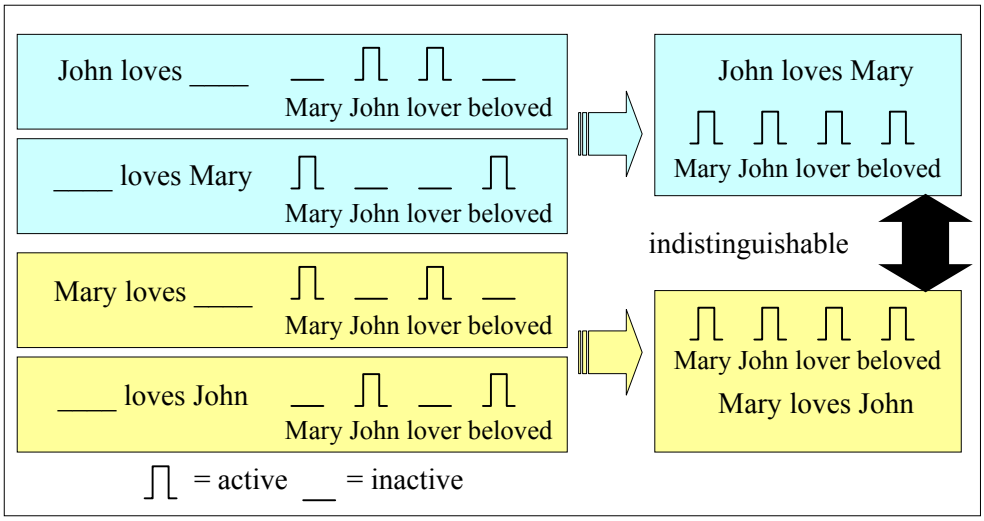


Figure 2: Superposition Catastrophe in a Semantic Representation.

representation and value representation. We should reject *multiplicative* representation, which depends on an activity appearing only on a certain binding of an attribute and a value. This is because a person can understand an unencountered sentence, which contains unknown bindings.

Especially the last point constrains a possible coding of semantic representation, which we pursue in the next section.

3 Complexity in Memory Coding

3.1 Necessity of Complexity

The additiveness requirement forces us to face with *superposition catastrophe* (Fujii et al, 1996), as illustrated in Figure 2. Binding of an attribute "lover" and a value "John" is represented as simultaneous activities of "lover" and "John." However, when we try to represent two binding relations, "John" = "lover" and "Mary" = "beloved," the activity becomes a mixture of "John," "Mary," "lover," and "beloved," which is indistinguishable from another set of binding "Mary" = "lover" and "John" = "beloved." Since a person rarely makes a mistake of dynamic binding, some inherent mechanism that solves this catastrophe should exist.

It seems that simple recurrent networks do not have such a mechanism. A context layer, which corresponds to a memory mechanism in a simple recurrent network, falls into the catastrophe if it represents the meaning "John loves Mary" in an additive way[1]. Thus, we can say that some sort of complexity is necessary to be incorporated into the coding of memory mechanism, in order to represent bindings.

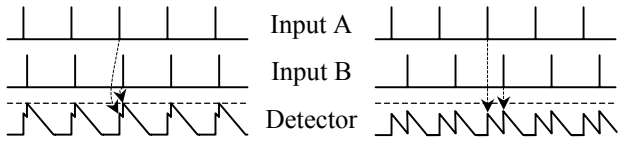
3.2 Possible Source of Complexity

Here we consider three possible sources of complexity to represent bindings: space, intensity, and time. Although the actual brain may have combination of them, we choose which should be the first one to be implemented.

The first candidate, *spatial complexity*, is to use more neurons and synapses to represent bindings. The easiest example is to introduce a neuron for each possible binding, such as 'John-lover' neuron, 'Mary-beloved' neuron and so on. However, this is obviously 'multiplicative' representation and violates additiveness requirement. We could not find a spatial extension that has a learnable semantic representation.

The second candidate, which we name *intensive complexity*, uses intensity (strength) of signals to store binding information. Sakurai (2001) pointed out that a neuron with infinite precision of signal levels can store arbitrary depth of nested information; such a neuron would be able to store binding information. However, he also proved that a sigmoid function neuron is unable to retrieve arbitrary depth of information, even with infinite precision. Moreover, actual neurons in

[1] This discussion is true on any coding with additiveness, such as distributed coding, although Figure 1 is illustrated with four 'grandmother' neurons for simplicity.



(a) Two inputs with almost coincident phases. The detector's signal level reaches the threshold and the detector fires periodically.

(b) Two inputs with different phases. Detector does not fire at all, since its signal level is always lower than the threshold.

Figure 3: Phase Coincidence detection by an integrate-fire neuron.

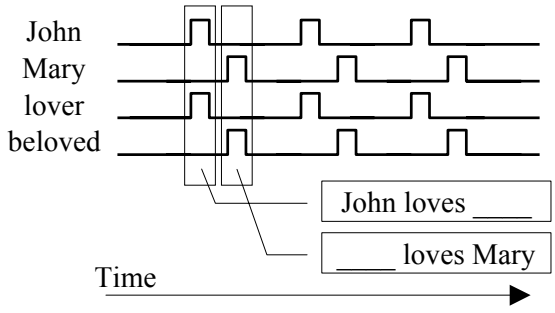


Figure 4: Synchrony-based Coding.

the brain have only finite precision of signal levels, which may be represented by the number of pulses and population rates in a neuron group. We decide not to pursue this approach here.

The last candidate, *temporal complexity*, uses temporal position of signals to represent binding information. This seems to violate memorability of semantic representation, since temporally transient activities of neurons cannot be kept over time. However, periodic activities such as oscillation can stay for a certain time on memory. Moreover, as illustrated in Figure 3, an integrate-and-fire neuron can detect coincidence of *phases* (temporal positions of periodic activity) among multiple neurons with high precision. Several studies suggest that temporal correlation of activities may be utilized as a coding in the brain in order to avoid superposition catastrophe (Fujii et al. 1996; Singer, 1994). From these arguments, we chose the temporal complexity for the first complexity to be implemented.

Actually, there are some implementations of the temporal complexity in the past studies. One of the simplest implementations of temporal coding on the ANN framework is a synchrony-based coding used in SHRUTI system (Shastri and Ajjanagadde, 1993). In their coding, a neuron oscillating by itself denotes either an attribute or a value, and synchronization of the oscillation denotes binding between them (Figure 4).

Henderson implemented a connectionist parser based on this coding (Henderson, 1994) and succeeded to make a neural network learn to parse by back-propagation through time (Henderson and Lane, 1998). His architecture, Simple Synchrony Network, is generally an extension of Simple Recurrent Network by the synchrony-based coding. He notes that the limitation of synchrony-based coding, e.g. capacity constraint caused by lack of time-slot, can predict human unacceptability of some sentences.

3.3 The Missing Link: Phase Arbitration

Although a network with temporal complexity looks quite promising, we found that our semantic representation cannot be applied directly to such a network: We have to answer how new items are memorized, and how unnecessary items are forgotten. This is because phases are limited resource, and an unused phase has to be allocated for each new binding to be memorized.

Current studies with temporal coding solve this problem artificially. The SHRUTI system determines every pulse phase by an artificial signal. Henderson's parser learns to use an unused phase for a new item, but it is based on the teacher signals in back-propagation. Moreover, both systems cannot forget items unless the systems are reset to original state. However, in our learning scheme, we cannot take such an artificial solution.

In this study, we name the allocation of an unused phase as *phase arbitration*, and pursue the way to implement phase arbitration on temporal-coding neural network. First, we have to determine whether a neural network can acquire phase arbitration through learning, or the neural network needs some inherent mechanism for phase arbitration. In the next section, we performed empirical experiments trying to simulate a neural network that learns phase arbitration.

4 Experiments: Network without Phase-Arbitration Mechanism

We tried to implement a minimized simulation

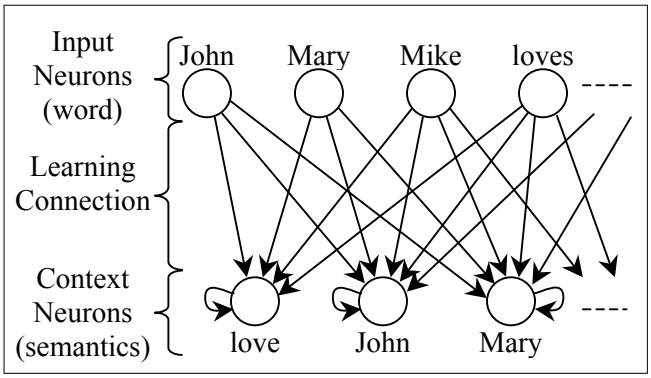


Figure 5: Architecture of the simulated model.

model of short-term memory mechanism on a pulse-based neural network simulator. Figure 5 shows the architecture of the network. Input to the memory is a sequence of words, where each word is represented as a pulse burst on an associated input neuron. The context neurons are hard-wired so that when a context neuron fires, the neuron oscillates by itself.

Figure 6 shows two stages of this simulation. We associated every 2-word sequence to an oscillation pattern on the memory neurons, so that every memory neuron corresponds to a single word, and the order of words is represented as a phase difference of two oscillations. At the training stage, both a sequence of words and its associated oscillation pattern on memory circuit are presented to the network. The network learns the association by an STDP, a spike-timing-dependent variant of Hebb's rule (Abbott and Nelson, 2000). At the test stage, the network is required to replay associated oscillation pattern from a given sequence of words.

Although our model was able to learn the association of a word and a memory neuron, we found that the model cannot learn the phase difference. The simulated model correctly activates the context neurons A and B at the input of a sentence AB, but the phase difference of two oscillations are determined by the interval length of two input pulse bursts, which is not stable in human environment. In other words, the oscillation pattern is indistinguishable from another sentence BA; the model cannot escape from superposition catastrophe.

This problem seems not only a problem of our simulated model but also a problem of any system depending on temporal coding. Even if there is a learning algorithm that can distinguish AB and BA, a network has to learn every possible relation between two entities, which is an im-

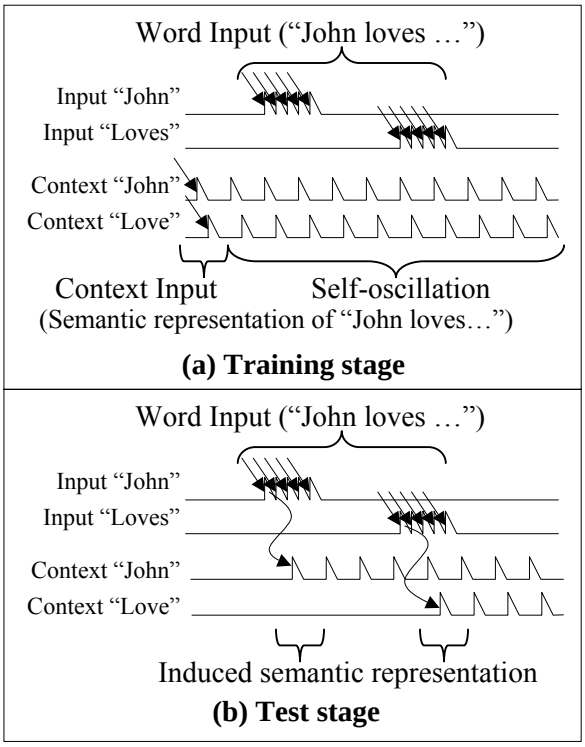


Figure 6: Two stages of the simulation.

practical solution. Thus, we conclude that some inherent mechanism that arbitrates pulse phases has to be introduced on the model of human memory mechanism. In the next section, we discuss a possible solution of phase arbitration mechanism.

5 Discussion

Phase arbitration mechanisms are classified into *local* and *global* mechanisms. In this section, these two possibilities are compared and discussed.

A *local* phase arbitration mechanism does not use any global signal to allocate a phase, and controls phase by only mutual connection between memory neurons. For example, excitatory and inhibitory connections from neuron A to neuron B can promote and suppress oscillation of neuron B in a specific phase difference from neuron A. However, when many activities are overlaid in an additive representation, such connections will induce or prohibit activity of unrelated neurons. It seems difficult to arbitrate phases only by local mechanisms.

On the other hand, a *global* phase arbitration mechanism uses some signal that represents

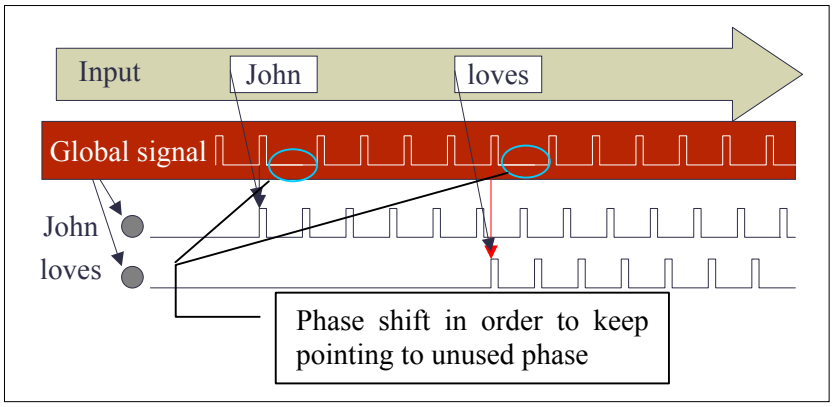


Figure 7: Example of Global Phase Arbitration.

global phase of a network[2]. As shown in Figure 7, each memory neuron uses this global signal to determine its oscillation phase. After that, a phase of either global signal or memory neuron shifts so that the global signal points to a new unused phase. In this way, the new items are stored on unused phases in order.

Such a mechanism suggests some structure exists on the memory mechanism. For example, we can easily add a mechanism that deactivates the last stored item to a memory; this formulates the memory as a *push-down stack*. Since a context-free language is easily parsed with a stack memory, such a structure in human memory may play an important role in sentence parsing and sentence understanding.

It should be noted that some mechanisms studied in brain sciences are similar to global phase arbitration. O'Keefe and Recce (1993) report that *phase precession* occurs in a rat hippocampus. Place-coding cells, which correspond to the current position of the rat, first become active in a specific phase to the Theta oscillation, and then shift their phase gradually to make phase difference to the next activation of other place-coding cells. This mechanism, which is supposed to provide short-term episodic memory, can also be regarded as a global phase arbitration mechanism using Theta oscillation as a global signal. It is possible that the phase arbitration for language is provided in such an episodic memory mechanism, since some research on neurolinguistics (Just and Carpenter, 1992) suggests the relation between sentence understanding and short-term memory capacity.

No study is known about memory deactivation mechanism in the brain, except old memories spilling out from the width of Theta oscillation. However, Ono (2000) reports that, in a mathematical model of phase precession (Lisman and Idiart, 1995), storage of multiple patterns sharing neurons to be active may cause interference between patterns to deactivate one of the patterns. In sentence parsing and understanding task, it is likely that a pattern of partial parsing result shares neurons with another partial result that covers the former result, thus this type of interference may occur on human memory. Since deactivation by interference suggests another memory structure different from stack, sentence parsing and understanding based on such a memory structure is worth to be studied in future.

6 Conclusion and Future Work

We explored a model of human working memory mechanism from a viewpoint of sentence understanding. We found that temporal complexity, which is necessary to avoid superposition catastrophe, poses a new problem to the memory model, i.e. phase arbitration. We discussed the mechanism of phase arbitration and suggested an existence of a global arbitration mechanism. This implies a structural operation can be formulated on human memory, possibly a push-down stack operation.

Future work is to construct a simulation of sentence understanding with global phase arbitration mechanism. We have already implemented a mathematical model of phase precession (Lisman and Idiart, 1995), which can be used as a memory with global phase arbitration

[2] This does not imply that every neuron is governed by some control center. Every neuron may control itself using a global signal to arbitrate phases.

mechanism. We are trying to attach hereto-associative mappings to the memory, which associate a sequence of words to a semantic representation. If a long sentence can be converted into semantic representation by several stages of the hetero-associative mappings, we can say that the mapping formulates a grammar.

Acknowledgements

This research is supported by CREST of JST (Japan Science and Technology Corporation) and JSPS Research Fellowships for Young Scientists.

References

Abbott, Laurence F., and Nelson, Sacra B. (2000). Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3:1178–1183.

Elman, Jeffrey L. (1990) Finding structure in time. *Cognitive Science*, 14:179–211.

Fujii Hiroshi, Ito Hiroyuki, Aihara Kazuyuki, Ichinose Natsuhiro, and Tsukada Minoru. (1996). Dynamical cell assembly hypothesis – Theoretical possibility of spatio-temporal coding in the cortex. *Neural Networks*, 9: |1303–1350.

Henderson, James. (1994). Connectionist syntactic parsing using temporal variable binding. *Psycholinguistic Research*, 23(5):353–379.

Henderson, James, and Lane, Peter. (1998). A Connectionist Architecture for Learning to Parse. *Proceedings of 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics* (COLING-ACL`98), pp. 531–537.

Just, Marcel A., and Carpenter, Patricia A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99:122–149.

Lisman, John E., and Idiart, Marco A. P. (1995). Storage of 7 ± 2 Short-Term Memories in Oscillatory Subcycles. *Science*, 267:1512–1515.

O'Keefe, John, and Recce, Michael. (1993) Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3(3): 317–330.

Ono Yasuhiro. (2000). *Research on neural circuit network model holding short-term memory*. Graduation thesis, Department of Mathematical Engineering and Information Physics, University of Tokyo. In Japanese.

Shastri, Lokendra, and Ajjanagadde, Venkat. (1993). From associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioural and Brain Sciences*, 16(3):417–494.

Sakurai, Akito. (2001). *Expressiveness of Recurrent Neural Networks for Grammars*. Personal Comminucation.

Singer, Wolf. (1994). Putative functions of temporal correlations in neocortical processing. In C. Koch, & J. L. Davis (Eds.), *Large-scale neuronal theories of the brain*, 201–238. MIT Press.