The acquisition of personal pronouns in English: Neural network simulations by Knowledge-Based Cascade Correlation

Yuriko OSHIMA-TAKANE (McGill University)
Shudan WANG (NY Presbyterian-Weill Cornell Medical Center)
Marina TAKANE (McGill University)
Yoshio TAKANE (McGill University)

Abstract

This paper presents computer simulation results and a network analysis of generative Knowledge-Based Cascade-Correlation networks modeling the acquisition of singular forms of personal pronouns in English. The results reveal 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, networks use prior knowledge of first and second person pronouns in order to learn third person pronouns even though they are not forced to do so. However, the type of prior knowledge the networks recruit depends on what they have learned previously and what they need to learn further.

1. Introduction

Learning the correct meaning and use of personal pronouns poses an interesting challenge for young children learning to speak. Unlike count nouns (e.g., man) and proper nouns (e.g., David), the referent of a personal pronoun shifts depending on the discourse situation, although it has a fixed meaning (Kaplan, 1978). Personal pronouns are a type of deixis for which the interpretation depends on the context of utterances (Clark, 1978). Children must discover the systematic relationship between pronouns and speech roles in order to acquire the deictic meaning of personal pronouns. In English, most typically-developing children show few production errors as they learn the deictic meaning of first and second person singular forms, generally by 3 years of age and that of third person pronouns by 5 years of age (Brenner, 1983; Oshima-Takane, 1985, 1999, 2014; Wells, 1985). However, a small number of children make persistent deictic errors, indicating that they have not yet understood that the interpretation of these pronouns depends on the context of utterances (Chiat, 1982; Clark, 1978; Evans & Demuth, 2012; Guerriero, 1998; Morgenstern, 2012; Oshima-Takane, 1998; Oshima-Takane, Cole, & Yaremko, 1993; Schiff-Myers, 1983).

Editor's Name, et al. (Eds.) *Studies in Language Sciences N*. pp. 1-pages.
Year, Kurosio Publishers, Tokyo, Japan.

Oshima-Takane (1985) has proposed a model which accounts for these individual differences in acquisition observed among children within the same pronoun learning mechanism. The basic premise of the model is that children first identify the referent of the pronoun in each instance in which they hear it in the input, and then induce its semantic rules. The type of semantic rules children induce for a particular pronoun depends on what type of input they have received, and what kind of constraints they bring to the learning situation (e.g., prior knowledge relevant to word learning, limited ability of attention, memory, or information processing). According to this model, overheard speech (speech not directed to the child) is essential for inducing the correct semantic rules because it allows children to observe two types of situations: one in which children can recognize that the same pronoun is used to refer to different persons, and one in which they can recognize that different pronouns are used to refer to the same person. The former situation helps children distinguish proper nouns from personal pronouns. The latter situation provides information for children to differentiate the meanings of first, second, and third person pronouns. This is because they are forced to search for any information in the input that can distinguish between the different types of pronouns (i.e., speech role information). This type of information is not available in child-directed speech. Therefore, it is difficult to learn the correct semantic rules for personal pronouns without exposure to some overheard speech.

This pronoun learning model was confirmed by experimental and observational studies showing the importance of overheard speech in pronoun learning. For instance, Oshima-Takane (1988) showed that only the children who had opportunities to observe the mother and the father playing me-you pointing games with each other (thereby simulating an overheard speech situation) were able to imitate parents' pointing gestures without errors when saying *me* or *you*. In contrast, children whose mother and father only played the game with them (in a child-directed speech situation) made errors in pointing gestures. Oshima-Takane, Goodz, & Derevensky (1996) found that second-born children who had plenty of opportunities to overhear conversations between the mother and the older sibling acquired first and second person pronouns earlier than first-born children.

The problem is, however, that the effect of child-directed speech cannot be tested empirically because real children cannot be raised in a pure child-directed speech or a pure overheard speech environment for obvious ethical reasons. It is here that computer simulation studies with neural networks become useful. Oshima-Takane and her collaborators (Oshima-Takane, Takane, & Takane, 1998) conducted a simulation study using a Cascade-Correlation (CC) learning algorithm (Fahlman & Lebiere, 1990) to examine the importance of overheard speech in pronoun learning. The results of their

simulation study confirmed Oshima-Takane's pronoun learning model. Pure non-addressee and mixed (addressee and non-addressee pattern) networks learned to use the correct pronouns from the other-speaking patterns with a little or no error-correcting feedback on the child-speaking patterns. The addressee networks, on the other hand, learned incorrect rules for pronouns from the other-speaking patterns and required extensive error-correcting feedback on the child-speaking patterns.

A primary motivation of the present study is to use a Knowledge-Based Cascade-Correlation network (KBCC) learning algorithm to confirm the mechanism by which children learn first, second and third person pronouns, as proposed in the previous CC network simulation study (Oshima-Takane, Takane, & Takane, 1998). Children use existing knowledge to speed up learning and to tackle more difficult tasks. However, CC networks always start learning new tasks from an initial minimal network configuration. Therefore, the previous simulation study used three training phases in order to force CC networks to use the prior knowledge of first and second person pronouns to learn third person pronouns. Unlike CC networks, however, KBCC networks can spontaneously use previously learned information (analogical learning) as well as new information (inductive learning) to acquire new knowledge, whenever the networks judge that the information is relevant to the new learning (Shultz, & Rivest, 2001). Thus, KBCC networks simulate pronoun learning in a more realistic way because they can capture children's ability to use relevant prior knowledge in learning. A second motivation of the present study is to use KBCC networks to confirm the previous finding that overheard speech is essential for learning the correct semantic rules of all three personal pronouns and to show that child-directed speech can be the source of pronominal errors (Oshima-Takane et al., 1998).

2. KBCC learning algorithm

Like CC neural networks, KBCC neural networks are based on a feed-forward, constructive learning algorithm that can grow to improve learning. The main difference lies in that while a CC network can only recruit single hidden units (Figure 1-a), a KBCC network can recruit previously learned networks (*source nets*) as well as single hidden units (Figure 1-b). Source nets or single hidden units are added to the network topology one at a time to minimize the error within a desirable range, at which point the learning is completed.

Initially, the input units and the bias unit are connected to the output units with random weights. During output phase training, the weights are adjusted to minimize the sum of squared errors between the network outputs and target outputs for all training patterns. If error reduction stagnates without reaching success, the algorithm switches to input phase training in

which the network chooses to recruit one unit from a pool of candidates. The candidate pool includes single units with sigmoid activation function and source nets. For each candidate, weights connecting non-output units

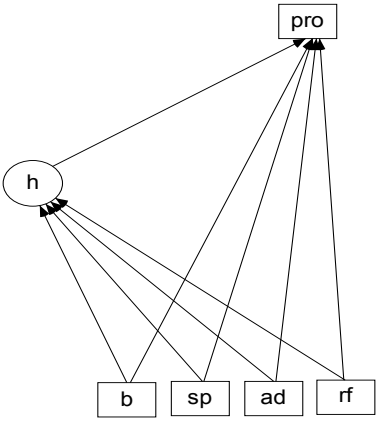

Figure1-a. An example of a CC network learning pronouns (pro) with four input units, bias (b), speaker (sp), addressee (ad), and referent (rf). One hidden unit (h) is recruited.

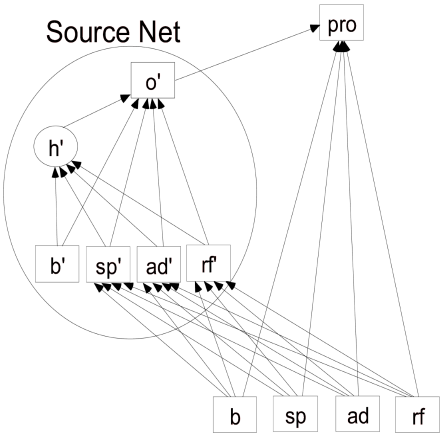

Figure 1-b. An example of a KBCC network learning pronouns. One CC network is recruited as a source net.

of the network to input units of the candidate are trained to maximize the correlation between candidate output and network error. The candidate with the greatest correlation would be selected and installed into the network. Then, the algorithm shifts back to output phase training to readjust the output weights for the newly recruited unit. The recruited unit's output can be viewed as an added 'input' to the network. The algorithm will move back and forth between output and input training until the error has been reduced to an acceptable level. The default score threshold was 0.4, meaning that the output values had to be within 0.4 distance from the targets to complete training (Shultz & Rivest, 2001).

3. Simulation

As in the previous simulation with CC networks (Oshima-Takane et al., 1998), a three-phase model was used to simulate the way children learn personal pronouns. Children learn how to produce personal pronouns by first listening to other people producing them. Phases I and II used *other-speaking patterns* to simulate this phenomenon. In Phase I training, networks learned other-speaking patterns for the first person pronoun *me* and the second person pronoun *you*. In Phase II, third person pronouns *he* and *she* training patterns were added to *me* and *you* training patterns. This phase simulates children's pronoun learning environment and their acquisition of third person pronouns at a later developmental stage, after having learned first and second person pronouns (Guerriero, 1998; Oshima-Takane & Derat, 1996). After a certain period of learning, children not only hear but begin to produce pronouns themselves. This was modeled in Phase III, by adding *child-speaking patterns*. Children may not have learned the correct semantic rules before they begin to produce pronouns. Therefore, they need to receive feedback from others when they produce pronouns incorrectly in order to correct their production errors. Phase III training simulated this error-correcting process by adding the child-speaking training patterns. If networks do not require any new learning in Phase III training, then they should recruit a single KBCC network from Phase II, but should not recruit any new hidden units. These networks are considered to have mastered the correct semantic rules in Phase III.

The learning of pronouns can be viewed as nonlinear function learning connecting inputs to outputs. There are three important input variables: who is talking to whom about whom. Thus, the KBCC networks in this simulation used 3 input variables: speaker (sp), addressee (ad), and referent (ref). A bias unit of value +1 was also included. In addition, previous

studies have shown that access to the kind PERSON[1] is important in pronoun learning, as it allows children to recognize that a pronoun refers to a member of the kind PERSON (Oshima-Takane, 1985, 1999; Oshima-Takane et al. 1999). Children must classify themselves and other people as members of the same kind PERSON before they can learn to use personal pronouns correctly. Therefore, analog coding was used in all the simulations to represent prior knowledge about the kind PERSON. Five number codes represent the five persons included in the input training patterns. The child was coded as 0, the mother as +2, and the father as -2. Two additional persons, coded as +1 and -1 were also added. Note that a positive code represents a female and a negative code represents a male. The child was treated as female in all the simulations. Four output units were used to distinguish between the four pronouns produced: *me* (+0.5, -0.5, -0.5, -0.5), *you* (-0.5, +0.5, -0.5, -0.5), *she* (-0.5, -0.5, +0.5, -0.5) and *he* (-0.5, -0.5, -0.5, +0.5). An output variable stipulates that *me* should be used when speaker and referent agree, *you* should be used when addressee and referent agree, and *he* or *she* should be used when the referent is neither the speaker nor the addressee. For example, when the father (-2) is speaking to the child (0) about the mother (+2), the output would be the pronoun *she* (-0.5, -0.5, +0.5, -0.5). Networks have to learn on which of the three input variables the values agree or do not agree. The function is simple but the type of function the networks learn depends on the type of input patterns networks receive according to Oshima-Takane's (1985) pronoun learning model.

To compare the effect of input on pronoun learning, networks were trained under three different conditions: pure addressee, pure non-addressee and mixed conditions. There were 20 networks for each condition. In Phases I and II, networks in the pure addressee group received training inputs in which the child is always the addressee. This simulates a hypothetical environment where the child only hears child-directed speech. Networks in the pure non-addressee group, on the other hand, were trained with patterns where the child is neither the speaker nor the addressee. This exposed the networks only to overheard speech. The mixed networks used equal numbers of addressee and non-addressee training patterns to simulate a more realistic learning situation. Since there were 20 possible addressee patterns and 60 possible non-addressee patterns, the number of training patterns was equalized across all three conditions by using repeated patterns. The previous studies show that repeated patterns did not affect learning time (e.g., Oshima-Takane et al., 1998). Addressee patterns were

[1] The term 'kind' used in this paper stands for a natural kind, a set of entities possessing properties bound by natural law. The kind PERSON refers to the set consisting of individual persons as its members.

repeated six times per epoch for addressee networks while non-addressee patterns were repeated twice per epoch for non-addressee networks. Networks in the mixed condition were trained with three repeats of addressee patterns and one repeat of non-addressee patterns in order to get equal exposure to both types of other-speaking patterns. The twenty child-speaking patterns were added for all networks in Phase III.

4. Results

Pronoun learning time is measured as the number of epochs required to reach success on all training patterns. An epoch is defined as one sweep through the training set. Table 1 summarizes the mean epochs, standard deviation, and range in the three different learning conditions. In Phase I, addressee networks needed significantly fewer epochs to train than non-addressee and mixed networks combined, $t(57)=-22.412$, one-tailed, $p<0.001$. However, there is no significant difference between training times for non-addressee and mixed networks. All networks recruited 1 hidden unit. In Phase II, addressee networks again took significantly fewer epochs to train than non-addressee and mixed networks combined, $t(57)=-9.218$, one-tailed, $p<0.001$. In addition, mixed networks took the most epochs to train, taking significantly greater time than non-addressee networks, $t(57)=3.905$, one-tailed, $p<0.001$. All networks recruited 1 CC network from Phase I. In addition, addressee, non-addressee, and mixed networks needed an average of 1.40 hidden units (range: 1-3), 2.40 hidden units (range 2-3) and 2.90 hidden units (range: 2-4) respectively to complete the task. Therefore, the number of epochs needed for training correlates positively with the number of hidden units recruited.

Unlike in Phases I and II, addressee networks needed significantly greater epochs to learn the child-speaking patterns in Phase III compared to non-addressee and mixed networks combined, $t(57) = 16.056$, one-tailed, $p<0.001$. The difference between non-addressee and mixed networks was not significant. Both non-addressee and mixed networks only needed to recruit 1 KBCC network from Phase II to master the child-speaking patterns. No hidden units were recruited by these networks, which confirmed that they have learned the correct functions by the end of Phase II. By contrast, addressee networks needed a substantial amount of training in Phase III as they attempted to unlearn the incorrect pronoun functions that were learned in previous phases. Each of the 20 addressee networks

Table 1. Mean epochs, mean hidden units, and the number of networks required in each phase by condition.

	Addressee *N* =20	Non-addressee *N*=20	Mixed *N* =20
Phase I			
Epoch	54.00 (4.77)	107.10 (11.13)	102.45 (7.67)
Hidden	1 (range:0)	1 (range:0)	1 (range:0)
Phase 2			
Epoch	196.85 (0.40)	308.20 (50.99)	380.30 (34.12)
Hidden	1.40 (range:1-3)	2.40 (range:2-3)	2.90 (range:2-4)
Network	1CC	1CC	1CC
Phase 3			
Epoch	306.30 (106.28)	37.25(3.65)	35.20 (2.86)
Hidden	1.20 (range:0-3)	0	0
Network	1CC, 1KBCC	1KBCC	1KBCC

Note. The numbers in parentheses indicate standard deviations of epochs.

recruited 1 KBCC network from Phase II , 1 CC network from Phase I and an average of 1.20 new hidden units (range: 0-3) in Phase III.

5. Network Analysis

Network analysis was performed to examine the function the networks have learned and their generalization ability. The four pronouns were graphically represented by two functions: r1 and r2. The r1 graph distinguishes between pronouns *me* and *you* whereas the r2 graph distinguishes between third person pronouns *he* and *she*.

5.1. Target function

The target function is the correct function connecting inputs to outputs that the network has to learn. In order to obtain the graphic representation of the 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 speaker, addressee, and referent respectively. The letter c represents some arbitrary large positive number (e.g., c=50000). Then, the calculated value within the parentheses undergoes sigmoid transformation $1/(1+e^{-x})$, where x is the value to be

transformed. Hence, y1 equals 1 when speaker and referent agree (i.e., *me* is produced) and equals 0 when they disagree. Similarly y2 equals 1 when addressee and referent agree (i.e., *you* is produced) and equal 0 when they disagree. We then define the following:

z1=y1-0.5
z2=y2-0.5
z3=y3(1-y)(1-y2)-0.5
z4=(1-y1)(1-y2)(1-y3)-0.5

The z1 function distinguishes *me* from *you* as it equals +0.5 when *me* is produced (y1=1) and -0.5 when all other pronouns are produced (y1=0). Similarly, z2, z3 and z4 functions serve to distinguish *you*, *she*, and *he* from all other pronouns respectively. Therefore, z1, z2, z3, and z4 are functions that produce +0.5 for their specific pronouns. Finally, we define the following to conserve space for graphic representation:

r1=0.5(z1-z2)
r2=0.5(z4-z3)

The r1 function combines the representations of *me* and *you* into one graph, where it equals +0.5 when *me* is produced and -0.5 when *you* is produced. At all other points where referent does not agree with speaker or addressee, 0 is the output as third person pronouns are produced. Similarly, the r2 function combines the representation of *she* and *he*, where it equals +0.5 when *he* is produced, and -0.5 when *she* is produced and 0 when *me* or *you* are produced. Figures 2 and 3 show the target function for the production of *me* and *you* (r1) and for the production of *he* and *she* (r2). Separate graphs were made by referent for the father (ref = -2), the child (ref = 0), and the mother (ref = +2) for r1 and r2. 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 the addressee are, ranging from -2 to +2. The vertical axis represents the output pronoun (prediction), ranging from -0.5 to +0.5.

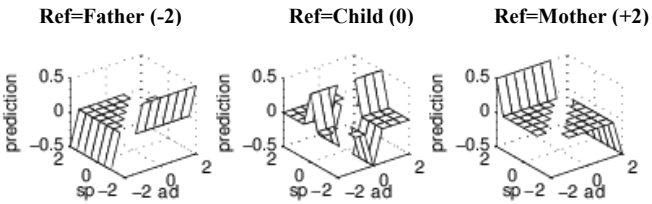

Figure2. Target function for correct production of *me* and *you* (r1). ref = referent, sp = speaker, ad = addressee, +0.5 = *me*, -0.5 = *you*, 0 = *he/she*.

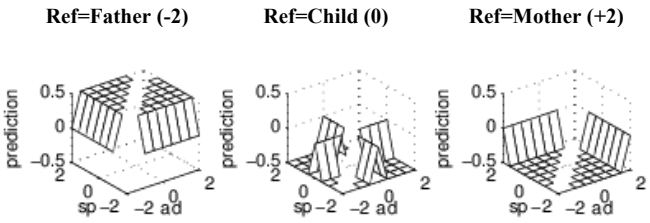

Figure 3. Target function for correct production of *he* and *she* (r2). ref = referent, sp = speaker, ad = addressee, +0.5 =*h*e, -0.5 = you, 0 = *me/you.*

5.2. Function approximation

Network analysis was performed to evaluate the networks' approximations of the function connecting inputs and outputs. The network's generalization test involved using learned training points to make interpolations to test points. A network's function approximation is indicative of how well it has learned the task using the training points. Therefore, networks that most resemble the target function represent those children who have learned the pronouns correctly.

5.2.1. First and second person pronouns, *me* and *you*

Figure 4 presents the function approximation of the production of first and second person pronouns *me* and *you* in three phases by an addressee network. As seen in Figure 4, the output value is -0.5 (*you*) when the referent is the child (ref=0) and +0.5 (*me*) when the referent is not the child (ref=-2 or ref=+2) in Phase I. Thus, addressee networks made *me-you* reversal errors initially: they always produced *you* to refer to the child and *me* to refer to everyone else. After adding third person pronoun training patterns in Phase II, networks learned to produce *me* whenever the speaker was either the mother or the father (ref=-2 or +2) but failed to produce *me* when the speaker was the child. They also failed to produce *you* when the addressee was a person other than the child and produced *he* or *she* instead as shown in the graphic representations of third person pronouns later in Figure 6. The exception to this pattern was when the child was the referent because the networks continued to misinterpret *you* as a proper name for the child in Phase II. Addressee networks finally corrected most errors in Phase III since the function approximation of graphs for the father (ref=-2) and the mother (ref=+2) resembled the target graph. The degree of correct generalization to untrained patterns was not as good as that of the non-addressee and mixed networks and varied across different addressee networks, especially when the child was the referent (ref = 0).

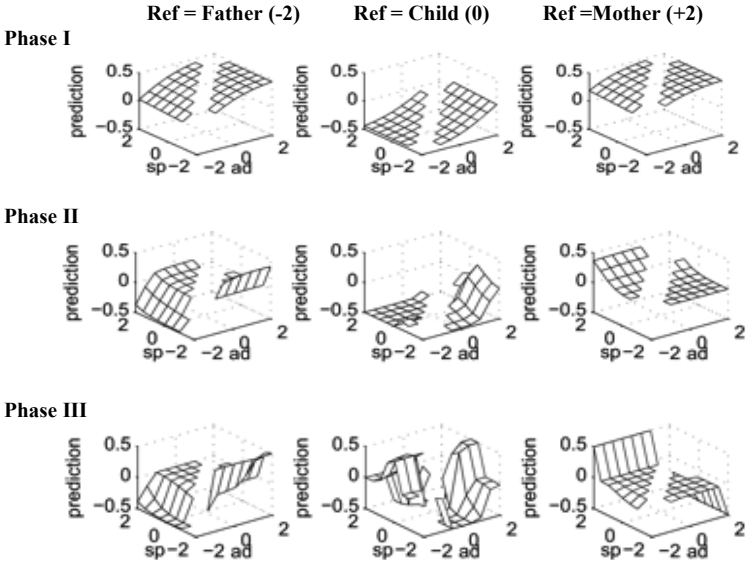

Figure 4. Function approximation for *me* and *you* in each phase by one pure addressee network.

Figure 5 presents the function approximation of a non-addressee network for *me* and *you*. In Phase I, the networks learned partially correct functions. Some networks such as the one shown in Figure 5 learned to produce *me* correctly but over-generalized the pronoun *you*. Other networks learned to produce *you* correctly but over-generalized the pronoun *me*. The difference in over-generalization error was random since each network was unique (i.e., the input and bias unit are initially connected to the output units with random weights). The over-generalization errors were all corrected in Phase II when networks learned to use *he* or *she* in reference to persons other than the speaker or the addressee. Like the function approximations in Phase II, the graphs from Phase III looked almost identical to the target functions. Thus, only slight weight adjustments were needed in Phase III since the networks had already learned the correct functions in Phase II. Function approximations by mixed networks are not presented here because it is very similar to networks trained with pure non-addressee patterns.

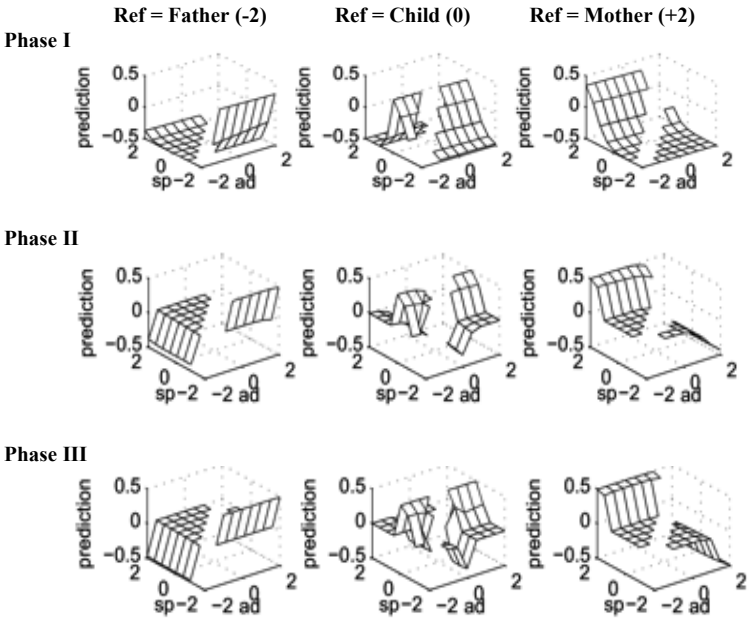

Figure 5. Function approximation for *me* and *you* in each phase by one pure non-addressee network.

5.2. 2. Third person pronouns, *he* and *she*

Figure 6 presents the function approximation for the production of third person pronouns *he* and *she* in three phases by one addressee network. The graphic surface after Phase I training is flat at the 0 level for all referents because no training patterns with third person pronouns were given to the networks until Phase II training. The graphic surface for father (-2) and for mother (+2) after Phase II training indicates that addressee networks learned to use third person pronouns in reference to non-speakers whether they were the non-addressee or the addressee. They made correct gender distinctions *he* and *she* according to the gender of the non-speakers. They also learned to use *me* in reference to speakers other than the child. However, addressee networks failed to learn third person pronouns referring to the child. The function approximation of graphs for the father and the mother resembled the target graph after Phase III training, indicating that addressee networks finally corrected overgeneralization errors and used third person pronouns to refer to the father and the mother as non-addressee. This was because addressee networks corrected me-you reversal errors by learning the child-speaking patterns in Phase III training. By learning to use *you* in reference

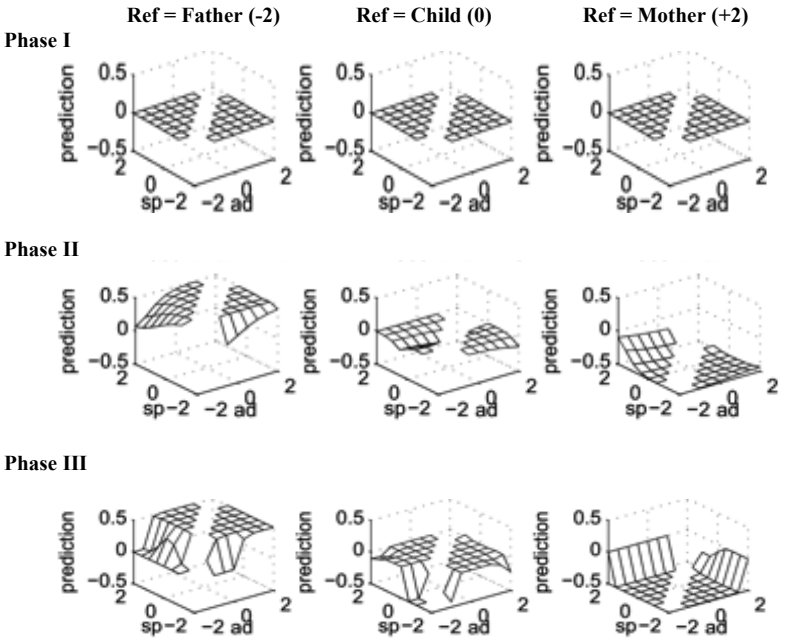

Figure 6. Function approximation for the production of third person pronouns *he* and *she* by one addressee network

to the addressee, they corrected third person pronoun overgeneralization errors to addressees. However, all addressee networks still failed to learn third person pronouns referring to the child. This is because the gender of the child was not explicitly coded and addressee networks did not have the opportunity to hear third person pronouns referring to the child

Figure 7 presents the function approximation by a non-addressee network. The graphic representations after Phase II training are similar to the target function, indicating that non-addressee networks learned to use *he* or *she* in reference to someone other than the speaker or the addressee. Like the function approximations in Phase II, the graphs from Phase III looked almost identical to the target functions. Thus, only slight weight adjustments were required in Phase III.

6. Discussion

Addressee networks completed the bulk of their learning in Phase III since they could not learn the correct semantic rules during the previous two phases. These networks made *me-you* reversal errors in Phase I and kept failing to produce *you* when referring to someone other than the child

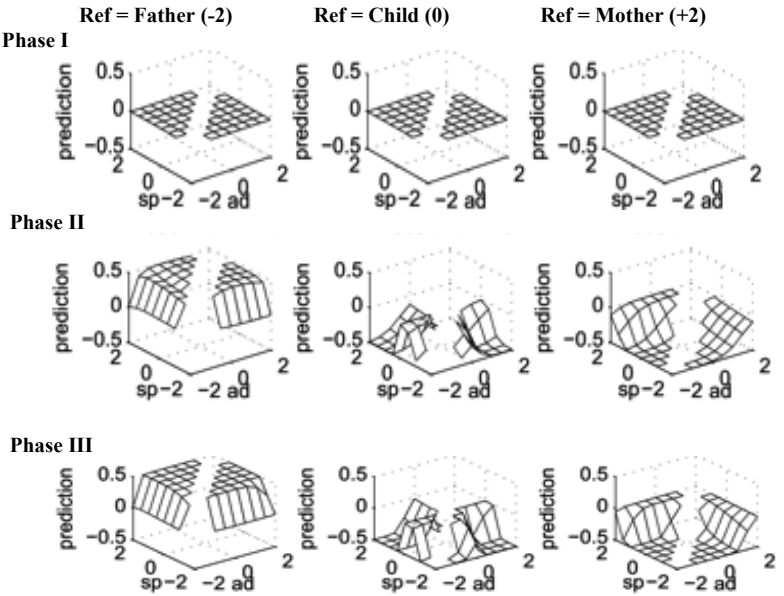

Figure 7. Function approximation for the production of third person pronouns *he* and *she* by one non-addressee network.

in Phase II. These errors suggest that children who are not exposed to overheard speech would make a persistent proper name interpretation for *you* because they did not have an opportunity to hear the shifting reference of *you* in child-directed speech. Extensive efforts were needed in the last phase to correct errors made in Phase I and II. Not only did these networks use many epochs in Phase III, they still had to learn new knowledge by recruiting several single hidden units.

On the other hand, non-addressee and mixed networks learned the correct semantic functions by Phase II. Network and epoch analysis reveal that these networks did most of their learning in Phase I and Phase II. Since the graphs of the Phase II function approximations were almost identical to the graphs of the target function, the networks only needed prior knowledge (1 KBCC net) to master the child-speaking patterns in Phase III. Overall, non-addressee and mixed networks were better at generalization to untrained patterns than addressee networks after all the training was completed. These results support Oshima-Takane's pronoun learning model that children with exposure to overheard speech learn the deictic meaning of all three personal pronouns with few production errors by listening to others using these pronouns. The non-addressee and the mixed network analysis

also confirm the results of the previous simulation studies with CC networks that knowledge of third person pronouns would help correct any over-generation errors made in Phase I (Oshima-Takane et al., 1998).

Previous results from KBCC simulation studies (Shultz & Rivest, 2001) have shown that the KBCC neural networks would recruit relevant source nets to help speed learning. Therefore, analysis of the type of units recruited can give clues as to the type of prior knowledge needed in each stage of pronoun learning, information that previous simulation studies with CC networks could not provide (Oshima-Takane et al., 1998). The finding that all three types of networks recruited 1 CC net in Phase II suggests that children use prior knowledge of first and second person pronouns in order to learn third person pronouns. Prior knowledge about the differences between speech roles (i.e., the speaker, the addressee and the non-addressee), and between individual persons as members of the same kind PERSON are just some information that could help children learn the semantic rules of personal pronouns. In addition, all networks recruited at least 1 KBCC net in Phase III. This indicates that KBCC networks attempt to apply the knowledge they have learned from listening to other people when learning the child-speaking patterns even though they are not forced to do so as they were in previous CC simulation studies (Oshima-Takane et al., 1998). Furthermore, the type of prior knowledge the KBCC networks recruit depends on what they have learned previously and what they need to learn further. The non-addressee and the mixed networks needed to recruit 1 KBCC network from Phase II in Phase III whereas most of the 20 addressee networks needed an additional CC network from Phase I and a few single hidden units to complete Phase III training. This finding suggests that children who are exposed mostly to child-directed speech would need new information to unlearn their incorrect pronoun productions and would not benefit as much from their prior knowledge.

7. References

Brener, R. (1983). Learning the deictic meaning of third person pronouns. *Journal of Psycholinguistic Research*, *12,* 235- 261.

Chiat, S. (1982). If I were you and you were me: the analysis of pronouns in a pronoun-reversing child. *Journal of Child Language, 9*, 359-379.

Clark, E.V. (1978). From gesture to word: on the natural acquisition. In J.S. Brunner & A. Garton (Eds.). *Human growth and development*: *Wolfson College Lectures*. Oxford: Oxford University Press.

Evans, K.E. & Demuth, K. (2012). Individual differences in pronoun reversal: Evidence from two longitudinal case studies. *Journal of Child Language, 39*, 162-191.

Guerriero, A.M.S. (1998). *Acquisition of deictic feminine third person pronouns*. Master thesis, McGill University.

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

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

Morgenstein, A. (2011). The self as other: self words and pronominal reversals in language acquisition. In Lorda & Zabalbeascoa (Eds.). *Spaces of Polyphony*. John Benjamin Publishing Company .

Oshima-Takane, Y. (1985). *Learning of pronouns*. PhD thesis, McGill University.

Oshima-Takane, Y. (1988). Children learn from speech not addressed to them: A case study. *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. (1999). The learning of first and second person pronouns in English. In R. Jackendoff, P.Bloom, & K.Wynn (Eds). *Language, Logic, and Concept: Essays in Memory of John Macnamara*. Cambridge, MA: MIT press.

Oshima-Takane, Y. (2014). Acquisition of personal pronouns. In P. Brooks, V. Kemp, & J.G. Golson (Eds.) *Encyclopedia of Language Development* (pp. 500-501). Los Angeles: Sage Publications.

Oshima-Takane, Y. Cole, E., & Yaremko, R. (1993). Semantic pronominal confusion in a hearing-impaired child: A case study. *First Language, 13*, 149-168.

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., Takane, Y., & Shultz, TR. (1999). The learning of first and second person pronouns in English: network models and analysis. *Journal of Child Language, 26*, 545-573.

Oshima-Takane, Y., Takane, M., & Takane, Y. (1998). Learning of first, second, and third person pronouns in English. *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (PP.800-805). Hillsdale, N.J.: Lawrence Erlbaum.

Schiff-Myers, M. (1983). From pronoun reversals to correct pronoun usage: a longitudinal study of a normally-developing child. *Journal of Speech and Hearing Disorders*, 48, 385-394.

Shultz, T. R., & Rivest, F. (2001). Knowledge-based cascade-correlation: Using knowledge to speed learning. *Connection Science, 13,* 43-72.

Wells, G. (1985). *Language Development in the pre-school years*. Language at home and at school, Vol. 2. Cambridge, UK: Cambridge University Press.

Acknowledgement

This research was supported by a grant from Natural Sciences and Engineering Research Council of Canada. The presentation of the paper at JSLS 2013 international meeting was supported by a paper presentation grant from McGill University.