Linguistic Rule Acquisition by Information Maximization:

Neural Networks Infer the Use of Donatory Verbs

Taeko Kamimura, †Ryotaro Kamimura and ‡Thomas R. Shultz
Department of English, Senshu University,
2-1-1 Higashimita, Tama-ku, Kawasaki, Kanagawa, 214-8580, Japan
taekok@isc.senshu-u.ac.jp
† Information Science Laboratory, Tokai University,
1117 Kitakaname, Hiratsuka, Kanagawa, 259-1292, Japan
ryo@cc.u-tokai.ac.jp
‡Department of Psychology, McGill University, 1205 Penfield Ave.,
Montreal, Quebec, Canada H3A 1B1
shultz@psych.mcgill.ca

Abstract

In this paper, we propose a new information theoretic method for a linguistic rule acquisition problem, and demonstrate that a linguistic rule acquisition process is an instance of realization of information maximization. The new method is based upon unsupervised competitive learning. The unsupervised learning is needed because children acquire rules without any explicit instruction. In the experiments concerning the acquisition of the donatory verb system, we could observe that by maximizing information we can generate internal representations leading to linguistic rule discovery.

Keywords: information maximization, competition, linguistic rules, donatory verbs

1 Introduction

In this paper, we will demonstrate that simple networks with an information maximization component can discover some linguistic rules in unsupervised ways. Many attempts have been made to model natural language acquisition processes by using neural networks [1], [2], [3]. The majority of the approaches were based upon supervised learning. However, it has been argued that one acquires one's first language without being given explicit target forms [4], [5]. It seems that we can acquire linguistic rules with partial information on target forms or even without targets. This is why some types of unsupervised or partially supervised learning methods are needed to build a realistic model of language acquisition.

Moreover, in past research, no guiding principle to govern overall acquisition has been postulated except some technical learning rules. In this paper, we propose a principle of information maximization by which networks eventually acquire linguistic rules. Information theoretic methods applied to neural computing have given promising results in various aspects of neural computing or pattern recognition. For example, Watanabe [6] pointed out in the early fifties that pattern recognition can be considered to be a process of entropy minimization. Linsker [7] applied information methods to perception systems. Roughly speaking, Linsker argues that networks should self-organize to preserve the maximum amount of information, and under the influence of this principle several attempts have been made to apply information theoretic methods to neural computing [8], [9], [10], [11]. However, these approaches are so limited that their applications have been confined to some research areas. In this context, we will attempt to show that the hypothesis of information maximization alone can model some complex aspects of language acquisition.

To demonstrate the performance of information maximization methods, we chose the complex problem of the use of donatory verbs in Japanese. As much linguistic research has pointed out, context plays an important role in comprehending and producing messages in Japanese. A group of Japanese donatory verbs provides a good example of contextual determination of verb selection by a speaker. In our experiments, the contextual factors are incorporated into input data to determine whether networks can discover these factors and produce appropriate linguistic rules.

2 Methods

For the inference of the donatroy verbs, we used a new unsupervised learning method with an information maximization component (see [12] for a detailed methodological explanation). We assumed that the information max-

imization method can detect or discover some features of input patterns not by imitating input patterns but by controlling directly competitive unit activation patterns. The information I in a neural system was defined by

$$I = -\sum p(j)\log p(j) + \sum\sum p(s)p(j \mid s)\log p(j \mid s), \quad (1)$$

where $f(x)$ is any differentiable function (the sigmoid function in the experiments), $p(j)$, $p(s)$ and $p(j \mid s)$ represent the probability of the jth competive unit, the sth input pattern and a conditional probability for j given s, respectively. All these probabilities were approximated by the normalized actual competitive units shown in Figure 1.

One of our basic premises was that information about input patterns must be maximized. Information was supposed to be represented in the form of competitive unit activation patterns.

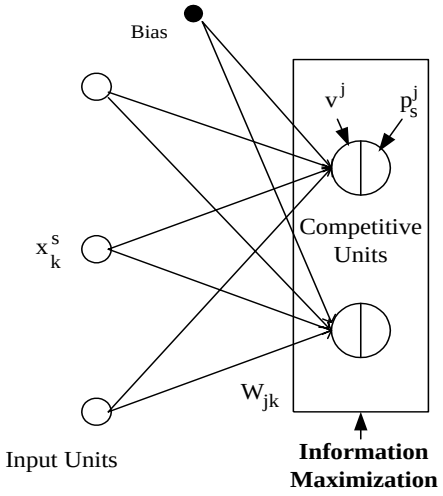


Figure 1: A network architecture that maximizes information about input patterns.

3 Inference of Japanese Donatory Verbs

3.1 Donatory Verb System

Any event of transfer can be looked at from two different points of view. When viewed from the giver's perspective, an event of transfer is a "giving action"; on the other hand, the same event becomes a "receiving action" if it is viewed from the receiver's perspective.

First, consider a "giving action." In describing a giving event, English has only one verb *give*. On the other hand, Japanese has a set of different verbs, and a speaker of Japanese selects an appropriate verb, depending on contextual factors that characterize the relationship between the giver and receiver. We consider the "in-group/out-group relationship" between the giver and the receiver and their relative social status to be the two contextual factors that determine the selection of appropriate donatory verbs in Japanese [13], [14].

Now consider a giving event where the second contextual factor, i.e., social status, is kept constant. This is a situation in which the giver and receiver are of the same social status. In this case, either *ageru* or *kureru* is used in Japanese. What determines a selection of these two verbs is the notion of "in-group/out-group." A giver and a receiver in a giving event can be located on the "in-group/out-group continuum." The speaker is the core of the in-group, and the in-group consists of those to whom the speaker feels psychological proximity. As a basic rule, when the speaker or an in-group member is a giver and the receiver is from the out-group member, *ageru* is chosen; on the other hand, when the giver belongs to the speaker's out-group and the receiver is from the in-group, *kureru* is used.

Next, consider a "receiving action." In describing a receiving event, English uses *receive*, while Japanese uses *morau*. Here again, the notion of in-group/out-group affects the use of the verb *morau*. As a basic principle, when using the verb *morau*, the grammatical subject of a sentence (which corresponds to the receiver in a receiving action) has to be the speaker or an in-group member. It is extremely rare to put an out-group member in the position of the grammatical subject [15].

3.2 Experiment No. 1

The objective of Experiment 1 is to show to what extent networks with an information maximization component can infer the correct use of the two verbs *ageru* and *morau*, depending upon the different relationships between various pairs of giver and receiver. In this experiment, four participants were considered ("I," "my wife," "my colleague," and "a stranger"). As Figure 2 shows, those four can be located on the in-group/out-group continuum. The giving actions conducted between an in-group giver and an out-group receiver can be depicted by the arrow (A), while those conducted between an out-group giver and an in-group receiver can be represented by the arrow (B) in Figure 2. In addition, the receiving actions taking place between an in-group receiver and an out-group giver can be illustrated by the arrow (C).

Because we used unsupervised learning, we incorporated some contextual information into the data. In the present study, we attempted to differentiate the distances between various pairs of giver and receiver on the in-group/out-group continuum for the data. For this purpose, "I", "my wife," "my colleague," and "a stranger" were encoded as $000, 001, 011, 111$, respectively. Because there were both a giver and a receiver, the number of input units was thus six. The number of competitive units was limited to two to simplify the experiments.

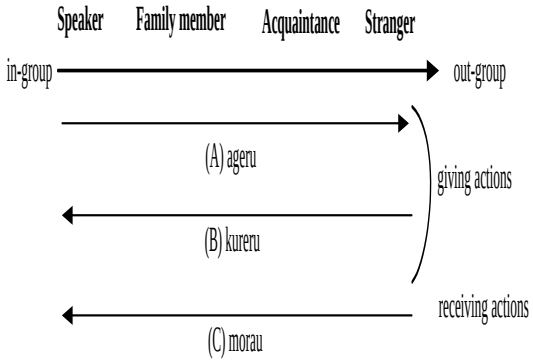


Figure 2: Use of *ageru*, *kureru*, *morau*,

Figure 3 shows the information I as a function of learning epochs. I was normalized for its values to range between 0 and 1. As can be seen in the figure, information rapidly increases, and reaches its maximum at about 600 epochs.

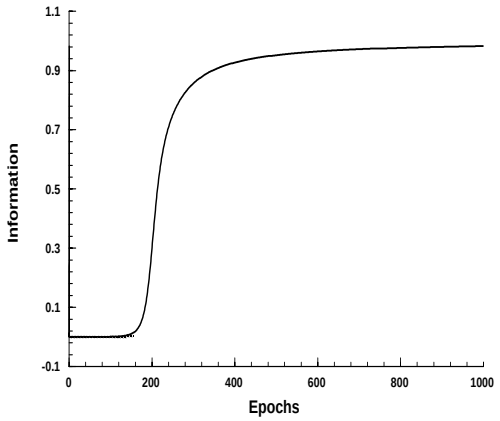
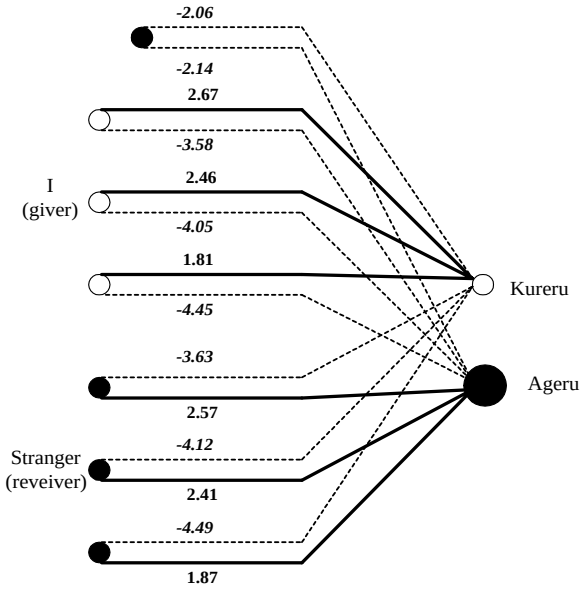
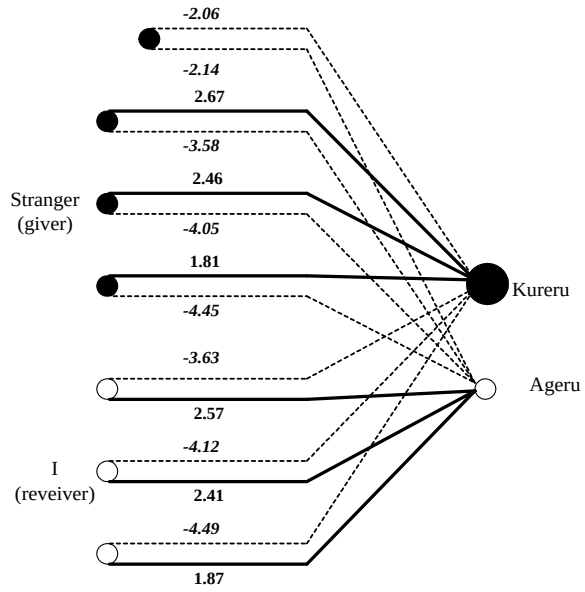


Figure 3: Information as a function of the learning epochs. for the verb *ageru*.

The networks could accurately classify the input data into *ageru* and *kureru*. Figure 4(a) shows that in the event where the speaker ("I") is the giver, and "a stranger" is the receiver, the verb to be used is *ageru*. The second competitive unit is turned on by the strong positive connections from the receiver units. On the other hand, Figure 4(b) illustrates the action in which "a stranger" is the a giver while the speaker "I" is the receiver. In this case, the positive connections from the giver units turn on the first competitive unit, and thus the verb *kureru* is chosen.



4(a) Ageru



4(b) Kureru

Figure 4: Internal Representations obtained by the parameter $\alpha = 0$. (a) and (b) show a case for the verb *Ageru* and *Kureru*.

3.3 Experiment No. 2

In the second experiment, the third verb *morau* was introduced. As mentioned before, when one expresses a certain action of transfer in a sentence in Japanese, the selection of the three verbs (*ageru*, *kureru*, and *morau*) is determined by (1) whether the transfer is a giving event or a receiving event, and (2) at what positions the giver

and the receiver are located on the in-group/out-group continuum. Our encoding system took these two factors into consideration. Specifically, we introduced an additional code (0 and 1), with 0 representing the receiver while 1 representing the giver. This code was prefixed to the code of each participant in order to differentiate a giving and receiving events in which the same participants are involved. Thus, for instance, a receiving action where the speaker ("I") is the receiver and "a stranger" is the giver was encoded as 00001111; on the other hand, a giving event in which the speaker ("I") is the giver and "a stranger" is the receiver was represented as 10000111.

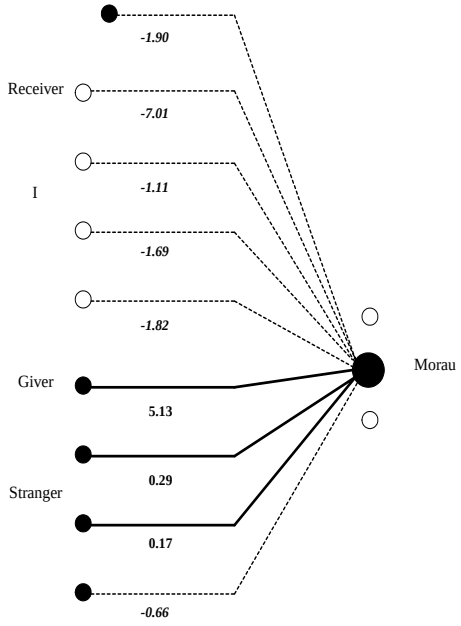


Figure 5: Internal representation for the verb *moraru*.

Figure 5 represents the receiving situation where the speaker ("I") is the receiver and "a stranger" is the giver. The strong positive connection from the giver unit turn on the second competitive unit for *morau*.

3.4 Experiment No. 3

In the previous section, we discussed the giving and receiving events where the giver and receiver are of the same social status. What happens if the two participants differ in social status? The basic framework explained in the previous sections also applies to such a situation; however, three others verbs, i.e., *sashiageru*, *kudasaru*, *itadaku* now come into the picture.

First, consider giving actions where the giver belongs to the speaker's in-group. In these actions, when the receiver is higher in social status than the giver, the verb *sashiageru* is used; when the receiver is lower in social status than the giver, the verb to be chosen is *ageru*. On the other hand, in giving events where the giver belongs to the speaker's out-group, if the giver is higher in social status than the receiver, *kudasaru* is used, while in a reverse situation where the giver has a lower social status than the receiver, *kureru* is chosen. Finally, in receiving events, if the giver occupies a higher status than the receiver, *itadaku* is used; if the giver has a lower social position than the receiver, the verb to be used is *morau*. *Sashiageru*, *kudasaru* and *itadaku* are honorifics that expresses politeness.

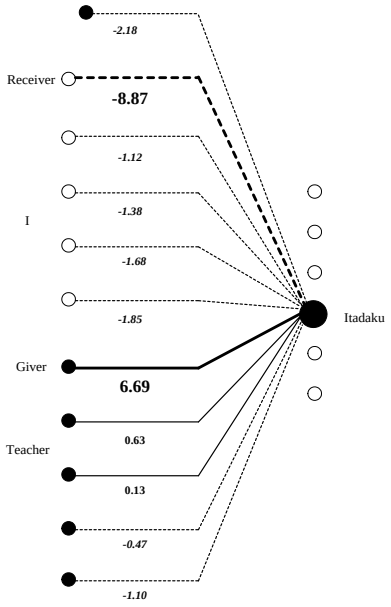


Figure 6: A representation for *itadaku*.

Taking the factors of social status into consideration, we used the six donatory verbs, *ageru*, *sashiageru*, *kureru*, *kudasaru*, *morau* and *itadaku*. In this experiment, we used five hypothetical persons, the speaker ("I"), his/her younger brother ("my brother"), his/her father ("my father"), his/her junior ("my junior") and his/her teacher ("my teacher"), and we examined giving and receiving actions among those five hypothetical participants.

It was found that the information maximization method can infer the correct use of the donatory verbs with about 90 percent accuracy (maximum value), while the traditional competitive learning predicts the use with 80 percent accuracy (maximum value). Figure 6 depicts the internal representation for a donatory verb *itadaku*. This was one of the most frequently observed patterns in the experiments. In the figure, we can find that the fourth competitive unit responds to the verb *itadaku*.

4 Conclusion

We showed that information maximization alone can allow unsupervised neural networks to acquire the complex linguistic rules underlying Japanese donatory verbs. The present results underscore the potential of information maximization in this domain.

Future research, however, is needed to further refine

this information maximization method. An example will be an attempt to enhance the networks' performance for the selection of the six different verbs dealt with in Experiment 3. As Maynard [16] points out, besides natural acquisition, conscious learning or training that involves focused attention to target forms are needed for honorifics such as *sashiageru*, *kudasaru*, and *itadaku*, as opposed to non-honorifics such as *ageru*, *kureru*, and *morau*. For this reason, we assume that a hybrid system will yield more positive results for the detection of the entire donatory verbs because the system incorporates both unsupervised and supervised learning components that correspond to natural acquisition and conscious learning, respectively.

References

[1] K. Plunkett and V. Marchman, "U-shaped learning and frequency effects in a multilayered perceptron: Implication for child language acquisition," *Cognition*, vol. 38, pp. 43–102, 1991.

[2] V. A. Marchman, "Constraints on plasticity in a connectionist model of the english past tense," *Journal of Cognitive Neuroscience*, vol. 5, no. 2, pp. 215–234, 1993.

[3] D. E. Rumelhart and J. L. McClelland, "On learning the past tenses of english verbs," in *Parallel Distributed Processing* (D. E. Rumelhart, G. E. Hinton, and R. J. Williams, eds.), vol. 2, pp. 216–271, Cambrige: MIT Press, 1986.

[4] R. Brown and C. Hanlon, "Derivational complexity and order of acquisition," in *Cognition and Development of Language* (R. Hayes, ed.), New York: Wiley, 1970.

[5] M. J. Demetras, K. N. Post, and C. E. Snow, "Feedback to first language learners: The role of repetitions and clarification questions," *Journal of Child Language*, vol. 13, pp. 275–292, 1986.

[6] S. Watanabe, *Pattern Recognition*. New York: John Wiley & Sons, 1985.

[7] R. Linsker, "Self-organization in a perceptual network," *Computer*, vol. 21, pp. 105–117, 1988.

[8] R. Linsker, "How to generate ordered maps by maximizing the mutual information between input and output," *Neural Computation*, vol. 1, pp. 402–411, 1989.

[9] J. J. Atick and A. N. Redlich, "Toward a theory of early visual processing," *Neural Computation*, vol. 2, pp. 308–320, 1990.

[10] S. Becker, "Mutual information maximization: models of cortical self-organization," *Network: Computation in Neural Systems*, vol. 7, pp. 7–31, 1996.

[11] S. Becker and G. E. Hinton, "Learning mixture models of spatial coherence," *Neural Computation*, vol. 5, pp. 267–277, 1993.

[12] R. Kamimura and T. Kamimura, "Structural information and linguistic rule extraction," in *Proceedings of ICONIP-2000*, pp. 720–726, 2000.

[13] A. M. Niyekawa, *Minimum Essential Politeness: A Guide to the Japanese Honorific Language*. Tokyo: Kodansha International, 1991.

[14] N. Tsujimura, *An Introduction to Japanese Linguistics*. Malden, MA: Blackwell, 1996.

[15] S. Makino, *Uchi to Soto no Genngobukagaku (A Linguistic-Cultural Study of Uchi and Soto)*. Tokyo: ALC, 1996.

[16] S. K. Maynard, *Japanese Communication: Language and Thought in Context*. Honolulu: University of Hawaii Press, 1997.