Learning of First, Second, and Third Person Pronouns

Yuriko Oshima-Takane (yuriko@hebb.psych.mcgill.ca)
Marina Takane (m.takane@marianopolis.edu)
Yoshio Takane (takane@takane2.psych.mcgill.ca)
Department of Psychology; McGill University; 1205 Penfield Avenue
Montreal, Quebec, Canada H3A 1B1

Abstract

This paper presents simulation results and network analysis of generative Cascade-Correlation (CC) networks which model the child's learning of English personal pronouns. The network analysis revealed that overheard speech is crucial in learning the correct semantic rules not only for first and second person pronouns but also for third person pronouns. In addition, in order to induce the fully correct semantic rules without error-correcting feedback, the networks need to learn all three personal pronouns. Network analysis techniques used in the present study proved to be a powerful tool for understanding of what the networks are actually learning.

Introduction

Recent studies have shown that the learning of personal pronouns is gradual and cannot be explained by any fast mapping mechanisms specific to word learning as proposed by the special lexical constraints approach (Markman, 1989; Oshima-Takane, 1985, in press). Unlike count nouns and proper nouns, the referent of a personal pronoun shifts systematically depending on the discourse situation, although each pronoun has a fixed meaning (Kaplan, 1978). In order to acquire the correct semantic rules for personal pronouns, children must discover the systematic relationship between pronouns and speech roles. First person pronouns refer to the person using the pronoun, while second person pronouns refer to the person addressed. Third person pronouns refer to a non-addressee who is neither speaker nor addressee, and have distinct masculine and feminine forms.

Oshima-Takane's (1985) model for pronoun learning has provided a detailed picture of the changing semantic representations children show in the course of acquisition. The basic assumption of the model is that children first determine the referent of a pronoun in each occurance of use and then induce its semantic rules. What type of semantic rules children induce depends on what type of input they have received and what kind of constraints children bring to the learning situation (e.g., prior knowledge relevant to the word learning, and limited attention, memory, or information processing). According to this model, whether the input contains the following situations is specifically important: (a) situations in which children can recognize that the same pronoun used in different discourse situations refers to different persons and (b) situations in which they can recognize that the different pronouns refer to the same person. The former situations (a) provide information necessary for distinguishing personal pronouns from proper names because children could avoid a proper name interpretation for a pronoun (e.g., "you" is another name of the child) if the pronoun is not limited to a specific individual. The latter situations (b) facilitate the learning of the distinctive meanings among different forms of personal pronouns (i.e., first, second, and third person pronouns). This is because when different forms refer to the same individual, children are forced to search any data that can distinguish them in the input. They are forced to utilize speech role information as well as referent information. As a result, children are more likely to learn the distinctive meanings among first, second, and third person pronouns.

In support of this model, Oshima-Takane and her collaborators have conducted a series of experimental and observational studies and have provided empirical evidence that overheard speech is an input necessary for inducing the correct semantic rules without error-correcting feedback because it provides children with these two situations, whereas child-directed speech does not (Oshima-Takane, 1988, 1992; Oshima-Takane, Goodz, & Derevensky, 1996). Recent computer modeling studies on the learning of first and second person pronouns with the Cascade-Correlation (CC) learning algorithm (Fahlman & Lebiere, 1990; Oshima-Takane, Takane, & Shultz, 1995) have provided converging evidence in support of the hypothesis that children induce the incorrect, me-you reversal rules by observing pronouns used in child-addressed speech, whereas they induce the correct semantic rules by observing pronouns used in non-addressed speech. In addition, the studies have shown that the prior knowledge that individuals appearing in training patterns were members of the same kind facilitates networks' generalization capability to untrained patterns (Oshima-Takane, 1985). Network analysis of non-addressee networks revealed, however, that the networks learned either one of the two possible partially correct functions (i.e, partially correct with regard to the correct semantic rules). They correctly produced a pronoun referring to a person who takes either a speaker or an addressee role (e.g., "me" referring to speaker) but overgeneralized the other pronoun referring to the non-addressee (e.g., "you" referring to addressee and non-addressee). Because no training patterns were given for the situations where neither speaker nor addressee is referent, it is quite natural that the networks showed overgeneralizations to those patterns. Brener (1983) reported that children, too, showed similar overgeneralization of the second person pronouns to non-addressee before they acquire the third person pronouns. Further, they show similar errors for third person pronouns, where third person pronouns are used to refer to both addressee and non-addressee before they understand that third person pronouns refer only to non-addressee.

A primary motivation for the present simulation study is to understand the mechanisms by which children learn to produce personal pronouns without explicit corrections. In particular, we investigated if, with the addition of third person pronouns in non-addressed speech, the CC networks can correct overgeneralization errors and can induce the correct semantic rule for first and second person pronouns. Further, we investigated if overheard speech is also crucial for learning the correct semantic rule for third person pronouns.

The CC learning algorithm

Unlike static feed-forward networks such as back-propagation networks, a CC network begins with a minimal network topology consisting of only the input and output units, to which hidden units are added and trained automatically to improve performance. Hidden units are added one at a time until error is within a range specified by the user, at which point learning has been accomplished.

Each input unit is connected to the output units by an adjustable weight. Initial weights are selected randomly, and are adjusted based on activations given in the training patterns. When performance cannot be improved any further by weight adjustments, a hidden unit with a sigmoid activation function is recruited, producing nonlinear interaction effects in the mapping of inputs to outputs. Incoming weights to this new unit are determined by maximizing correlation between the unit's activation and network error, and are fixed throughout the remainder of the training period. Thus error is not propagated back across different levels of the network, resulting in quicker, more stable convergence. After the hidden unit has been recruited, output weights are readjusted to optimize performance. This cycle of error reduction is repeated until an acceptable range is reached.

Simulation

The present simulation consists of three training phases to simulate how children learn to produce all pronoun forms by listening to other persons producing them. In Phase I training, networks learn other-speaking patterns with a first person pronoun "me" and a second person pronoun "you". In Phase II training they learn other-speaking patterns with the third person pronouns "he" and "she" added to Phase I training patterns. In Phase III training, child-speaking patterns with all pronoun forms are added to Phase I and Phase II training patterns. Third person pronouns were presented to the networks after exposure to first and second person pronouns in order to simulate the child's pronoun learning environment (Oshima-Takane & Derat, 1996).

The initial CC networks in this simulation had three input units representing speaker, addressee, and referent. In addition, there was a bias unit having a value of +1 on the input side. Five persons, child, mother, father, and two additional persons, one female and one male, appeared in the training patterns. Analog coding was used in order to implement prior knowledge that individuals appeared in the training patterns were in the same kind PERSON. The child was coded as 0, the mother as +2, the father as -2, and the other two persons were coded as +1 and -1. Positive values represented female persons and negative values represented male persons, to code for gender in this system. The gender of the child was not explicitly coded and must be derived from the regularities of pronoun use in the training patterns. In this simulation, the child was treated as female.

Localist coding was used to code for the output pronoun. Of the four units, three had a negative value, while the remaining unit was positive. The position of this positive unit determined the pronoun. The pronouns "me", "you", "she", and "he" were represented when the first, second, third and fourth units were positive, respectively. Each input unit was connected to four output units. Figure 1 depicts a network after the recruitment of two hidden units.

Networks were trained under three different conditions: pure addressee, pure non-addressee, and mixed. In Phase I and Phase II training, networks in the pure addressee condition were trained with the addressee patterns in which the child was the addressee, whereas those in the pure non-addressee condition were trained with the non-addressee patterns in which the child was neither the addressee nor the speaker. In the mixed condition, networks were trained with a combination of equal numbers of addressee and non-addressee patterns. As there were 20 addressee, 60 non-addressee, and 20 child-speaking patterns, the number of training patterns was equalized across conditions. In each phase, the addressee patterns were given six times per epoch for networks in the addressee condition, while non-addressee patterns were given twice for those in the non-addressee condition. Networks in the mixed condition were given the addressee patterns three times and the non-addressee patterns once per epoch, for an equal number of patterns across conditions. It should be noted that repeated patterns did not

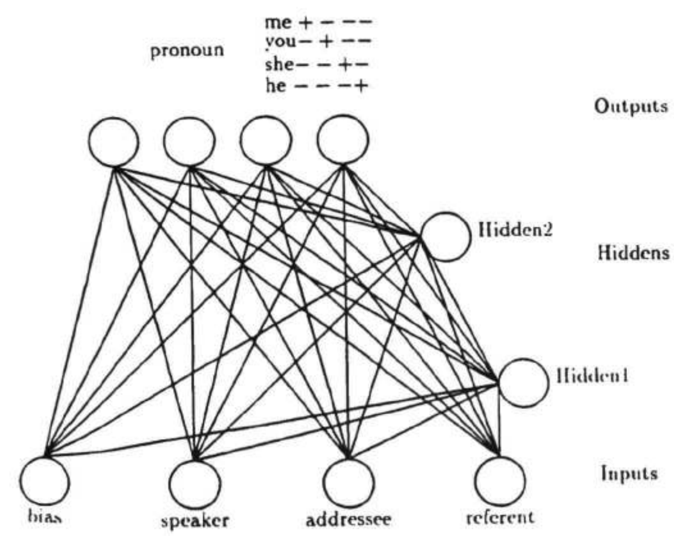

Figure 1: Pronoun network after recruitment of two hidden units.

affect learning time significantly. Child-speaking patterns where the child was the speaker were added in the final phase of training for all conditions.

It was expected that networks in the addressee condition would learn incorrect, me-you reversal rules during Phase I training and would produce "you" in reference to the child exclusively and "me" in reference to any other person. In Phase II training, four persons other than the child were referred to by more than one pronoun, specifically "me" and "she/he". Thus, the addressee networks were expected to produce "me" in reference to another person as speaker, and "she/he" in reference to another person as non-addressee. However, no third person pronouns would be produced in reference to the child as non-addressee, because the gender of the child could not be derived from the addressee patterns.

The networks in the non-addressee condition and those in the mixed condition were expected to induce either one of the two possible partially correct functions in Phase I training and would show overgeneralization errors of either "me" or "you" to non-addressee. However, with the addition of third person pronouns in Phase II training, these networks would learn the fully correct semantic rules for all pronoun forms. This is because all four other persons were referred to by three pronouns in the training patterns. Therefore, the networks were forced to utilize the speech role information, not just referent information. The performance of the mixed networks was expected to be slightly better than that of the non-addressee networks, since the only difference between them is that the child is referred to by both "you" and "she" in the mixed condition as opposed to only "she" in the non-addressee condition.

Results

Table 1 summarizes the mean epochs required for learning training patterns by phase and by condition. Non-addressee networks took a significantly greater number of epochs to train in Phase I compared to mixed networks, $t(57)=5.46$, one-tailed, $p<0.001$. In addition, the number of epochs for the addressee networks was significantly fewer than those in the non-addressee and mixed condition combined, $t(54)=-55.92$, one-tailed, $p<0.001$. All networks recruited one hidden unit. In Phase II, the difference in number of epochs between the non-addressee and mixed networks was not significant; however, addressee networks again took significantly fewer epochs compared to those in the non-addressee and mixed conditions combined, $t(52)=-26.15$, one-tailed, $p<0.001$. Networks recruited an average of 1.95 hidden units (range: 1 - 3) in the addressee condition, 3.20 (range: 3 - 4) in the non-addressee condition, and 3.95 (range: 3 - 5) in the mixed condition. In Phase III training, non-addressee networks required significantly more epochs to learn than mixed networks, $t(26)=2.54$, the separate variance estimate, one-tailed, $p<0.01$. Addressee networks required significantly more epochs to learn the child-speaking patterns compared to the non-addressee and mixed networks combined, $t(20)=12.27$, the separate variance estimate, one-tailed, $p<0.001$. Addressee networks recruited an average of 2.65 hidden units (range: 2 5), while none of the non-addressee and mixed networks required any additional hidden units in Phase III. One non-addressee network and four mixed networks did not require any Phase III training at all.

Table 1: Mean epochs required for the three phases of learning by condition.

		Condition		
		Pure addressee n=20	pure non-addressee n=20	mixed n=20
Phase I	*Mean*	96.8	266.8	243.9
	SD	6.15	16.21	15.16
	Range	82-107	239-310	218-285
Phase II	*Mean*	281.2	742.8	771.2
	SD	47.98	73.95	108.70
	Range	225-391	637-955	445-888
Phase III	*Mean*	458.2	31.2	10.8
	SD	158.35	33.08	13.84
	Range	237-766	0-127	0-65

Network Analysis

Network analysis was conducted in order to examine the function the networks have learned and their generalisation capabilities. In order to depict the network representation, separate graphs were made for the father, mother, and child as referents. In addition, two graphs were presented for each referent: r1 which distinguishes between the pronouns "me" and "you", and r2 which distinguishes between the male and the female third person pronouns "he" and "she". In both, the left horizontal axis represents the speaker dimension, and the right horizontal axis represents the addressee dimension. Numbers on these dimensions represent who the speaker and addressee are, ranging from -2 to +2. The vertical axis represents the output pronoun.

Target function

We first define

$$y1=\text{sigmoid}(-c\{(S-R)^2-0.125\})$$
$$y2=\text{sigmoid}(-c\{(A-R)^2-0.125\})$$
$$y3=\text{sigmoid}(c(R+0.25))$$

where S, A, and R represent the values for the speaker, addressee, and referent, and where c is some large positive value (e.g., c=5000). The calculated value then undergoes the sigmoid transformation $1/(1+e^{-x})$, where x is the value to be transformed. Thus, y1 will equal 1 if the speaker and referent agree (ie. "me" is produced) and will equal 0 if they disagree. The same can be said about y2, but with the addressee instead of the speaker. In y3, female referents are distinguished from male referents, as 1 is produced when the referent is non-negative (female) and 0 is produced when the referent is negative (male).

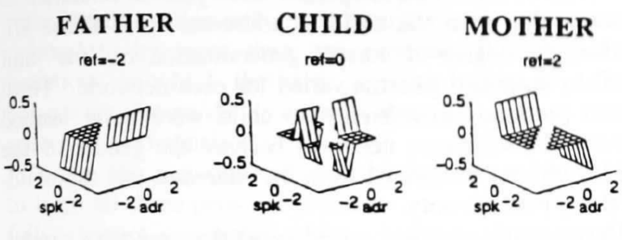

Figure 2: Target representation for correct use of first and second person pronouns (r1). me=+0.5, you=-0.5, he/she=0

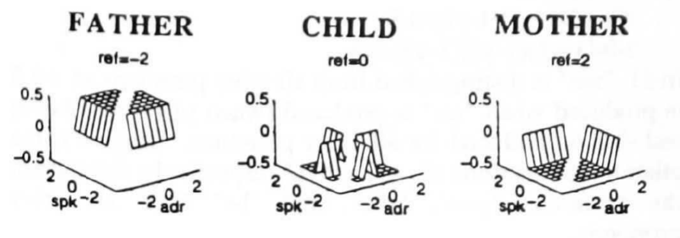

Figure 3: Target representation for correct use of third person pronouns (r2). he=+0.5, she=-0.5, me/you=0

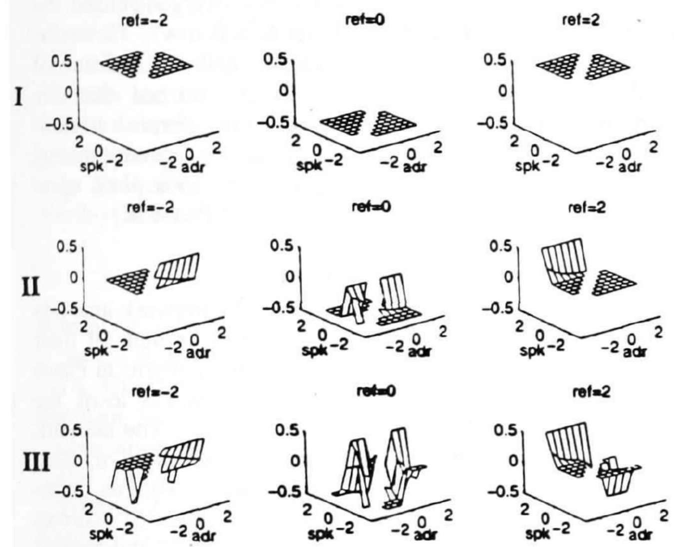

Figure 4: Addressee network representations after Phase I, Phase II and Phase III training for use of first and second person pronouns (r1).

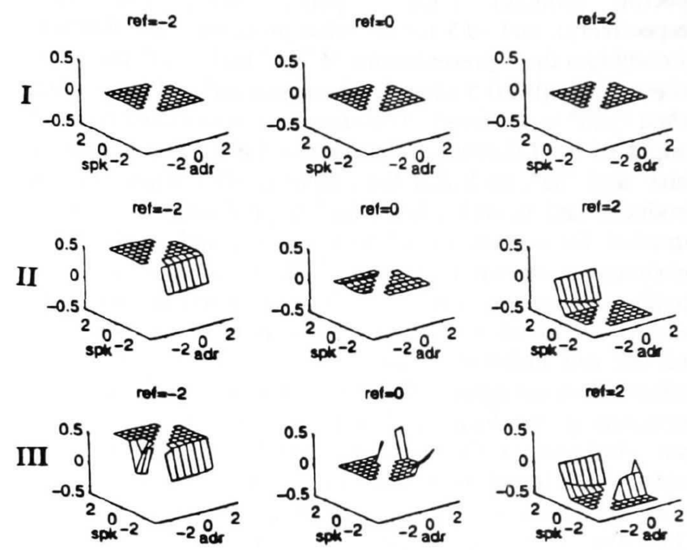

Figure 5: Addressee network representations after Phase I, Phase II and Phase III training for use of third person pronouns (r2).

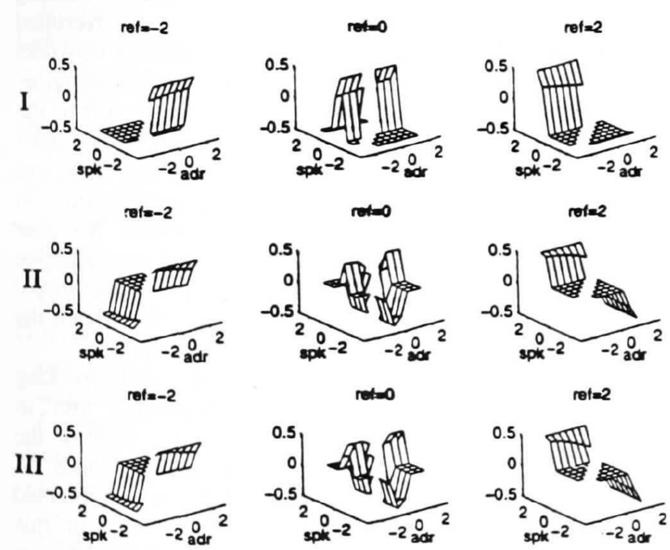

Figure 6: Non-addressee network representations after Phase I, Phase II and Phase III training for use of first and second person pronouns (r1).

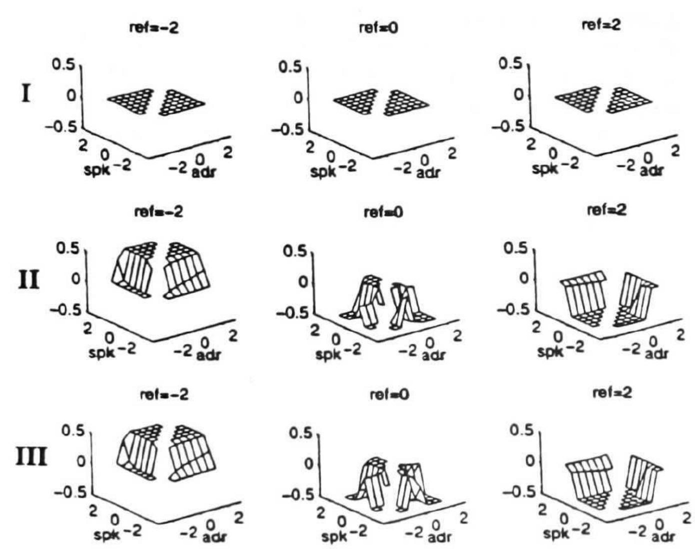

Figure 7: Non-addressee network representations after Phase I, Phase II and Phase III training for use of third person pronouns (r2).

We then define

$z1=y1-0.5$

$z2=y2-0.5$

$z3=y3(1-y1)(1-y2)-0.5$

$z4=(1-y1)(1-y2)(1-y3)-0.5$

In z1, "me" is distinguished from all other pronouns as +0.5 is produced when "me" is produced (when y1=1, see above) and -0.5 is produced for all other pronouns. Similarly, the other three functions z2, z3, and z4 respectively distinguish the pronouns "you", "she", and "he" from all other pronouns.

To conserve space, we further transform z's into:

$r1=0.5x(z1-z2)$

$r2=0.5x(z4-z3)$

where z1, z2, z3, and z4 are functions producing +0.5 for a specific pronoun ("me", "you", "she", and "he", respectively), and -0.5 for all other pronouns. The function r1 combines the representations of "me" and "you" into one, where r1 equals +0.5 when "me" is produced and equals -0.5 when "you" is produced. The other pronouns would have an output of 0. Likewise, r2 combines the representations of "she" and "he", such that the output is +0.5 when "he" is produced and is -0.5 when "she" is produced. The target function for correct use of first, second, and third person pronouns is shown in Figures 2 & 3. In Figure 2, the graphical representation of r1, "me" is produced when the referent and speaker agree, and "you" is produced when the referent and addressee agree. At all other points, where referent does not agree with speaker or addressee, third person pronouns are produced, and therefore the output should be zero. In Figure 3, the representation of r2, "he" is produced when the referent is a male third person, and "she" is produced when the referent is a female third person. Otherwise, first or second person pronouns are produced, and the output is zero.

Function Approximations

Network analysis examined the graphical representations of network approximations at training points and the generalisation to test points. Generalisation tests consisted primarily of interpolation between training points. In general, the closer the networks' approximations were to the target function, the greater their generalisation capability.

Figures 4 & 5 present the function approximation of a network in the pure addressee condition by phase. As seen in Figure 4, when the child is the referent (ref=0), the graphical surface after Phase I training is at the -0.5 level, indicating that "you" is produced. Also, when the referent is the mother (ref=2) or father (ref=-2), "me" is produced, as the graphical surface is at +0.5. Thus, networks learned the incorrect, me-you reversal rule during Phase I training. During Phase II training, the network learned to differentiate between the referents as speakers and non-speakers: "me" was produced when the referent is the speaker, and third person pronouns were produced elsewhere. The exception to this was when the child was the referent, who was referred to almost exclusively by the pronoun "you", except when taking on the role of speaker.

The result seemed to indicate that the networks were able to discern that "you" was not the correct pronoun to produce in this situation but were unable to discover which pronoun should be produced. Networks did not learn to produce "me" in reference to the child as speaker and "you" in reference to persons other than the child as addressee until Phase III. Further, the degree of correct generalisation of "me" and "you" to untrained patterns varied for each network. Third person pronouns in reference to child were never learned correctly by addressee networks because the gender of the child could not be derived from the addressee and the child-speaking patterns only.

The non-addressee and mixed networks were very similar, and therefore only one set of figures for a non-addressee network (Figures 6 & 7) are presented. As the network trained with me-you patterns only in Phase I, a partially correct function was formed. In this case, the network learned the pronoun "me" correctly, but overgeneralised the use of "you" to non-addressee (Fig. 6, top row). However, this was corrected in Phase II training, with the addition of third person pronouns, as the network learned that the pronouns "he" or "she" should be used in reference to non-addressees. Network analysis of more detailed developmental data indicated that correct learning of "you" took place upon recruitment of one or more hidden units in Phase II.

Discussion

Both the analysis of learning time and the network analysis indicate that addressee networks undergo the bulk of their learning after the addition of child-speaking patterns in Phase III, whereas non-addressee and mixed networks learn the correct semantic rules in the first two phases. The network analysis clearly indicated that with the addition of third person pronouns to the Phase I training patterns, non-addressee and mixed networks were able to correct overgeneralization errors of either "me" or "you" and learned the fully correct semantic rules for all personal pronoun forms. In fact, some of the mixed and non-addressee networks showed perfect generalization to child-speaking patterns without Phase III training. The remaining mixed and non-addressee networks needed some Phase III training for simply adjusting weights, as none of them recruited hidden units. On the other hand, all the addressee networks needed Phase III training to correct the me-you reversal errors and never learned the third person pronoun in reference to the child. When the representations of the networks after completion of all the training are examined, non-addressee and mixed networks showed better generalisations to untrained patterns compared to addressee networks. No clear difference was observed between mixed and non-addressee networks except that there were slightly more mixed networks than non-addressee networks which could learn the correct semantic rules without Phase III training.

An interesting finding is that without child-speaking patterns, addressee networks were unable to produce "you" in reference to a person other than the child. Instead, the addressee networks produced third person pronouns in reference to others as addressee, although the networks could have produced "me" rather than "he" or "she" in this situation. The consistent overgeneralization of third person pronouns to others as addressee suggests that addressee networks were influenced by the fact that a person other than the child was referred to by "he" or "she" (12 patterns) more frequently than by "me" (4 patterns) in the addressee training patterns. This, in turn, suggests that the networks used referent information only, even though the same person was referred to by more than one pronoun. It appeared that situations in which the same pronoun refers to different persons are also needed in order to force the networks to use speech role information.

In sum, the network analysis revealed that networks need to learn all three personal pronouns in order to induce the fully correct semantic rules without error-correcting feedback. In addition, the results confirmed that overheard speech is crucial in learning the correct semantic rules not only for first and second person pronouns but also for third person pronouns.

Acknowledgements

This research was supported by grants from Natural Sciences and Engineering Research Council of Canada to Yuriko Oshima-Takane and to Yoshio Takane and by a team grant from Fonds pour la Formation de Chercheurs et l'Aide a la Recherche to Thomas Shultz, Yoshio Takane, and Yuriko Oshima-Takane.

References

Brener, R. (1983). Learning the deictic meaning of third person pronouns. *Journal of Psycholinguistic Research, 16*, 330-352.

Fahlman, S.E., & Lebiere, C. (1990). The cascade correlation learning architecture. In D.S. Touretzky (Ed.), *Advances in neural information processing systems* 2 (pp. 524-532). San Mateo:Morgan Kaufmann.

Kaplan, D. (1978). On the logic of demonstratives. *Philosophical Logic*, 81-98.

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

Oshima-Takane, Y. (1985). *Learning of pronouns.* Unpublished Ph.D. thesis submitted to McGill University.

Oshima-Takane, Y. (1988). Children learn from speech not addressed to them: The case of personal pronouns. *Journal of Child Language, 15*, 94-108.

Oshima-Takane, Y. (1992). Analysis of pronominal errors: A case study. *Journal of Child Language, 19*, 111-131.

Oshima-Takane, Y. (in press). The learning of first and second person pronouns in English. In R. Jackendoff, P. Bloom, & K. Wynn (Eds), Language, Logic, and Concept: In Memory of John Macnamara. Cambridge, MA: MIT Press.

Oshima-Takane, Y. & Derat, L. (1996). Nominal and pronominal references in maternal speech during the later stage of language acquisition: A longitudinal study. *First Language, 16*, 319-338.

Oshima-Takane, Y., Goodz, E., & Derevensky, J.L. (1996). Birth order effects on early language development: Do second born children learn from overheard speech? *Child Development, 67*, 621-634.

Oshima-Takane, Y., Takane, Y., & Shultz, T.R. (1995). *Learning of personal pronouns: Network model and analysis.* Technical Report 2095, McGill Papers in Cognitive Science, McGill University, Montreal.