In S. J. Hanson, T. Petsche, M. Kearns, & R. L. Rivest (Eds.) (1994). *Computational learning theory and natural learning systems, Vol. 2: Intersection between theory and experiment* (pp. 347-362). Cambridge, MA: MIT Press.

A Connectionist Model of the Learning of Personal Pronouns in English

Thomas R. Shultz, David Buckingham, and Yuriko Oshima-Takane
Department of Psychology
McGill University

Both experimental and observational psycholinguistic research have shown that children's acquisition of first and second person pronouns is affected by the opportunity to hear these pronouns used in speech not addressed to them. These effects were simulated with the cascade-correlation connectionist algorithm. The networks learned, in effect, to produce the pronouns *me* and *you* depending on identification of the speaker, addressee, and referent. Analysis of network performance and structure indicated that generalization to correct pronoun production was aided by listening to non-addressed speech and that persistent pronoun errors were created by listening to directly addressed speech. It was noted that explicit symbolic rule models would likely have difficulty simulating the pattern frequency effects common to the present simulations and to the natural language environment of the child.

20.1 Introduction

Learning the semantics of personal pronouns such as *me* and *you* poses interesting problems for young children. The problems stem initially from the fact that the referent of these pronouns is not fixed, but rather shifts with conversational role. For example, the child's mother calls herself *me* and the child *you*; but these terms need to be reversed when the child uses them. Not only does the referent of personal pronouns shift with conversational role, but the model for correct use of these pronouns is not ordinarily provided in speech that is addressed to children. As in the above example, the child must learn to reverse these pronouns to use them correctly. If children simply imitated what they heard in speech addressed to them, they would incorrectly refer to themselves as *you* and to the mother as *me*.

Yet children typically master the correct use of personal pronouns in English by about 3 years of age (Clark, 1978). Furthermore, the majority of children do so with few reversal errors of the type cited above (Charney, 1980; Chiat, 1981). Some children do make such errors, however, and these errors often persist for months (Clark, 1978; Oshima-Takane, 1992; Schiff-Meyers, 1983). Researchers occasionally refer to these errors as non-reversal errors because children fail to reverse pronouns they hear (Oshima-Takane, 1988; Tanz, 1980). In this paper, we call them reversal errors in order to be consistent with the idea that the child has reversed the correct use of the two pronouns. The semantic rules underlying correct use of first and second person pronouns have been analyzed by philosophers of language who have emphasized the importance of context, speech roles, and discourse situations (Barwise & Perry, 1983; Kaplan, 1977). Essentially, the correct semantic rules specify that a first person pronoun refers to the person who uses it and that a second person pronoun refers to the person who is addressed when it is used.

Whereas many researchers and theorists (Ervin-Tripp, 1971; de Paulo & Bonvillian, 1978) assume that speech not addressed to children is unimportant in language acquisition, Oshima-Takane (1988) has hypothesized that children may be able to learn the semantic rules for pronouns primarily from speech not addressed to them, that is, from overheard speech. In overheard speech, children often observe that second person pronouns refer to a person other than themselves and that first and second person pronouns reciprocate each other. In speech addressed to children, on the other hand, children simply observe that second person pronouns always refer to themselves and that first person pronouns refer to the person who is speaking to them. Thus, the relationship between pronouns and speech roles should be better understood when children hear other people talking to each other.

Oshima-Takane's theoretical analysis of pronoun acquisition stands in contrast to the analyses of previous theorists who had focused exclusively on children's errors (Bartak & Rutter, 1974; Fay & Shuler, 1980; Simon, 1975) or on the absence of children's errors (Charney, 1980; Shipley & Shipley, 1969). Those theorists who had focused only on errors failed to explain the early errorless acquisition of pronouns characteristic of most children, and those who had focused only on correct pronoun use failed to explain the persistent reversal errors found in some children. Oshima-Takane's analysis was the first to explain the variation in pronoun acquisition across different types of children.

In support of her theoretical analysis, Oshima-Takane (1988) conducted a training experiment with 19-month-olds who were about to learn personal pronouns and found that children profited more in pronoun production from overheard speech than from speech directly addressed to them. In fact, only those children who had opportunities to hear pronouns in overheard speech could produce pronouns without errors. Extending this argument to a naturalistic setting, she predicted and found that second-borns acquired these pronouns earlier than first-borns, even though these children did not differ on other language measures such as mean length of utterance (Oshima-Takane & Derevensky, 1990). Presumably, second-born children have relatively more opportunities to hear pronouns used in speech that is not addressed to them, that is, in conversations between their parent and older sibling.

Extreme versions of Oshima-Takane's (1988) addressee and non-addressee training conditions are illustrated in the top row of figure 20.1. In this figure, *Fa* refers to the father, *Mo* to the mother, and *Ch* to the child. The arrows originate from the speaker. The original direction of the arrow indicates the addressee; the point of the arrow indicates the referent. In the addressee condition, each parent in turn addresses the child. The father points to the child and says *you*, and points to himself and says *me*. The mother does the analogous thing, saying *you* when pointing to the child and *me* when pointing to herself. In the non-addressee condition, the parents address each other and the child merely listens. Each parent points to him- or herself and says *me*, and points to the opposite parent and says *you*. The critical tests are those involving the child's productive use of these personal pronouns. As illustrated in the bottom of figure 1, what does the child say as he or she joins in the game, when pointing to a parent or to the self?

We report two connectionist simulations on learning of the personal pronouns *me* and *you* under addressee and non-addressee conditions. In the first experiment, we contrasted the pure versions of these two conditions. In the second, we created a series of conditions in which addressee and non-addressee patterns were mixed in varying, more realistic degrees. In both experiments, after the networks learned the parent-speaking patterns, we added test patterns in which the child spoke, either referring to himself or to a parent. The key question was how long it would take to learn in this second phase.

20.2 The Learning Algorithm and the Initial Network

We used a relatively new connectionist learning algorithm called cascade-correlation (Fahlman & Lebiere, 1990). Cascade-correlation builds its own network topology by recruiting new hidden units as it learns. Thus, it affords a more principled approach to network construction than, say, back-propagation and other algorithms that require a full specification of the network. Cascade-correlation begins with a minimal topology containing only the input and output units defined by the programmer. In what is called the output phase, connections to the output units are trained until error can no longer be reduced. Then, in the input phase, a pool of candidate hidden units receives trainable input from the input units and any existing hidden units. Outputs from the candidate hidden units are not connected to the output units in the input phase. The purpose of the input phase is to identify the candidate unit whose activations correlate best with the output errors. This best candidate unit is then installed in the network, receiving input from all input units and any hidden units already in place and sending output to all of the output units. Once installed, the input side weights to the new hidden unit are frozen, and its output side weights are allowed to change with learning in the next output phase. Because cascade-correlation uses second order error minimization techniques in computing weight changes and learns only one level at a time, it is typically 10-50 times faster at learning than is back-propagation. We used default parameter settings for all of the present simulations and always trained to victory. Cascade-correlation declares victory and stops learning when all of the outputs are within score-threshold of their targets on all of the training patterns. Score-threshold is a parameter that reflects the allowable difference between output and target activations. Normally, learning continues until all output activations are within score-threshold of their targets for all training patterns.

The cascade-correlation algorithm has been successfully applied to a variety of problems in cognitive development, including balance scale phenomena (Shultz & Schmidt, 1991) and certain aspects of causal reasoning (Shultz, Zelazo, & Strigler, 1991). As far as we know, cascade-correlation has not yet been applied to psycholinguistic phenomena, but a number of other connectionist algorithms have successfully captured selected aspects of language development, including acquisition of the German article system (MacWhinney, Leinbach, Taraban, & MacDonald, 1989), U-shaped development in past-tense English morphology (Plunkett & Marchman, 1991), semantic over- and under-extension (Chauvin, 1989), the mutual exclusivity constraint in word learning (Schyns, 1991), and the importance of starting small in syntactic acquisition (Elman, 1993). Theoretical relations between connectionism and cognitive development have been discussed by Plunkett and Sinha (1992) and by Shultz (1991).

The initial cascade-correlation network used in our pronoun simulations employed distributed representations on both input and output units, as portrayed in figure 20.2. There were three pairs of input units, two for the speaker, two for the addressee, and two for the referent. The inputs were coded as 01 for the mother, 10 for the father, and 11 for the child. The two output units were coded as +- for *me*, and as -+ for *you*. All of the input units were connected to all of the output units. In addition, there was a bias unit, which was always on, connected to all of the output units. The arrow in figure 20.2 indicates full connectivity between input and output layers.

20.3 Experiment 1: Pure Conditions

The phase 1 training patterns in the addressee condition consisted of the four utterances portrayed in the top left of figure 20.1. They are listed more formally in the top half of table 20.1. The training patterns for the non-addressee condition consisted of the four utterances in the top right of figure 20.1; they are listed in the bottom half of table 20.1. There were 20 runs in each condition. In the first phase, each network learned the initial four training patterns associated with the appropriate condition. Then, in the second phase, the four test patterns illustrated in the bottom of figure 20.1 were added to the training sets of both conditions and learning resumed. Thus, in phase 2, there were eight training patterns in each condition. The new patterns, along with their target pronouns, are listed in table 20.2.

The mean epochs to learn in phase 1 were 13.80 in the addressee condition and 47.25 in the non-addressee condition, $F(1, 38) = 2609.28$, $p < .001$.[1] No hidden units were recruited by any networks in the addressee condition, but each network in the non-addressee condition recruited one hidden unit. The recruited hidden unit is presumably necessary to encode the shifting pronominal reference in the non-addressee condition. In the second phase, as predicted, it took more epochs to learn in the addressee (M = 110.90) than in the non-addressee (M = 11.55) condition, $F(1, 38) = 2512.65$, $p < .001$. For the second phase, one hidden unit was recruited by each network in the addressee condition, but no hidden units were recruited in the non-addressee condition.

The configuration of a network after recruiting a hidden unit is illustrated in figure 20.3. The arrows in figure 3 indicate full connectivity between the input and output layers, between the input and hidden layers, and between the hidden and output layers. Once a hidden unit has been installed, weights coming into it are frozen. These weights have already been adjusted before installation so that the output of the hidden unit correlates highly with the network's existing error. Weights coming into the output units continue to be adjusted with further learning.

Plots of the output activation values for individual networks over epochs in phase 2 can reveal what is happening in the networks. Figure 20.4 shows such a plot for network 13, simulating a child speaking to the father and pointing to herself. In these graphs, triangles refer to the activation level on the output unit that should be positive when the target is *me* and negative when the target is *you*. Circles refer to the activation level on the output unit that should be positive when the target is *you* and negative when the target is *me*. With the score-threshold parameter set to the default value of 0.4, the two output units need to be on opposite sides of the dashed lines, drawn at ordinate values of +0.1 and -0.1, in order to be interpreted as uttering a particular pronoun. By epoch 2, network 13 is correctly saying *me* when pointing to itself, indicating a very good generalization from the non-addressee training. Not all of the non-addressee networks did this well, but errorless generalization to previously unseen problems did occasionally occur. More typically, a few more epochs of phase 2 learning were required to produce correct performance.

Networks in the addressee condition often showed persistent reversal errors coupled with confusion (i.e., producing neither pronoun) before finally getting the pronouns correct. An example is provided in figure 20.5 from a network (number 9) simulating speaking to the mother. This network started out saying *you* when pointing to itself, persisted in this error for around 15 epochs, then went into a long period of not clearly producing either pronoun, before finally sorting it out at around 53 epochs. Children often avoid pronoun errors by using proper names (Oshima-Takane & Oram, 1991); one can perhaps imagine a network doing that when it fails to produce a clear pronoun. This could be explored in future simulations in which proper names are included in the training patterns.

We chose a few networks haphazardly and drew Hinton diagrams of their network structure at the completion of phases 1 and 2. Each Hinton diagram shows the size and sign of incoming weights. Size of weight is indicated by the size of the corresponding square; white squares indicate positive weights and black squares indicate negative weights. Hinton diagrams for two representative networks are presented in figure 20.6. For network 10 in the addressee condition, there were enormous changes in both the sign and size of the weights from the end of phase 1 to the end of phase 2. In contrast, for network 13 in the non-addressee condition, the weight adjustments from the end of phase 1 to the end of phase 2 were fairly minor.

To examine this finding more systematically, we computed the mean absolute weight changes from the end of phase 1 to the end of phase 2. These differences were obtained separately for each of the two output units and averaged across seven output connections in the addressee condition and eight output connections in the non-addressee condition. This difference in number of output connections is due to the fact that a hidden unit was recruited during phase 1 in the non-addressee condition but not in the addressee condition. The mean absolute weight changes were subjected to a two-way analysis of variance in which condition served as a between network factor and output unit served as a repeated measures factor. This analysis yielded only a main effect of condition, $F(1, 38) = 8.26$, $p < .01$. There was more weight change in the addressee (M = 1.20) than in the non-addressee (M = 0.86) condition. This is quite a conservative method of assessing the amount of required change since it ignores the fact that an additional hidden unit always gets recruited during phase 2 in addressee networks, but not in non-addressee networks. The addressee networks must, of course, also learn the connection weights entering and exiting this new hidden unit during phase 2.

---

[1]All of the statistical tests reported in this paper are analyses of variance. One of several good background sources for this technique is Winer (1971).

20.4 Experiment 2: Mixed Conditions

The pure addressee and non-addressee conditions in Experiment 1 are interesting for theoretical reasons, but are too extreme to simulate the child's natural language learning environment, which undoubtedly involves some mixture of addressee and non-addressee material. Variation in such mixtures was achieved by including the four addressee and four non-addressee patterns from table 20.1 in various frequency multiples. There were five conditions, with the frequency multiples of addressee:non-addressee of 9 : 1, 7 : 3, 5 : 5, 3 : 7, and 1 : 9. For example, in the 7 : 3 condition, the four addressee patterns occurred seven times and the four non-addressee patterns occurred three times. Thus, there were 40 phase 1 training patterns in each condition.

The 9 : 1 condition could be regarded as roughly equivalent to the speech heard by a first-born child with no siblings; mostly addressee while one parent is away at work plus a bit of non-addressee in the evening when the working parent returns. The 5 : 5 condition could be regarded as roughly equivalent to the speech heard by a second-born child; addressee and non-addressee speech all day in about equal measure. Once the 40 patterns in phase 1 had been learned, the 4 child-speaking patterns from table 20.2 were added to the training patterns and phase 2 learning resumed until completion.

The mean epochs to learn in phase 1 are presented in figure 20.7. A one-way analysis of variance revealed a main effect of condition, $F(4, 95) = 149.01$, $p < .001$, with a strong negative linear trend, $F(1, 95) = 529.44$, $p < .001$. This negative linear trend reflects the fact that, with less non-addressee material, it takes longer to discover the necessity of adding a hidden unit. Note that this is opposite to the phase 1 trend in Experiment 1, where it was not necessary to add a hidden unit in the addressee condition in the total absence of non-addressee patterns. In the present mixed experiment, there was one hidden unit recruited in each run during phase 1.

The mean epochs to learn in phase 2 are presented in figure 20.8. A one-way analysis of variance yielded a main effect of condition, $F(4, 95) = 12.44$, $p < .001$, with a strong negative linear trend, $F(1, 95) = 41.83$, $p < .01$. As in Experiment 1, this negative linear trend reflected the fact that more non-addressee material in phase 1 enabled better generalization to the child-speaking patterns in phase 2. One network in the 3 : 7 condition recruited no hidden units in phase 2; otherwise each network in all conditions of phase 2 recruited one hidden unit.

The network configuration after recruiting a second hidden unit is portrayed in figure 20.9. The thick arrows represent full connectivity between network layers; the thin arrow indicates a single connection between the two hidden units.

20.5 Discussion

Comparison of addressee vs. non-addressee conditions in both pure and mixed form qualitatively simulated the results of both experimental and naturalistic studies with children. In general, the results were consistent with Oshima-Takane's (1988) hypothesis that children learn correct usage of personal pronouns from speech not addressed to them, whereas they learn incorrect use from speech directly addressed to them. This interpretation of the condition effects was supported by analyses of output activations and weight changes. Output activation diagrams indicated fairly rapid acquisition of pronouns with more non-addressee training and persistent pronoun errors, including reversals, with more addressee training. Analyses of weight changes between phases revealed that, with greater non-addressee material in the training patterns, the network weights after phase 1 training were closer to those required for mastery of the phase 2 child-speaking patterns. Thus, these initial models are capable of simulating some of the wide variation that exists in the pronoun errors children make during acquisition and relating this variation to the same factors, i.e., addressee vs. non-addressee speech, that affect such variation in children. The simulations are still incomplete since they do not yet contain proper names nor any trends for earlier acquisition of first person than second person pronouns (Chiat, 1981; Clark, 1978; Oshima-Takane & Oram, 1991). Because the present simulations focus on situations in which the referent is clearly specified, in this case by pointing, they also neglect the learning of referents in other situations. Planning for experiments on these open issues is currently underway. Even with such additions, these sorts of models will still be greatly simplified compared to the child's actual social environment and computational resources.

It is not yet clear whether our phase 2 learning is a realistic simulation of the child's natural language environment or whether it is merely a diagnostic device to assess the effectiveness of phase 1 learning. It is clear that direct feedback on correct pronoun use is often difficult for the child to understand. Oshima-Takane (1992) has reported that a boy who was making reversal errors completely misunderstood parents' corrections and did not correct his errors for a long period of time. It is possible, however, that other sorts of feedback, such as parental comprehension failures, are effective in correcting the child's pronoun errors. Cascade-correlation networks can also be temporarily quite resistant to error feedback, at least in terms of classifying output activations, so temporary resistance in children should not at all disqualify network models.

Discussing cognitive development in general, Siegler concluded " . . . existing knowledge facilitates new acquisitions when, and only when, directly relevant learning experiences arise. Thus, transfer of existing knowledge may be more apparent in savings of time and effort in future learning, or in the range of experiences that produce learning, than in direct extension of the knowledge to new domains without intervening experience in those domains" (Siegler, 1991, p. 354). If this were true of pronoun acquisition, the necessity of phase 2 learning in our simulations would be justified. However much phase 2 learning is required in future models, the main thrust of current psychological and simulation results is that the principal determinant of correct personal pronoun use is the opportunity to hear these pronouns in speech not addressed to the child.

The prevailing view of language acquisition has focused on speech addressed to children as the primary linguistic input, and correspondingly little theoretical attention has been paid to overheard speech. Many researchers believe that young children do not attend to most of utterances they overhear because such utterances are insensitive to the limited linguistic abilities of the children. For example, overheard utterances are generally longer and more complex than those used in speech addressed to children. But it is clear from the present simulations as well as from earlier psychological studies that hearing pronouns in speech not addressed to the child is essential for producing correct pronouns. These results suggest that overheard speech is an important source of input for early language development.

Although no other detailed models of pronoun acquisition exist, perhaps the most plausible alternative to the present model would be an explicit symbolic rule model, such as could be developed in a self-modifying production system like Soar (Newell, 1990). We have little doubt that Soar could be made to learn these pronoun rules, but we wonder whether it would be sufficiently sensitive to the frequency effects we observed in the mixed patterns experiment and which are presumably responsible for the experimental and observational results with human children. Soar would likely learn the correct rules immediately on presentation of a minimal number of patterns, ignoring the effects of frequencies of the patterns. Production system interpreters that are quantitatively sensitive to frequency effects are certainly possible, but to our knowledge no one has shown how such characteristics can be smoothly integrated with other rule-learning constraints.

The present results are relevant to issues of constructive induction and the effect of prior knowledge on learnability. The cascade-correlation learning algorithm spontaneously constructs representations of shifting pronominal reference by the dual process of weight adjustment and recruitment of hidden units that correlate well with existing error. The learnability of correct pronoun production is shown to depend on previous learning of pronoun patterns in non-addressee speech. Coupled with the connectionist models reviewed earlier, the present simulation results suggest that human psychological development can be modeled and elucidated by connectionist techniques.

Acknowledgments

This research was supported by a grant to the first author from the Natural Sciences and Engineering Research Council of Canada and to the third author from the Social Sciences and Humanities Research Council of Canada. We are grateful to Scott Fahlman for providing code for the cascade-correlation algorithm and for comments on a previous draft, and to Chris Schunn for providing code for automatic drawing of Hinton diagrams.

References

Bartak, L., & Rutter, M. 1974. Use of personal pronouns by autistic children. *Journal of Autism and Childhood Schizophrenia* 4:217-222.

Barwise, J. & Perry, J. 1983. *Situations and Attitudes*. Cambridge, MA: MIT Press.

Charney, R. 1980. Speech roles and the development of personal pronouns. *Journal of Child Language* 7:509-528.

Chauvin, Y. 1989. Toward a connectionist model of symbolic emergence. *Proceedings of the Eleventh Annual conference of the Cognitive Science Society* :580-587. Hillsdale, NJ: Lawrence Erlbaum.

Chiat, S. 1981. Context-specificity and generalization in the acquisition of pronominal distinctions. *Journal of Child Language* 8:75-91.

Clark, E. V. 1978. From gesture to word: On the natural history of deixis in language acquisition. In *Human Growth and Development*, edited by J. S. Bruner and A. Garton. Oxford: Oxford University Press.

de Paulo, B. M. & Bonvillian, J. D. 1978. The effect on language development of the special characteristics of speech addressed to children. *Journal of Psycholinguistic Research* 7:189-211.

Elman, J. L. 1991. Incremental learning, or the importance of starting small. Technical Report 9101, Center for Research in Language, University of California, San Diego.

Ervin-Tripp, S. 1971. An overview of theories of grammatical development. In *The Ontogenesis of Grammar: Theoretical Symposium*, edited by D. I. Slobin. New York: Academic Press.

Fahlman, S. E. & Lebiere, C. 1990. The Cascade-Correlation learning architecture. Technical Report CMU-CS-90-100, School of Computer Science, Carnegie-Mellon University, Pittsburgh, PA.

Fay, W. H. & Shuler, A. L. 1980. *Emerging Language in Autistic Children*. Baltimore: University Park Press.

Kaplan, D. 1977. Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. Department of Philosophy, UCLA.

MacWhinney, B., Leinbach, J., Taraban, R., & MacDonald, J. 1989. Language learning: Cues or rules? *Journal of Memory and Language* 28:255-277.

Newell, A. 1990. *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Oshima-Takane, Y. 1988. Children learn from speech not addressed to them: The case of personal pronouns. *Journal of Child Language* 15:95-108.

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

Oshima-Takane, Y. & Derevensky, J. L. April, 1990. Do later-born children delay in early language development? Paper presented at the International Conference on Infant Studies, Montreal.

Oshima-Takane, Y. & Oram, J. July, 1991. Acquisition of personal pronouns: What do comprehension data tell us? Paper presented at the International Society for the Study of Behavioural Development, Minneapolis, MN.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition*, **38**, 43-102.

Plunkett, K., & Sinha, C. (1992). Connectionism and developmental theory. *British Journal of Developmental Psychology*, **10**, 209-254.

Schiff-Meyers, N. 1983. From pronoun reversals to correct pronoun usage: A case study of a normally developing child. *Journal of Speech and Hearing Disorders* 48:385-394.

Schyns, P. 1991. A modular neural network model of concept acquisition. *Cognitive Science* 15:461-508.

Shipley, E. F., & Shipley, T. E. 1969. Quaker children's use of *Thee*: A relational analysis. *Journal of Verbal Learning and Verbal Behaviour* 8:112-117.

Shultz, T. R. 1991. Simulating stages of human cognitive development with connectionist models. In *Machine learning: Proceedings of the Eighth International Workshop* :105-109, edited by L. Birnbaum and G. Collins. San Mateo, CA: Morgan Kaufman.

Shultz, T. R. & Schmidt, W. C. 1991. A Cascade-Correlation model of balance scale phenomena. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* :635-640. Hillsdale, NJ: Lawrence Erlbaum.

Shultz, T. R., Zelazo, P. R., & Strigler, D. April, 1991. Connectionist modeling of the development of the concepts of potency and resistance in causal prediction. Paper presented at the meeting of the Society for Research in Child Development, Seattle.

Siegler, R. S. 1991. *Children's Thinking*. Englewood Cliffs, NJ: Prentice-Hall.

Simon, N. 1975. Echolalic speech in childhood autism: Consideration of possible underlying loci of brain damage. *Archives of General Psychiatry* 32:1439-1446.

Tanz, C. 1980. *Studies in the Acquisition of Deictic Terms*. Cambridge: Cambridge University Press.

Winer, B. J. 1971. *Statistical Principles in Experimental Design*. New York: McGraw-Hill.

Table 20.1

Phase 1 Training Patterns

Condition	Speaker	Addressee	Referent	Pronoun
addressee	father	child	father	me
	father	child	child	you
	mother	child	mother	me
	mother	child	child	you
non-addressee	father	mother	father	me
	father	mother	mother	you
	mother	father	mother	me
	mother	father	father	you

Table 20.2

Training Patterns Added in Phase 2

Speaker	Addressee	Referent	Pronoun
child	father	child	me
child	father	father	you
child	mother	child	me
child	mother	mother	you

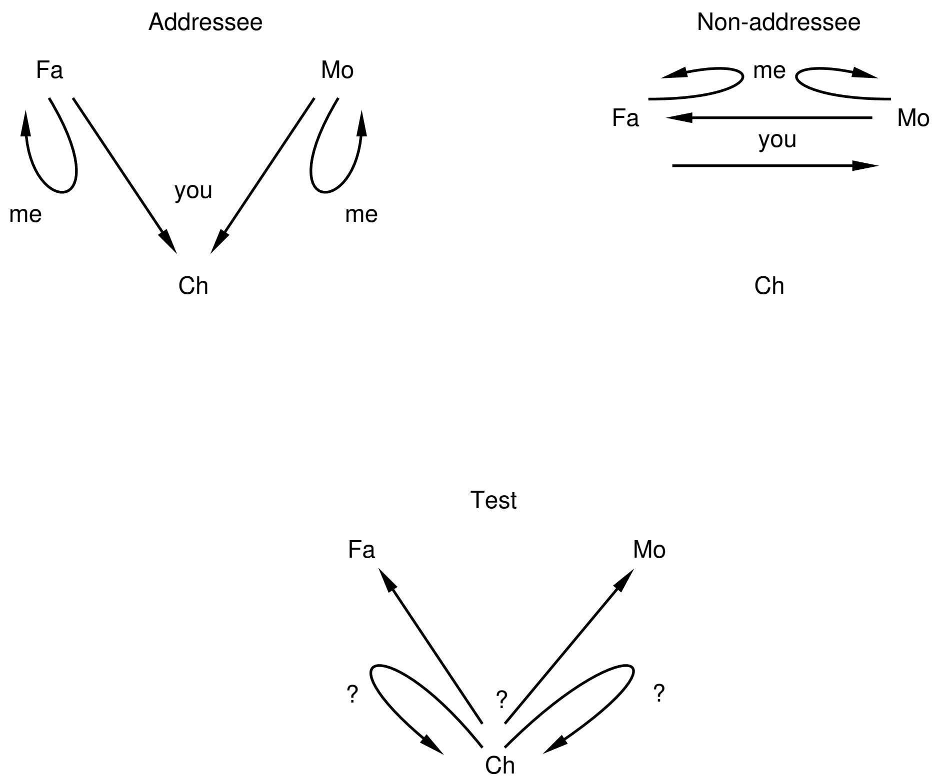

Figure 20.1. Schematic representation of training patterns.

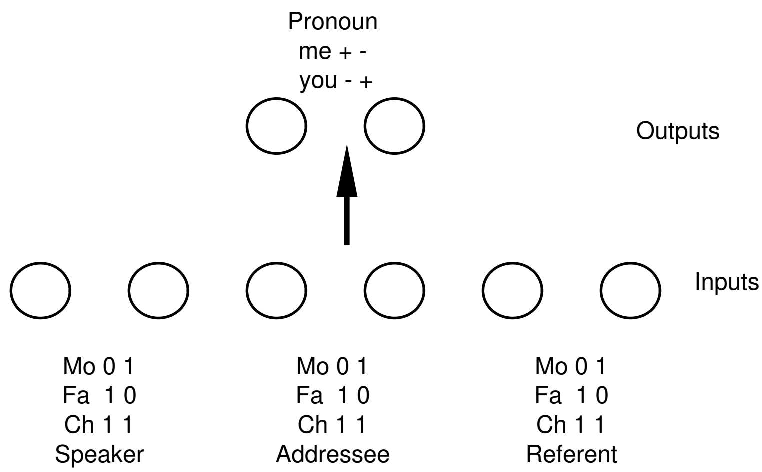

Figure 20.2. Initial pronoun network before recruitment of hidden units.

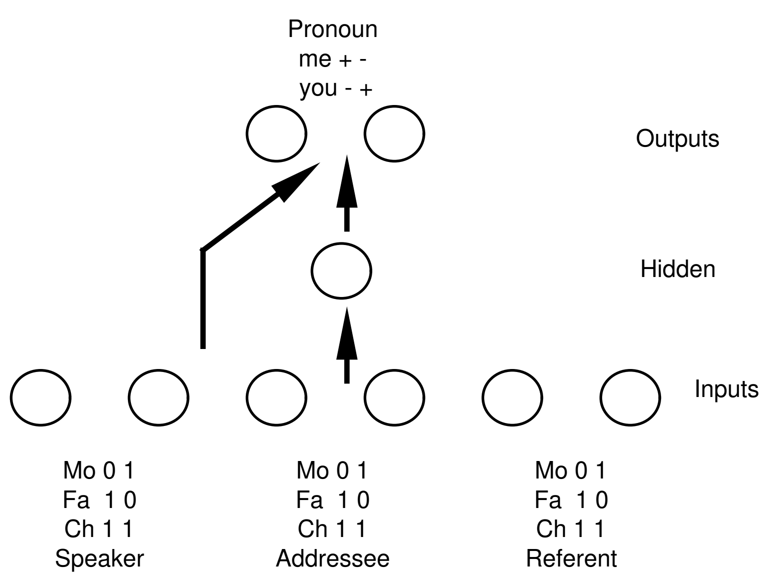

Figure 20.3. Pronoun network after recruitment of a single hidden unit.

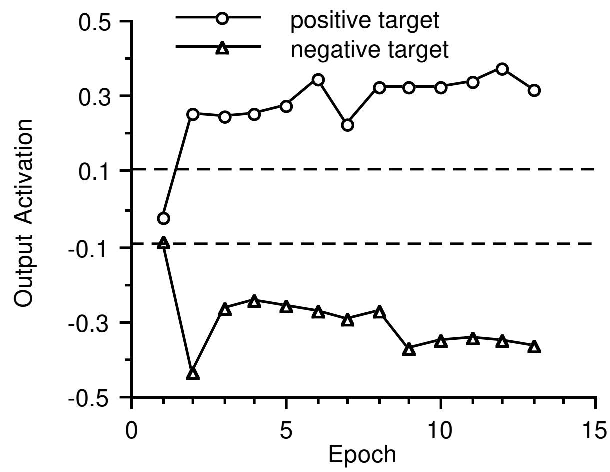

Figure 20.4. Activation diagram for Network 13, non-addressee condition.

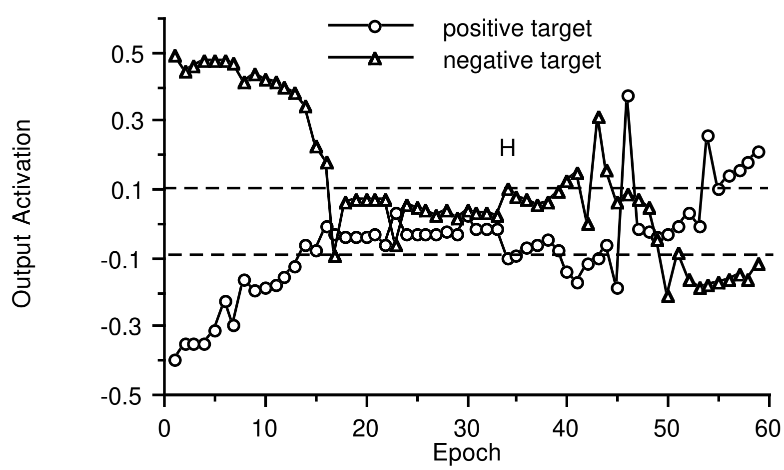

Figure 20.5. Activation diagram for Network 9, addressee condition.

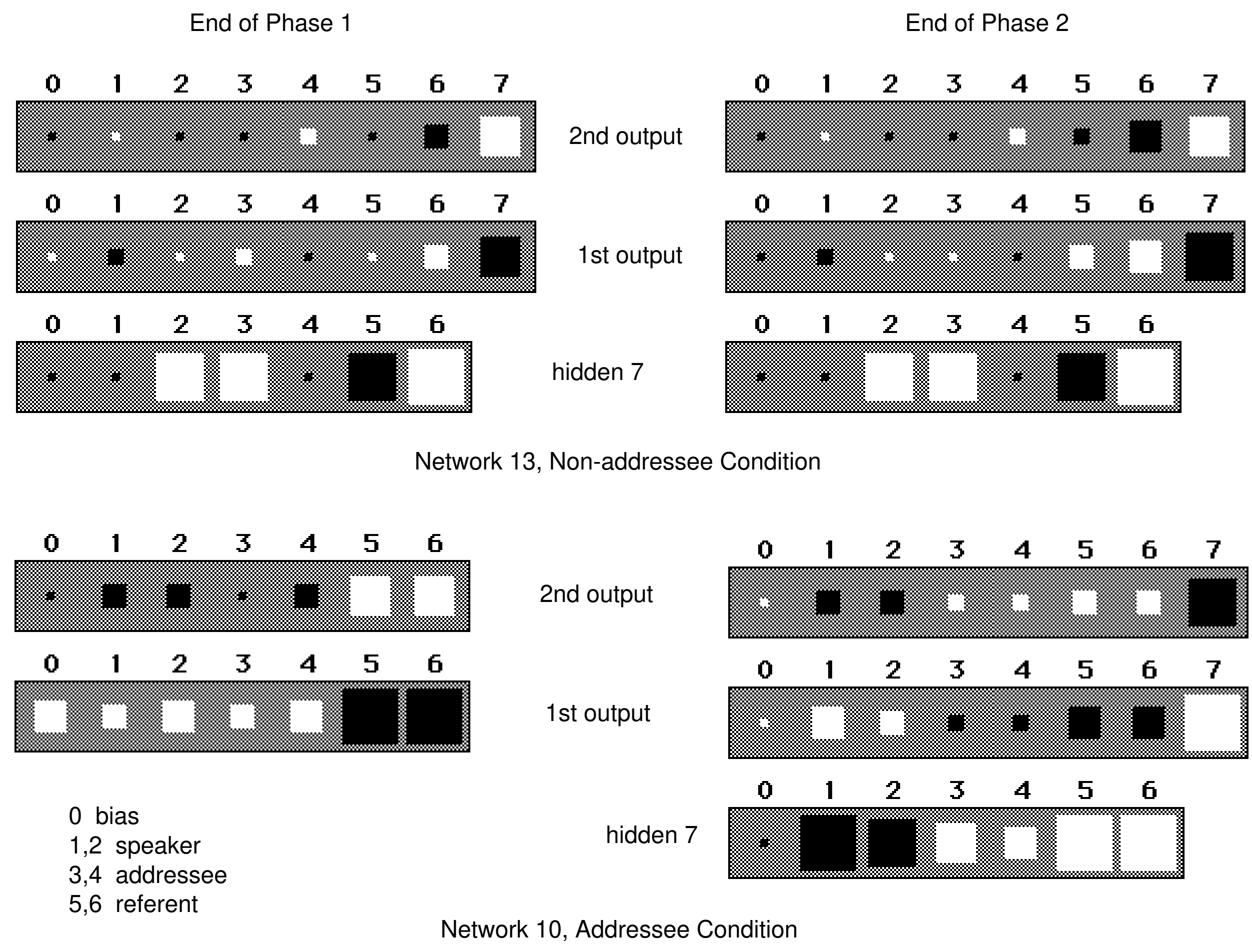

Figure 20.6. Hinton diagrams of incoming weights for two networks.

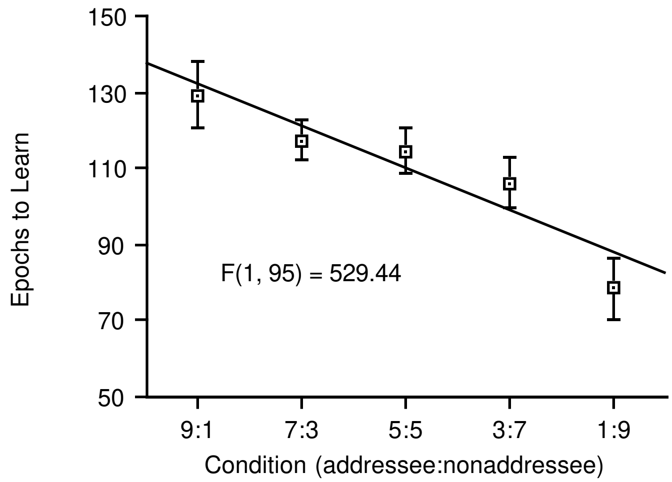

Figure 20.7. Mean epochs to learn in phase 1.

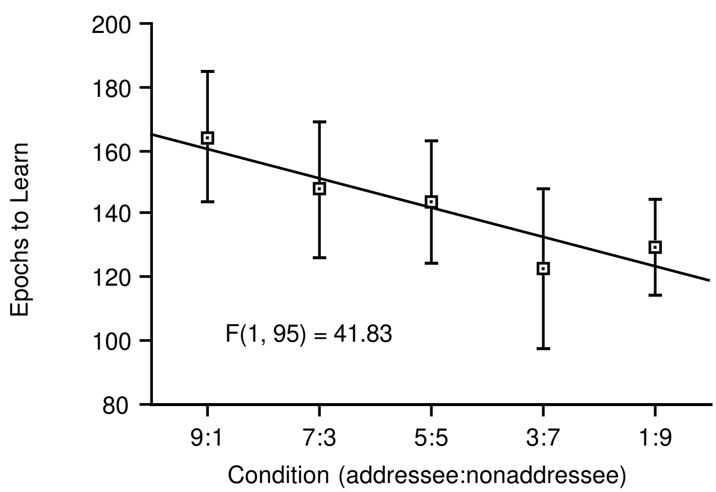

Figure 20.8. Mean epochs to learn in phase 2.

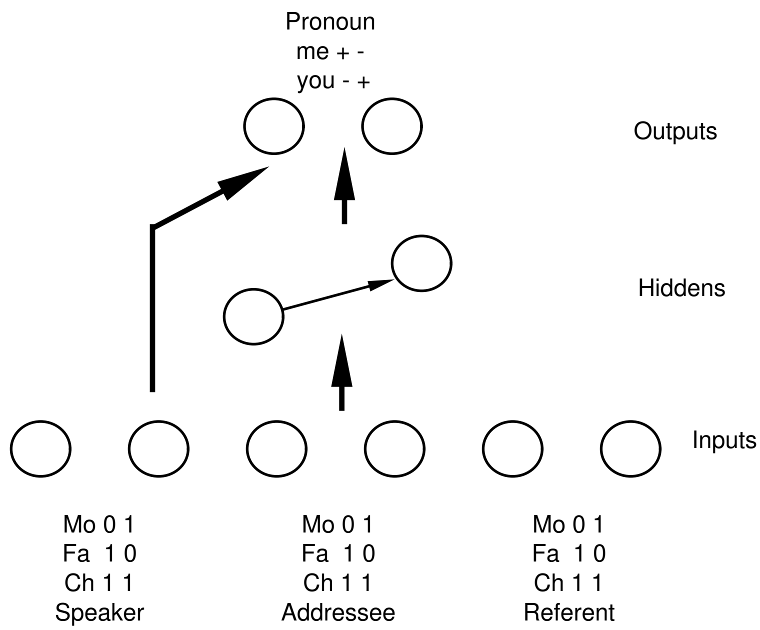

Figure 20.9. Pronoun network after recruitment of a second hidden unit.