言語理解の計算理解が神経回路網モデル開発に及ぼす影響

Impact of Computational Theory of Language Understanding for Development of Neural Network Model

牧野 貴樹　**MAKINO Takaki**†‡ (PY)　　　合原 一幸　**AIHARA Kazuyuki**†*

†東京大学大学院 新領域創成科学研究科 複雑理工学専攻

Department of Complexity Science and Engineering, University of Tokyo

‡理化学研究所 脳科学総合研究センター (RIKEN BSI)

*科学技術振興事業団 戦略的基礎研究推進事業 (CREST, JST)

mak@sat.t.u-tokyo.ac.jp　　　**aihara@sat.t.u-tokyo.ac.jp**

Abstract

This paper shows how the computational theory of language understanding can affect the design of language-understanding neural networks. We propose some basic properties of a computational theory of language understanding, and we show that the generalizability requirement in the property forces us to face with the feature binding problem. Based on the advantage of the temporal-coding, we discuss the necessity of global phase control for language understanding. We also show simulation results with a simple neural network model for sentence understanding on the bases of the theory.

1 Introduction

Language facility is one of the most uncharted fields in the brain. One of the reasons is the limitation for measurement method for brain activities during language processing. Since only humans can use languages, we cannot resort to single neuronal recording in the living brain.

An effective research methodology for such a field is the approach from the viewpoint of computational neuroscience (Marr, 1982). When a computational theory for the problem domain is given, we can suggest representations and algorithms as an information processing problem. We can treat these representations and algorithms as a hypothesis of the process in the brain, to help investigation of the mechanism in the brain.

We need a computational theory for language understanding to take this approach. However, the mainstream of computational theories for languages, such as Chomsky's theory (Chomsky, 1965), has concerned only grammars. Based on the grammar theory, we will be able to obtain representations and algorithms suitable for grammatical decision; we need, however, another theory, which concerns understanding of language, to elucidate the mechanism of understanding.

This paper proposes some basic properties in the computational theory of language understanding, and shows an impact of the theory on artificial neural network models. Especially, we discuss the advantage of temporal coding, and discuss the necessity of global phase control for language understanding. We also show a simple neural network model for sentence understanding on the basis of the theory.

In Section 2, the sentence-understanding theory is outlined. Then, in Section 3, we explain feature binding problem and possible solutions, and point out the necessity of the global phase control, which is called *phase arbitration*. Finally, in Section 4, we show our simulation model based on the theory, and discuss the future work.

2 Language-Understanding Theory

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

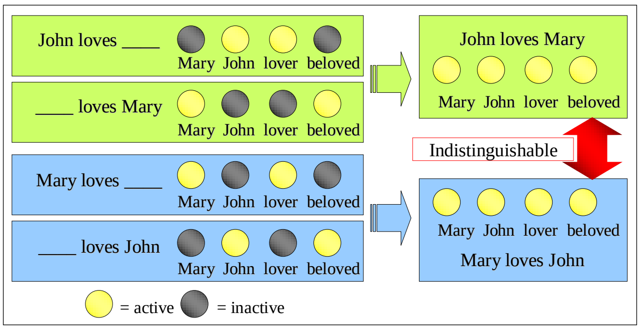
When a person understands a sentence, his/her brain is activated in some pattern. For example, if the sentence describes some feeling, a part of the pattern will match to the activity caused by experiencing the feeling. In other words, the matching part of pattern provides meaning of the sentence in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

Keywords : Language understanding, Temporal coding, Phase arrbitration, Semantic representation



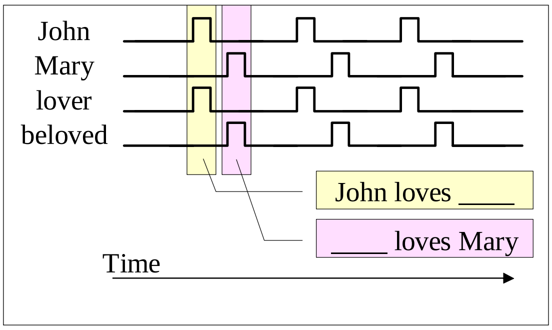
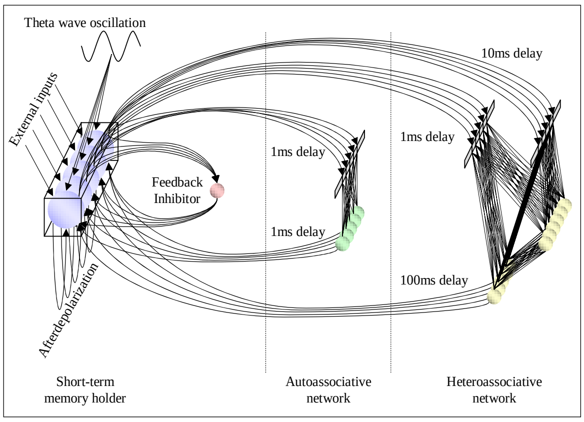


Figure 1: Feature binding problem.

Figure 2: Synchrony-based Coding.



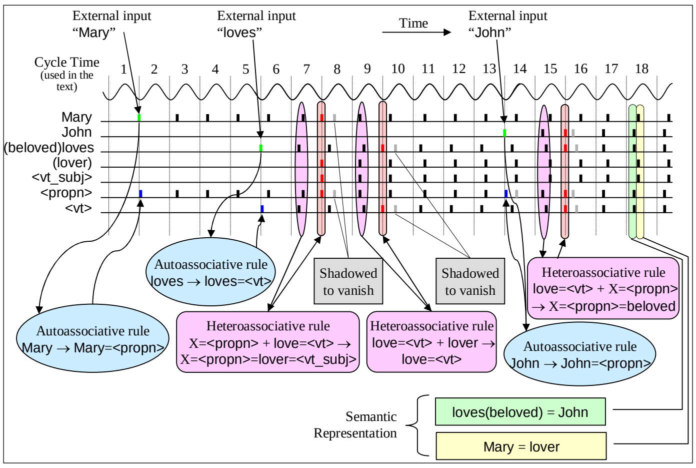


Figure 3: Network Structure of the Simulation model.

Figure 4: Process of the simulation model.

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 **generalizable**, in a sense that mapping from a sentence to semantic representation can be learned by the brain. Especially, the representation must be able to hold unencountered bindings; otherwise the language losts.its capability to convey an idea.

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

3 Complexity in Memory Coding

3.1 Necessity of Complexity

The generalizability requirement forces us to face with *feature binding problem* (von Malsburg, 1981), or *superposition catastrophe* (Fujii et al, 1996), as illustrated in Figure 1. 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 problem 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 problem 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, with the generalizability requirement, every possible binding must have its associated neuron, which is not practical as a model of the brain

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 the brain have only finite

[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.

precision of signal levels, which may be represented by the number of pulses and population rates in a neuron group. Here we decide not to pursue this approach.

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, an integrate-and-fire neuron can detect coincidence of *phases* (temporal positions of periodic activity) among multiple neurons with high precision (Singer, 1994). It is suggested that temporal correlation of activities may be utilized as a coding in the brain in order to avoid feature binding problem (Fujii et al. 1996). 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 artificial neural network 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 2).

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 Elman's 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 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 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.

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 (Makino, 2001) uses some signal that represents global phase of a network[2]. 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.

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.

4 Simulation Model

We developed a simulation model of artificial neural network, which satisfies the properties of the computational theory of language understanding. We adapted a leaky-integrate-and-fire neuron model and continuous simulation time.

Figure 3 shows the architecture of the simulation model. As a memory neuron with global phase arbitration, we adapted a mathematical model of phase precession (Lisman and Idiart, 1995). Two associative mappings are attached to the memory neurons: Autoassociative mapping associates words to the part-of-speech information, while heretoassociative mapping correlates two consecutive memory items, through two different delays from memory neuron, into a new memory item, which may contain bindings. Figure 4 illustrates the process of the two associative memories. For detail, please refer (Makino, 2002).

Figure 5 shows the actual result of the simulation on Punnets (Makino, 2002b). When a signal corresponding to the word Mary is input to the system (the sign "+Mary"), the associated neuron is activated along corresponding part-of-speech neuron (<propn> = proper noun). Because of the Lisman's memory model, these activities sustain over oscillations, and it uses a different phase from the activity corresponding to the next signal ("+loves"). At

[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.

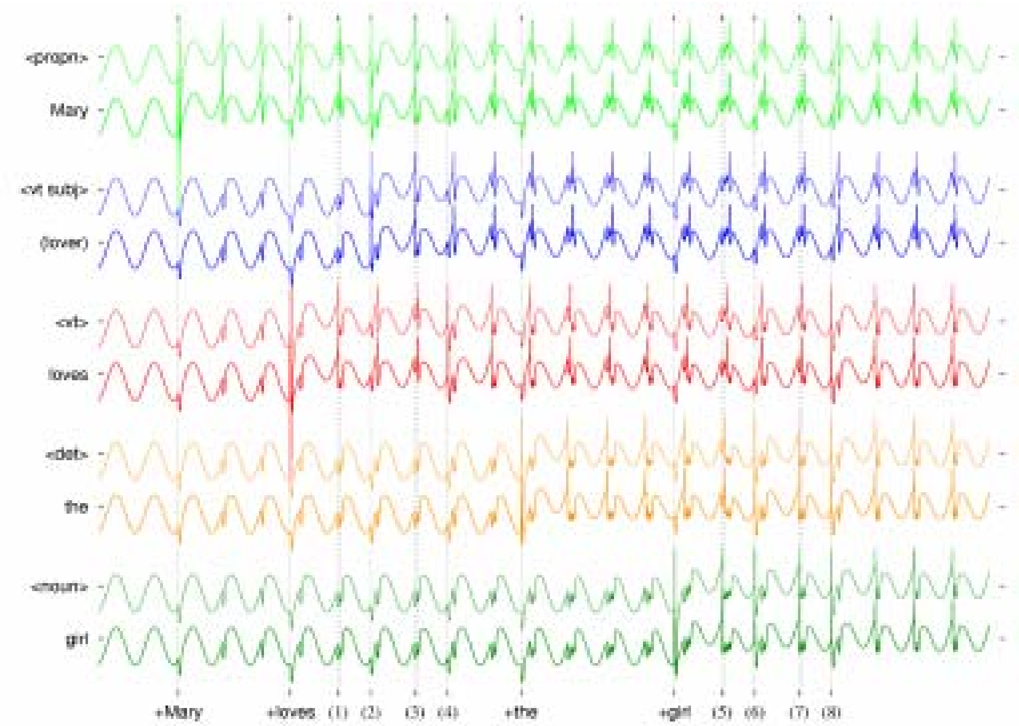


Figure 5: Result of the simulation model.

time (1), two memory items ("Mary" and "loves") falls into a time interval that activates heteroassociative network, and the result ("Mary = lover") is returned to the memory neurons at time (2). Repeating this process, the memory neurons finally holds two bindings "Mary = lover" and "the = girl = beloved", which correctly represents the meaning of the input sentence, "Mary loves the girl". Since this neural network model satisfies the computational theory of language understanding, we regard this model can be a start point to the more sophisticated "understanding" computation.

5 Conclusion

We explored a computation theory of language understanding and its impact to the design of neural network model. We showed that the property of meaning representation causes the feature binding problem to the classical neural network model. We found that temporal complexity, which is suitable for avoiding the problem, 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. Based on these discussions, we built a neural network simulation model, which satisfies the computational theory of language understanding.

References

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

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.

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

Fujii, H., Ito, H., Aihara, K., Ichinose, N., and Tsukada, M. (1996). Dynamical cell assembly hypothesis–Theoretical possibility of spatio-temporal coding in the cortex. *Neural Networks*, 9: 1303–1350.

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

Henderson, J., and Lane, P. (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, M. A., and Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99:122–149.

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

Makino, T., Aihara, K., and Tsujii, J. (2001) Towards sentence understanding: Phase arbitration in temporal-coding memory mechanism. In *The Second Workshop on Natural Language Processing and Neural Networks (NLPNN'2001)*, pages 46 – 52,

Makino, T. (2002) *A Pulsed Neural Network for Language Understanding: Discrete-Event Simulation of a Short-Term Memory Mechanism and Sentence Understanding.* Ph.D. dissertation, Department of Information Science, Graduate School of Science, Tokyo University,

Makino, T. (2002b) Discrete-event network simulation of arbitrary spiker-esponse model. in submission to Neural Computing and Applications.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman.

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

Shastri, L., and Ajjanagadde, V. (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, A. (2001). *Expressiveness of Recurrent Neural Networks for Grammars*. Personal Comminucation.

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