

SENSORIMOTOR MEANING REPRESENTATIONS IN LANGUAGE
ACQUISITION

SENSORIMOTOR REPRESENTATIONS OF MEANING IN EARLY
LANGUAGE ACQUISITION

By

STEVE R. HOWELL, BA

A Dissertation

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree

Doctor of Philosophy

McMaster University

© 2004 Steve R. Howell, July 2004

Doctor of Philosophy (2004)

(Psychology)

McMaster University

Hamilton, Ontario

TITLE: SENSORIMOTOR REPRESENTATIONS OF MEANING IN EARLY
LANGUAGE ACQUISITION

AUTHOR: Steve R. Howell, BA (York University)

SUPERVISOR: Professor Suzanna Becker

NUMBER OF PAGES: xiv, 204

Abstract

Evidence suggests that children’s prelinguistic conceptual knowledge significantly influences the course of language acquisition. In a series of nine experiments we investigate this influence. We begin with two experiments using adult human subjects, in which we develop an analogue of children’s early sensorimotor semantic representations and demonstrate that we have captured important aspects of meaning. We then use these sensorimotor semantic representations in simulation experiments using neural network models of language acquisition. First, we provide evidence that having these sensorimotor representations improves grammatical learning. Then we demonstrate that with these rich semantic representations there are strong correlations between the time course of lexical and grammatical learning analogous to those found in children. We suggest that this supports the position that grammar emerges from the formation of a rich lexicon. Finally, we show that it is not necessary to provide these sensorimotor representations for all words. We provide evidence that, given a directly grounded foundation of children’s earliest words, the model can indirectly acquire grounded, embodied semantic representations for novel ungrounded words. Our results thus provide evidence that the initial structure of children’s conceptual or semantic ‘space’ provides an important constraining and simplifying foundation that influences the course of later language acquisition.

Acknowledgements

I am sincerely grateful to my supervisor, Suzanna Becker, for her thoughtful advice, patience, and guidance. She has provided essential motivation and encouragement, in some particularly difficult times. Her confidence in my ability to do independent research has motivated me throughout my graduate career. I thank her as well for supervising my work even when it diverged from her primary interests. I would like to thank Lee Brooks, who has guided me in many moments of difficulty and whose advice on matters of cognitive psychology and general scholarship is greatly appreciated. I have learned much from him in many areas, not limited to those relevant to this thesis. I would also like to thank Betty-Ann Levy, whose casual wit and friendliness enlivened the early years of my dissertation work. I owe her for giving me the opportunities and the confidence to teach university courses at McMaster and hence round out my skills as an academic and future professor. I appreciate her abilities to put the esoterica of connectionist research into a broader psychological perspective.

I would also like to thank Damian Jankowicz for his constant encouragement and friendship over the years. He has been my most important sounding board and provided valuable feedback on many topics. Patrick Byrne, Melissa Dominguez, and Chris Gilbert have also been good friends and listeners whose input has at times improved this work, and at other times merely maintained my sanity.

I would like to thank Jerome Feldman for supervising me as a visiting student at UC Berkeley, and George Lakoff for inspiring my use of sensorimotor features while I was at Berkeley.

Financial support for this work was provided by departmental scholarship, a National Science and Engineering Research Council (NSERC) postgraduate scholarship and two Ontario Graduate Scholarships.

Finally, I would like to thank my family who helped me through difficult times and who always supported my many years of education. Similarly, I would like to thank Catherine McKee for her support in many trying times, and for helping to support me financially during this research. I am truly sorry that my lengthy doctoral education became an issue between us.

Table of Contents

Abstract.....	iii
Acknowledgements	iv
List of Figures	ix
List of Tables.....	xii
Preface	xiii
1 - Introduction.....	1
1.1 The Simple Recurrent Network	4
1.2 The Prediction Task and Gold’s Proof.....	8
1.3 Reference, Meaning, and Embodiment	10
2 - A Model of Grounded Language Acquisition: Sensorimotor Features Improve Grammar Learning	19
2.1 Preface	19
2.2 Abstract.....	19
2.3 Introduction.....	20
2.4 Experiment 1 - Generation of Noun Sensorimotor Features	30
2.4.1 Method.....	31
2.4.2 Results.....	34
2.4.3 Discussion.....	36
2.5 Experiment 2 - Generation of Verb Sensorimotor Features.....	38
2.5.1 Method.....	38
2.5.2 Results.....	40
2.5.3 Discussion.....	42
2.6 Experiment 3 - A Model of Grounded Lexical Acquisition.....	43
2.6.1 Method.....	44
2.6.2 Results.....	48
2.6.3 Discussion.....	49
2.7 Experiment 4A – A Large Corpus Model of Lexical Acquisition.....	51
2.7.1 Method.....	53
2.7.2 Results.....	55
2.7.3 Discussion.....	56
2.8 Experiment 4B – Pause Markers Removed.....	57
2.8.1 Method.....	57
2.8.2 Results.....	58
2.8.3 Discussion.....	59
2.9 Experiment 4C – Reduced Hidden Layer.....	59

2.9.1	Method.....	60
2.9.2	Results.....	61
2.9.3	Discussion.....	63
2.10	General Discussion and Conclusions.....	65
3 - Grammar from the Lexicon: Evidence from Neural Network Simulations of Language Acquisition..... 69		
3.1	Preface.....	69
3.2	Abstract.....	69
3.3	Introduction.....	70
3.4	Simulation Experiment 1.....	72
3.4.1	Method.....	73
3.4.2	Results.....	80
3.4.3	Discussion.....	82
3.5	Simulation Experiment 2.....	85
3.5.1	Method.....	85
3.5.2	Results.....	86
3.5.3	Discussion.....	89
3.6	General Discussion.....	90
4 - Grounding Words in Meaning Indirectly – A Computational Model of the Propagation of Grounding..... 93		
4.1	Preface.....	93
4.2	Abstract.....	94
4.3	Introduction.....	94
4.4	Method.....	100
4.4.1	Corpus and Training Schedule.....	104
4.4.2	Control Network.....	105
4.4.3	Experimental Networks.....	105
4.4.4	Categorical Analysis.....	108
4.5	Results.....	109
4.5.1	Categorical Analysis.....	116
4.5.2	Verb Group.....	117
4.6	Discussion.....	118
5 - General Discussion..... 127		
5.1	Sensorimotor Feature Representations.....	128
5.2	Sensorimotor Representations’ Effect on Grammar Learning.....	131
5.3	Lexicon to Grammar Effects and Emergent Grammar.....	132
5.4	Propagation of Grounding.....	134
5.5	“Facilitative Interference”.....	136
5.6	Model Limitations.....	136
5.7	Conclusion.....	140

References.....	142
Appendix A: Forms and Instructions for Experiment 1, Chapter 2	153
Appendix B: Cluster Analysis of 352 Nouns from Chapter 2	164
Appendix C: Forms and Instructions for Experiment 2, Chapter 2.....	173
Appendix D: Cluster Analysis of 90 Verbs from Chapter 2.....	188
Appendix E: SRNEngine - A Windows-based neural network simulation tool for the non-programmer	191

List of Figures

- 1.1: A simple recurrent network in which activations are copied from hidden layer to context layer on a one-for-one basis, with fixed weights of 1.0. Dotted lines represent trainable connections.....7
- 2.1: Self-organizing Feature Map of Experiment 1 Feature Vectors – Each concept is written on the unit that responded most highly to presentation of that concept after training. Note the grouping of similar concepts on nearby units, as well as the overall topography of similarity.....37
- 2.2: Self-Organizing Map of the Verb Feature Ratings - Note the grouping together of words involving similar motor activities such as drink/lick/taste and listen/say/talk as well as modes of locomotion such as slide/jump/go/walk/hurry.....42
- 2.3: CMU Phonemes and their compressed 14-bit Representations. The bits represent articulatory features such as voiced/unvoiced, place and manner of articulation, etc.....45
- 2.4: The network used in Experiment 1. Note the use of two different inputs per word, one containing the phonemic representation of the word, the other the real-valued noun features of the word.....46
- 2.5: Graph of mean prediction accuracy of Experimental and Control Networks averaged across 6 runs starting from random initial weights. Error Bars are standard errors.....49
- 2.6: Modified SRN architecture, including standard SRN hidden layer and context layer, standard linguistic (word) prediction output, and novel noun feature output and verb feature output. The linguistic input is a whole-word phonetic representation of up to 10 phonemes. The Noun and Verb feature targets are meant to be an abstract representation of pre-linguistic sensory and motor-affordance semantics.....52
- 2.7: Mean grammatical prediction performance for a large naturalistic corpus (10,742 words) which includes pauses/periods. Number of networks in each condition is 10, and error bars indicate standard error.....54
- 2.8: Mean grammatical prediction performance for a large naturalistic corpus (8328 words) which *excludes* pauses/periods. Number of networks in each condition is 10. Error bars indicate standard error.....58
- 2.9: Mean grammatical prediction performance for a large corpus (8328 words) which excludes pauses/periods, with a reduced hidden/context layer (size 10). The number of networks in each condition is 10. Error bars indicate standard error.....62
- 2.10: Noun and Verb feature encoding accuracy from Experiment 4C. These two output layers were performing a recoding from the phonetic features of a word to the semantic features of a word. The noun feature encoding is significantly different across the two conditions, as measured by t-test at the

terminal point, but the verb features are not. The number of networks in each condition is 12. Error bars indicate standard error.....	64
3.1: Modified SRN architecture, including standard SRN hidden layer and context layer, standard linguistic (word) prediction output, and novel noun feature output and verb feature output. The linguistic input is a whole-word phonetic representation of up to 10 phonemes. The Noun and Verb feature targets are meant to be an abstract representation of pre-linguistic sensory and motor-affordance semantics.....	74
3.2: Noun Lexicon to Grammar Correlations in all three conditions of Experiment 1. Curves are noun to grammar correlations from the Epoch 20 noun reference point to all grammar points (Epochs 1 to 500). Simultaneous correlations occur when the Epoch is equal to the Noun reference point (Epoch 20). Correlations before that point are from earlier grammar to later lexical learning, correlations after are lexical learning to later grammatical learning.....	81
3.3: Verb Lexicon to Grammar Correlations in all three conditions of Experiment 1. Curves are verb to grammar correlations from the Epoch 20 verb reference point to all grammar points (Epochs 1 to 500). Simultaneous correlations occur when the Epoch is equal to the Verb reference point (Epoch 20). Correlations before that point are from earlier grammar to later lexical learning, correlations after are lexical learning to later grammatical learning.....	82
3.4: Noun Lexicon to Grammar Correlations in both conditions of Experiment 2. Curves are noun to grammar correlations from the Epoch 40 noun reference point to all grammar points (Epochs 1 to 500). Simultaneous correlations occur when the Epoch is equal to the Noun reference point (Epoch 40). Correlations before that point are from earlier grammar to later lexical learning, correlations after are lexical learning to later grammatical learning.....	87
3.5: Verb to Grammar Correlations in all three conditions of Experiment 2. Curves are verb to grammar correlations from the Epoch 40 verb reference point to all grammar points (Epochs 1 to 500). Simultaneous correlations occur when the Epoch is equal to the Verb reference point (Epoch 40). Correlations before that point are from earlier grammar to later lexical learning, correlations after are lexical learning to later grammatical learning.....	88
4.1: Modified SRN architecture, including standard SRN hidden layer and context layer, standard linguistic (word) prediction output, and novel noun feature output and verb feature output. The linguistic input is a whole-word phonetic representation of up to 10 phonemes. The Noun and Verb feature targets are meant to be an abstract representation of pre-linguistic sensory and motor-affordance semantics.....	101
4.2: CMU Phonemes and their compressed 14-bit Representations. The bits represent articulatory features such as voiced/unvoiced, place and manner of	

articulation, etc. This representation is not meant to make any claims as to the relevance of these features, it was chosen only for practical purposes of compressing the number of bits required to represent a phoneme.....103

4.3: The hierarchical cluster analysis dendrogram of the Noun master features, as generated by human raters.....110

4.4: The hierarchical cluster analysis dendrogram of the Noun Feature layer output prototypes for the Noun Control network. Note the correspondence to the master feature dendrogram in Figure 3.....111

4.5: An example of a noun hierarchical cluster analysis dendrogram for the Exact Match results of the Noun Group.....114

4.6: An example of a noun hierarchical cluster analysis dendrogram for the Near Match results of the Noun Group.....115

4.7: An example of a noun hierarchical cluster analysis dendrogram for the Incorrect results of the Noun Group.....116

List of Tables

2.1: Noun Category Agreement Results - Feature Vectors compared to Centroids of Categories Drawn from MCDI.....	35
2.2: Table 2.2 - Verb Category Agreement Results	41
2.3: Output Accuracy from sample network at 500 epochs, during training.....	56
4.1: A summary of the results of the hierarchical cluster analysis on the Noun Group's Noun Feature prototypes.....	113
4.2: The results of the categorical analysis on the Noun Group's Noun Feature prototypes. Centroids were calculated by averaging the feature vectors of its members, and then each word's features were compared to the category centroids. The word was assigned to the closest category by Euclidean distance measure.....	117

Preface

This dissertation is divided into three main sections, each of which represents material that has been submitted for publication and is now under revision. Since each of these articles have multiple authorship, my contribution to each is explained here. I am first author on all articles.

Howell, Becker and Jankowicz (submitted) reports 6 experiments. All six experiments represent my contribution to the paper, and thus are all relevant to this dissertation. We create novel semantic representations of words based on human subject ratings along sensorimotor feature dimensions, then provide experimental evidence that these features capture important aspects of conceptual meaning. Furthermore, we provide simulation evidence that the incorporation of these semantic representations improves another aspect of language acquisition, word-sequence (grammatical) learning. All experiments provide a novel contribution to the literature. Experiments one and two were conducted between September 2001 and September 2002. The remaining experiments were conducted between September 2002 and October 2003.

Howell and Becker (submitted) reports two experiments. Both experiments represent my contribution to the paper, and thus are both relevant to this dissertation. We provide evidence that lexical learning that is grounding in pre-linguistic sensorimotor features causes a distinct correlation between lexical and grammatical learning analogous to that found in children. Control experiments of acquisition that lack this grounded sensorimotor meaning do not

show this high correlation. Both experiments provide a novel contribution to the literature. These experiments were conducted between April 2003 and December 2003.

Howell and Becker (submitted b) reports one several-part simulation experiment. All parts represent my contribution to the paper, and thus are relevant to this dissertation. We provide evidence on a possible mechanism for the indirect acquisition of grounded meaning for words, a process we call ‘propagation of grounding’. This process is argued to be similar to how children learn the meanings of novel words encountered in the absence of their physical referents, such as in conversation or reading. This experiment provides a novel contribution to the literature. These experiments were conducted between November 2003 and December 2003, although extensive pilot work investigating this phenomenon took place between September 2002 and November 2003, which is not reported herein.

Finally, Howell and Becker (Submitted c) is a methods paper describing a computational modeling software system that I created for the purposes of this dissertation. It is attached in Appendix E. The unique properties of this simulation platform have allowed me to process the large, naturalistic simulations contained in the dissertation in manageable amounts of time. This is a novel contribution to the literature. This simulation environment was created (programmed and tested) by me between September 1999 and December 2003, on an ongoing basis.

Chapter 1

Introduction

The study of language using the methodology of modelling is an interesting and difficult challenge, and one that has been addressed in a number of different ways by different researchers. This is due in part to the inherent difficulty of the study of language. Language is composed of multiple different interacting levels of processing, dealing with initial perception, semantics, syntax and grammar, pragmatics or content, and more, all operating together. Attempts have been made to model different parts of this multi-level process. Examples include Weibel et al. (1989, 1989a) on phoneme recognition, McClelland and Elman (1986) on phoneme and word recognition, McClelland and Rumelhart (1981) on speech perception, Mozer (1987) with his Blirnet model of word recognition, and Seidenberg and McClelland (1989) on the pronunciation of written words and simulated lesioning. Hinton and Shallice (1989) and Hinton and Sejnowski (1986) also addressed the lesioning of trained networks in order to study the performance of the disrupted net in comparison to such human disorders of language as dyslexia. Rogers & McClelland (1999) provide an interesting model

of semantic featural representation. Landauer et al. (1997) demonstrates the usefulness of pure semantic analysis with his Latent Semantic Analysis, while Elman and colleagues (1990, 1991, 1993) deal with pure grammar learning. While some of these simulations involved more than one 'level' of processing, some of the most impressive models (Landauer, Elman) incorporate only a single aspect of language. I believe that some of the most promising directions for future research may lie in extending these excellent single-level models to encompass more of the language task. Specifically, working within the Elman-Jordan Simple Recurrent Network (SRN) tradition (Elman, 1990; Jordan, 1986), I shall examine extensions and modifications of the basic model, including the incorporation or examination of constraints and structures from computer science (structured connectionism, see Feldman, 1989), psycholinguistics (e.g. Lakoff and Johnson, 1980, 1999; Goldberg, 1995, 1999) and constrained connectionism (Regier, 1996).

While it follows a completely different methodology than the SRN tradition, Landauer's simulation of semantic knowledge, based on what he calls Latent Semantic Analysis (LSA), is perhaps the most impressive work on language. It is relevant to any work that intends to model semantics since it provides a sense of what is possible to date in this area. Cast in the framework of a neural network model, LSA could be viewed as a method for training a network that associates two classes of events reciprocally, by linear connections through a single hidden layer. Landauer's model was trained to learn and represent relations

among very large numbers of words (20k - 60k) and very large numbers of natural text passages (1k-70k) in which those words occurred. The result was 100-350 dimensional 'semantic spaces' in which any word or passage could be represented as a vector. Similarities could then be measured between any two vectors (words), usually via the cosine of the angle between the two vectors. Landauer reports very impressive results on a variety of human tasks, such as multiple-choice vocabulary and domain knowledge tests, emulation of expert essay evaluations, and in a number of other ways (Landauer, Laham, and Foltz, 1998).

There are several impressive aspects to Landauer's model. First, the vocabulary size used is immense, at twenty to sixty *thousand* words. This is a vocabulary comparable, or at least on the right order of magnitude, to that of a human adult. Few other models exceed more than a few hundred word vocabulary, which is too small even to compare to a child's. Second, Landauer trains his model purely by the input of natural language texts, and the statistical processing that the model performs upon it. This process of discovering the semantic relationships between words by exposure seems analogous to the way infants acquire semantics (word knowledge) just by exposure.

Still, despite the model's ability to do such things as grade student essays (with up to 80% correlation with expert human evaluation!), it treats language as nothing more than a "bag of words" (Landauer, 1998, p. 48). That is, the words are in an unordered collection, there is no 'grammar' or regularity to them. Furthermore, it is not apparent how this architecture could be extended to include

grammar, being founded as it is solely on word-to-text correlations. Including grammar would necessitate some role for context or order of the input words. Of course, the order that is present in language is first and foremost a *temporal* order. As Becker (1999) demonstrated, incorporating temporal context can improve object recognition in a sequence of visual images. How much more effect must it have on language, which is a *strictly* serial process, happening one word after another in time? Which words come before and after others is dictated by the grammatical ‘rules’ of the language¹. Thus it seems as though any more comprehensive model of language, one which includes grammar, would have to incorporate temporal context of some sort.

1.1 The Simple Recurrent Network

While there are a variety of ways of incorporating temporal context, the SRN architecture has emerged as the leading candidate in this sort of language modelling effort. Simple Recurrent Networks (SRNs), as advanced and popularized by Jeff Elman, to the point that they are often known as Elman nets, work admirably well². The SRN architecture possesses a number of desirable

¹ Note that I when I refer to the “rules” of grammar, I mean this only heuristically, and am not claiming that there are rules for grammar embodied anywhere in the brain.

² It is important to note that Simple Recurrent Networks, as can be inferred from the name, are simplified versions of *fully* recurrent networks. In a fully recurrent network, the feedforward sort of architecture does not necessarily hold, and units can be connected to themselves or to other units in a multitude of ways,

temporal properties, and has yielded many interesting results for Elman and others, in such language-related areas as grammar learning (Elman, 1990, 1991, 1993, 1995) and grammar generalization (Elman, 1998, 1998a, 1999), English past-tense learning, (Elman et al., 1996; Hare, Elman, & Daugherty, 1995) morphological drift in English past-tense (Hare & Elman, 1994), processing of recursive sentences (Weckerly & Elman, 1992), counting (Wiles & Elman, 1995), the mathematical properties of dynamical systems such as language (Elman, 1995), and neo-Piagetian developmental thought (Bates & Elman, 1993).

No matter the application, the power of the SRN (Figure 1) is that it represents time indirectly, through its effects on processing, not by any sort of explicit repetition or re-representation at the input. The system possesses dynamic properties that are responsive to temporal sequences. In short, the network has *memory*.

As Elman (1990) notes, a number of possible ways of giving a network memory have been suggested. However, perhaps the most promising one, and the method used by SRNs, was suggested by Jordan (1986). Jordan defined a

with different amounts of *memory* being involved. I will not consider them further here, but a full discussion of one of several commonly used learning algorithms for training them, known as backprop-through-time, is available in Williams and Zipser (1989). Simple recurrent networks are essentially a truncated version of this algorithm, which consider the slope of the error at only one time step back, and not any farther, when calculating the net's error signal. Elman and others have been able to achieve significant results with the truncated version, but if necessary a more accurate model could easily be achieved by using the full backprop-through-time algorithm, at the expense of being more computationally intensive.

network containing recurrent connections, in which the output could be associated, not just with the current input, but with the network's previous output. Simple recurrent networks operate in much the same way; they use a layer of recurrently connected “context” units to store a memory of the network’s internal state at previous time steps.

The architecture of an SRN is straightforward. There is a layer of input units, which receive input from the environment. These feed forward into a layer of hidden units, with each input unit connected to every hidden unit. These hidden units are similarly fully connected to the output units, but also connected to a set of *context units*. The connections of the hidden units to the context units are of a one-to-one nature, that is, each hidden unit feeds forward into one and only one context unit. Furthermore, the weights on each hidden unit to context unit connection are 1.0. Thus the context units receive an exact copy, at each time step, of the hidden units' output. The context units themselves feed forward, in a fully connected fashion, back to the hidden units. Thus what the context units are supplying to the hidden units is the hidden unit’s own output from the *previous* time step.

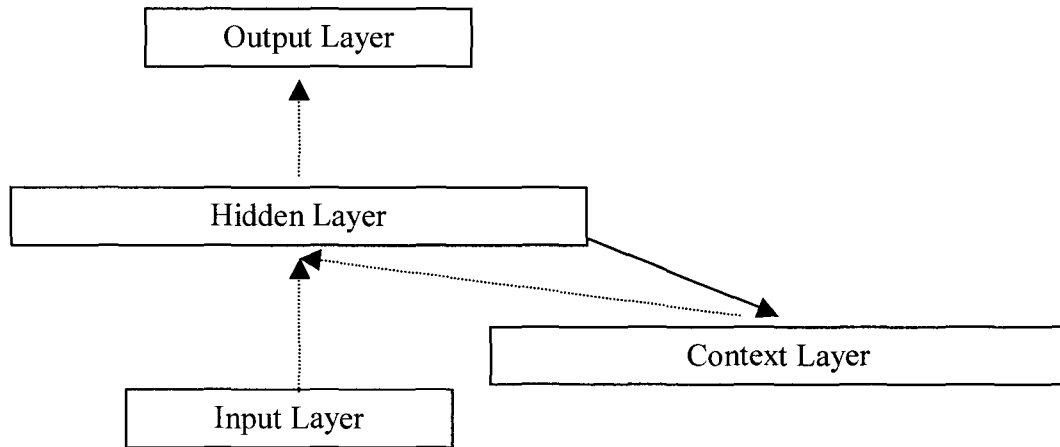


Figure 1.1 - A simple recurrent network in which activations are copied from hidden layer to context layer on a one-for-one basis, with fixed weights of 1.0. Dotted lines represent trainable connections.

It is important to note that the memory of these networks is not for a single time step only. Rather, the activation at each time step gets fed back into the processing of the next, and the combined processing of that input gets fed back on the next time step, and so on. The memory is thus ‘smeared-out’, with the effect of any given preceding time step decreasing every new time step. This effect is the motivation for adding parameters to the context layer to control the rate of this memory decay, in effect specifying the duration of that memory trace. This addition also makes useful the incorporation of more than one hidden layer, each with a different rate of decay of context, as has been explored by Howell and Becker (2000).

1.2 The Prediction Task and Gold's Proof

Also important, however, is the task performed by the SRN, which in most language work to date has been the *prediction* task. Since SRN's are trained with the supervised backpropagation algorithm (Rumelhart, Hinton, & Williams, 1986), some training signal is of course required. The use of the prediction task is at least partially to avoid the implications of Gold's proof (Gold, 1967; Rohde & Plaut, 1999). Gold demonstrated that if a language learner is presented with positive-only data, that only *regular* languages can be learned. Unfortunately, natural languages belong to a more powerful, and thus more complex, class than these, and thus cannot be learned solely on the basis of positive evidence. However, perhaps natural language learners (i.e. children and neural nets) receive more than just positive data during language acquisition. That is, perhaps in addition to exposure to the proper uses of language (positive evidence) they also receive negative evidence in the form of explicit corrections and modifications from mature language users (i.e. adults). Unfortunately, there is no good evidence that children receive or use negative data during grammar learning (Brown & Hanlon, 1970), although they may during word learning (Bloom, 2002). Gold suggests several possible explanations for the fact that children somehow do manage to learn language, in spite of his findings. Gold suggests that while children may not make use of explicit negative evidence, they may in fact make use of several forms of indirect negative evidence. One form is innate knowledge or constraints on the language mechanism. While certainly true to some extent, the nature of

these innate constraints is very much an open question to date (Elman, 1993; Pinker, 1995). Thus connectionist research has focused on the second form of indirect negative evidence, the violation of expectations.

Imagine a child who is beginning to put sentences together. He or she begins by consistently using non-grammatical utterances. However, the speech that he or she hears from others is typically of a different form (grammatical) and the non-grammatical structure is rarely heard. The child might be *expecting* to hear similarities to his or her own constructions in the speech of others. In spite of not being told directly that his or her utterances were not correct, we might expect that the child would *learn* that they were not correct from the failure of his or her predictions of other's speech. This is the form of indirect negative evidence that is used in the prediction task. The network makes a prediction about, for example, what word will be 'heard' next in the input sequence, and this is compared against the actual next word in the stream. If the prediction is accurate, the network will alter its weights to make that response more likely in the future. If the prediction is not confirmed, then the weights will be altered slightly to make that response less likely. Prediction-task SRNs thus don't use an external 'teaching signal' to provide the supervision for the SRNs, they use their own input at the next time step. This 'self-supervising' form of training signal makes the use of the algorithm more biologically plausible.

1.3 Reference, Meaning, and Embodiment

However plausible the learning mechanism of the SRN, there remains a crucial disadvantage. It has been said that an SRN's operation is analogous to "learning a language by listening to the radio" (McClelland, quoted in Elman, 1990). The suggestion here is that an SRN is simply manipulating words and other symbols, and while it may be able to learn rules for ordering these symbols in a stream, it will be incapable of learning what the symbols *mean*. This criticism can be interpreted as a demand for the grounding of the symbol system in reality, which to a linguist is the problem of *embodiment*. Obviously this criticism is not a fatal flaw, since Landauer's LSA is capable of making semantic comparisons with surprising accuracy. In fact, LSA can serve as a baseline level beyond which embodiment may indeed be necessary to ground semantics. There is considerable evidence, however, that for full language understanding embodiment is necessary, perhaps especially at the earliest points of learning. Among others, a community of researchers at the University of California at Berkeley have addressed this issue. Their group is called the Neural Theory of Language (formerly L0), and incorporates psycholinguists (George Lakoff, Adele Goldberg), artificial intelligence researchers (Jerome Feldman) and structured or constrained connectionists (Feldman, Terry Regier, Srinivasan Narayanan). All of these researchers have contributed to an understanding of the process of embodiment of meaning (semantics).

Perhaps most far-reaching is the long-standing work of Lakoff and Johnson on metaphor. Lakoff & Johnson (1980, 1999) offer detailed analysis of metaphor in everyday and even scientific language that, according to them, influences the way all but the most concrete sentences are perceived. Sentences such as “He’s crazy about her” or “She’s steaming mad” can only be understood using metaphor, according to Lakoff & Johnson. The metaphors in these two cases would be Love is a Sickness and Anger is a Hot Fluid, respectively. Lakoff and Johnson examine a great many metaphors in their work, convincingly demonstrating that metaphor is more than just poetic, that it is in fact essential to any understanding of abstract concepts. Indeed, any concept that is not directly grounded in bodily experience (e.g. through perception) is indirectly grounded through metaphor, they claim. These metaphors thus conventionalize the extension of some concrete experience to an abstract domain, rendering that domain more comprehensible. As mentioned, this is applicable to scientific as well as everyday language, including such abstract domains as economics, where Srinivasan Narayanan (1995) has used structured connectionism to build a system that could understand economics newspaper headlines in light of embodying metaphor (e.g. “France *crawled* out of a recession.”). Still, since metaphor builds on concrete concepts, even if we accept that metaphor may underlie our understanding of abstract concepts, we must first address the acquisition of those concrete concepts.

Relational words, such as those encompassing the spatial relationships, are more concrete than metaphor, labeling as they do physical concepts that must be learned fairly early in life. Regier (1996) has simulated, using connectionist methods, the acquisition of spatial prepositions like under, on, and into from simulated perceptual information (2D movies). Regier's model consists primarily of a standard feedforward (PDP) neural network which performs the learning of the spatial terms. Importantly, however, his model also incorporates detailed, hand-wired neurologically-motivated 'structures' that are viewed analogously to various aspects of the visual system. Specifically, he incorporates orientation-comparing nodes (orientation tuned cells), perceptual 'filling-in' structures (centre-surround cells), and boundary maps (the hypothesized 'feature maps' of attentional researchers like Treisman). The structures process the 2D visual information and feed into and constrain the operation of the standard PDP net, making the problem easier to learn than if the net were 'searching' without constraint for the solution in an effectively much larger potential problem space. Regier's model thus demonstrates, quite successfully, that linguistic terms for spatial relations emerge from the interaction of simple perceptual structures (biological constraints such as orientation tuning in the visual system) and perceptual and linguistic experience. Furthermore, he offers possible extensions to his model that could account for other forms of relational words.

Regier's model handles relational words, which serve both a concrete and a grammatical function, and are learned at a later point by children, but what

about the earliest and arguably most concrete part of language acquisition, noun learning? Smith (1999) argues that learned attentional biases in children facilitate the acquisition of new words, specifically nouns. She proposes that children generalize from the form of the utterances typically used by adults to instruct them or speak to them³, and acquire a frame or template for noun acquisition: the shape bias. The count noun frame “This is a ____” is often used when introducing a new object to a child, and Smith summarizes evidence that attention to this learned frame results in much better noun acquisition (Smith, 1999). Other frames offer advantages for other words, such as the adjective frame (“This is a ____ thing.”) or the mass noun frame (“This is some ____”), but Smith focuses on the count noun frame and the way it arises out of a learned cue (the frame) and an attentional bias to shape that is learned along with it. That is, when the count noun frame is used, children learn that shape is what is relevant to the distinction of a new word for this new object. In fact, when asked to categorize objects, children who have learned this bias will categorize differently based on whether the experimenter phrases the request with a count noun frame (attention to shape) or a mass noun frame (attention to substance & texture). Evidence also exists that

³ It is important to note that this is not contradicting the evidence that shows that parents tend not to explicitly correct their children’s grammar. That evidence is from a later stage in children’s development, during the ‘grammar burst’, while Smith’s argument pertains to children who are learning their first words. Typically around 300 words are learned at the single word stage before the grammar burst begins. (Bates and Goodman, 1999). Although for a differing view see Bloom (2002).

this is indeed an attentional bias that emerges with word learning; it does not exist in children who have fewer than about 50 count nouns in their vocabulary, but is present in those who do, and can be learned by children during experimentation (Smith, 1999).

For an SRN language modeling effort, what is important about this research is its emphasis on the importance of some early learning for later learning to be facilitated. This “bootstrapping” phenomenon seems to occur at various stages of language acquisition. For example, available evidence (reviewed in Bates & Goodman, 1999) suggests it is likely that a certain critical mass of words is necessary before beginning to acquire grammar, and this fact, along with the manner in which those words are learned by children, may inform the design of extensions to the SRN architecture, as we shall see.

We have briefly examined aspects of the learning of concrete nouns, spatial prepositions, and the metaphorical processes that may be necessary to extend them to deal with abstract concepts. Verbs and verb phrases, on the other hand, turn out to be perhaps the most complex, and the most related to what we consider to be grammar. Verbs have been argued to determine the lexical structure of utterances, for example, by determining the argument ‘slots’ required in the sentence, into which nouns and relational words are ‘dropped’. For our purposes, we can consider that this is what SRNs are learning when they acquire aspects of the syntax of a language from its serial order. Beyond even that, however, Goldberg and others (Goldberg, 1995, 1999) have argued that there

exists a level, *constructions*, above that of verbs that actually determines sentence structure more than verbs do. These constructions can actually take different verbs as fillers of sentential ‘slots’ and change the verbs’ allowable uses and requirements for arguments. Thus a verb, such as *sneeze*, that normally isn’t allowed a direct object or patient (“He sneezed”) can be used in a construction, e.g. the caused-motion construction, and suddenly is allowed to take them (e.g. “He sneezed the foam off the cappuccino.”) Goldberg examines the acquisition of these constructions as an emergent property of the early use of light verbs. Specifically, Goldberg shows that these ‘light’ verbs (such as *go*, *do*, *make*, *give*, and *put*) are the most frequent verbs in children’s early language. They are also, she argues, the most generally useful verbs, not too specific and applicable to a wide range of instances. For example, ‘*go*’ can be used in any sentence involving motion, and while it will often significantly under-specify the true actions (e.g. “He drives the car to the store” vs. “He goes to the store”) the central meaning of the events is still conveyed. In chapter two I take advantage of the centrality and wide applicability of these light verbs by using them in the verb feature generation pilot section of experiment two.

The frequency and applicability of light verbs, Goldberg argues, gives them a central status in the acquisition of types of sentence ‘frames’, such that they actually come to define and characterize the frames. Then later, as additional words are learned that specify the same general event with greater specificity, these other words (*run*, *drive*, etc) can likewise be fitted into the frame. When a

frame becomes highly well-learned over time (such as the caused motion construction in the sneeze example above) it can actually *impose* its structure on novel verbs in the frame, allowing them to be used (and importantly, *understood*) in ways that would be ungrammatical otherwise. This aspect of language acquisition, like that of metaphorical processing mentioned above, is among the more complex language processes that one could attempt to model. While our SRN models of language acquisition are presently not powerful enough to address these sorts of processes, I do in later chapters discuss some ways in which these might eventually be addressed. Furthermore, we again see the idea of language “bootstrapping”, with existing knowledge providing constraints that makes possible more difficult learning.

While the preceding discussion has addressed more aspects of language acquisition than can presently be addressed in a single model, the general themes are valuable to keep in mind as we think about what *can* be modeled successfully. Starting with the most readily accessible, I have chosen to address both concrete and abstract noun learning, verb learning, and then the beginnings of syntactic and grammatical learning. Specifically, I conduct experiments using SRN models of language acquisition that incorporate lexical learning of both nouns and verbs, along with aspects of syntactic learning (grammar learning). In the experimental evidence that follows, grammar learning will refer exclusively to syntactic learning, specifically sequence learning, leaving out any inflectional or other aspects of grammar for present purposes.

The central question addressed in this dissertation is thus: How do children acquire language? In particular, I focus on factors contributing to the acquisition of vocabulary and early grammar, especially how prelinguistic sensorimotor learning affects the developmental trajectory of this language acquisition.

I begin in chapter two by exploring the hypothesis that sensorimotor grounding of meaning facilitates grammar development. First I develop a semantic representation of words that allows me to address the phonology-to-meaning mapping that is necessary for lexical learning in children. I investigate these representations in a variety of ways, including incorporating them into variants of the SRN neural networks introduced above. These simulation experiments indicate that having pre-existing grounded semantic representations facilitates the process of grammatical learning.

In chapter three I investigate the relationship between lexical learning and grammatical learning in a different way, by examining the correlations between lexical learning and grammatical learning in the presence of different amounts of semantic grounding of word forms. This investigates the hypothesis that the correlations found between lexical and grammatical status in children's development are due to the semantic grounding of the words and the effect of that grounding on grammar development.

In chapter four I explore the hypothesis that possessing word meanings grounded in sensorimotor features allows the language learner to readily infer the

meanings of novel, ungrounded words. Specifically, I examine how the incorporation of the sensorimotor features (which accomplish the phonology-to-meaning mappings throughout this research) actually allows the network to indirectly acquire grounded meanings for novel words not previously linked to a meaning, without any specific training as was done in chapter two.

Finally, in chapter five, I briefly summarize what I have demonstrated in the work contained in the previous chapters, consider implications and limitations, and draw conclusions.

Chapter 2

A Model of Grounded Language Acquisition: Sensorimotor Features Improve Grammar Learning

2.1 Preface

This chapter is reproduced from Howell, Becker, and Jankowicz, (submitted).

This paper was first submitted to the Journal of Memory and Language in November, 2003 and is currently under revision for resubmission. In this paper we explore two hypotheses: first, that sensorimotor features can capture significant aspects of the intuitive meanings and similarity structures of words; and second, that sensorimotor grounding of meaning facilitates grammar (syntax) acquisition.

2.2 Abstract

It is generally accepted that children have sensorimotor mental representations for concepts even before they learn the words for those concepts. We argue that

these prelinguistic and embodied concepts direct and ground word learning, such that early concepts provide scaffolding by which later word learning, and even grammar learning, is enabled and facilitated. We gathered numerical ratings of the sensorimotor features of many early words (352 nouns, 90 verbs) using adult human participants. We analyzed the ratings to demonstrate their ability to capture the embodied *meaning* of the underlying concepts. Then using simulation experiments we demonstrated that with language corpora of sufficient complexity, neural network (SRN) models with sensorimotor features perform significantly better than models without features, as evidenced by their ability to perform word prediction, an aspect of grammar. We also discuss the possibility of indirect acquisition of grounded meaning through "propagation of grounding" for novel words in these networks.

2.3 Introduction

Considerable evidence suggests that, by the time children first begin to learn words around the age of 10-12 months, they have already acquired a fair amount of sensorimotor (sensory/perceptual and motor/physical) knowledge about the environment (e.g. Lakoff, 1987, Lakoff & Johnson, 1999; Bloom, 2000; Langer, 2001), especially about objects and their physical and perceptual properties. By this age they are generally able to manipulate objects, navigate around their environment, and attend to salient features of the world, including parental gaze

and other cues important for word learning (Bloom, 2000). These cues help them to learn their first words, which correspond to the most salient and imageable (Gillette, Gleitman, Gleitman, & Lederer, 1999) objects and actions in their environment, the ones they have the most experience with physically and perceptually. Generally speaking, the more "concrete" or "imageable" a word, the earlier it will be learned. This helps to explain the preponderance of nouns in children's early vocabularies. The meanings of verbs are simply more difficult to infer from context, as demonstrated by Gillette et al. (1999). Only the most clearly observable or "concrete" verbs make it into children's early vocabularies. However, later verbs are acquired through the assistance of earlier-learned nouns. If a language learner hears a simple sentence describing a real-world situation, such as a dog chasing a cat, and already knows the words dog and cat, the only remaining word must be describing the event, especially if the learner already has built up a pre-linguistic concept of "dogs chasing cats" at the purely observational level. As Bloom (2000) describes, the best evidence for "fast-mapping" or one-shot learning of words in children comes from similar situations in which only one word in an utterance is unknown, and it has a clear, previously unknown, physical referent present.

These very first words that children learn thus help constrain the under-determined associations between the words children hear and the objects and events in their environment, and help children to successfully map new words to their proper referents. This happens through the use of cognitive heuristics such

as the idea that a given object has one and only one name (Smith, 1999), or more basic object-concept primitives (Bloom, 2000) such as object constancy. With a critical mass of some 50 words, children begin to learn *how to learn* new words, using heuristics such as the count-noun frame, or the adjective frame (Smith, 1999). These frames are consistent sentence formats often used by care-givers that enable accurate inference on the part of the child as to the meaning of the framed word, e.g. "This is a ____". These factors combine to produce a large increase in children's lexical learning at around 20 months. As they begin to reach another critical mass of words in their lexicon (approaching 300 words), they start to put words together with other words - the beginnings of expressive grammar (Bates & Goodman, 1999). Around 28 months of age children enter a "grammar burst" in which they rapidly acquire more knowledge of the syntax and grammar of their language, and continue to approach mature performance over the next few years.

By this account, conceptual development has primacy; it sets the foundation for the language learning that will follow. Words are given meaning quite simply, by their associations to real-world, perceivable events. Words are directly *grounded* in embodied meaning, at least for the earliest words. Of course, it seems clear that the incredible word-learning rates displayed by older children (Bloom, 2000) indicate that words are also acquired by linguistic context, through their relations to other words. Children simply are learning so many new words each day that it seems impossible that they are being exposed to the referents of

each new word directly. The meanings of these later words, and most of the more abstract, less imageable words we learn as adults, must clearly be acquired primarily by their relationships to other known words. It may in fact be true that these meanings can *only* be acquired indirectly, through relationships established to the meanings of other words.

Evidence for the indirect acquisition of meaning is not limited to children's behavior. The work of Landauer and colleagues (e.g. Landauer, Laham, & Foltz, 1998; Landauer & Dumais, 1997) provides perhaps the clearest demonstration that word "meanings" can be learned solely from word-to-word relationships (although see Burgess & Lund, 2000, for a different method called HAL). Landauer's Latent Semantic Analysis (LSA) technique takes a large corpus of text, such as a book or encyclopedia, and creates a matrix of co-occurrence statistics for words in relation to the paragraphs in which they occur. This yields a very high-dimensional vector representation for each word. This high-dimensional representation is then reduced via the statistical technique of singular value decomposition to a more manageable number of dimensions, usually 300 or so. The resulting compressed meaning vectors have been used by Landauer et. al. in many human language tasks, such as multiple choice vocabulary tests, domain knowledge tests, or grading of student exams. In all these cases, the LSA model demonstrated human-level performance.

While these high-dimensional models of meaning such as LSA and HAL perform well on real world tasks, using realistically-sized vocabularies and

natural human training corpora, they do have several drawbacks. First, they lack any consideration of syntax, since the words are treated as unordered collections (a 'bag of words'). Second, LSA and HAL meaning vectors lack any of the grounding in reality that comes naturally to a human language learner.

Experiments by Glenberg and Robertson (2000) have shown the LSA method to do poorly at the kinds of reasoning in novel situations that is simple for human semantics to resolve, due largely to the embodied nature of human semantics.

So it seems that there are two sources of meaning, direct embodied experience, and indirect relations to other words. However, there is an infinite regress in the latter. If words are only ever defined in relation to other words, we can never extract meaning from the system. We have only a recursive system of self-defined meaning, symbols chained to other symbols. To avoid this dilemma, at least some of the words in our vocabularies *must* be defined in terms of something external. In children, at least, the earliest words serve this role. They are defined by their mappings to pre-linguistic sensory and motor experience, as discussed above. They do not require other words to define their meaning. The most imageable words are thus directly grounded, while the less imageable and more abstract the words that are encountered during later learning, the more indirectly grounded they are. At some point, we argue, the adult semantic system begins to look much like the LSA or HAL high-dimensional meaning space, with our many abstract words (e.g. love, loyalty, etc) defined by relations among words themselves. However, the mature human semantic system is superior to the high-

dimensional models, since it can trace its meaning representations back to grounded, embodied meaning, however indirectly for abstract words.

Intuitively, this is something like trying to explain an abstract concept like "love" to a child by using concrete examples of scenes or situations that are associated with love. The abstract concept is never fully grounded in external reality, but it does inherit some meaning from the more concrete concepts to which it is related. Part of the concrete words' embodied, grounded, meaning becomes attached to the abstract words which are often linked with it in usage. The grounded meaning 'propagates' up through the syntactic links of the co-occurrence meaning network, from the simplest early words to the most abstract. Thus we have chosen to call this the "propagation of grounding" problem. We argue that this melding of direct, embodied, grounded meaning with high-dimensional, word co-occurrence meaning is a vital issue in understanding conceptual development, and hence language development. We believe it is essential to resolving the disputes between embodied meaning researchers and high-dimensional meaning researchers.

In previous work (Howell and Becker, 2000, 2001; Howell, Becker, & Jankowicz, 2001) we began developing what we consider to be a promising method for modeling children's language acquisition processes using neural networks. In this work, we continue this effort, emphasizing the inclusion of pre-linguistic sensorimotor features that will ground in real-world meaning the words

that the network will learn. This is a necessary precursor to addressing the "propagation of grounding" problem itself.

Our goal is to capture with one model the process by which children learn their first words *and* their first syntax or grammar. As mentioned above, this is a period stretching from the earliest onset of the first true words (10-12 months), through the “lexical-development burst” around 20 months up to the so-called “grammar burst” around 28 months. Developing a network that attempts to model the language acquisition that is happening during this period in children is, of course, an ambitious undertaking. However, given the discussion on propagation of grounding above, this sort of developmental progression may actually be *necessary* not just for children learning language, but also for any abstract language learner such as a neural network or other computational model. A multi-stage process of constrained development may be necessary to simplify the problem and make it learnable, with each ‘stage’ providing the necessary foundation for the next, and ensuring that meaning continues to be incorporated in the process. As such, we seek to develop and extend a single model that can progress through these ‘stages’ of language acquisition, from initial lexical learning, through rapid lexical expansion, to the learning of the earliest syntax of short utterances. Developing a model that fits developmental behavioural data on child language acquisition is one way to ensure that this process is being followed. For the simulations reported here, we have adopted and extended the Simple-Recurrent Network architecture that has been shown many times to be

capable of learning simple aspects of grammar, namely basic syntax (e.g. Elman, 1990, 1993; Howell & Becker, 2001). Furthermore, SRN's have been shown to be able to produce similar results to high-dimensional meaning models. Burgess and Lund (2000) point out that their HAL method using their smallest text window produces similar results in word meaning clustering to an Elman SRN. Also, they state that the SRN is somewhat more sensitive to grammatical nuances. SRN's may be able to model the acquisition of meaning *and* grammar, unlike the high-dimensional approaches.

The present emphasis of our model is on the inclusion of sensorimotor knowledge of concepts or words (for clarity, in what follows we use the term “concept” to mean the mental representation of a thing or action, and the term “word” to mean merely the linguistic symbol that represents it). This pre-linguistic sensorimotor knowledge (following Lakoff, 1987) is represented by a set of features for each word presented to the network, features that attempt to capture perceptual and motor aspects of a concept, such as “size”, or “hardness”, or “has feathers”. If a word that the network experiences is accompanied by a set of values or ratings on these feature dimensions, then the network should be able to do more than just manipulate the abstract linguistic symbol of the concept (the word itself). Like a child learning the first words, it should then have some access to the *meaning* of the concept. The network’s understanding would be grounded in embodied meaning, at least at the somewhat abstracted level available to a model without any actual sensory abilities of its own.

Unlike most existing language models that employ semantic features (e.g. Hinton & Shallice, 1991, McRae, de Sa, & Seidenberg, 1997) our sensorimotor feature set has been designed to be pre-linguistic in nature. That is, features that derive from associative knowledge about which words occur together or other language-related associations are excluded. Only features that a preverbal child could reasonably be expected to experience directly through his or her perceptual and motor interactions with the world are included. As discussed above, while children's first words are obviously learned without any knowledge of language-related word associations, children quickly begin to incorporate linguistic associative information into the semantic meanings of concepts. Certainly, at some point words begin to acquire associative meaning not only from the sensory properties of the concept, but from the linguistic contexts in which the word has been experienced. We take the conservative stance herein of excluding any linguistic associative influences on sensorimotor meaning; the sensorimotor feature representations do not change with linguistic experience. The network is capable of learning these associations, but they do not affect the sensorimotor features directly.

Whereas most language models employ binary features, our features are real-valued (range 0-1), allowing a network to make finer discriminations than merely the binary presence or absence of a feature. For example, two similar items (for example, two cats) may be perceived, but they are not identical; one is larger. Our dimension of size would differentiate the two, with one receiving a

rating of 0.2, one of 0.3. Binary features cannot easily make such fine distinctions. Finally, inspired by the work of McRae et al. (1997) on human-generated semantic features, the feature ratings that we use are all derived empirically from human participants.

One of the advantages of the neural network model of child language development that we present below is the ability to measure word-learning performance using analogues of lexical comprehension tasks that have been used with children. Since the network learns to associate the sensorimotor features of each concept with a separate phonemic representation of the word, it is possible to examine the strength of the associative connection in either direction. Thus, given the phonemes of the word, we can measure the degree to which the network produces the appropriate sensorimotor meaning vector. This we refer to as the ‘grounding’ task, analogous to when a child is asked questions about a concept and must answer with featural information, such as “What kind of noise does a dog make?” or “Is the dog furry?” Similarly, we can also ask if, when given the meaning vector alone, the network will produce the proper word. This is an analogue to the ‘naming’ task in children, where a parent points to an object and asks “What is that?” In the network, if the completely correct answer is not produced, we can still measure how close the output was to the correct answer. For example, we can check whether the answer was a case of “cat” produced in place of ‘dog’, two concepts with a high degree of featural overlap, or whether it was a complete miss. These measures can be used singly or together to assess

lexical comprehension. In this paper, we address the grounding task, but not the naming task, although the model can account for both. However, the central aim of this paper is to investigate the contribution of the sensorimotor features to improving the model's grammar learning.

In Experiments 1 and 2, we describe the empirical collection of feature ratings for nouns and verbs respectively, and describe the results of several analyses performed to verify that they are capturing an abstract representation of the words' meanings. In Experiment 3, we describe simulations of a simple neural network model using these features with a small test corpus, to demonstrate the utility of feature grounding for language acquisition. In Experiment 4A, we use these feature ratings in a larger model with a naturalistic corpus, and examine the extent to which features improve grammar learning over a control condition. In Experiment 4B and 4C, we address the issues discovered in Experiment 4A, and try to clarify the contribution of sensorimotor features to grammar learning. It is important to note that in referring to “grammar learning” we are in fact only considering the simplest aspects of grammar, namely basic syntax or sequence learning.

2.4 Experiment 1 - Generation of Noun Sensorimotor Features

Developing a set of sensorimotor dimensions that are plausible for 8 - 28 month old infants was an important first stage of this research effort. In our previous models of lexical grounding and acquisition of grammar (Howell & Becker, 2001,

Howell, Becker, & Jankowicz, 2001) we used a more simplistic semantic feature representation of words (Hinton & Shallice, 1991) that was both artificial and confounded word's conceptual semantics with "associative semantics", the linguistic relationships between words. We needed a more child-appropriate set of semantic features.

2.4.1 Method

In order to avoid having artificial, experimenter-created semantic feature representations, we investigated the McRae et al. empirically generated feature set (e.g. McRae et al., 1997). However, of the thousands of features contained in that set, many were non-perceptual (e.g. linguistically associative), and few were common across many concepts. To obtain an appropriate set of input features for a neural network model of child language acquisition, we required a more compact, concrete set of features that are perceptual and motor in nature, and could reasonably capture purely pre-linguistic knowledge. Thus we narrowed down the McRae et al. feature list to some 200 common and widely-represented features. This list was further condensed by converting each set of polar opposites and intermediate points to a single set of 19 polar-opposite dimensions. For example, “small” and “large” became a single continuous dimension of size, ranging from small (0) to large (10), and eliminating the need for “tiny”, “medium”, “huge”, etc. The remaining 78 features which could not be unambiguously converted to a set of polar opposites were retained as a condensed list of real-valued dimensions, such as color (is_red) or texture (has_feathers),

where the real value indicated the probability of possession of that feature by that concept.

This resulting list of features was then reviewed by a developmental psychologist, for accessibility to children of the age range in question (8-28 months), and any features that were not considered developmentally appropriate were removed. For example, “age” is not reliably perceived by children beyond simply “young” or “old” (Dr. Laurel Trainor, private communication, 2001) and so was removed.

The final list of sensorimotor feature dimensions was small enough to be feasible as input for our neural network models, and broad enough to be applicable to many concepts. Given this set of feature dimensions, it was next necessary to obtain ratings of the early concepts along each feature dimension. We used a large sample of human raters to generate the featural ratings for our early words. Our raters were undergraduates at McMaster University who participated in this experiment for course credit in an introductory psychology course.

Participants were presented with the concepts and the list of feature dimensions along which to rate them on a computer screen. The display was presented via a web browser, and responses were entered by filling in response boxes on the display (See Appendix A). Participants were given detailed instructions as to how to make judgments, and which anchoring points to use in assigning numerical values. For example, in rating the size of an object, the

smallest item a child might know about might be ‘pea’ for example, while for adults it might be something microscopic like ‘virus’. Thus participants were specifically instructed to make judgments taking into account the limited frame of reference that a pre-school child would have, especially relevant for polar-opposite dimensions such as “size”. Participants entered their data as numbers between 1 and 10, which were later scaled down to the 0 - 1 range for easier presentation to neural network models.

The rating forms were administered over the Internet as web forms. The data was checked carefully for outliers. Three participants’ data were excluded due to obvious response patterns (all 0’s, all 10’s, 1-2-3’s, etc.), indicating insufficient attention given to the task. Ratings were collected for 352 noun concepts from the MacArthur Communicative Development Inventory (MCDI - Fenson et al., 2000) in 38 separate phases with approximately 10 concepts each during winter, 2002. The first two phases had 10 participants each; the rest had 5 participants each. Participants received course credit for participation so long as the data was not obviously invalid as discussed above. The resulting ratings were then averaged across participants yielding a single feature vector of size 97 for each concept, 352 in all.

Three forms of analysis were performed on these newly created feature representations, in order to demonstrate that they do capture important aspects of the meanings of the words represented: a hierarchical cluster analysis, a Kohonen

self-organizing map, and a Euclidian-distance-based categorical membership analysis.

2.4.2 Results

We analyzed the 352 averaged feature vectors in a hierarchical cluster analysis using SPSS version 11.5, to see whether our features captured our intuitive sense of word similarity. The 352 concepts clearly clustered by meaning, with subcategories merging nicely into superordinate categories (See Appendix B). Animals are separated from people, people and animals are separated from vehicles and inanimate objects, etc. Thus while the high degree of variability between participants' ratings was originally a concern, after averaging, the regularity inherent in the feature vectors is quite reassuring. To provide another view on the ratings, the ratings vectors were fed into a Self-Organizing Map (Kohonen, 1982; 1995) neural network, which sought to group the concepts topographically onto a two-dimensional space based on their feature similarity. The resulting topographic organization respects the semantic similarities between concepts, showing intuitive groupings based only on the sensorimotor features of concepts (See Figure 2.1). Note, for example, the grouping of "creatures that fly" in the top left corner, and the grouping of parts of the body in the middle-left.

A more clearly-defined measure of success is provided by the categorical analysis. We formed category centroids for each of the pre-existing categories of nouns on the MCDI form from which the words were originally drawn. This was done by taking all of the words that belonged to that category and averaging

together their feature vector. Then each and every word's feature vector was compared to the centroids of each of the 11 categories represented, and the closest match indicated into which category the word should fall. This was done both with the target word included in the centroid generation process, and with it excluded (a more conservative approach). Results are very good, at 92.8% and 88% accuracy respectively (Chance performance would be 9.1%). See Table 2.1 for details.

**Table 2.1 – Noun Category Agreement Results
Feature Vectors compared to Centroids of Categories Drawn from MCDI**

Category Number	Category Name	Inclusive Accuracy	Exclusive Accuracy
2	Animals	0.8205128	0.8205128
3	Vehicles	0.9166667	0.75
4	Toys	0.9166667	0.8333333
5	Food & Drink	1	1
6	Clothing	0.9285714	0.8928571
7	Body Parts	1	0.862069
8	Small Household Items	1	1
9	Furniture and Rooms	0.8484848	0.7878788
10	Outside Things	0.9333333	0.8666667
11	Places to Go	0.8636364	0.6363636
12	People	0.8461538	0.8461538
	Overall	0.9283668	0.8796562

2.4.3 Discussion

We believe all three analyses indicate the success of the experiment. The hierarchical clustering analysis, while vast and somewhat difficult to interpret, shows many clear separations of concepts, and consistent local clusters of meaning. The SOM representation shows clear clustering by meaning, with both fine-grained and broader similarity structures across the map. Finally, the categorical analysis provides a clear numerical measure of the goodness of fit of our features to the preexisting categorizations of these nouns, with 93% accuracy of word to category. The sensorimotor feature ratings thus capture much of the meaning of the concepts, definitely enough to be useful as inputs to our language learning model, and they certainly capture what's important for categorical reasoning.

Noun Feature SOM

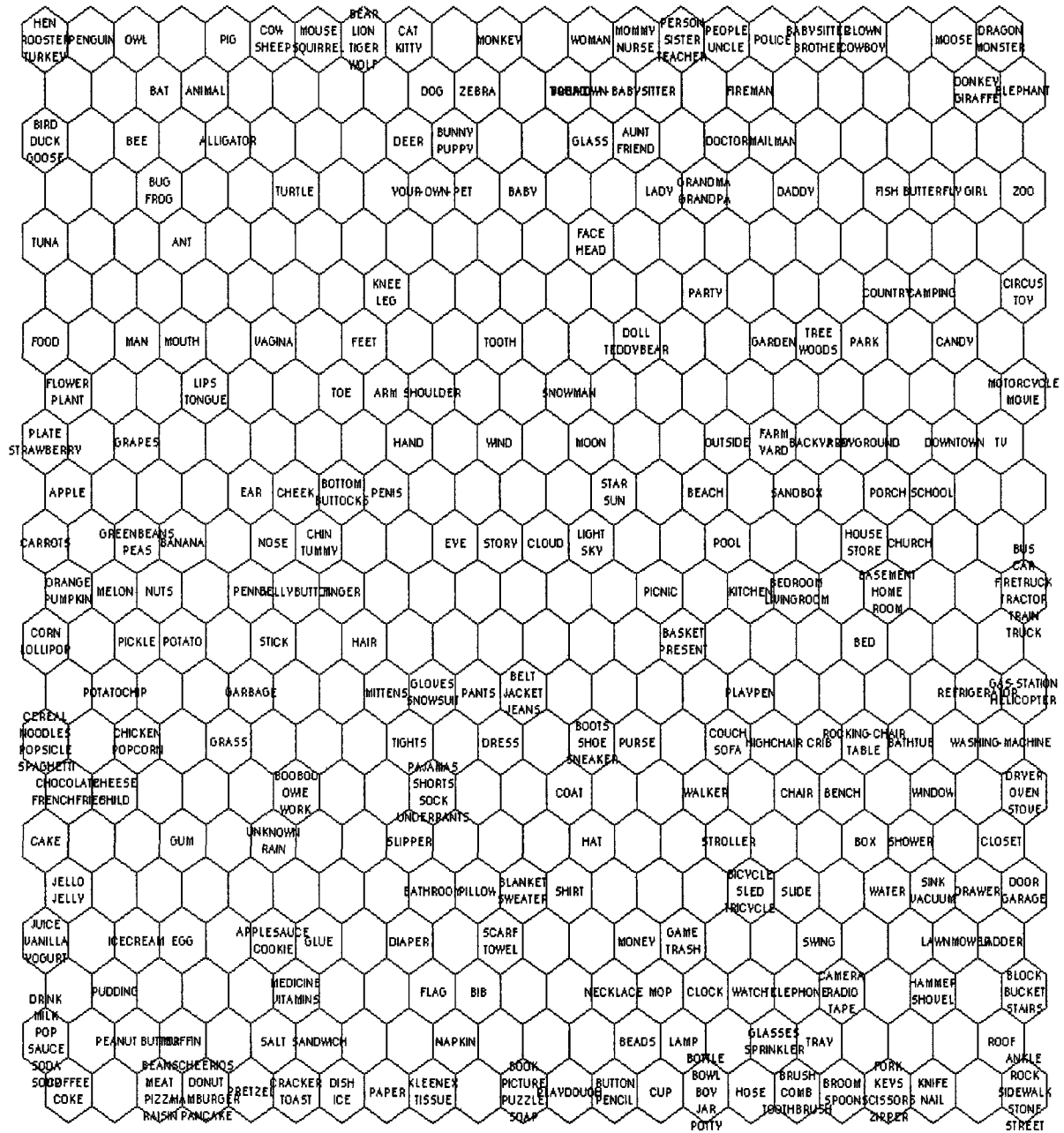


Figure 2.1: Self-organizing Feature Map of Experiment 1 Feature Vectors
 Each concept is written on the unit that responded most highly to presentation of that concept after training. Note the grouping of similar concepts on nearby units, as well as the overall topography of similarity.

2.5 Experiment 2 - Generation of Verb Sensorimotor Features

Verbs are more difficult to develop representations for than are simple, concrete nouns, so our method had to be slightly more exploratory.

2.5.1 Method

In this experiment we followed much the same methodology as for *Experiment 1*, this time for verb features. However, given that verbs correspond to events in the world rather than to objects, the nature of verb features was expected to be different from that for nouns. Also, there was no pre-existing taxonomy of verb features readily accessible in the literature, as there had been for nouns.

Therefore, our collection of verb features proceeded in two steps. First we conducted a pilot experiment in verb feature generation with human participants, and from that we created a set of verb feature dimensions to be rated in an online phase of the experiment exactly as in *Experiment 1* (See Appendix C for forms and instructions used in both the pilot and the experiment). The pilot experiment was conducted with 12 undergraduate participants at McMaster University.

Participants completed a feature generation form for some of the earliest (MCDI - Fenson et al., 2000), and most prototypical (Goldberg, 1999) verbs, with the objective being not complete characterization of any given verb but rather the creative generation of a set of feature dimensions which might be common to many verbs.

While fully half of the features generated were unusable due to contamination by functional relationships with corresponding nouns, associational relationships, etc., there were sufficiently many perceptual and motor features identified to allow us to create an initial set of feature dimensions. From this beginning, we were able to fill in missing complements of existing dimensions. For example, several participants focused on limb movement to define verbs, which is in line with some existing models of verb definition in computer science (see for example Bailey, Feldman, Narayanan, & Lakoff, 1997). From this and considerations of bodily motion and proprioceptive constraints in humans we were able to generate a large primary set of joint-motion dimensions. We also included some other features that had been identified by pilot participants, which brought the total to 84 feature dimensions (See Appendix C)

A second group of participants participated in the rating phase of the verb experiment. As in Experiment 1, they rated each verb on the list with a value between 0 and 10 on the 84 feature dimensions. We then converted these ratings to the 0-1 range, which became the feature representations for verbs used in the Experiments below. We analyzed the results of the experiment (the feature ratings) in the same three ways as in Experiment 1: A hierarchical cluster analysis (see Appendix D), a self-organizing map (Kohonen, 1982; 1995), and a Euclidian-distance-based categorical membership analysis. The categories used in the latter analysis were drawn from Levin, 1993, and grouped together into superordinate categories with the assistance of linguists Anna Dolinina of

McMaster University, and Sylvia Gennari, of the University of Wisconsin – Madison. Nine categories were used, as can be seen in Table 2.2 below. Only the inclusive methodology was used to create the category centroids, based on the results from experiment 1.

2.5.2 Results

Overall, the verbs do not perform as well as the nouns. Still, as can be seen from the SOM, similar verbs do group together in space (See Figure 2.2). Note the grouping of "tongue-verbs" in the top left, and movement verbs in the bottom right, for example. Major trends in the cluster analysis for verbs are less clear than for nouns, although the analysis does find many intuitively reasonable groupings, such as take, bring, push, put, and move, for example. (See Appendix D).

Finally, the categorical agreement analysis, while not as clear as that for the nouns shown previously, still demonstrates a 70% overall accuracy of the target words to their correct category. Chance performance would be 11.1%. Categorical performance by category is shown in table 2.2 below.

Table 2.2 - Verb Category Agreement Results
Feature Vectors compared to Centroids of Categories Drawn from MCDI

Verb Category	Percentage Correct
Body-Movements	60%
Motion	83%
Creation/Destruction	64%
Food-Related	83%
Possession & Relocation	78%
Change of State	55%
Statives	78%
Communicative	67%
Perception	75%
Overall	70%

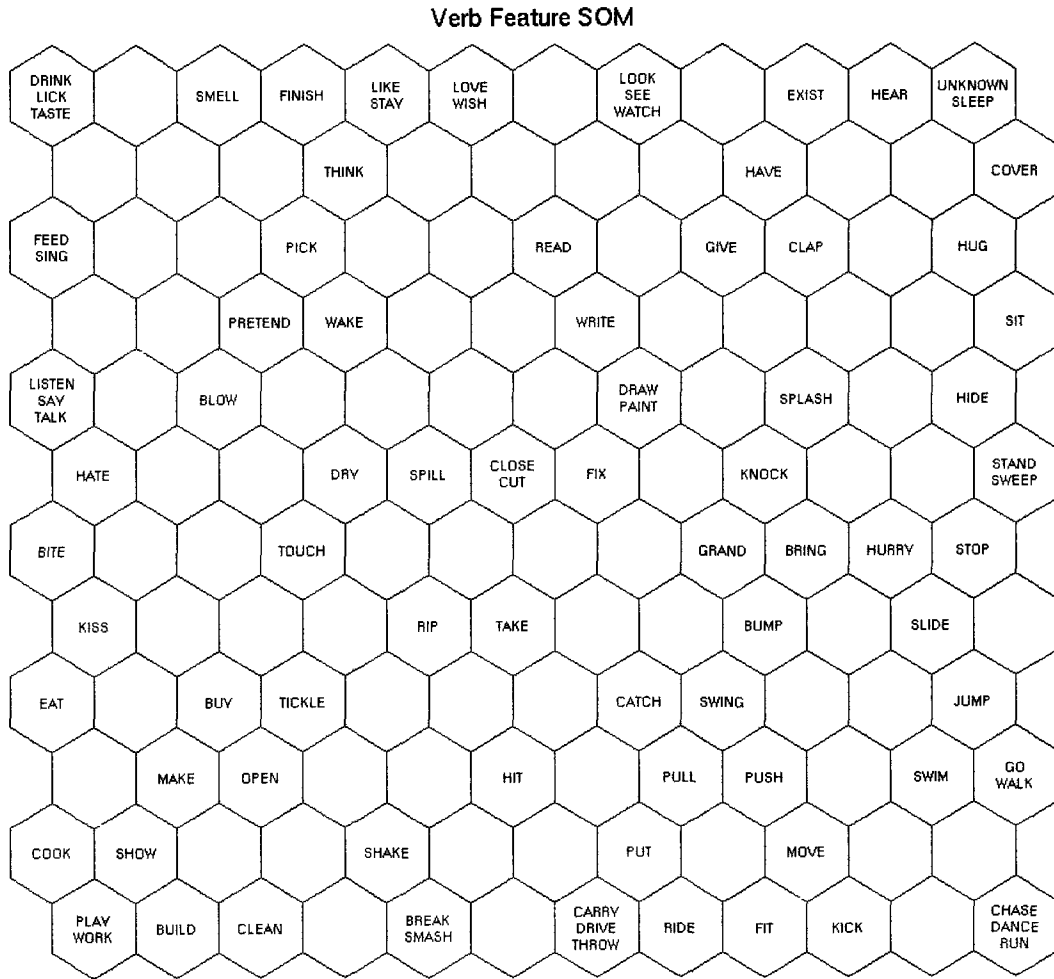


Figure 2.2: Self-Organizing Map of the Verb Feature Ratings

Note the grouping together of words involving similar motor activities such as drink/lick/taste and listen/say/talk as well as modes of locomotion such as slide/jump/go/walk/hurry

2.5.3 Discussion

The somewhat weaker clustering of our verb features is consistent with the results of Vinson and Vigliocco (2002), who also show that verbs generally do not cluster very well. Their verb features were also human-generated, but they placed

no developmental or sensorimotor restrictions on the form of those features as we did in this experiment. Nonetheless, it seems that verbs, or pre-linguistic verb concepts, simply do not share as tight a similarity space as nouns do, although the fact that there were fewer verbs in Experiment 2 than there were nouns in Experiment 1 may have an effect, as there is less opportunity for featural similarity to become apparent. However, our features are still capturing important aspects of the meanings of verbs, as can be seen qualitatively in the hierarchical cluster analysis and SOM, and quantitatively in the Category Agreement analysis. An agreement rating of 70% is more than sufficient for us to wish to use these features in further experiments.

In experiments 3 and 4, we investigated the contributions of sensorimotor feature grounding to language learning in a series of neural network simulations. In experiment 3 we used a small training corpus and only the noun features. In experiment 4 (parts A, B, and C) we used a larger corpus and both noun and verb features, and made the language learning task progressively more difficult.

2.6 Experiment 3 - A Model of Grounded Lexical Acquisition

In this first simulation experiment, using only our noun feature set (the verbs were not yet complete at the time of this experiment), we used a small corpus and simple vocabulary to test whether extending the SRN architecture to include feature input would assist in grammar learning (basic syntax learning), as evidenced by performance on the word prediction task. In previous work (Howell

& Becker, 2001) we determined that adding an artificial set of semantic features improved word prediction dramatically (18.5% to 37.1%). However, in that experiment the word representations were localist (a series of zeroes with a single 1), while the feature representations were binary distributed codes (a sequence of zeros and ones). It was impossible to determine how much of the improvement in word prediction was due to the simple increase in the information content of the combined input representation, rather than the inter-word similarity structure inherent in the semantic features.

In contrast, in this experiment, the word representation is a very long (140 elements) distributed representation of phonemic features, described below. The feature representations are smaller, 97-element vectors of real-valued features. A significant benefit is thus expected to be due to the statistical regularities inherent in the sensorimotor feature information, beyond a simple increase in the information content due to the use of distributed representations.

2.6.1 Method

We used the Simple-Recurrent Network (SRN) architecture, performing the word prediction task: predicting from the current input word what the next word would be. Each word was encoded as a set of up to 10 phonemes using 140 input units. The 140-element word inputs represented 10 phonemic slots each of 14 phonemic feature bits, without representation of word boundaries. The Carnegie Mellon University (CMU) machine-readable phonetic transcription system and pronouncing dictionary was used to generate our phonetic representations of

words (available at: <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>). Each phoneme was uniquely mapped to a set of 14 bits (See Figure 2.3), representing articulatory dimensions of the phonemes. Words shorter than 10 phonemes had their rightmost slots padded with 14 zeros, while longer words were truncated.

"AA"	"1,0,0,0,0,1,0,0,0,1,0,0,0,0"	"L"	"0,0,1,0,1,0,0,0,0,0,1,0,0,0"
"AE"	"1,0,0,0,1,0,0,0,0,0,0,0,1,0"	"M"	"0,0,0,1,0,1,0,0,0,1,0,0,0,0"
"AH"	"1,0,0,0,0,1,0,0,0,0,1,0,0,0"	"N"	"0,0,0,1,0,1,0,0,0,0,1,0,0,0"
"AO"	"1,0,0,0,0,0,1,0,0,0,0,1,0,0"	"NG"	"0,0,0,1,0,1,0,0,0,0,0,1,0,0"
"AW"	"0,1,0,0,0,0,0,0,0,0,0,1,0,0"	"OW"	"0,1,0,0,0,0,0,0,0,0,0,0,0,1"
"AY"	"0,1,0,0,0,0,0,0,0,1,0,0,0,0"	"OY"	"0,1,0,0,0,0,0,0,0,0,1,0,0,0"
"B"	"0,0,0,1,1,0,0,0,0,0,1,0,0,0"	"P"	"0,0,0,1,1,0,0,0,0,1,1,0,0,0"
"CH"	"0,0,0,1,0,0,0,0,1,0,1,0,0,0"	"R"	"0,0,1,0,0,1,0,0,0,1,0,0,0,0"
"D"	"0,0,0,1,1,0,0,0,0,0,0,1,0,0"	"S"	"0,0,0,1,0,0,0,1,0,1,0,0,1,0"
"DH"	"0,0,0,1,0,0,0,1,0,0,0,1,0,0"	"SH"	"0,0,0,1,0,0,0,1,0,1,0,0,0,1"
"EH"	"1,0,0,0,1,0,0,0,0,0,0,1,0,0"	"T"	"0,0,0,1,1,0,0,0,0,1,0,1,0,0"
"ER"	"1,0,0,0,0,1,0,0,0,0,0,1,0,0"	"TH"	"0,0,0,1,0,0,0,1,0,1,0,1,0,0"
"EY"	"0,1,0,0,0,0,0,0,0,0,0,0,1,0"	"UH"	"1,0,0,0,0,0,1,0,0,0,1,0,0,0"
"F"	"0,0,0,1,0,0,0,1,0,1,1,0,0,0"	"UW"	"1,0,0,0,0,0,1,0,0,1,0,0,0,0"
"G"	"0,0,0,1,1,0,0,0,0,0,0,0,1,0"	"V"	"0,0,0,1,0,0,0,1,0,0,1,0,0,0"
"HH"	"0,0,0,1,0,0,1,0,0,1,0,0,0,0"	"W"	"0,0,1,0,1,0,0,0,0,1,0,0,0,0"
"IH"	"1,0,0,0,1,0,0,0,0,0,1,0,0,0"	"Y"	"0,0,1,0,0,1,0,0,0,0,1,0,0,0"
"IY"	"1,0,0,0,1,0,0,0,0,1,0,0,0,0"	"Z"	"0,0,0,1,0,0,0,1,0,0,0,0,1,0"
"JH"	"0,0,0,1,0,0,0,0,1,1,0,0,0,0"	"ZH"	"0,0,0,1,0,0,0,1,0,0,0,0,0,1"
"K"	"0,0,0,1,1,0,0,0,0,1,0,0,1,0"	Pause	"0,0,0,0,0,0,0,0,0,0,0,0,0,0"

Figure 2.3: CMU Phonemes and their compressed 14-bit Representations. The bits represent articulatory features such as voiced/unvoiced, place and manner of articulation, etc.

In addition to the phonemic word features, a set of noun features was also input to the network, simulating the co-occurrence of sensorimotor information and phonological information in the child's environment. A single hidden layer and a single context layer (both of 30 units) were used (See Fig. 2.4). The output

layer was the same size and used the same representation as the Word Phonemic representation. A continuous stream of sentences was input to the network, one word at a time. The task of the network was to predict the phonemic features of the next word in the input stream. Training used the back-propagation of error learning algorithm (Rumelhart, Hinton, & Williams, 1986).

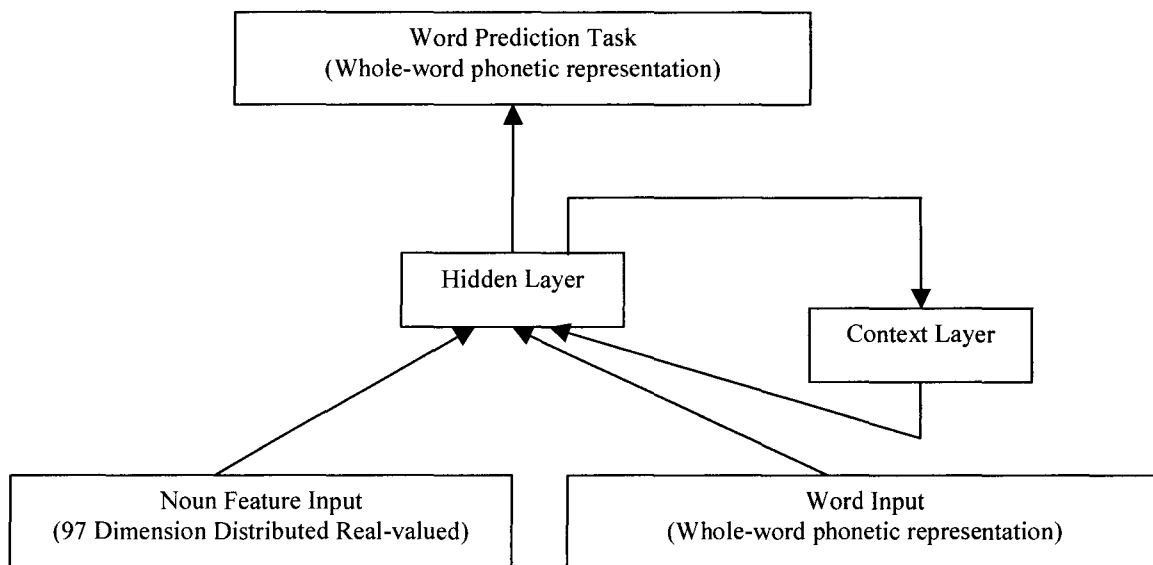


Figure 2.4: The network used in Experiment 3. Note the use of two different inputs per word, one containing the phonemic representation of the word, the other the real-valued noun features of the word.

Corpora and Training Schedule

We used a very small 60 word corpus of text consisting of three word subject-verb-object (SVO) sentences with a vocabulary of only 13 words. The text corpus

was presented to the network as a continuous stream of words with no breaks between sentences and the network was trained to predict the next word in the stream. It was hypothesized that word prediction, a measure of syntactic learning which is one part of grammar (Elman, 1990), would improve with sensorimotor grounding of nouns. While we expect to find more support for our hypothesis in larger corpora, using a smaller corpus allowed many runs of the network to be processed in several experimental conditions. Semantic relations holding between subjects, verbs, and objects in the text were not random but obeyed physical constraints (only MAN can HOLD something, whereas DOG or CAT cannot).

Three variants of the network were run, to simulate one experimental condition and two control conditions. The Experimental condition used the full network as described above. The SRN-only condition used a similar SRN, with an input layer containing word (phonemic) features but lacking the noun sensorimotor features. The Random Control network used the same architecture as the Experimental condition, but replaced the human-generated (and meaningful) noun features with randomized permutations of that same set of features. This condition is intended to control for sheer number of connections and input vector magnitudes, aspects in which the Experimental and the SRN-only conditions differ dramatically. The randomization was performed by iteratively swapping the value at each position on the 97 element vector with that of another random position. When all words' representations had been randomized, each word's entire randomized feature representation was then

swapped with another word's representation. This manipulation minimizes any featural similarity between related words.

Six networks were run in each of the three conditions, for a total of 18 networks. Each network was run for 500 epochs using the SRNEngine simulation package (Howell & Becker, submitted). The networks' grammatical accuracy, as measured by word prediction accuracy, was recorded at the output layers. The network used a Euclidian distance based output rule to convert its output activations to a word label; thus every time step resulted in a discrete word prediction, as opposed to any sort of phonological blend state. Comparison of this word to the target word produced the accuracy measure.

In preliminary simulations, randomly selected representative networks from each condition were run up to 8000 epochs, and it was found that no substantial alteration in the network behavior occurred beyond 500 epochs.

2.6.2 Results

Accuracy curves for the three conditions are shown in Figure 2.5. Only one difference is clear from this Experiment. The learning trajectory of the SRN-only condition is clearly different from the other two (See Figure 5). The accuracy in the SRN-only condition stays low much longer, then catches up to the accuracy of the other two conditions around the half-way point of training (250 epochs), but seems to have reached an asymptote around that point, while the accuracy for the other two conditions continues to rise. The Experimental and Random conditions are indistinguishable from one another at all points during the training.

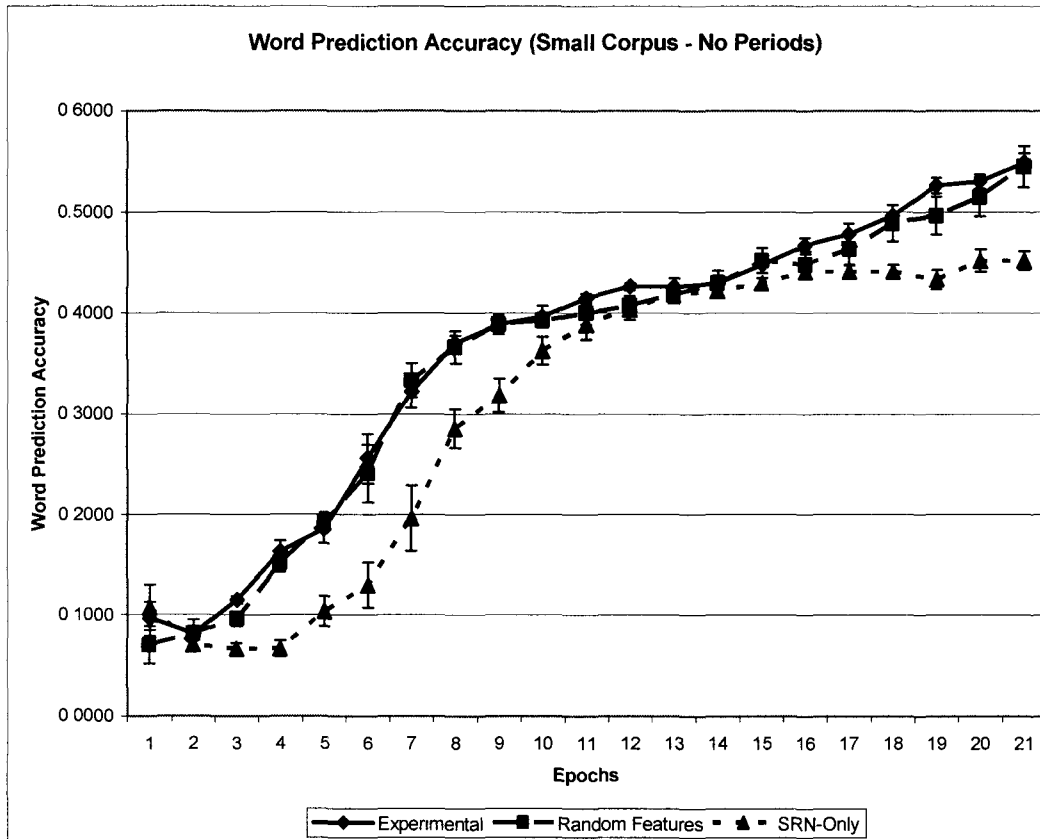


Figure 2.5: Graph of mean prediction accuracy of Experimental and Control Networks averaged across 6 runs starting from random initial weights. Error Bars are standard errors.

2.6.3 Discussion

Augmenting the input with additional features has a significant effect on word prediction (grammar) performance, but it appears to be solely due to the extra input information, rather than the structure present in the sensorimotor feature

inputs. This extra information could be helping the network to learn more distinct representations for individual words in the hidden layer. One way around this would be to use the features as output targets instead of inputs. This seems reasonable in light of our assumption that children have already formed internal representations of the features of a concept by time of initial language learning. By using these features as output targets, rather than inputs, the network is forced to focus on the mapping of words to meanings and therefore learning the associations between words and existing concepts, as well as how those words predict each other in the speech stream.

However, why was there no difference between networks with real features and randomized features? While both provide extra bits of information at input, only the real features convey information about interword similarity. The fact that some concepts are perceived by people as similar is at least partially due to their similar perceptual qualities and leads to similar sensorimotor feature ratings for them, and hence a high degree of overlap on the 97 dimensions defining them. Thus we predicted that these features would tend to be learned as being linked together due to their pattern of covariation (e.g. Rogers & McClelland, 2003), thereby facilitating the learning of word "meanings". However, the size of this covariance effect on learning will depend on how many of these concepts there are, and how often they are used. In the small training corpus used in this experiment, there may have been insufficient covariation of feature representations for the network to learn to take advantage of this structure.

In the remaining experiments, we therefore used a large, naturalistic corpus of child-parent speech, one in which many of the words (but not all) had corresponding sensorimotor feature representations from our MCDI list. We also modified the architecture slightly in ways that should both improve performance and make it possible to perform additional analyses on the results later.

2.7 Experiment 4A – A Large Corpus Model of Lexical Acquisition

In Experiment 4A, we modified the Simple-Recurrent Network (SRN) architecture to perform three separate tasks simultaneously, in three separate pools of output units (See Figure 2.6) A small common hidden layer and context layer of 30 units each were used, to force the network to develop an integrated internal representation common to the three tasks. A single input layer presented whole-word phonetic representations of words (as in Experiment 3), in serial order through the corpus.

The Linguistic Predictor output layer performed the same word-prediction task as in Experiment 3. At each time step, its task was to predict the phonemic representation of the input word at the next time step. The task for the remaining outputs was to produce the sensorimotor features of the current word. The Noun Features layer had output targets that represented the sensorimotor features for the current word, as created in Experiment 1. The Verb Features layer had output targets that represented the sensorimotor features for the current word, as created

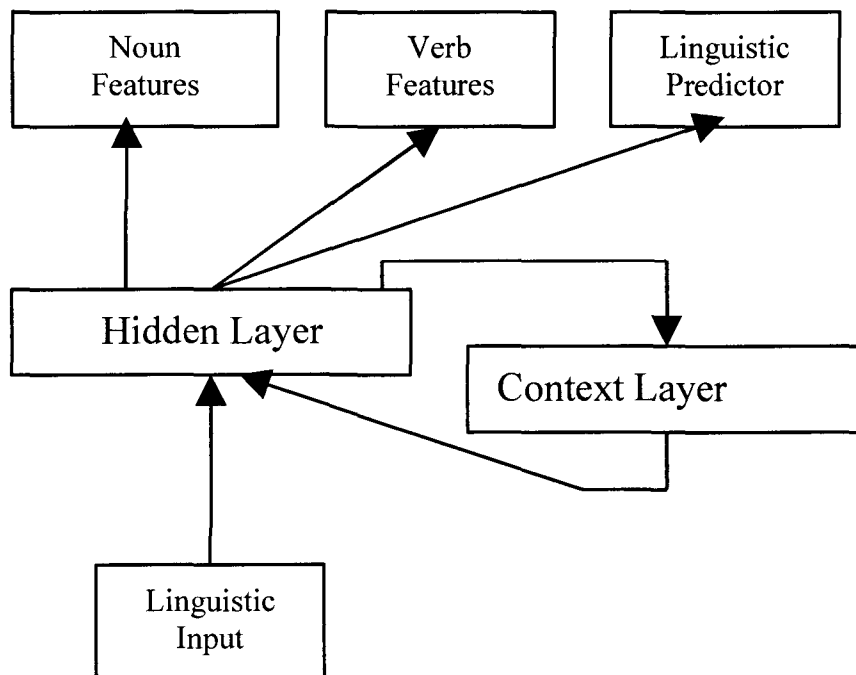


Figure 2.6: Modified SRN architecture, including standard SRN hidden layer and context layer, standard linguistic (word) prediction output, and novel noun feature output and verb feature output. The linguistic input is a whole-word phonetic representation of up to 10 phonemes. The Noun and Verb feature targets are meant to be an abstract representation of pre-linguistic sensory and motor-affordance semantics.

in Experiment 2. When the current input was not a noun or a verb (respectively), a vector input of all 0's was presented at that layer, and no backpropagation of error was performed for that layer.

Employing the sensorimotor features as output targets was designed to eliminate the confound of representational richness involved in using additional inputs, as discussed in Experiment 3 above. Also, the fact that the network is

producing sensorimotor noun and verb features at the output means that we can examine the ability of the network to generate the correct features for any given word. This gives us a measure of vocabulary acquisition both during learning and when testing generalization performance on novel words presented at the input. See the general discussion for further comments on this.

2.7.1 Method

We used a large (10,742 word) selection of speech drawn from the ChildDES database (McWhinney, 2000) transcribed from mother-child playtime interactions. This corpus was created by appending all of the Bates FREE20 data sets (Bates, Bretherton, & Snyder, 1988; Carlson-Luden, 1979) from the ChildDES database into a single body of text. Any pauses, periods, etc. in the original corpus were replaced with a generic pause marker, intended simply to assist in defining clause boundaries in this corpus of more complex sentences.

Since simulations of networks trained on these large corpora take a long time to run, we eliminated the SRN-only control condition (supported by the clear difference between both the experimental and random conditions and the SRN-only condition from Experiment 3) and focused on examining the difference in word prediction performance between the Experimental Condition (meaningful features) and Random Control Condition. 10 networks were run in each condition, for a total of 20. Each network was run for 200 epochs using the SRN*Engine* simulation package (Howell & Becker, submitted). Rather than running these larger networks to asymptotic performance, we simply ran them for

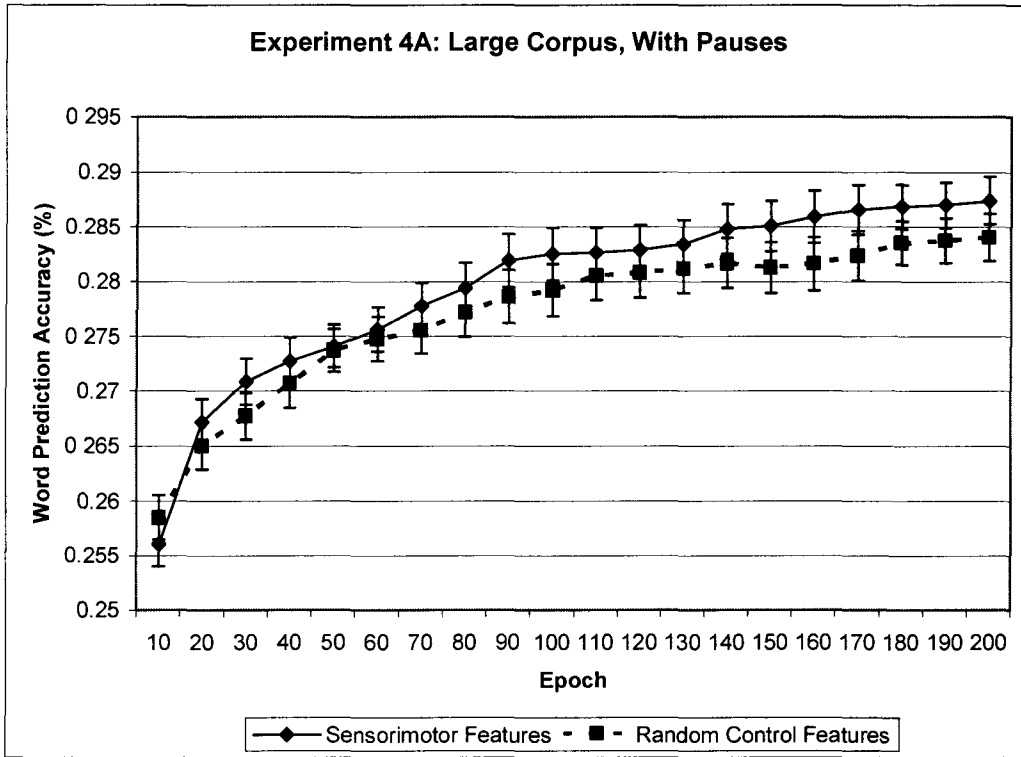


Figure 2.7: Mean grammatical prediction performance for a large naturalistic a fixed period (200 Epochs) within which grammatical prediction began to approach reasonable levels of performance. The networks' word prediction (grammatical) accuracy was recorded over the course of training. The network used a Euclidian distance based output rule to convert its output activations to a word label; thus every time step resulted in a discrete word prediction, as opposed to any sort of phonological blend state. Comparison of this word to the target word produced the accuracy measure.

We also analyzed the networks' ability to produce the correct sensorimotor features at the output layers for each phonetically-presented input word, including any relationships to the frequency of the word.

2.7.2 Results

The grammatical prediction performance of the networks in the Experimental Condition tends to be higher than that of the Random Condition, but the effect size is quite small (1.2% difference) with only a trend towards significance (t-test at Epoch 200, $p = 0.161$, $df = 18$) (See Figure 2.7).

We also analyzed the performance of the network on the sensorimotor-feature grounding task. That is, how accurate is the network at producing the sensorimotor features when it experiences the phonemes of the word? We ran one of the Experimental condition networks above (chosen at random) for a total of 500 epochs. At this point, noun and verb grounding were quite good, as can be seen from Table 1 below, although based on past experience accuracy could rise as high as 90% with further training. The network did not learn to produce sensorimotor features for any noun that occurred fewer than 4 times in the corpus, nor for any verb that occurred fewer than 5 times. Feature production accuracy for both nouns and verbs was correlated highly with frequency (nouns, $r = 0.7353$, verbs, $r = 0.6828$). Word Prediction accuracy was also highly correlated with frequency of the target word ($r = 0.6266$) (See Table 2.3).

Table 2.3: Output Accuracy from sample network at 500 epochs, during training

	Noun Features Encoding	Verb Features Encoding	Word Prediction
Accuracy	65.535%	75.251%	28.030%
Number of Items	60 grounded nouns in this corpus	49 grounded verbs in this corpus	529 words in this corpus total

2.7.3 Discussion

The network is clearly able to learn to produce sensorimotor features at output that correspond to the meaning of the word presented phonetically at input. This is not critical to our present analysis, and so will not be considered further, although it will prove important in further studies of "propagation of grounding" - the ability of the network to learn to produce features for novel words that have never had features. (see Howell, Becker, & Jankowicz, 2001).

We expected that the modified architecture and larger corpora would be sufficient to demonstrate the advantage of including meaningful features in the word learning and word prediction process. While the results show a trend in this direction, the difference is not large enough to be truly convincing.

With the richness and complexity of the corpora seeming to be sufficient, what other factors could have swamped our ability to detect an advantage for sensorimotor features in word prediction? We hypothesized that it is the numerous pauses that occur in natural speech and are included in this corpora that obscure the advantage for meaningful features over random features. The pauses

provide extra information that may make feature information less important (as well as artificially inflating the accuracy figures for input-symbol prediction by being very frequent). For example, after a sentence boundary, the next word will most likely be an animate noun (an agent). In simple (SVO) sentences, after an object noun, a sentence boundary is likely to follow. In Experiment 4B we repeated the methodology of Experiment 4A on the same corpus, but after the removal of the markers at clause and sentence boundaries.

2.8 Experiment 4B – Pause Markers Removed

Removal of the pauses from the training corpus reduced the size of the corpus to 8,328 words. However, it also increased the difficulty of the word prediction task. Periods or other sentence boundaries serve as indicators that one thought or message is complete and another is beginning, and this information affects the predictions that can be made. Thus we expect to see a larger difference between the Experimental Condition and the Random Control Condition in this experiment than in Experiment 4A.

2.8.1 Method

The network had exactly the same architecture as that used in Experiment 4A (See Figure 2.6). The control condition and training parameters were likewise identical, and the training corpus was identical except for the removal of all pause markers. We ran ten networks in each condition.

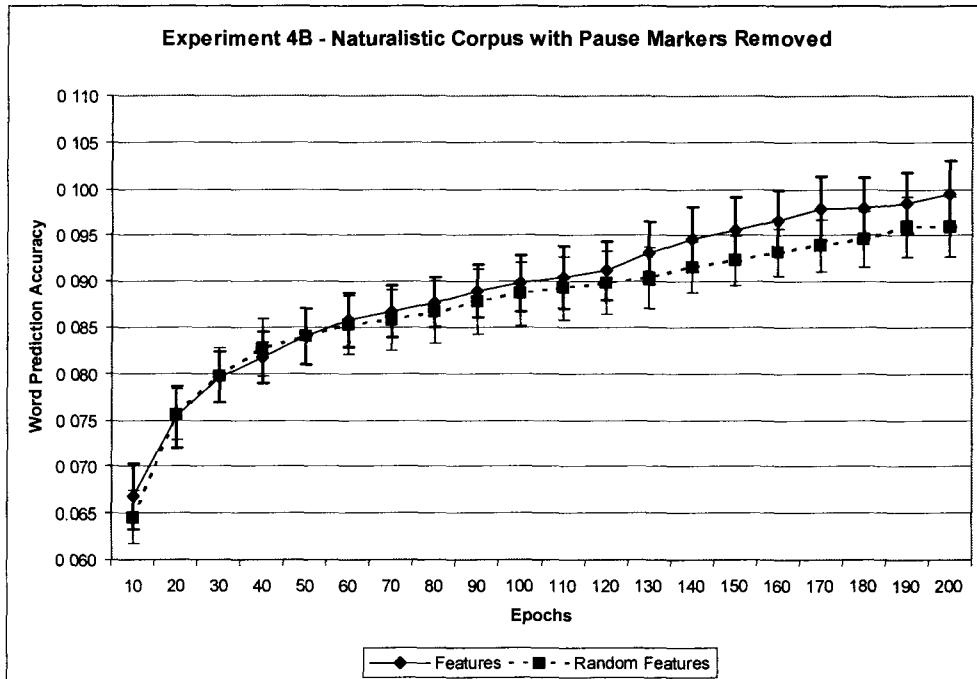


Figure 2.8: Mean grammatical prediction performance for a large naturalistic corpus (8328 words) which *excludes* pauses/periods. Number of networks in each condition is 10. Error bars indicate standard error.

2.8.2 Results

The results are slightly clearer than in Experiment 4A. The mean performance of networks in the Experimental condition is higher than that of the Random condition (See Figure 2.8), but the difference is not significant (t-test at Epoch 200, $p = 0.236$, $df = 18$). The gap is wider in the later epochs than in the earlier ones, implying a possible divergence in the two conditions with further training. The difference between the two conditions is approximately 3.6% at the end of training.

2.8.3 Discussion

Removing the pauses in the training corpus made the learning task more difficult, resulting in a greater reliance by the network on the structured information in the sensorimotor features. However, while the difference between the two conditions has increased by a factor of 3 from Experiment 4A to Experiment 4B, it is still not very dramatic. However, in light of the relative success of Experiment 4B, the inability of Experiment 4A to demonstrate a clear advantage for sensorimotor features might now be viewed as a sort of ceiling effect, due to the powerful ability of standard SRN's to learn the word prediction task. However, as we are making the task more difficult, we are seeing a larger advantage for the sensorimotor features materialize. This implies that if we make the task yet more difficult for the network, we may see an even larger advantage of using sensorimotor features. In our final experiment, Experiment 4C, we do just that by reducing the size of the network's hidden layer.

2.9 Experiment 4C – Reduced Hidden Layer

Experiment 4C was identical to Experiment 4B except that the size of the hidden layer and context layer was reduced. By limiting the network's resources for forming internal representations, we intended to make the task harder. This may be a better analogue to children's early learning, when attentional and other resources are immature and very limited. We predicted that this would magnify the advantage for the Experimental Condition over the Control Condition.

2.9.1 Method

The network had the same architecture and training corpus as that used in Experiment 4B (See Figure 2.6). The only exception is the size of the hidden layer and context layer, which have been reduced from 30 units to 10 units. In addition, we used two different grammatical error criteria in our analysis. The first, the ‘exact match’ criterion, is quite conservative, as already mentioned. The predicted word has to be the exact target word expected. The second is a more generous “categorical match” criterion, where the predicted word only had to be in the same grammatical category as the target word. All words in the corpus were divided into 1 of 12 grammatical categories, which included: adjective, adverb, conjunction, determiner, other, noun, possessive, preposition, pronoun, meaningless, and verb. The inclusion of this measure is to test the possibility that our exact match criterion is too conservative to have enough power to detect a difference between the Experimental and Control Conditions.

Also, for the first time we analyzed the data from the other two output layers, the Noun Feature encoding accuracy and the Verb Feature encoding accuracy, to see if there was any difference in the accuracy between Experimental and Control conditions. To speed processing of the simulations in this experiment, accuracy data was logged at only 50 epoch intervals instead of 20 epoch intervals. We ran 10 networks in each condition.

2.9.2 Results

The results show a larger gap in word prediction accuracy between the two conditions than in Experiment 4A or 4B (See Figure 2.9). Using the exact match error criterion, the difference between the two conditions at epoch 200 is significant (t-test at Epoch 200, $p = 0.017$, $df = 18$). The percentage difference between the two conditions is 13%. Also, the gap is wider in the later epochs than in the earlier ones, implying a possible divergence in the two conditions with further training. Indeed, a repeated measures ANOVA on the data from epochs 150 and 200 yields a significant interaction effect of training by condition ($p = 0.034$, $df = 18$).

Using the categorical match error criterion, the mean accuracy of the Experimental group rises to 0.185, the control group to 0.171. The size of the difference is 0.014, or an 8.2% difference between the two groups. The difference under this error criterion is also significant, (t-test at Epoch 200, $p = 0.035$, $df = 18$). Due to the processing demands of calculating this error criterion, it was only calculated for the final epoch of training.

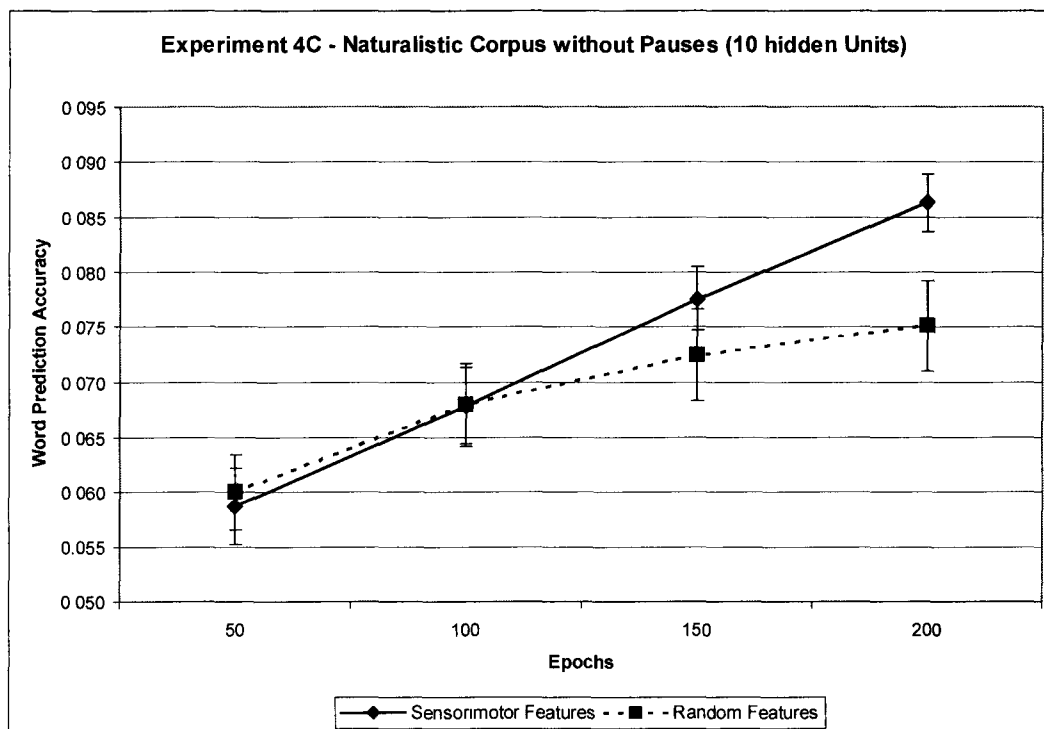


Figure 2.9: Mean grammatical prediction performance for a large corpus (8328 words) which excludes pauses/periods, with a reduced hidden/context layer (size 10). The number of networks in each condition is 10. Error bars indicate standard error.

Noun encoding accuracy is also significantly different between the two conditions at the 200 epochs (t-test at 200 epochs, $p = 0.0344$, $df = 18$), with the sensorimotor feature condition being superior to the random features condition (see Figure 2.10). The difference in verb encoding accuracy was not significant, however ($p = 0.120$, $df = 18$).

2.9.3 Discussion

Increasing the difficulty of the task for the network has magnified the advantage of the sensorimotor features in the Experimental condition. Using the exact match error criterion, a significant difference of 13% in word prediction accuracy (a simple measure of grammar learning abilities) is evident at the final point of training, and the difference between the two conditions' average accuracy curves is increasing with the amount of training.

The categorical match error criterion produces a similar result (8.2% difference) at the final epoch of training, and is also significant. However, given that using the exact match measure is much easier to calculate than the categorical match measure, and does not involve issues such as the choice of the right level of grammatical categories to use, etc., it seems appropriate that we have been using the more conservative exact match grammatical accuracy measure throughout these experiments. Still it is interesting to see that the results do not depend on the choice of grammatical accuracy error criterion.

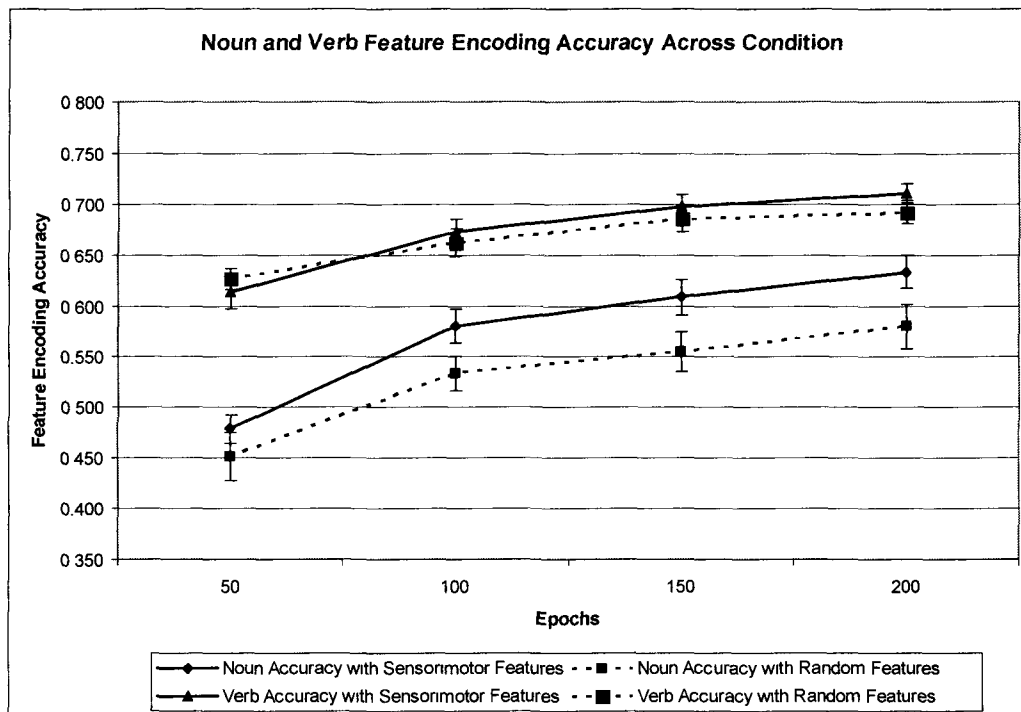


Figure 2.10: Noun and Verb feature encoding accuracy from Experiment 4C. These two output layers were performing a recoding from the phonetic features of a word to the semantic features of a word. The noun feature encoding is significantly different across the two conditions, as measured by t-test at the terminal point, but the verb features are not. The number of networks in each condition is 12. Error bars indicate standard error.

Also, the ability of the network to re-code the phonetically presented word inputs as semantic features is significantly different between the two conditions, at least for nouns. The fact that this effect was not significant for verbs may be due to the fact that fewer of them were grounded in our training corpus (60 nouns versus 49 verbs) and the fact that the network has more exposure to nouns (since most simple sentences contain only one verb, but several nouns).

These results demonstrate the ability of sensorimotor features to improve both the grammar learning process, and the process of mapping words to conceptual features.

2.10 General Discussion and Conclusions

From the preceding sequence of experiments we can see that while the influence of sensorimotor features on language acquisition is expected to be large, trying to explore it using only the word prediction task is tricky. As other work (Elman, 1990, 1993) has shown, a normal SRN is already quite good at learning sequences of words; demonstrating even better performance with the inclusion of sensorimotor features is therefore not easy. Also, it is important to note in all of these experiments that even with this child-directed corpus, only 60 nouns and 49 verbs were actually represented in our vocabulary of 352 grounded early nouns and 90 grounded early verbs. Had more of the corpora's vocabulary of 529 words been grounded, the effects of including the sensorimotor features might have been larger and easier to detect. However, in several simulation experiments, and most strongly in Experiment 4C, we were able to demonstrate the advantage of using sensorimotor features in the word prediction task, with features yielding up to a 13% improvement over the control condition on this measure of grammatical performance for a language acquisition network. We also showed that the networks in the Sensorimotor Feature condition were better able to learn to encode the relationships between phonetically input words and their

corresponding conceptual features. Both of these results demonstrate that having sensory and motor knowledge of objects and events in the environment is a significant advantage when trying to acquire language for the first time, for networks and presumably for children.

There are numerous other advantages in using networks with sensorimotor features. One is that we could simulate an analogue of word comprehension, by assessing whether a word label (the sound of a word) has been linked to a meaning representation. Data was briefly presented herein to demonstrate that these networks perform quite well at learning by this measure. Another advantage relates to word sense disambiguation: these sort of meaning representations may be used to disambiguate multiple senses of a word encountered in text, through the operation of feature prediction in concert with word prediction. That is, if the network is predicting the word "bank" next, by examining the features it is predicting at the same time we might be able to tell whether it means to output "a place to store money" or "the edge of a river".

Once we have a set of explicitly grounded sensorimotor features for the earliest words, a question then naturally arises: do we need to derive featural ratings for every concept that the network is exposed to? Fortunately, that should not be necessary. As discussed previously, evidence indicates that only the child's earliest words are fully grounded in sensory experience (Gillette et al., 1999); in fact it is the early words' very imageability and accessibility to observation that leads them to *be* the first words generally learned by children.

As lexical learning progresses, less and less imageable (i.e. more abstract) words are experienced and learned. Also, the learner is exposed to novel words in speech or text that are not directly grounded in immediate sensory experience. Both of these sorts of words can be grounded only indirectly by association with other more imageable words in the context. Therefore, if we empirically generate the sensorimotor features for the most imageable, earliest words in children's lexicons, we can reasonably expect that later words will be effectively grounded via their relationships to these earlier words. In the neural network model, novel words presented to the model without accompanying sensory input should begin to elicit the appropriate sensorimotor features due to similarities to other concepts or words that share context or usage (See Howell, Becker & Jankowicz, 2001, for a discussion). This is our "propagation of grounding" process.

We have not yet addressed directly with these networks this question of whether novel "ungrounded" words introduced to a trained network will automatically develop sensorimotor representations, via word co-occurrence relations to similar words which *are* grounded in sensorimotor features. The present demonstration of the contribution of sensorimotor features to lexical and grammatical learning was a necessary first step. We are now experimenting with networks designed to investigate this propagation of grounding more directly.

Finally, sensorimotor features can also be particularly useful in modelling in detail the process of word learning. If as Bloom (2000) suggests, children learn the meanings of words through attention to what the caregiver is attending to, then

combining feature representations with phoneme-by-phoneme speech representations might be a network analogy. This would help the network to learn to bind individual phonemes into words, using the constancy of sensorimotor features (as an analogue to focused joint attention with a caregiver) to determine that all these phonemes apply to the same perceived object. In unpublished work, we have begun to examine exactly this.

In summary, since children seem to use preexisting sensorimotor concepts to bootstrap the language learning process, neural network models of language should benefit from including them too, in a variety of ways. We have demonstrated herein three of the ways in which they provide an advantage to a language learning network, specifically, the improvement of the network's word prediction (an aspect of grammar), improvements in noun feature encoding, and the ability to assess word comprehension. More work is clearly needed, however, to explore the capabilities gained by adding sensorimotor grounding to a neural network model of language acquisition.

Chapter 3

Grammar from the Lexicon: Evidence from Neural Network Simulations of Language Acquisition

3.1 Preface

This chapter is reproduced from Howell and Becker (submitted). The paper was first submitted in December 2003, and is currently under revision. We provide evidence that lexical learning that is grounded in pre-linguistic sensorimotor features causes a distinct correlation between lexical and grammatical learning analogous to that found in children. Control experiments of acquisition that lack this grounded sensorimotor meaning do not show this correlation. This furthers our research program into the effects of children's pre-linguistic embodied knowledge on language acquisition.

3.2 Abstract

Previous evidence indicates that there is no dissociation between lexical learning and grammatical learning in children's language acquisition. Rather, the acquisition of words is the driving force behind the acquisition of early grammar,

with grammatical performance being strongly correlated with earlier measures of lexical performance. We suggest that this is due to children's rich knowledge of the sensorimotor content of early words: knowledge about basic semantic features such as motion or size. This semantic knowledge helps to constrain the ways in which those words will be expected to occur in sentences heard by children, and eventually produced by them. We present neural network simulations of language acquisition that show strong correlations between early lexical performance and later grammatical performance. Importantly, this relationship greatly diminishes as the semantic content of the word representations is reduced. This supports the hypothesis that early grammar is emergent from children's previous lexical learning.

3.3 Introduction

There are several theories in the language acquisition literature concerning the origins of grammar in children. For example, some have argued for innate components to grammar, either as dedicated hard-wired neural structures devoted to syntax or at least as a set of constraints imposed by neural architecture (e.g. Chomsky, 1988; Pinker, 1994). This often includes the notion of a specific grammar module. Others, especially recently in linguistics, have suggested a more constructivist perspective, a 'lexicalized' grammar, or an 'emergent' grammar (e.g. Bates & Goodman, 1999). From this point of view early grammar is thought to

be built upon the acquisition of a lexicon of words, a lexicon rich in semantic detail indicating properties such as agency, motion, etc.

Bates and Goodman (1999), among others, have suggested that should the former position be correct in child language, we should expect to see signs of a dissociation in lexical and grammatical learning or performance. They argue that such evidence is not convincing. They present data that contradicts commonly held ideas about the relative sparing of grammar or lexicon in various disorders, such as William's syndrome for grammar. Furthermore, they present data showing that children's lexical acquisition status, as measured by vocabulary size, is strongly correlated with their later grammatical performance, as measured by their mean length of utterance (MLU). Specifically, the correlation between lexical status at 20 months and grammatical status at 28 months is between $r = 0.7$ and $r = 0.84$. These correlations are as high as those between separate measures of grammatical status.

In previous work, we investigated this lexicon to grammar correlation using neural network models of language acquisition (Howell and Becker, 2001). We reasoned that if a simple neural network architecture could demonstrate the same sort of lexicon to grammar relationship, this would provide support for the constructivist explanation. Our neural network was designed to map linguistic forms of words to semantic meanings of words, while simultaneously learning to predict which word should come next, a grammatical task. The former was our measure of lexical performance, the latter our measure of grammatical

performance, and indeed, we did find that the highest correlation ($r = 0.8$) between the two was at a moderate learning lag, from earlier lexical learning to later grammatical learning (Howell and Becker, 2001).

However, the similarities between that work and the child data were somewhat limited, in that it used a very small vocabulary, somewhat unrealistic localist representations of words, and an artificial set of feature dimensions that was partially contaminated by linguistically associative features. The feature representations of pre-linguistic children, as we have argued elsewhere (Howell, Becker and Jankowicz, 2001), should be restricted to those features directly available to the child through sensory-motor interactions with the world, such as size, or texture (see Lakoff, 1987). We have since created such a feature set (Howell, Becker and Jankowicz, submitted), by having human participants rate nouns and verbs on a set of sensorimotor feature dimensions (97 for nouns, 84 for verbs). Thus, we were able to attempt a clearer demonstration of the lexicon-to-grammar effect with a new set of simulation experiments. The experiments described below use both simple and complex training corpora with our sensorimotor training set and phonemic rather than localist word encoding, all in an attempt to more closely match the measurement conditions used with children.

3.4 Simulation Experiment 1

In this experiment, we trained a neural network model of language acquisition under several different conditions to investigate the relationship between lexical

learning and grammatical learning. Specifically, by grammar learning we are referring to basic syntax, or sequence learning.

3.4.1 Method

We used an extended Simple-Recurrent Network (SRN) architecture as our language-acquisition model in this experiment (See Figure 1). These networks have been shown (e.g. Elman, 1990, Howell & Becker, 2001) to be capable of learning aspects of syntax or grammar. Elman (1990) first demonstrated this by training a network to perform the word prediction task: predicting from the current input word what the next word would be. In order to do this, the network must develop an internal representation of how words relate to each other sequentially; in other words, it must learn simple elements of grammar.

Elman's original SRN architecture employed a simple form of recurrence in the hidden layer to maintain context. We extended this SRN model in an important way. In addition to performing word prediction, the network was trained to produce the sensorimotor features of the input word. We hypothesized that by using a common pool of hidden units for both word prediction and word recognition, the network would make use of its semantic knowledge of words in learning word-word relationships. Our prediction was that given enough structured knowledge about words, the network would behave similarly to children: grammatical performance (basic syntactic learning) would be highly correlated with earlier lexical performance.

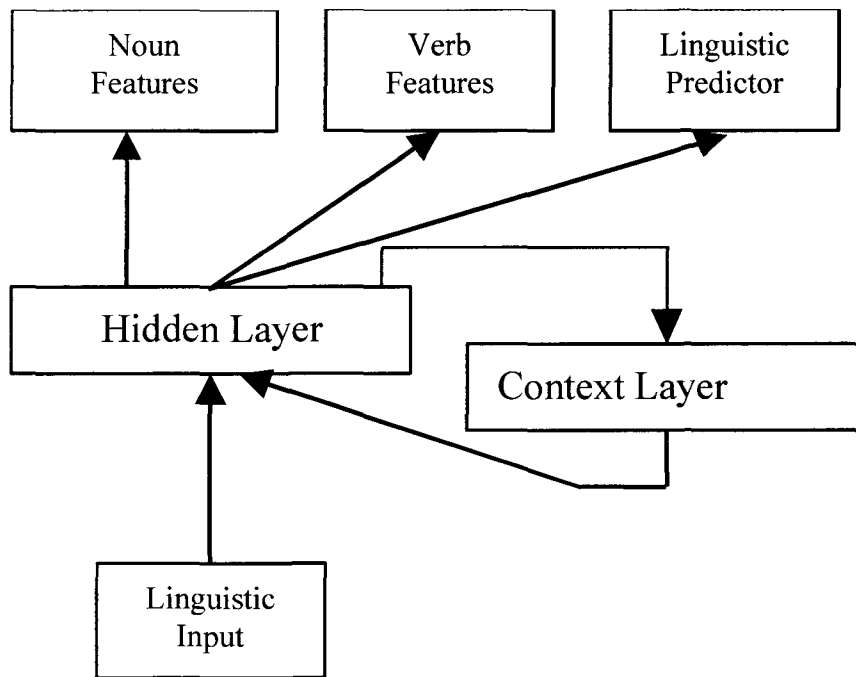


Figure 3.1: Modified SRN architecture, including standard SRN hidden layer and context layer, standard linguistic (word) prediction output, and novel noun feature output and verb feature output. The linguistic input is a whole-word phonetic representation of up to 10 phonemes. The Noun and Verb feature targets are meant to be an abstract representation of pre-linguistic sensory and motor-affordance semantics.

Thus, our network was trained to perform two tasks simultaneously: 1) word prediction and 2) sensorimotor feature retrieval. This was done using three separate pools of output units (See Figure 1). The Linguistic Predictor output layer performed the word-prediction task. At each time step, its goal was to predict the phonemic representation of the input word at the next time step. The task for the remaining two outputs was to produce the sensorimotor features of the current word, or essentially, to map the sounds of the word to its 'meaning'. The Noun Features layer and Verb Features layer had output targets that represented

the sensorimotor features for the current word, either noun or verb (see Howell, Becker & Jankowicz, submitted). When the current input was not a noun or a verb (respectively), a vector input of all 0's was presented at that layer, and no backpropagation of error was performed for that layer.

At input, whole-word phonetic representations of words were presented in serial order through the corpus. Each input word was encoded as a set of up to 10 phonemes using 140 input units. The 140-element word inputs represented 10 phonemic slots each containing 14 phonemic feature bits. The Carnegie Mellon University (CMU, 1995) machine-readable phonetic transcription system and pronouncing dictionary was used to generate our phonetic representations of words. Each phoneme was uniquely mapped to a set of 14 bits, representing articulatory dimensions of the phonemes (See Howell, Becker, and Jankowicz, submitted). Words shorter than 10 phonemes had their rightmost slots padded with zeros, while longer words were truncated. Thus while we are using a phonetic encoding scheme for words, the network's input is not a sequence of phonemes and pauses but a continuous series of phonemic "chunks" pre-segmented into words.

A small common hidden layer and context layer of 30 units each were used. Whatever internal representation of grammar that the network develops is encoded in the weights to and from the hidden layer. Given a large enough hidden layer, the three pools of output units could recruit a separate portion of the hidden units for their own purposes. Thus, to force the network to develop an integrated

internal representation common to both tasks, we keep the hidden layer resources limited.

The structure of our model allows us to measure both lexical and grammatical performance. Lexical performance can be inferred from the ability of the network to generate the correct sensorimotor features for any given word. This gives us a measure of vocabulary acquisition both during learning and when testing generalization performance on novel words presented at the input. Grammatical performance, on the other hand, can be inferred from the network's word prediction accuracy. Thus we have measures of lexical performance (features encoding accuracy) and grammatical performance (word prediction accuracy), which are obviously necessary before we can examine the relationship between them.

The training input for the networks in this experiment consisted of a small corpus (390 words) of two and three word subject-verb (SV) and Subject-verb-object (SVO) sentences. These sentences were an excerpt from a larger corpus of sentences which was created via a simple random generation of allowable sentences using a lexicon of 29 words (18 nouns and 11 verbs). Three classes of verbs were included, transitive, intransitive, and optionally transitive. This produced telegraphic sentences like "Dog Chase Cat", "Man Eat Sandwich", or "Cat Sleep". Training used the back-propagation of error learning algorithm (Rumelhart, Hinton, & Williams, 1986).

In addition to the network model described above, which we refer to as the Experimental condition, two control conditions were run to evaluate the contribution of semantic meaning (structured sensorimotor features) to the lexical-grammatical relationship. The Experimental condition used human-generated meaningful noun and verb features. These scalar-valued features corresponded to human ratings of dimensions such as size (see Howell, Becker, and Jankowicz, submitted, for details). This condition has maximum semantic content.

The Random Control condition used the same architecture as the Experimental condition, but replaced the human-generated (and meaningful) noun features with randomized permutations of that same set of features. This condition is intended to provide a minimal baseline and control for input vector magnitudes. The randomization was performed by first swapping each word's feature representation with the representation of another word in its class (either noun or verb). Secondly, within each of these vectors, the representations were further randomized by iteratively swapping the value at each position on the 97 element vector with that of another random position. These manipulations minimize any featural similarity between related words, while maintaining the overall magnitude and range distribution of the feature vectors.

The Swapped Control Condition falls somewhere between the other two. The existing, meaningful, feature representations of the Experimental condition were randomized using only the first part of the above-described randomization

scheme. That is, the featural representations of each word were swapped with another random word. Within a featural representation, however, the position of each individual bit did not change, and so neither did its meaning. In most cases, this would mean that the featural information for a word no longer agreed with the usage of the word. Thus a word like "Book" with a speed feature of zero might now have the features of "Dog", giving it a speed rating of 0.5. This should cause the network to expect to see it in sentences involving running or chasing, when in fact the same sentences will be presented for "Book", in spite of its changed 'meaning' representation. Of course, in some cases similar representations might be maintained, e.g. if "Cat" were swapped with "Dog". Minor meaning representation differences would be present, but at the level of complexity of our corpus this might not make a noticeable difference. Thus we expect the relationship between lexical semantics and grammar to be preserved to a greater degree in this condition than in the case of the Random Control condition, but less than the Experimental condition.

Eight networks were run in each of the three conditions, for a total of 24 networks. Each network's weights were initialized randomly and trained for 500 epochs using the SRNEngine simulation package (Howell & Becker, submitted). The networks' output accuracy was measured using the Euclidean distance from the actual output to the target output. We measured accuracy by counting as correct only those outputs that produced a smallest distance measure to the correct target. Near-misses, such as "Dog" in place of "Cat" or "Move" in place of "Run"

were counted as failures. There may be reason to count near misses like the above as accurate, especially for word prediction, when predicting any valid subject as the third word of a sentence might be considered accurate, for example. However, for the present experiments we measured accuracy in the more conservative way. Accuracy was recorded in this fashion in all three pools of output units.

Measurement

In order to calculate a lexical-grammatical correlation we need to specify points on the learning trajectories of both tasks at which to take measurements. For lexical status we simply took the peak performance point of the lexical accuracy curve (approximately Epoch 20) as our point of reference. This was roughly the point at which the steeply rising curve began to plateau. We also took several points on either side of this one, for comparison purposes, to ensure that the choice of lexical reference point was not critical. Grammatical accuracy, however, does not 'peak' during our training range. Thus we decided to avoid the necessity of picking a single point, and calculated the correlation from each of our several noun points to grammatical accuracy at every epoch during the training range. Thus, instead of looking at the value of any particular correlation coefficient, we examined the overall pattern of correlations. We predicted that this overall pattern of correlations would be higher and more stable for the Experimental condition than the Random condition, and that furthermore the Swapped condition should be intermediate between them.

3.4.2 Results

Figures 2 and 3 show plots of the lexicon to grammar correlation across condition, for nouns and verbs, respectively.

In the Experimental Condition, the lexical-grammatical correlations start extremely low at lexical to grammatical lags of zero or negative values, but at positive measurement lags, the correlations climb quickly and stay very high. The approximate sustained correlation is $r = 0.9$ for the nouns, and $r = 0.8$ for the verbs, throughout the training range. Furthermore, the difference between the correlation lines for the different lexical performance measurement points (discussed above) is quite small; they parallel each other closely. This is true for both noun and verb correlations. Thus, the choice of noun and verb reference point is not a critical factor in our results.

In the Random Condition, the noun lexical-grammatical correlations start negative to low, and rise slowly over time, staying below $r = 0.4$. They are somewhat stable. The Verb lexical correlations are wildly varying, with a mild central peak midway through learning of about $r = 0.3$

In the Swapped Condition, the noun lexical-grammatical correlations are quite variable, with (in the best case) a correlation that jumps up to around $r = 0.8$ fairly quickly and then tails off to the $r = 0.4$ level. The verb lexical-grammatical correlations are even worse, varying a great deal, and in the best case peaking at around $r = 0.75$ after very short lags and then dropping negative.

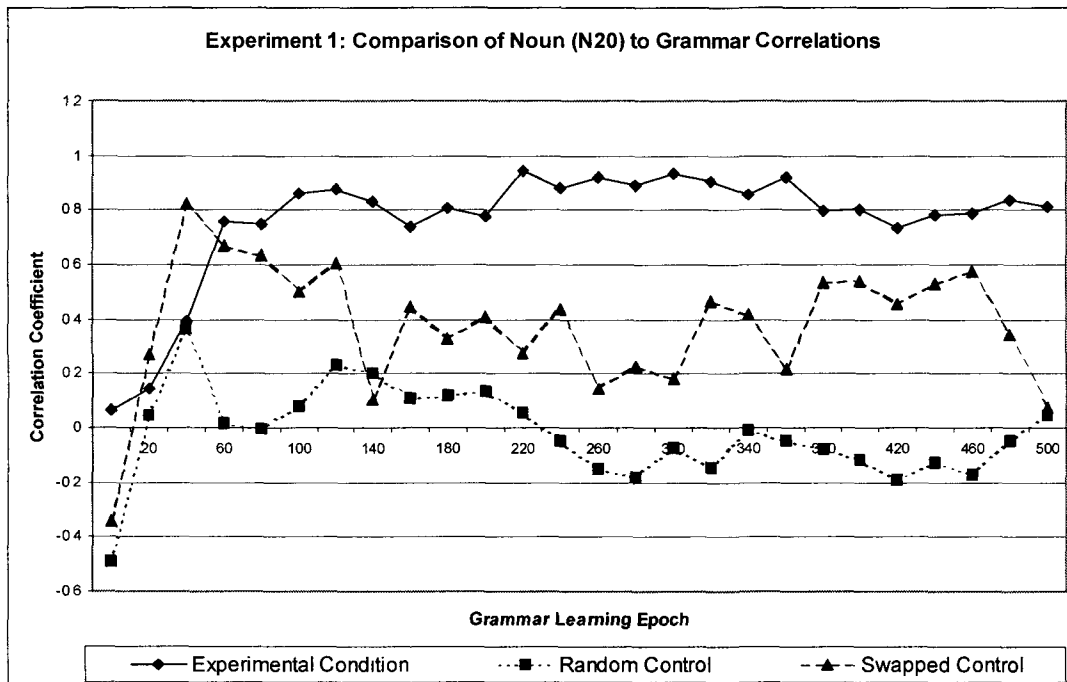


Figure 3.2: Noun Lexicon to Grammar Correlations in all three conditions of Experiment 1. Curves are noun to grammar correlations from the Epoch 20 noun reference point to all grammar points (Epochs 1 to 500). Simultaneous correlations occur when the Epoch is equal to the Noun reference point (Epoch 20). Correlations before that point are from earlier grammar to later lexical learning, correlations after are lexical learning to later grammatical learning.

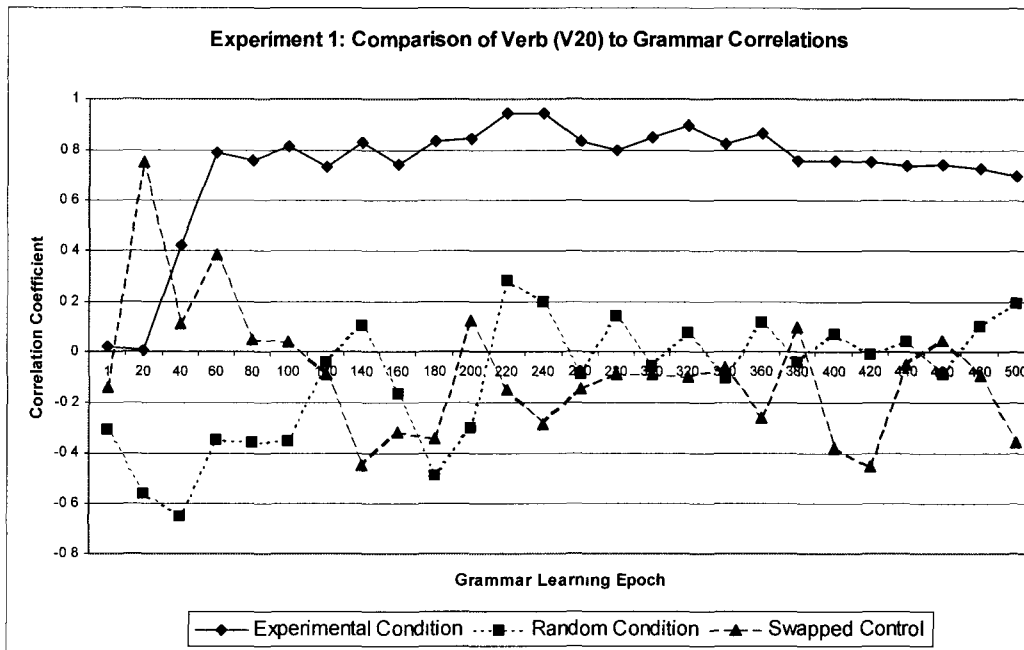


Figure 3.3: Verb Lexicon to Grammar Correlations in all three conditions of Experiment 1. Curves are verb to grammar correlations from the Epoch 20 verb reference point to all grammar points (Epochs 1 to 500). Simultaneous correlations occur when the Epoch is equal to the Verb reference point (Epoch 20). Correlations before that point are from earlier grammar to later lexical learning, correlations after are lexical learning to later grammatical learning.

3.4.3 Discussion

The difference between the sustained high correlations of the Experimental condition and the lower or varying correlations of the other two are dramatic, both for nouns and verbs. Overall, the Random condition clearly has a lower correlation. The Swapped condition, while occasionally demonstrating higher correlations, also demonstrates large negative correlations. This variability may be explained by the differing forces at work on the learning of the networks in the Swapped condition, with sensorimotor feature information being sometimes

opposed to syntactic information provided by word order and sometimes in concert with it.

While trying to make too close an examination of the correlations in the Random and Swapped conditions is pointless due to the high degree of variability, it seems obvious that the connection between lexical learning and grammatical learning in the random condition is somewhat more stable than the swapped condition. The lexical grammatical correlations in the Random condition are consistently lower than in the Experimental condition, and indeed, at whichever lexical to grammatical correlation point we examine, the Experimental condition is a great deal higher than the Random condition.

The implications of this finding for theories of lexicalized grammar or emergent grammar are interesting. Networks that have access to informative representations about the sensory and functional semantics of words show a high lexical-grammatical correlation, while networks that have less semantic information do not. Thus, in our networks at least, the more that is known about the semantics of the words in the lexicon, the greater will be grammatical knowledge (in this case, basic syntactic knowledge) at a slightly later point in learning. The explanation for this is straightforward. The kinds of knowledge that our sensorimotor features represent are information about concrete properties of objects and events. Those concrete properties are the same things that dictate in the environment how those objects and events can interact. For objects, some of these properties would be related to the objects' affordances, the list of things that

it is possible to do with an object, thanks to its physical properties (Glenberg & Robertson, 2000). Most language is used to discuss or describe the real world and possible events in it. Thus, having prior knowledge about how the physical referents of the new and to-be-learned word symbols can interact or appear would simplify the language learning task. Understanding a sequence of words about these objects or events becomes a lot easier when the learner has strong, extra-linguistic expectations as to the allowable combinations of referents, and hence their word symbols. That is, knowing that dogs often move while books don't makes it easier to predict a sentence like "dog move" than one like "book move". Some of the work of learning syntactic rules has been already done for the network, or the child, by the operation of semantic knowledge and expectancies that has been incorporated into the mental lexicon, and is not inherent in the syntax of the speech or text stream.

The above effects were produced by training our model on a relatively small and simple corpus of artificially generated sentences with a small, fully-grounded lexicon. The sentence structures involved were somewhat limited as well (only SV and SVO structures). Natural language is much more complex and less rigidly structured. Further, spoken language frequently contains ungrammatical utterances. We therefore wanted to determine whether the same effect would be apparent in a larger, more naturalistic corpus with a more realistically-sized lexicon.

3.5 Simulation Experiment 2

In this experiment we used a much larger, naturalistic corpus of text with the same network as in Experiment 1, to ascertain whether similar effects are visible in its lexical-grammatical relationships.

3.5.1 Method

We trained on a large (10,742 word) selection of speech drawn from the ChildDES database (McWhinney, 2000) transcribed from mother-child playtime interactions. This corpus was created by appending all of the Bates FREE20 data sets (Bates, Bretherton, & Snyder, 1988; Carlson-Luden, 1979) from the ChildDES database into a single body of text. Any pauses, periods, etc. in the original corpus were replaced with a generic pause marker, intended simply to assist in defining clause boundaries in this corpus of more complex sentences. This corpus had a vocabulary size of 529 words. Of these 529 words, a sizable minority of the content words were grounded in our sensorimotor feature representations (60 nouns, 49 verbs). The grounded words were those that are represented earliest in children's vocabularies (Fenson et al, 2000).

Due to time and processing constraints with these larger networks, we eliminated the swapped control condition and compared only the Experimental and the Random conditions. 11 networks in each condition were trained for 200 epochs using the SRNEngine simulation package (Howell & Becker, submitted). Rather than running these larger networks to asymptotic performance, we simply

ran them for a fixed number of epochs (200) within which grammatical prediction began to approach reasonable levels of performance. The networks' word prediction (grammatical) accuracy was recorded over the course of training, as well as its noun and verb feature encoding accuracy. We analyzed the noun-to-grammar and verb-to-grammar relationships in the same way as in Experiment 1.

3.5.2 Results

The epoch at which a 'peak' of lexical accuracy was reached was approximately epoch 40 for both nouns and verbs. Thus we calculated the lexicon-to-grammar correlations from both noun and verb performance at Epoch 40 to every epoch of grammar learning. We show the comparison of the Experimental to the Random condition for nouns in Figure 3.4, and for verbs in Figure 3.5.

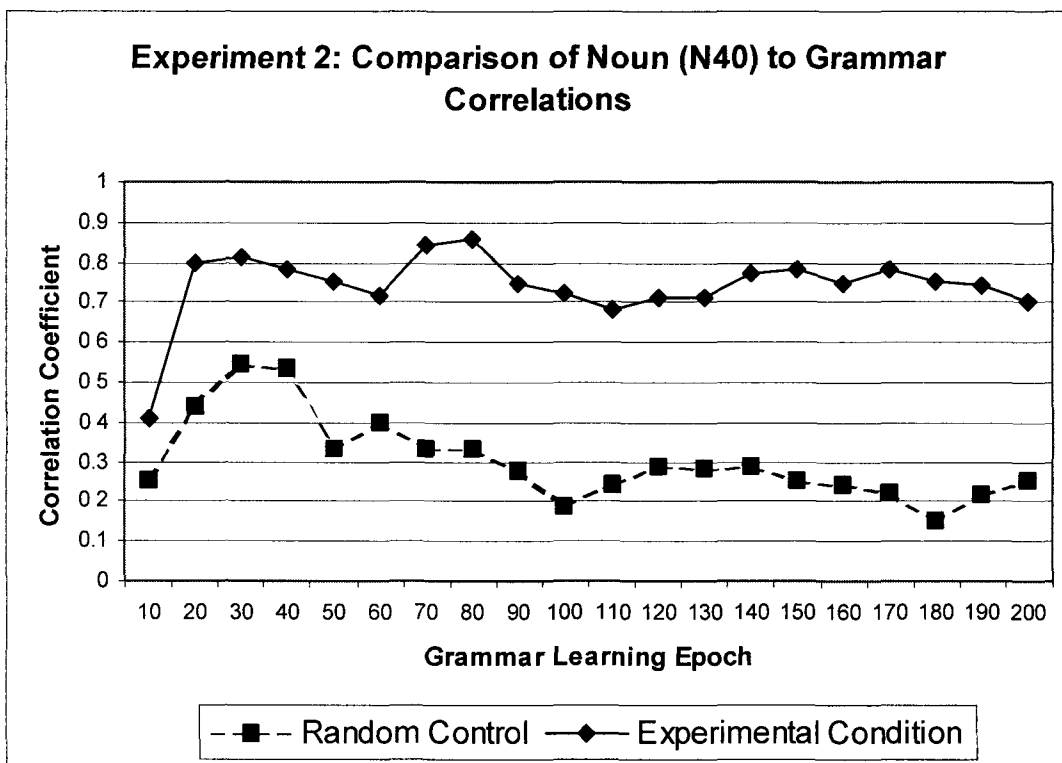


Figure 3.4: Noun Lexicon to Grammar Correlations in both conditions of Experiment 2. Curves are noun to grammar correlations from the Epoch 40 noun reference point to all grammar points (Epochs 1 to 500). Simultaneous correlations occur when the Epoch is equal to the Noun reference point (Epoch 40). Correlations before that point are from earlier grammar to later lexical learning, correlations after are lexical learning to later grammatical learning.

For nouns, the Experimental condition correlations are everywhere substantially higher than the Random condition (averaging around $r = 0.75$ compared to $r = 0.3$), and they rise slightly from the contemporaneous Epoch 40 point to a peak of $r = 0.86$ at a point 40 epochs later. The Random condition correlations steadily decline. For verbs there is also an increase, from Epoch 40 to a correlation of $r = 0.88$ at a point about 40 epochs later (grammar learning at Epoch 80) in the Experimental condition. There is a steady decline in the

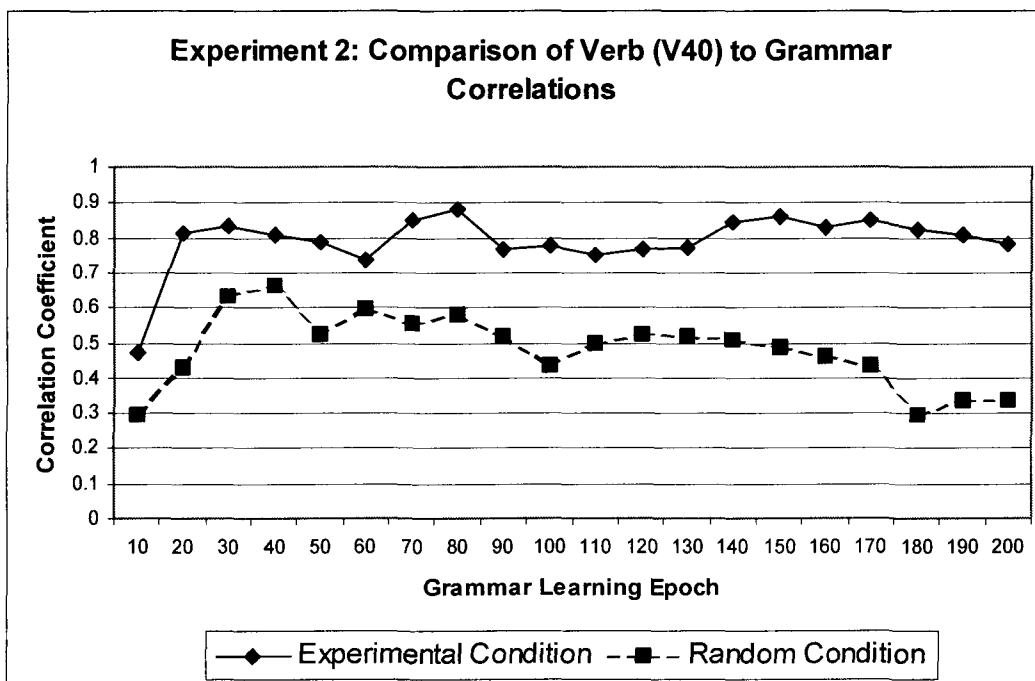


Figure 3.5: Verb to Grammar Correlations in all three conditions of Experiment 2. Curves are verb to grammar correlations from the Epoch 40 verb reference point to all grammar points (Epochs 1 to 500). Simultaneous correlations occur when the Epoch is equal to the Verb reference point (Epoch 40). Correlations before that point are from earlier grammar to later lexical learning, correlations after are lexical learning to later grammatical learning.

Random condition from the Epoch 40 point. The verb-to-grammar correlation stays high and relatively stable ($r = 0.75$ to $r = 0.85$) throughout, while the Random condition is lower and declining (max $r = 0.65$, average around $r = 0.5$).

Choosing different noun and verb lexical accuracy reference points makes little difference to the correlations in this Experiment, as correlation curves calculated from neighboring points are relatively close to the curve from the Epoch 40 point, although the Epoch 40 correlations are generally the highest. These additional curves are not shown.

3.5.3 Discussion

In this experiment, the Experimental and Random conditions are clearly different throughout all of the training period, for both noun-to-grammar relationships and verb-to-grammar relationships. The verb-to-grammar correlations remain slightly higher at large epoch lags, at around $r = 0.85$, while the noun-to-grammar correlations are sustained at approximately $r = 0.75$. We cannot make too much of such a small difference, but there is the possibility that verb semantics might be somewhat more important to sentence meaning (and hence grammatical performance) than noun meaning, especially in the more complex sentences of this Experiment. This would certainly be consistent with the literature on the importance of early verbs as prototypes for basic argument structure constructions (e.g. Goldberg, 1999).

Interestingly, in this experiment the lexical to grammar correlation is already quite high at grammar points before the epoch 40 noun and verb reference points. These correlations would represent a correlation between earlier grammatical and later lexical learning, which is not what we are expecting in these experiments. Of course, these correlations are not as high as those observed at later grammatical points (representing the expected lexical-to-grammatical lags) but they are still unexpected.

Once possible reason for these first several points is that they are spurious correlations. In drawing our conclusion about the lexical-to-grammatical correlation pattern at positive time lags, we have a sequence of many points, all

showing the pattern in question, which indicates that the pattern is not due to chance alone. The two data points supporting the pre-lexical correlation, on the other hand, are not as convincing evidence for a meaningful pattern. Furthermore, the sets of correlations taken from different noun reference points (which normally show a close correspondence, taken above as an indication that the exact noun reference point is unimportant) show a higher range of variability in the neighborhood of the pre-reference point correlations. The lexical-to-grammatical relationship for those first several points thus seems to be highly variable, and for the particular data points that were sampled and plotted, happened to be high. The post-reference point correlations, on the other hand, are more regular, and do not depend as much on which data points are used or plotted

3.6 General Discussion

In these experiments a similar pattern of lexicon-to-grammar relationships was found for both nouns and verbs, in both a small, simple, fully grounded text corpus and a large, naturalistic, partially grounded speech corpus. The Experimental condition, in which some or all of the words experienced are grounded in sensorimotor feature representations, always showed a higher correlation between earlier lexical performance and later grammatical performance than conditions without these meaningful feature representations. This is an effect analogous to the relationship found in children, where their

lexical performance at 20 months is highly correlated with grammatical performance at 28 months.

In essence, what these simulations indicate is that a language learner possessing rich representations for the concepts which will become the basis for the first words has an easier time learning how those words can be used in sentences. The featural information that is provided to our networks, or via the children's senses, helps to constrain the ways the word representing that concept can be used in sentences. A rich lexicon that contains these kinds of low-level semantic features allows the similarities between related words to be detected and taken advantage of during learning. Thus grammar learning is influenced in two ways in our networks, the straightforward statistical learning about which words occur in which syntactic roles, and the learning of semantic constraints of what words can fit together in a sentence based on the relationships that hold between the underlying concepts. Thus grammar, at least partially, emerges from the lexicon.

Of course, the analogy between the lexicon to grammar relationships displayed in these simulations and that found in children has certain limits. For example, in children lexical performance is measured at a pre-determined age for these calculations, 20 months, and grammatical performance is assessed based on a mean-length of utterance (MLU) measurement at 28 months, once grammatical production has taken off. We have an analogous lexical reference point, but not any single logical grammatical reference point, hence our method of looking at

patterns of correlations across the entire trajectory of grammar learning. Also, we are still only able to use grammatical comprehension as our measure of grammatical performance, not grammatical production. This may not be a problem, however. The reason that production is measured in children is simply because it *is* measurable, whereas comprehension, which necessarily precedes production, is more difficult to assess. In the networks, this situation is reversed; comprehension is easier to assess than production. However, the design of our networks does allow us to force them to produce sequences of text. In future work, we plan to investigate measuring MLU from the networks' production behaviour. This will allow us to more closely approach the way the child correlation is measured, and provide even stronger support for the view that early grammar emerges from the lexicon.

Chapter 4

Grounding Words in Meaning Indirectly – A Computational Model of the Propagation of Grounding

4.1 Preface

This chapter is reproduced from Howell and Becker (Submitted b). This paper is in preparation, to be submitted shortly to the Journal of Memory and Language. This paper is written partly in response to the debate between proponents of the grounded, embodied meaning of words in children and proponents of high-dimensional models of conceptual meaning. This paper provides evidence for a process by which the former can, with development and experience, build up to the latter. We believe this process has the potential to demonstrate a way to merge the two opposing theoretical viewpoints. This work also provides simulation evidence of a candidate cognitive process for the statistical inference of the meaning of novel words in children. Specifically, we explore the hypothesis that possessing word meanings grounded in sensorimotor features

allows a language learner to readily infer the meanings of novel, ungrounded words.

4.2 Abstract

The problem of mapping the forms of words to their meanings is a long-standing one, and in connectionist models of language generally requires training the direct pairing of word and meaning. We describe a neural network that begins with this direct pairing but also generalizes to more indirect acquisition of word meanings. The network had the dual task of producing the meaning representations for each word, as well as predicting the next word. Most words had training targets provided for their meaning representations ('grounded'), and the network quickly learned these. However, the network also produced the correct meaning representations for many words which had never had meaning representation targets (ungrounded). We argue that this "propagation of grounding" is due to the overlapping task demands that allowed syntactic and semantic word co-occurrence information to influence the word meanings being learned, much as children might learn novel words from context while reading.

4.3 Introduction

Considerable evidence indicates that children learn the meanings of many of their early words from direct sensory evidence (e.g. Bloom, 2000, Lakoff, 1987). The actual sequence of this learning is informative, however. Children start doing most of their word learning during the second year (Bloom, 2000). However, by

the end of the first year, they have already learned a great deal about navigating the world, interacting with objects, and comprehending events in it (e.g. Piaget, 1952; Spelke, 1994; Baillargeon, Spelke, and Wasserman, 1985; Mandler, 1992). Furthermore, they have already learned a great deal about the sounds of language, through listening to the verbalizations of those around them, and through their own babbling (e.g. Saffran, Aslin, and Newport, 1996; Vihman, 1996). Thus in the second year, they are ready to put together the non-linguistic conceptual knowledge that they have developed about the world, and the knowledge of phonology that they have learned, to begin acquiring stable word representations that they can use (Bloom, 2000).

In previous work (Howell, Jankowicz, and Becker, submitted), we addressed this stage of the language acquisition process in children. We created a set of pre-linguistic conceptual representations that were designed to be as representative as possible of the sensory and motor (sensorimotor) representations of concepts that pre-linguistic children might have, for all of children's earliest words as represented on the MacArthur Communicative Development Inventory (MCDI – Fenson et al, 2000). We demonstrated that our sensorimotor feature representations captured much of the similarity structure of children's early words, while using only feature dimensions that are either sensory or motor in nature. By sensorimotor we mean aspects of meaning that might be captured in cortical representations, that are derived from either direct sensory perception of the world (and hence sensory cortex) or from manipulation and motor-affordance

experience with objects and events in the world (Glenberg and Robertson, 2000) that might involve representations in motor cortex. This idea of concept formation through mental imagery is consistent with the viewpoints of Barsalou (1999) and Lakoff (1987). Finally, we used these sensorimotor feature representations in neural network models of language acquisition to study the process of the phonology to meaning mapping, and its implications for other aspects of language, such as grammar learning.

We used a modified SRN architecture (Elman, 1990) which had the dual task of mapping input phonology of words to sensorimotor meaning representations of words, as well as predicting the next word in the input stream. We trained the networks on a stream of language taken from actual mother-to-child speech (the ChildDes Corpus, McWhinney, 2000; Bates, Bretherton & Snyder, 1988; Carlsen-Luden, 1979) and compared the experimental networks to control networks which had random meaning representations rather than sensorimotor representations. We demonstrated that having this pre-linguistic conceptual information, and simultaneously performing the two tasks, makes it easier to learn syntactic information (word sequence information). Essentially, having prelinguistic conceptual representations makes it easier to learn aspects of grammar. This supports the importance of the early grounding of words (in physical, embodied, meaning) for later learning.

However, children certainly do not learn all words, not even all nouns and verbs, from direct physical experience. This cannot be the case, given the

incredible word learning rates that occur in the school-age child (Bloom, 2000). These older children must be learning the meanings of some of their new words indirectly, by inferring meanings from the context, from the relationship of any given novel word to nearby words, sentence structures involved, or verb constructions (see Goldberg, 1999), and to other, related words that are not present but that typically occur in the same ways as the word in question (for a similar argument, see Landauer & Dumais, 1997). If so, how do they develop a rich representation for the word? Is it *possible* for them to do so for such words, or is the conceptual representation of such words eternally more vague, abstract and ill-defined than the more grounded and embodied conceptual representations held for words learned through direct sensory experience?

Presumably, children's later learning of novel words in the above-described manner is not destined to be eternally vague. Many of the sorts of words that children learn as they expand their vocabulary will be simply rarer words that nonetheless do have direct physical referents, and which children may eventually be exposed to directly, and be able to physically interact with. For example, children may have a strong, embodied concept for the word 'dog', based on their direct sensory experience with the family pet, and on their physical interactions with it. When a new word such as 'wolf' or 'fox' is encountered, it may have a vague conceptual representation at first, based solely on contextual similarity to the usage of the word 'dog', but over time the child may be taken to a

zoo, and may gain direct sensory or even motor experience⁴ with the animal (or to a lesser extent through watching television or in picture books). The vague conceptual representation built up solely through contextual co-occurrences will eventually be supplemented and solidified by direct physical experience. In the meantime, however, the vague conceptual representation derived from context allows the young reader (or listener) to continue to comprehend the text or speech, in at least a superficial way.

Of course, the novel word in question could be a more abstract word that does not have a direct physical referent, such as ‘love’, ‘economics’, or ‘value’. What happens in this case? The word cannot be learned through direct physical experience, but perhaps as some have suggested it is learned through experience of multiple exemplars of subcomponents of the word, such as their effects, or persons engaged in the field, or examples of things of value, respectively. Or, as Lakoff and colleagues have suggested (Lakoff, 1987; Narayanan, 1995) these abstract concepts may be understood in terms of analogies to more basic, embodied concepts. For example, a headline about economics such as “France

⁴ For the importance of motor experience in addition to sensory experience in concept formation and retention, an example from neuropsychology is relevant. Some lesion patients who are suffering anomia are left with particular deficits, unable to name any animals except for those with which they have had direct physical contact (petting, riding, etc) (Dr. George Lakoff, private communication, 2000). This seems to support the inclusion of motor-area cortex in the representations of concepts, and may indicate that when lesions of more central semantic association areas are suffered, that motor-cortex components of the distributed meanings of words may be spared, and be sufficient cue to provide entry to the word’s meaning.

crawls out of recession” maps abstract concepts directly onto more embodied concepts in a metaphorical way, and yet it is more than just a colourful metaphor. The text above is a legitimate way of conceptualizing the meaning of an abstract concept in terms of more concrete concepts, with France visualized as a person laboriously extricating himself or herself from a hole in the ground into which he or she had fallen. The essential aspects to understanding the meaning are present in the metaphor.

In either case discussed above, the process of understanding the meaning of the novel word involved is one of indirect mapping from the occurrences of the novel word to the occurrences of known words, whether by similarity of sentence structure, verb construction, word co-occurrence, or metaphoric imagery. The grounded, embodied meaning of the directly-learned early words serve as a foundation for the learning of the later words. Of course, this argument may not apply to non-conceptual closed-class or functional words like prepositions, whose meaning is much more syntactically linked than present in the external environment, and we make no claims regarding these words.

So this argument about word learning, if true, eliminates one objection to the use of sensorimotor features, namely that it may be impossible to represent all concepts in this fashion. Certainly not all concepts are nouns that have direct physical referents, but whether abstract nouns, verbs, or whatever, perhaps all can be represented indirectly in terms of simpler, basic words or concepts that are grounded directly in embodied experience.

In the present work, we investigate this hypothesis. Using neural networks like those in our previous work discussed above, we attempt to train a set of networks in which ungrounded words (words without a sensorimotor meaning representation) acquire a meaning representation without any explicit training for those words. Will grounded meaning essentially ‘propagate’ through the network from grounded words to ungrounded ones?

4.4 Method

Our neural network model of language acquisition is relatively simple. We modified the common Simple-Recurrent Network (SRN) architecture to perform three separate tasks simultaneously, in three separate pools of output units (See Figure 4.1). A small common hidden layer and context layer of 30 units each were used, to force the network to develop an integrated internal representation common to the three tasks. A single input layer presented whole-word phonetic representations of words, in serial order through the corpus.

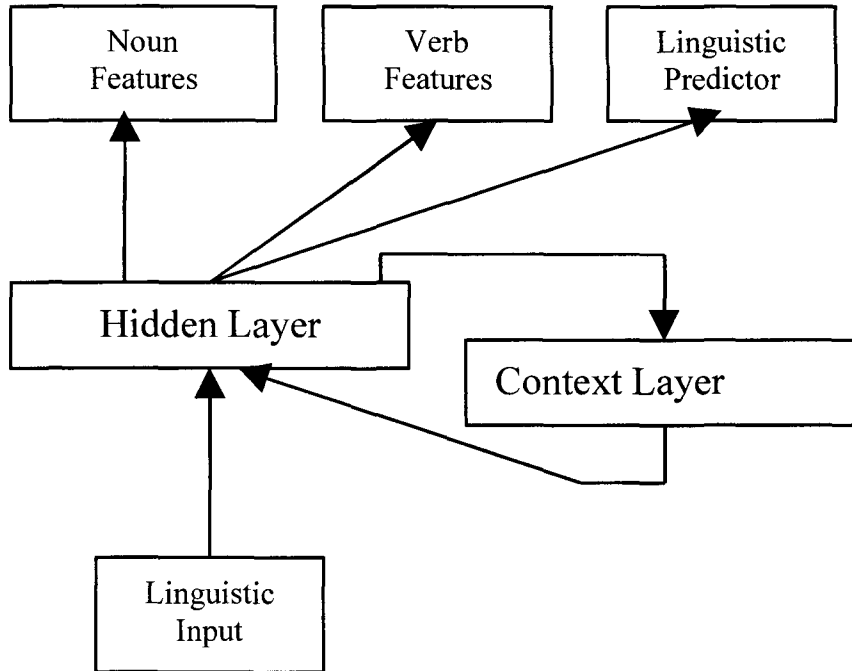


Figure 4.1: Modified SRN architecture, including standard SRN hidden layer and context layer, standard linguistic (word) prediction output, and novel noun feature output and verb feature output. The linguistic input is a whole-word phonetic representation of up to 10 phonemes. The Noun and Verb feature targets are meant to be an abstract representation of pre-linguistic sensory and motor-affordance semantics.

Each word was encoded as a set of up to 10 phonemes using 140 input units. The 140-element word inputs represented 10 phonemic slots each of 14 phonemic feature bits, without representation of word boundaries. The Carnegie Mellon University (CMU) machine-readable phonetic transcription system and pronouncing dictionary was used to generate our phonetic representations of words (available at: <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>). Each phoneme was uniquely mapped to a set of 14 bits (See Figure 4.2), representing articulatory dimensions of the phonemes. This articulatory representation is primarily meant simply to reduce the number of units required to represent a

phoneme. Localist phoneme units would have worked just as well, but made the network larger and slower to simulate. Words shorter than 10 phonemes had their rightmost slots padded with 14 zeros. Longer words, had any existed in our test corpus, would have been truncated.

The Linguistic Predictor output layer performed the word-prediction task that is commonly used in SRN's. By forcing the network to attend to word-order, this sort of layer has been shown to enable the network to learn simple aspects of syntax or grammar (Elman, 1990; Howell and Becker, 2001). At each time step, its task was to produce the phonemic representation of the input word from the next time step. Thus, the size of this layer is the same as the input layer, 140 units. The task for the remaining outputs was to produce the sensorimotor features of the current word. These sensorimotor semantic features are drawn from Howell, Jankowicz, and Becker (submitted), and are distributed, real-valued feature representations of semantic meaning derived from human raters, which emphasize the sensory and motor-affordance properties of objects and events.

"AA"	"1,0,0,0,0,1,0,0,0,1,0,0,0,0"	"L"	"0,0,1,0,1,0,0,0,0,0,1,0,0,0"
"AE"	"1,0,0,0,1,0,0,0,0,0,0,0,1,0"	"M"	"0,0,0,1,0,1,0,0,0,1,0,0,0,0"
"AH"	"1,0,0,0,0,1,0,0,0,0,1,0,0,0"	"N"	"0,0,0,1,0,1,0,0,0,0,1,0,0,0"
"AO"	"1,0,0,0,0,0,1,0,0,0,0,1,0,0"	"NG"	"0,0,0,1,0,1,0,0,0,0,0,1,0,0"
"AW"	"0,1,0,0,0,0,0,0,0,0,0,1,0,0"	"OW"	"0,1,0,0,0,0,0,0,0,0,0,0,0,1"
"AY"	"0,1,0,0,0,0,0,0,0,1,0,0,0,0"	"OY"	"0,1,0,0,0,0,0,0,0,0,1,0,0,0"
"B"	"0,0,0,1,1,0,0,0,0,0,1,0,0,0"	"P"	"0,0,0,1,1,0,0,0,0,1,1,0,0,0"
"CH"	"0,0,0,1,0,0,0,0,1,0,1,0,0,0"	"R"	"0,0,1,0,0,1,0,0,0,1,0,0,0,0"
"D"	"0,0,0,1,1,0,0,0,0,0,0,1,0,0"	"S"	"0,0,0,1,0,0,0,1,0,1,0,0,1,0"
"DH"	"0,0,0,1,0,0,0,1,0,0,0,1,0,0"	"SH"	"0,0,0,1,0,0,0,1,0,1,0,0,0,1"
"EH"	"1,0,0,0,1,0,0,0,0,0,0,1,0,0"	"T"	"0,0,0,1,1,0,0,0,0,1,0,1,0,0"
"ER"	"1,0,0,0,0,1,0,0,0,0,0,1,0,0"	"TH"	"0,0,0,1,0,0,0,1,0,1,0,1,0,0"
"EY"	"0,1,0,0,0,0,0,0,0,0,0,0,1,0"	"UH"	"1,0,0,0,0,0,1,0,0,0,1,0,0,0"
"F"	"0,0,0,1,0,0,0,1,0,1,1,0,0,0"	"UW"	"1,0,0,0,0,0,1,0,0,1,0,0,0,0"
"G"	"0,0,0,1,1,0,0,0,0,0,0,0,1,0"	"V"	"0,0,0,1,0,0,0,1,0,0,1,0,0,0"
"HH"	"0,0,0,1,0,0,1,0,0,1,0,0,0,0"	"W"	"0,0,1,0,1,0,0,0,0,1,0,0,0,0"
"IH"	"1,0,0,0,1,0,0,0,0,0,1,0,0,0"	"Y"	"0,0,1,0,0,1,0,0,0,0,1,0,0,0"
"IY"	"1,0,0,0,1,0,0,0,0,1,0,0,0,0"	"Z"	"0,0,0,1,0,0,0,1,0,0,0,0,1,0"
"JH"	"0,0,0,1,0,0,0,0,1,1,0,0,0,0"	"ZH"	"0,0,0,1,0,0,0,1,0,0,0,0,0,1"
"K"	"0,0,0,1,1,0,0,0,0,1,0,0,1,0"	Pause	"0,0,0,0,0,0,0,0,0,0,0,0,0,0"

Figure 4.2: CMU Phonemes and their compressed 14-bit Representations. The bits represent articulatory features such as voiced/unvoiced, place and manner of articulation, etc. This representation is not meant to make any claims as to the relevance of these features, it was chosen only for practical purposes of compressing the number of bits required to represent a phoneme.

The Noun Features layer had output targets that represented the sensorimotor features for the current word if it was a noun, in 97 feature dimensions (units). The Verb Features layer had output targets that represented the sensorimotor features for the current word if it was a verb, in 84 feature dimensions (units). When the current input was not a noun or a verb (respectively), a vector input of all 0's was presented at that layer, and no backpropagation of error was performed for that layer. This separation of nouns and verbs is not central to our model, but is rather simply a practical

simplification. It would be possible to interleave noun and verb features in a single feature layer, which would simply be larger, but would provide fewer intrinsic cues to syntactic category.

The fact that the network is producing sensorimotor noun and verb features at the output means that we can examine the ability of the network to generate the correct features for any given word. This gives us a measure of vocabulary acquisition both during learning and when testing generalization performance on novel words presented at the input.

4.4.1 Corpus and Training Schedule

We used a large (27, 494 word) selection of two and three word sentences created by a simple grammar which obeyed semantic constraints, based on Elman, 1990. There were 29 unique words in the corpus, 18 nouns and 11 verbs. Words could occur as subject or object, as semantically and syntactically appropriate, and verbs could take an object or not. Sentences were thus either SVO or SV in format (a very basic level of syntax), with no attempt to provide for agreement in order to keep the network as simple as possible. Each network was run for 50 epochs using the SRNEngine simulation package (Howell & Becker, submitted). Rather than running these large networks to asymptotic performance, we simply ran them for a fixed period (50 Epochs) by which point the accuracy of phoneme to meaning mapping had peaked and grammatical prediction accuracy had begun to rise.

4.4.2 Control Network

We ran one simulation of this network with semantic feature targets provided for each and every one of the 29 words in this corpus. This was our fully grounded control network, and it served as a comparison for each of the other simulations to follow. It was run for 50 epochs. At this point, word prototypes were taken, as described in Elman (1990), but at the output layer, not the hidden layer. That is, we were not interested in what hidden layer representation the network was developing (at least not at this time) but rather in what semantic features the network was producing for this word. Thus, for every instance of the word in the corpus, the output activation at the corresponding output layer was recorded (noun or verb). These were averaged together over all occurrences of a given word in the corpus to produce an averaged meaning representation for that word (this is analogous to the way a human language learner encountering a new word for the first time might slowly accumulate evidence as to its meaning from the contexts it occurs in), and the results were subjected to a hierarchical cluster analysis using SPSS. The results were compared to a similar hierarchical cluster analysis performed on the raw feature vectors produced by human raters for validation of the network's prototyping process.

4.4.3 Experimental Networks.

While we used the same corpus for each experimental run of our network, the vocabulary changed slightly each time. There were 29 words in the vocabulary

for this corpus, and so we ran 29 different simulations, grouped into the Noun Group (the 18 nouns) and the Verb Group (11 verbs). For each different simulation, the input representation of the word being tested remained the same, but its output targets at the noun or verb sensorimotor feature layer were removed. The target for that word was replaced with a marker that indicated to the network that it should not do any learning on this word on this time step (no backpropagation of error, no weight updates). However, statistical learning of the word order relationships through the Linguistic Prediction layer continued to take place, and backpropagation of error and weight updates from that source continued for all words throughout training. Effectively, while the subject word continued to be present in the input stream with all other words, it was not linked to a meaning. This is meant to be analogous to a child hearing (or reading) a word for which a referent is not known; the word symbol may be learned, the pattern of its occurrence may be learned, but no direct mapping to meaning (e.g. via direct physical perception of the word's referent) takes place. Meaning must be inferred from context and patterns of word co-occurrence.

Despite not having been trained in a word to meaning mapping for the subject word, the network will still produce an output at the meaning layer, and so some meaning representation will be indirectly acquired for the subject word, through random fluctuations in the output activations if nothing else. However, any useful aspects of meaning that are acquired can only be based on the learning of the other words' mappings and the influence from the learning done by the

Linguistic Prediction layer. So, at the completion of training, word prototypes were taken as in the Control network above, and the prototypes were subjected to a hierarchical cluster analysis, to see just how much meaning for the subject word the network was able to learn.

We examine the nature of this meaning by where it groups in the hierarchical cluster dendrogram. For our present purposes, we can classify this into one of three results. The most accurate result would be that the ungrounded subject word clusters exactly where it should according to the control network. This would clearly indicate that the subject word has acquired a correct meaning representation. The next most accurate would be a close match, a meaning representation that lets the subject word fall somewhere close to where it should according to the control network, but not exactly in the right cluster. For example, boy clustering with the animals “dog” “cat” etc. This is somewhat accurate, since there is more similarity between ‘boy’ and these animals than there is to other concepts in the vocabulary, such as car or book. So this would indicate that there has been at least partial learning of the word-meaning mapping for the subject word. Finally, the subject word could fall in a cluster that is completely dissimilar to it (within the bounds of the corpus) such as in the ‘boy’ clustering with ‘book’ example. ‘Boy’ is animate, large, intelligent, etc., while ‘book’ is immobile, nonintelligent, etc. Examining the actual features produced by the network can elucidate exactly what aspects of the word’s meanings were indirectly learned in this manner, but this detailed analysis is not required at this

point for this simple corpus. The hierarchical cluster analysis is sufficient to capture the overall characterization of the meaning.

Of course, the network could arrive at the correct meaning vectors for the subject word solely by chance, although this is unlikely. We could examine how likely it is that a given clustering would occur by chance, and then run a given subject word's simulation for enough times to allow us to distinguish its mean cluster performance from chance. Or, since we are not interested in the acquisition of individual words' meanings at this time, but rather in the entire process of propagation of grounding, we can simply consider the number of correct matches achieved across the 18 nouns, and 11 verbs, and compare that to the chances of that result occurring by chance. The latter is the approach we take here.

4.4.4 Categorical Analysis

We were also able to perform an additional analysis on the Noun Group that we could not on the Verb Group, since the nouns were originally grouped into six categories in Elman, (1990) while the verbs had no a priori categorical distinction, only syntactic distinctions. We used the six noun categories, formed a category centroid for each by averaging the individual meaning representations of their members, and then compared each individual word to each of the categories. The category to which each word was closest was compared to the category to which it is a priori supposed to belong, and an accuracy measure across words was calculated. A word was considered correct if it was found to be in the correct

category, incorrect otherwise. This accuracy count was compared to that which might be expected due to chance (Binomial test, $P = 1/6$, $q = 5/6$, $n = 18$).

One criticism of this approach is that there is sometimes the possibility of bias when the word being tested has been included in forming the centroid vector for one of the categories. This has not been an issue in previous work with these sensorimotor feature representations, however. In Howell, Jankowicz and Becker (submitted), we found very high (92.8%) agreement between the category memberships calculated via the above method (inclusive calculation), and a very similar agreement score (87.9%) using the more conservative method of removing the item under consideration from the calculation of its a priori category's centroid (exclusive method). This difference has been so small that we consider it fair to use the inclusive method in the present work, which is computationally simpler.

4.5 Results

We will discuss the results of our simulations according to their group: Control, Noun Group, or Verb Group. First, the Control network performed well, learning how to map the input words to their sensorimotor features after only 50 epochs of training. The accuracy of feature encoding was 99.98% at the noun feature output, and 100% at the verb feature output. The hierarchical cluster analysis of the original noun features is provided in Figure 4.3, and that of the Noun feature layer prototypes is shown in Figure 4.4. Note the close correspondence,

demonstrating that the output prototype technique is capturing the feature representation accurately. The control group had feature targets provided for all words (fully grounded) so no analysis of indirect grounding is necessary. In the rest of the networks, however, we examined how well the particular ungrounded word in each simulation acquires a featural representation indirectly.

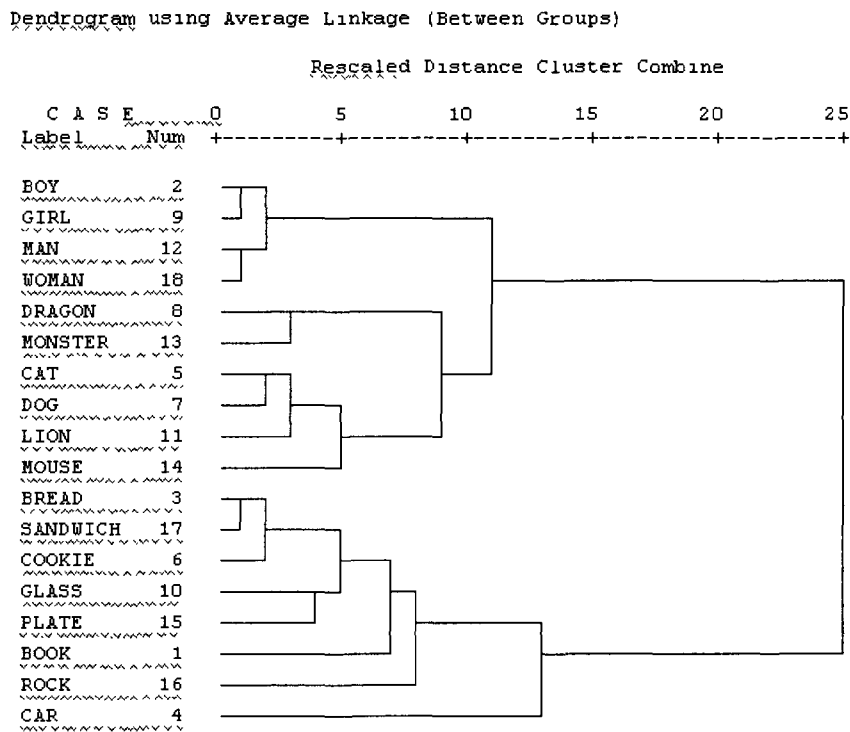


Figure 4.3: The hierarchical cluster analysis dendrogram of the Noun master features, as generated by human raters.

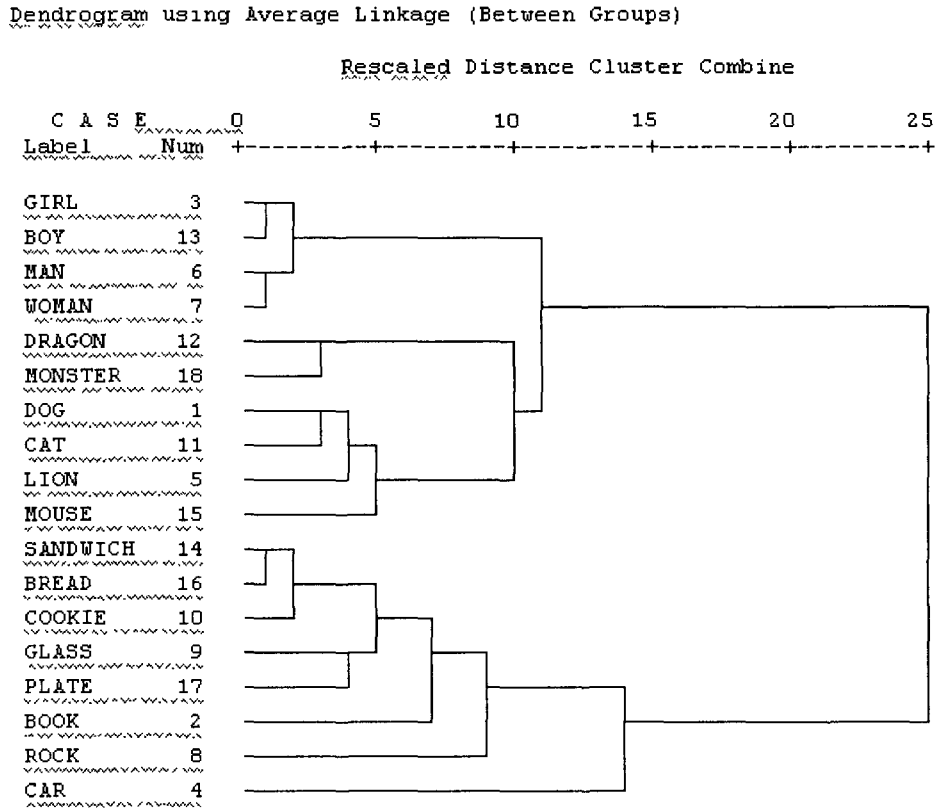


Figure 4.4: The hierarchical cluster analysis dendrogram of the Noun Feature layer output prototypes for the Noun Control network. Note the correspondence to the master features

In the Noun group, 18 simulations were run, with a different noun excluded from phoneme-to-features training each time. The results showed a mix of feature-encoding performance across words, including 7 exact matches, 5 near matches, and the rest counted as incorrect matches. Chance performance for the dendrogram grouping is somewhat difficult to estimate. A word is counted as an exact match if it is in the same subordinate cluster as it should be according to the control network. There are somewhere between 9 and 12 clusters into which any

item could fall, depending on how many subordinate clusters occur in the analysis and how many words link in at a higher level. Taking the most conservative estimate, this would result in a chance level of performance of $1/9$ (0.111), which is substantially lower than the $7/18$ (0.389) exact match performance, or the $12/18$ (0.667) exact + partial match performance. The exact match performance is significantly different from chance using the binomial test ($p=1/9$, $x=7$, $n = 18$, $p = 0.005$). A summary of the hierarchical clustering results is shown in Table 4.1 below. Representative hierarchical clustering dendrograms are also shown in each of the three conditions in figures 4.5, 4.6, and 4.7.

Table 4.1: Noun Propagation of Grounding – Hierarchical Cluster Summary

Book Propagation Test	Near Match
Boy Propagation Test	Match
Bread Propagation Test	Partial Match
Car Propagation Test	Incorrect
Cat Propagation Test -	Incorrect
Cookie Propagation Test	Near Match
Dog Propagation Test	Incorrect
Dragon Propagation Test	Exact Match
Girl Propagation Test	Exact Match
Glass Propagation Test	Near Match
Lion Propagation Test	Near Match
Man Propagation Test	Far Match
Monster Propagation Test	Near Match
Mouse Propagation Test	Exact Match
Plate Propagation Test	Exact Match
Rock Propagation Test	Exact Match
Sandwich Propagation Test	Incorrect
Woman Propagation Test	Exact Match

Incorrect: 4 (4/18) = 22.2%

Car, Cat, Dog, Sandwich

Far Match (Correct Superordinate distinction): 2 (2/18) = 11.1%

Bread, Man

Near Match (Correct Subordinate Distinction): 5 (5/18) = 27.8%

Book, Cookie, Glass, Lion, Monster

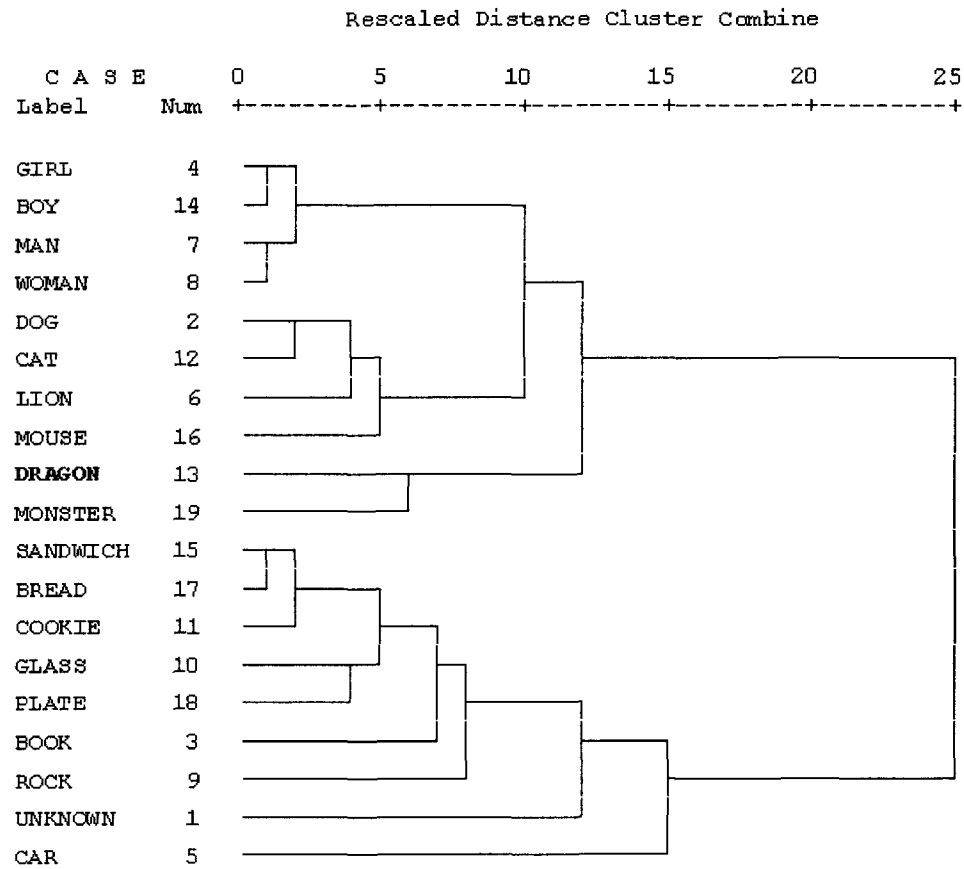
Exact Match (Proper place in Grouping): 7 (7/18) = 38.9%

Boy, Dragon, Girl, Mouse, Plate, Rock, Woman

Good Match (Near Match plus Exact Match): (12/18) = 66.7%

Figure 4.5: An example of a noun hierarchical cluster analysis dendrogram for the Exact Match results of the Noun Group.

Dendrogram using Average Linkage (Between Groups)



Dendrogram using Average Linkage (Between Groups)

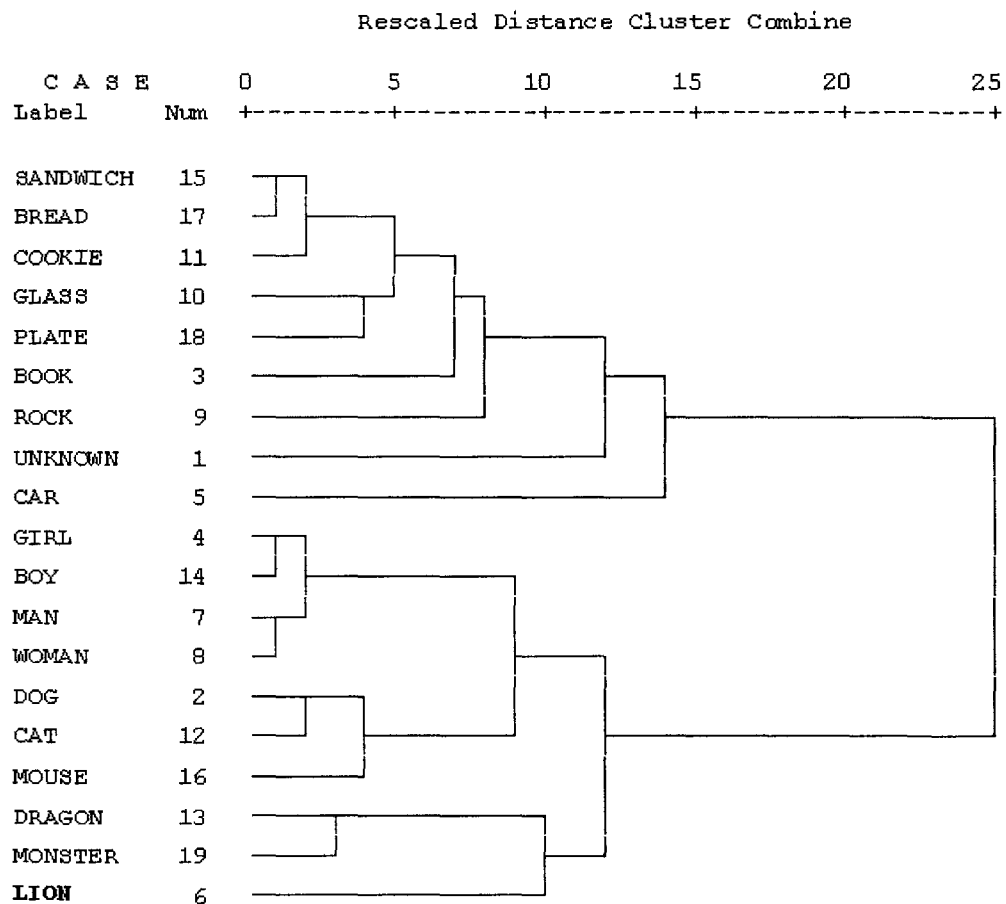


Figure 4.6: An example of a noun hierarchical cluster analysis dendrogram for the Near Match results of the Noun Group.

Dendrogram using Average Linkage (Between Groups)

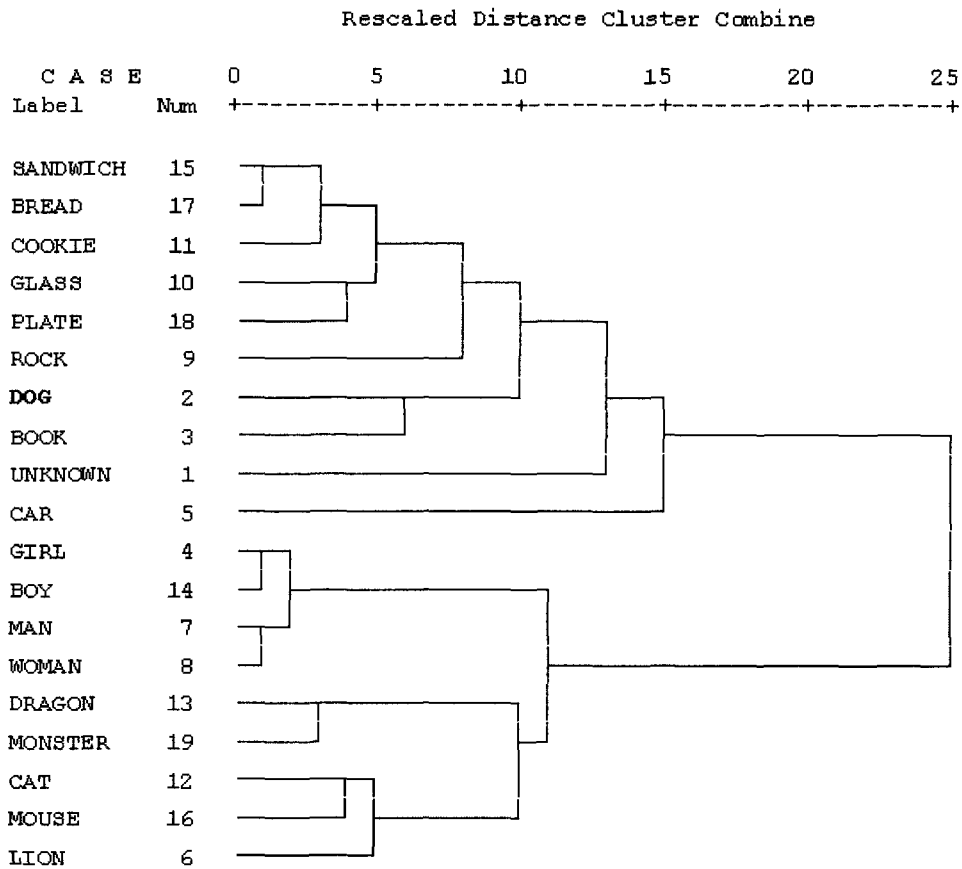


Figure 4.7: An example of a noun hierarchical cluster analysis dendrogram for the Incorrect results of the Noun Group.

4.5.1 Categorical Analysis

The results of the Noun group categorical analysis are shown in Table 4.2. The subject word was closest to its correct a priori category in 7 out of 18 trials. This is significantly different from what one might expect due to chance (binomial test, $p = 1/6$, $q = 5/6$, $X = 7$, $n = 18$, $p = 0.024$). We consider this test to be particularly

useful at assessing the partial learning that is occurring during propagation of grounding, since with limited training data we do not expect the network to learn the exact meaning vector for the subject word, and in fact they do not. Rather, the overall character of the subject word’s meaning vector should tend to group with other words with similar meanings, as shown in this analysis.

Table 4.2: Categorical Analysis of the Noun Group

	Nominal Category	Actual Category	Nominal Label	Actual Label	Accuracy
Book	6	4	Inanimate	Edibles	
Boy	1	1	Human	Human	1
Bread	4	4	Edibles	Edibles	1
Car	6	3	Inanimate	Animals	
Cat	3	4	Animals	Edibles	
Cookie	4	6	Edibles	Inanimate	
Dog	3	6	Animals	Inanimate	
Dragon	2	2	Dangerous	Dangerous	1
Girl	1	1	Human	Human	1
Glass	5	4	Fragile	Edibles	
Lion	3	2	Animals	Dangerous	**
Man	1	3	Human	Animals	
Monster	2	1	Dangerous	Human	
Mouse	3	3	Animals	Animals	1
Plate	5	5	Fragile	Fragile	1
Rock	6	6	Inanimate	Inanimate	1
Sandwich	4	1	Edibles	Human	
Woman	1	3	Human	Animals	
				Number	
				Correct:	7/18

4.5.2 Verb Group

The generalization performance of the Verb Group is substantially lower than the noun group. Hierarchical cluster analyses show only 3 correct matches out of 11,

with the rest incorrect. This is not clearly different from chance performance. As explained above, categorical analysis is not possible for the verb group in this corpus, due to an absence of preexisting categories and a small sample size of verbs.

4.6 Discussion

Our results indicate that ‘propagation of grounding’, or context-mediated acquisition of semantic features, can and does take place. Many of the test words across the various networks acquired a sensorimotor representation *without direct training* that was sufficiently accurate that the word grouped precisely where it should have according to semantic similarity to other words. This occurred despite the fact that the training corpus involved was a small, not particularly rich sequence of short sentences. According to Landauer (Landauer & Dumais, 1997) this sort of statistical learning about word co-occurrences, which is similar to that involved in high-dimension models of meaning such as Latent Semantic Analysis (LSA) or Hyperspace Analogue to Language (Burgess & Lund, 2000), only becomes effective with large, complex corpora that provide a large number of relatively weak word-to-word connections as influences. Thus, larger, richer corpora, which use the words involved in more varied ways, should allow even more effective propagation of grounding than small, simple corpora like that used herein.

We can view these large number of weak word-to-word relationships as probabilistic constraints, possibly operating at many levels of linguistic complexity (word to word, word to sentence structure, word to verb construction, etc.) and the operation of the network as a constraint-satisfaction process that arrives at correct meanings for the new words (following Seidenberg and MacDonald, 2001). This is another reason to expect that this effect will scale well with larger corpora, since those more complex corpora should have more varied sentences, which will put words into a wider array of linguistic situations, and provide more simultaneous probabilistic constraints that will allow their meanings to be more fine-tuned. Of course, our current process, and our current feature dimension representations, are necessarily somewhat crude. We probably are not currently capturing with this architecture the more advanced forms of context effects such as verb construction similarity, or metaphoric imagery. The network is simply not that powerful at present. However, it suggests some directions for further modeling work.

It is important to note that the small size and lack of complexity of the current corpus might also contribute to the lack of significance of the propagation of grounding effect that we found for the verb group. While there are only a small number of nouns in this corpus, there are even fewer verbs, and fewer exemplars for each ‘category’ of verb. This may indicate that the size and complexity of the noun vocabulary and noun usage is a lower limit on the type of corpus which will exhibit the propagation of grounding effect. On the other hand,

we have already seen in previous work with these verb feature representations that while they do cluster by featural similarity and meaning, they do not cluster as well as nouns (Howell, Becker, and Jankowicz, submitted). It is possible that our verb feature representations are still not capturing enough of the meaning of verbs to allow indirect grounding of verbs to occur. One way of addressing this that we plan to implement in future work with these features is to integrate the noun and verb feature layers, such that the primarily sensory noun features and the primarily embodied and affordance-related verb features are combined. This would allow us to investigate the learning of both sorts of features by propagation, especially the learning of novel verb meanings from their occurrences with known nouns, as demonstrated in humans in Gillette et al. (1999). If nouns begin to acquire verb features due to their occurrences with known verbs, when a novel verb is encountered it may acquire these verb features from the nouns, in an indirect verb to noun to verb process.

Informal examination of the feature dimensions themselves indicates that in some cases, the individual values are a close match to those of the directly grounded word, but of course this does not happen for all features or for all words. A more rigorous analysis of the individual feature dimensions, how well each is learned through propagation of grounding, and what elements of the corpus affect which dimensions are learned is an important next step of this work, following along the constraint satisfaction viewpoint discussed above. For example, if a word occurs in sentences where it is the agent, or in which animacy

is clearly indicated, then propagation of grounded might be more effective for features related to animacy, such as speed, or intelligence. This would be due to the word occurring repeatedly in sentence structures in which many other animate words also occur, and so the novel word should acquire features like those words. If we bias a corpus to contain only sentences that provide certain linguistic contexts and constraints, then we should expect to see that only some aspects of featural meaning propagate to novel words introduced to that corpus.

This brings us to a possible criticism of our networks, namely the introduction of novel words. For simplicity's sake in the simulations herein, what we call a novel word was merely a word from the existing corpus which did not participate in the phoneme to meaning mappings. It existed in the corpus from the beginning, and was not introduced to a partially-trained network midway through training. We do not consider this to be a serious problem, since if anything, adding the novel word at a later point in training, when the network has already been able to learn the phoneme to meaning mappings for the fully grounded words, will only increase the speed with which the network is able to generalize features to this new word. The word sequence learning will be further along as well, allowing the network to “pay attention” to the novel word more, as it would be well on its way to learning the initial word's usage. This is actually reminiscent of experiments by Gillette, Gleitman, Gletman and Lederer (1999) regarding “bootstrapping” in language learning. They showed subjects sensory information about a scene (such as movies) along with various degrees of

utterances pertaining to that scene. Target words in that utterance, such as the verb for example, that subjects found impossible to identify at first, became clear when other words in the utterance were unmasked. For example, knowing which nouns occurred in an utterance allowed subjects to infer the verb that was taking place in the scene. A similar process should take place in our networks, where the earlier learning means that the network is more skilled at recognizing and predicting the early-learned words, and these ‘known’ words make it possible for the network to infer the meaning of a novel word occurring with them.

Another hypothesis regarding the importance of the early grounding of words relates to the possible attractor-dynamics or ‘foundational’ aspects of grounded conceptual representations, as we have suggested in previous work (Howell, Becker and Jankowicz, 2001; See also a discussion of attractor dynamics in language comprehension examined in detail in Tabor and Tannenhaus, 2001). It might be the case that the reliance of more abstract concepts and words on the dynamical attractors developed in previous learning of more embodied and grounded words is partially responsible for the critical period effect of language. Children who do not develop proper conceptual representations of the earliest, most basic words early in life, when cortex is most plastic, might not form sufficiently well-organized attractors in their cortical representations. Then later learning does not have accessible linking and ‘scaffolding’ points or a solid foundation on which to build, resulting in a relatively low level of overall linguistic attainment. Of course, the simulations in the present work are not

sophisticated enough to examine whether the learning of the meanings for the earliest words, or even pre-linguistic conceptual representations, are responsible for critical period effects in language, but it is an interesting avenue for future work.

On a technical note, the demonstration that propagation of grounding occurs has practical consequences for connectionist language researchers as well. If we want to gather sensorimotor feature representations for more or even all of the directly-grounded words that children can learn, do we have to create them manually for every word? The process of gathering human ratings of words along feature dimensions can be laborious (for details on different versions of the process see McRae, de Sa, & Seidenberg, 1997; Howell, Jankowicz and Becker, submitted; Vinson and Vigliocco, 2002). Happily this does not seem to be the case, since various methods of incrementally building a grounded lexicon in networks such as those discussed herein can be imagined, and we are presently implementing one. The general idea is to train a network on a progressive corpus of text (similar to Elman's (1993) 'Starting Small' argument) with new, novel words being introduced in much the same fashion they might in children's exposure to language. The new words will develop grounded featural representations through the process of propagation of grounding. When sufficiently well-learned (through having been experienced many times, and in many different contexts) the word's interim representation can be essentially 'added' to the directly grounded lexicon, and new novel words introduced into the

input stream. The new words will now have the previous novel words' grounded representations to learn with as well, and in such manner meaning can propagate to all words in the training corpus. This bootstrapping approach is another example of the potential importance of developmental constraints and experiential timing on learning (Elman, 1993), as they allow the learner to focus on the current crop of novel words, learn them, and then use them to learn future novel words. Learning of all of the words at once would be impossible, since there would be insufficient grounded words for novel words to borrow meaning from.

Finally, these results may be relevant to the debate between proponents of high-dimensional models of meaning (e.g. Landauer, & Dumais, 1997; Landauer, Laham & Foltz, 1998; Burgess & Lund, 1996) and those that advocate very embodied approaches to conceptual knowledge (e.g. Glenberg & Robertson, 2000, Barsalou (1999), Lakoff, 1987). The high-dimensional models of meaning such as LSA and HAL are impressive, having been shown to be able to perform human-like tasks such as similarity judgments, essay grading, etc. However, their 'features' are not clearly defined, having no meaning per se on their own. Thus examining the meaning of a word through its LSA or HAL representation is impossible, except by comparing it to other words. Clearly-defined feature representations like ours, however, have the advantage of providing information about the components of the meaning of a word. This componential meaning might be important in allowing for the types of affordance judgments that Glenberg and Robertson use to critique HAL and LSA, since it could allow

various different aspects of the meaning of a word to be highlighted in different contexts and situations, allowing for situation-appropriate judgments about affordances to be made on the fly.

Additionally, as Burgess and Lund point out, their HAL model using its smallest window size is essentially an SRN, like our architecture (Burgess & Lund, 2000). This implies that as the power and complexity of our simple network is increased somewhat, it may be able to reach approximately the same levels of performance as the high-dimensional models, while of course retaining the transparency of the individual feature meanings, and the ability to address questions of syntax (while LSA has been referred to as a ‘Bag of words’ approach). Furthermore, while LSA and HAL may be functional as models of adult processing, they do not explain how the system develops to its adult state. Our approach has the advantage of being rooted in the very basic, grounded learning of children’s first words, and in possibly providing a developmental mechanism, through the idea of propagation of grounding, for how the adult ‘steady-state’ may be reached, while retaining the ability to allow for judgments based on individual meaning components (i.e. affordances) of the meanings of the words.

In conclusion, we have demonstrated a simple version of the propagation of grounding phenomenon, and suggest a number of ways in which it might be expanded, and the implications those expansions could have for the modeling, and

the understanding, of children's and adult's language acquisition. We are presently investigating several of these areas.

Chapter 5

General Discussion

The goal of this dissertation has been to investigate the contribution of prelinguistic conceptual knowledge to the processes of language acquisition. The results from the six experiments reported in Chapter 2 suggest that we can create useful representations of this prelinguistic conceptual knowledge for nouns and verbs, and that having this knowledge improves sequence learning, an aspect of grammar. The results from the two experiments reported in Chapter 3 suggest that it may be the rich prelinguistic semantics that children possess that causes grammar to develop the way it does, which is evidence for an “emergent grammar” viewpoint. Finally, the results of Chapter 4 demonstrate that even words for which this prelinguistic conceptual knowledge is not available (ungrounded) can acquire this knowledge indirectly from relationships with other, grounded, words over the course of learning.

5.1 Sensorimotor Feature Representations

In chapter 2, I developed using human raters a set of sensorimotor feature representations intended to capture the same kind of knowledge that might be found in children's prelinguistic conceptual knowledge. Using hierarchical clustering analysis, self-organizing maps, and categorical analysis using Euclidian distance, I demonstrated that both noun and verb representations were capturing a meaning structure that corresponded with what we would expect based on human knowledge. However, this demonstration was clearer for nouns than for verbs. There are several possible reasons for this.

One is that the similarity spaces for nouns and verbs are inherently of differing complexity. Given the presence of similar results in the literature (Vinson and Vigliocco, 2002), and the contrast between nouns as having object referents and verbs as having event referents (which we would expect to be a more complex mapping), there is at least some validity to this explanation. However, a simpler explanation may be that the lesser clarity in the clustering analysis of verbs is due to the fact that there are fewer of them in our corpus. Since there were 352 grounded nouns, but only 90 grounded verbs, the similarity spaces are of different densities.

Of course, it could also be the case that the feature dimensions that were chosen to represent the range of verb meaning are themselves flawed. In creating these dimensions, I drew upon existing successes representing verb meaning, specifically representations of hand-action verbs (Bailey, 1997). However, the

usefulness of each verb-meaning dimension is not as well confirmed for these purposes as are the well-used noun-meaning dimensions, which are drawn from McRae, De Sa, & Seidenberg (1997). Also, it would be desirable to perform feature-by-feature statistical analyses, confirming or disconfirming the usefulness of each feature dimension to the distinctions of meaning that are required to be made.

However, this sort of analysis is a lengthy undertaking (Dr. Ken McRae, private communication, 2002) and since the creation of these features was a necessary first step to the remainder of this research program, lengthy delays could not be afforded. However, if these sensorimotor feature representations, or a later incarnation, are to be accepted and widely used, this sort of in-depth analysis of their dimensions and ratings will need to be performed. In fact, with the assistance of an undergraduate summer student, Yue Wang, I have recently completed the process of cross-comparing and adjusting all of the noun and verb feature values to compensate for any unusual subject ratings, and have begun the statistical analysis of the feature dimensions themselves, studying how much variance in the meaning categorization each accounts for. The completion of this process may tell us how much, if any, improvement is necessary or desirable in our verb features. Whatever the case, development of a set of featural representations is an ongoing, fine-tuning process, and to my knowledge I am the first to pursue this with the goal of capturing pre-linguistic knowledge.

Eventually, it would be very useful to validate these pre-linguistic features directly against children's knowledge, but that is another task entirely.

Another concern regarding the sensorimotor features is the duration of their applicability. It is important to note that the use of a static set of sensorimotor feature dimensions, while eminently reasonable when considering pre-linguistic knowledge and the earliest stages of language acquisition as discussed herein, might begin to be too limiting at later stages of language learning. That is, once an older child is learning words indirectly from context, as in Chapter 4, and is processing enough speech and text (as perhaps in school age children) to be making the kinds of direct linguistic-linguistic relationships that we specifically restricted in our prelinguistic sensorimotor features, the dimensions of semantic meaning may need to be expanded. While sensorimotor dimensions may remain the core of a broader semantic representation, additional frequent linguistic or contextual relationships may become common enough to be “featurized”, and be able to take place in processes such as the propagation of grounding. The process of extension of basic sensorimotor features of meaning to a broader, mature semantic representation is a significant part of our call for future work into bridging the embodied-meaning/high-dimensional meaning gap, as discussed in chapter 4.

5.2 Sensorimotor Representations' Effect on Grammar

Learning

Also in chapter two, I used the above-discussed features in neural network models of language acquisition, to study the contribution of these features to simple aspects of grammatical learning. I found a significant effect of the inclusion of these features; they assist grammar learning by 17% over a control condition with random features (equated for range and distribution of values). These experiments were performed using the most naturalistic corpus I could find, a concatenation of mother-to-child speech taken from the ChildDes database (Bates, et al., 1997; MacWhinney, 2000; Carlsen-Luden, 1979). This naturalistic training corpus was intended to allow for good input representativeness (Christiansen & Chater, 2001) on the part of the model, the idea being that it is receiving the same sort of input that the child might, and so the model's results would be extendable more readily to the case of child language acquisition. This goal certainly still holds; we *can* be confident that this effect of pre-linguistic conceptual knowledge on grammatical learning should appear in children as well as the network. At an extreme level, it is obvious that it has to. If a child has no meaning representations for any of the words that he or she is hearing in speech, then the grammar of the language will be impossible to learn. This is McClelland's "learning a language by listening to the radio" criticism (Elman, 1990). Thus, the more words whose meaning is known that occur in the speech stream, the more the grammar is inferable, and the more easily that novel words can be understood.

Experimental evidence of this process has been discussed earlier (Gillette, Gleitman, Gleitman, and Lederer, 1999) and is addressed directly for novel words in chapter 4 of this dissertation.

However, the naturalistic corpus that I used for these simulations was *not* a very grammatical corpus! Upon examination, the mother-to-child speech contains very few proper sentences, and very many partial sentence fragments, repetitions of words, attention-eliciting verbal behaviors (e.g. “look at this, what is this”, etc.), and as mentioned previously, only a minority of the words were grounded in sensorimotor features. While this may make us confident in generalizing from the network’s behavior to that of children, it makes it very difficult to produce that network behavior in the first place, hence the small sizes of the effects found in Experiment 4 of chapter 2. In fact, by including only mother-to-child speech, this training corpus is in fact much more impoverished than children would be exposed to, since they also overhear more grammatical adult-to-adult speech. I would thus expect larger differences between experimental and control conditions to be evident with training corpora that were more grammatical and more completely grounded.

5.3 Lexicon to Grammar Effects and Emergent Grammar

In chapter three I continue an investigation into the well-known phenomenon in children of the lexical-grammatical correlation over time. I have addressed this previously in a paper published in a conference proceedings (Howell and Becker,

2001). In the chapter I investigate this further, using techniques increasingly more analogous to those used with children. Specifically, the simulations used in chapter three incorporate phonemic input representations, similar to those children would experience, rather than localist lexical representations. Also, the sensorimotor features that are used in the chapter are far more appropriate to children's experience than were the simpler, more artificial features used in the previous work.

The results reported in chapter three demonstrate that the richness of the word representations in children's lexicons is directly related to their later grammatical competence. Specifically, when the lexical representations are meaningful sensorimotor semantic ones, much higher correlations (0.7 in Experiment 1, 0.8 in Experiment 2) between lexical accuracy and later grammatical accuracy are found than in conditions where the lexical representations are less structured and more random. These correlations are very high, and provide evidence for the position that grammar emerges from the operation of a semantically rich lexicon. There are several possible criticisms to address, however. As discussed previously, the measure of lexical and grammatical performance used with the networks differs from that used in measuring children, being accuracy of performance on a limited lexicon rather than overall size of the lexicon. However, in fact these are really just transformations of one another. Parents typically provide the measures of the sizes of children's lexicons, and they do so by recalling via a checklist which

words their child uses properly. Our measure of lexical accuracy is essentially an assessment of how properly the child uses a word. The likelihood that a parent would rate a child as knowing the word will be a function of how properly that child uses it. Thus the two measures will be closely related, and should not hinder generalization of the network results to children.

Furthermore, in children the grammatical measure most commonly used is a measure of production (mean length of utterance), since it is impossible to get inside the child's head and assess his or her grammatical competence. In network simulation this situation is reversed, such that it is easier to get a measure of grammatical comprehension than production. However, the two measures will of course be closely related, and in fact the comprehension measure may well be the purer of the two, as it is uncontaminated by behavioral variance that may mask the expression in production of grammatical comprehension competence. Again, this should not hinder generalization of the network results to children.

5.4 Propagation of Grounding

Chapter four contains perhaps the most interesting results of the dissertation. Specifically, in chapter 4 I demonstrate that when the network is exposed to a novel word occurring in a well-structured context, it is able to 'infer' the meaning of that novel word from its context, and produce the correct sensorimotor features for the novel word. This is very similar to the way that the meanings of novel words might be learned by children while reading or listening to speech, and is to

my knowledge the first demonstration of a candidate process for the indirect inference of the meanings of novel words. Of course, in the network's case, this process is purely a statistical one, with word co-occurrence statistics slowly building up a similarity between known, grounded words, and novel words, until values on the feature dimensions of the novel word come to be similar to those on the known, similar word. A full discussion of the impact of this process is provided in the chapter, but there are additional criticisms that need to be addressed. First, it might be suggested that it is not the overlap of the two tasks (lexical and grammatical learning) through a common hidden layer that causes the propagation of grounding. Perhaps the prediction task that is being performed to learn grammar is irrelevant to this effect, and a network without the Word Prediction output layer would still show this generalization from known word to novel word, based on some other factor? One factor might be simple similarity of the meaning of a novel word to the grand mean of all the words in the lexicon. That is, if most words are animals, and an unknown word is presented, it is more likely that the features corresponding to an animal will be randomly activated as an output activation than anything else. An examination of which words propagate and which do not from Table 4.2 will show that this is not the case.

Another possibility is that it is something about the overlap of the phonemic representations that causes the generalization to novel words (Dr. Art Glenberg, private communication, 2004). Again, Table 4.2 does not show any clear relationship between the phonology of any test word and the rest as an

influence on when propagation will occur. However, to investigate both this criticism and the preceding one more closely, I ran a number of pilot control simulations of words that *did* show propagation of grounding when used as novel words in chapter 4. In these controls, I completely eliminated the Word Prediction Layer, ran the network for the same number of epochs (50) and then ran the hierarchical cluster analysis. None of the tests showed any hint of propagation of grounding for the new words.

5.5 “Facilitative Interference”

Thus, it does seem to be the overlap and facilitation found between the lexicon to meaning mapping Noun Feature output layer and the Word Prediction output layer that is responsible for the propagation of grounding effect. This constitutes, if anything, the reverse of the usual catastrophic interference effect that critics of connectionist modeling often raise as a limit on their usefulness. Rather than one task interfering with the other, there is mutual facilitation (Experiment 4, Chapter 2) allowing learning to generalize (Chapter 4). Further investigation of these kinds of effects in models of language acquisition may be most fruitful.

5.6 Model Limitations

The network architecture used in the simulation experiments herein, while successful at modelling these phenomena, does have other limitations. First, the learning algorithm that it uses, backpropagation of error, is not considered to be very biologically plausible. That is, we do not think that neurons in the brain

actually learn in this fashion, nor that they receive explicit teaching signals.

However, these same sorts of architectures can be implemented in more biologically plausible ways, by algorithms based on the more biologically plausible Hebbian learning principle.

Related to this is the issue of neural representation. How exactly does the brain represent and process the information and mechanisms that our neural network simulations are analogues to? Our simulations are too abstract at this point, too behaviourally oriented, to be able to make claims about how or where in the brain these processes occur in humans. Perhaps the closest to a neurological claim we can make relates to the representation of sensorimotor information. As Lakoff (1987), Barsalou (1999), and others have suggested, the meanings of words may involve not just semantic association areas. Rather, the meanings of words are represented in a distributed fashion across a wider area of cortex, including sensory areas and motor areas. This, as alluded to in Chapter 2, is part of the motivation for our use of sensorimotor features specifically, as opposed to any other sort of meaning features more generally.

Still, whether something like the prediction task is actually represented in the brain, or how exactly the brain handles sequential grammar-like learning, awaits future research with more complex methods. Generally speaking, our models operate as behavioral analogues, not direct neurological analogues, although we do try to make as close a fit as we are able at this time, in general terms.

However, one interesting line of neurological work does support the use of the “violation of expectations” sort of learning that is incorporated in the prediction task. Connolly and colleagues (e.g. D’Arcy et al, 2004; Newman et al, 2003) have identified, using brain imaging techniques, the time course of semantic and phonological processing. Using event related potential methods (ERP) in particular, they have shown that the brain produces a spike of activity for violations of phonological and semantic expectations. The two are doubly dissociable, with the phonological signal occurring earlier than the semantic. The phonological mismatch signal is known as the PMN, or phonological mismatch negativity, and occurs at about 250 ms post-stimulus. It has a frontal, right-asymmetrical scalp topography. The semantic mismatch is an n400 wave, a negative spike at about 400 ms post-stimulus, which has a centro-parietal scalp topography. The fact that two such mismatch signals have been identified in human language processing thus makes the use of violation of expectations prediction learning much more plausible, whether we are predicting phonology or semantics.

Of course, there are many aspects of language acquisition that are important, but outside the scope of this research, as must be the case when studying a phenomenon as complex and multi-faceted as language. Acquisition of phonology is a subfield of its own, and one that we do not address in this work, except to borrow a set of idealized phonemes for use as input representations. The aspects of grammar that we are able to address with these networks are very

simple ones, essentially word-to-word correlations and sequence learning. The degree to which these sorts of models can capture more complicated aspects of syntax and grammar is a topic for future investigation.

Also, issues such as varying presentation rates of input are abstracted away in our models, which enforce a discrete timestep-by-timestep rate of presentation, rather than a more realistic continuous-time representation. Algorithms for continuous time models are now available, but do require additional computer processing power to employ. Still, this is a valuable avenue for future work. Even with discrete-time networks like those used herein, however, there is an issue of what level of language input will be represented at each time step. In our simulations, each time step corresponds to one word, and the phonemes of that word are presented all at once. It is entirely possible to choose to have each time step represent one phoneme, and thus a word would take a series of time steps to be represented via a sequence of its phonemes. In unpublished work we have used networks like these to investigate word segmentation and intraword and between-word errors. However, when we are interested in both lexical *and* grammatical phenomenon, as we are in the simulations described herein, it is more practical to use the word as the level of input representation. It would be useful to develop a multi-leveled network which used inputs which were a sequence of phonemes (as would be encountered in speech), and yet was able to develop a word level representation, and from that

perform the word-prediction task necessary to study grammar. This, however, is a more difficult modelling task that remains to be solved adequately.

Importantly, even if such a hierarchical, multi-leveled model were developed, it would still be an abstraction in its own way, since the idealized phonemes used at input are not the sort of coarticulated and acoustically varying phonemes that children actually encounter. Unless we attempt to model an entire brain, something that is presently far beyond our computational power as well as our understanding of neuroscience, our neural network models will always have to abstract away some part of the complexity of the task. However, so long as we abstract away the least relevant aspects for the task at hand, our models should still generalize to human behaviour and human neural processing.

5.7 Conclusion

The results reported in this dissertation provide evidence for the importance of pre-linguistic conceptual knowledge (herein operationalized as sensorimotor semantic features) in facilitating language acquisition. I report converging evidence from several lines of investigation, including examination of the effect of sensorimotor representations on simple grammar learning, investigations of the strong correlations between early lexical learning and later grammatical learning, and demonstrations of how grounded meaning can be acquired indirectly for novel words via a process of ‘propagation of grounding’ from known, grounded

words. Sensorimotor, embodied concepts thus facilitate and ground the processes of word and grammar learning, throughout the course of language acquisition.

References

- Bailey, D. R. (1997). When push comes to shove: A computational model of the role of motor control in the acquisition of action verbs. Ph. D Thesis, University of California at Berkeley
- Bailey, D., Feldman, J., Narayanan, S, and Lakoff, G. (1997). Modelling Embodied Lexical Development. Proceedings of the Cognitive Science Society, 1997.
- Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in 9 month old infants. Cognition, 20, 191-208.
- Barsalou, L.W. (1999). Perceptual symbol systems. Behavioral and Brain Sciences, 22, 577-609.
- Bates, E., Bretherton, I., & Snyder, L. (1988). From first words to grammar: Individual differences and dissociable mechanisms. Cambridge, MA: Cambridge University Press.
- Bates, E., & Elman, J.L. (1993). Connectionism and the study of change. In M.H. Johnson (Ed.), Brain Development and Cognition: A Reader. Oxford: Blackwell Publishers. Pp. 623-642.
- Bates, E., & Goodman, J. C. (1999). On the emergence of grammar from the lexicon. In MacWhinney, B. (Ed.) (1999). The Emergence of Language. New Jersey: Lawrence Erlbaum Associates.

- Becker, S. (1999) "Implicit learning in 3D object recognition: The importance of temporal context". Neural Computation, 11(2): 347-374.
- Bloom, P. (2000). How children learn the meanings of words. Cambridge: Cambridge University Press
- Brown, R., & Hanlon, C. (1970). Derivational complexity and the order of acquisition in child speech. In R. Brown (Ed.), Psycholinguistics. New York: Free Press.
- Burgess, C., & Lund, K. (2000). The dynamics of meaning in memory. In Dietrich & Markham (Eds.) Cognitive Dynamics: Conceptual change in humans and machines.
- Carlson-Luden, V. (1979). Causal understanding in the 10-month-old. Unpublished doctoral dissertation. University of Colorado at Boulder.
- Chomsky, N., (1988). Language and problems of knowledge: The Managua Lectures. Cambridge, MA: MIT Press.
- Christiansen, M. H. & Chater, N. (2001). Connectionist Psycholinguistics. Westport, CT: Ablex Publishing
- CMU (1995). Carnegie Mellon Pronouncing Dictionary. Available at <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>).
- D'Arcy R. C., Connolly J. F., Service, E., Hawco, C. S., Houlihan, M. E. (2004). Separating phonological and semantic processing in auditory sentence processing: a high-resolution event-related brain potential study. Human Brain Mapping, 114(4), 662-672.

- Elman, J. L. (1990). Finding structure in time. Cognitive Science, 14, 179-211.
- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. Machine Learning, 7, 195-224.
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. Cognition, 48, 71-99.
- Elman, J. L. (1995). Language as a dynamical system. In R.F. Port & T. van Gelder (Eds.), Mind as Motion: Explorations in the Dynamics of Cognition. Cambridge, MA: MIT Press. Pp. 195-223.
- Elman, J.L. (1998). Generalization, simple recurrent networks, and the emergence of structure. In M.A. Gernsbacher and S.J. Derry (Eds.) Proceedings of the Twentieth Annual Conference of the Cognitive Science Society. Mahwah, NJ: Lawrence Erlbaum Associates.
- Elman, J. L. (1998a). Connectionism, artificial life, and dynamical systems: New approaches to old questions. In W. Bechtel and G. Graham (Eds.) A Companion to Cognitive Science. Oxford: Basil Blackwood.
- Elman, J.L. (1999). The emergence of language: A conspiracy theory. In B. MacWhinney (Ed.) Emergence of Language. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D. and Plunkett, K. (1996). Rethinking Innateness. Boston: MIT Press

- Fenson, L., Pethick, S., Renda, C., Cox, J.L., Dale, P.S., & Reznick, J.S. (2000). Short form versions of the MacArthur Communicative Development Inventories. Applied Psycholinguistics, 21, 95-115.
- Feldman, J. A. (1989). Neural representation of conceptual knowledge. In Lynn Nadel et al., Neural Connections, Mental Computation, pp 68-103. Cambridge, Mass.: MIT Press.
- Gillette, J., Gleitman, H., Gleitman, L., Lederer, A. (1999). Human simulations of vocabulary learning. Cognition, 73, 135-176.
- Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. Journal of Memory and Language, 43, 379-401.
- Gold, E.M. (1967). Language identification in the limit. Information and Control, 10, 447-474.
- Goldberg, A. (1995). Constructions: A construction grammar approach to argument structure. Chicago and London: University of Chicago Press.
- Goldberg, A. (1999), The emergence of argument structure semantics, in B. MacWhinney, ed., The Emergence of Language, New Jersey: Lawrence Erlbaum Associates.
- Hare, M., & Elman, J.L. (1994). Learning and morphological change. Cognition.
- Hare, M., Elman, J.L., & Daugherty, K.G. (1995) Default generalization in connectionist networks. Language and Cognitive Processes, 10, 601-630.

Hinton, G. E. & Shallice, T. (1991). Lesioning a connectionist network:

Investigations of acquired dyslexia, Psychological Review, 98, 74-75.

Hinton G. E. & Sejnowski, T. J. (1986). Learning and re-learning in Boltzmann machines. In J. L. McClelland, D. E. Rumelhart and the PDP Research Group, Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations (pp. 282-317). Cambridge, MA: The MIT press.

Howell, S. R. & Becker, S. (2000). Modelling language acquisition at multiple temporal scales. Proceedings of the 22nd Annual Conference of the Cognitive Science Society, 2000, 1031.

Howell, S. R., & Becker, S. (2001) Modelling language acquisition: Grammar from the lexicon? Proceedings of the 23rd Annual Conference of the Cognitive Science Society Conference, 2001, 429-434

Howell, S. R. & Becker, S. (Submitted). Grammar from the Lexicon: Evidence from Neural Network Simulations of Language Acquisition, Cognitive Science.

Howell, S. R. & Becker, S. (Submitted b). Grounding Words in Meaning Indirectly – A Computational Model of the Propagation of Grounding, Journal of Memory and Language.

Howell, S. R. & Becker, S. (Submitted c). SRNEngine: A Windows-based neural network simulation tool for the non-programmer. Behavior Research Methods, Instruments, and Computers.

- Howell, S. R., Becker, S., & Jankowicz, D. (2001). Modelling language acquisition: Lexical grounding through perceptual features, Proceedings of the 2001 Developmental and Embodied Cognition Conference, July 31, 2001.
- Howell, S. R., Jankowicz, D., & Becker, S. (Submitted). A Model of Grounded Language Acquisition: Do Sensorimotor Features Improve Grammar Learning? Journal of Memory and Language.
- Howell, S.R., Schmidt, L. A., Trainor, L.J., and Santesso, D.L. (2002). Neural Network Categorization of Infant Emotional States., Presented at the 13th Biennial International Conference on Infant Studies, Toronto, Ontario.
- Howell, S.R., Trainor, L.J., and Sonnadara, R. (2002). Neural Network Categorization of Infant-directed and Adult-directed Emotional Speech, Presented at the 13th Biennial International Conference on Infant Studies, Toronto, Ontario.
- Jordan, M.I. (1986). Serial order: A parallel distributed processing approach. Institute for Cognitive Science Report 8604. University of California, San Diego.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. Biological Cybernetics, 43, 59-69.
- Kohonen, T. (1995). Self-organizing maps. Berlin: Springer-Verlang.
- Lakoff, G. (1987). Women, fire and dangerous things: What categories reveal about the mind. Chicago and London: University of Chicago Press.

- Lakoff, G., and Johnson, M. (1980). Metaphors we live by. Chicago and London: University of Chicago Press.
- Lakoff, G. and Johnson, M. (1999). Philosophy in the flesh: The embodied mind and its challenge to western thought. New York: Basic Books.
- Landauer, T. K. & Dumais, S.T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. Psychological Review, 104, 211-242.
- Landauer, G. T., Laham D. & Foltz, P. (1998). Learning Human-like knowledge by singular value decomposition: A progress report.
- Lang,, K. J., Waibel, A. H., and Hinton, G. E. (1990). A time-delay neural network architecture for isolated word recognition, Neural Networks, vol. 3, no. 1, 33-43.
- Langer, J. (2001). The mosaic evolution of cognitive and linguistic ontogeny. In Bowerman, M., & Levinson, S. C. (2001). Language acquisition and conceptual development. Cambridge: Cambridge University Press.
- Levin, B. (1993). English Verb Classes and Alternations. Chicago and London: The University of Chicago Press.
- Macho, S. (2002). Cognitive modeling with spreadsheets. Behavior Research Methods, Instruments, & Computers, vol. 34, no. 1 (February), 19-36.
- MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Mandler, J. M. (1992). How to build a baby II: Conceptual primitives.

Psychological Review, Vol. 99, No. 4, 587-604.

McClelland, J.L. & Elman, J.L. (1986). Interactive processes in speech

perception: The TRACE model. In J. L. McClelland, D. E. Rumelhart and

the PDP Research Group, Parallel Distributed Processing: Explorations in

the Microstructure of Cognition. Vol. 2: Psychological and biological

Models (pp. 122-169). Cambridge, MA: The MIT press.

McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of

context effects in letter perception: part 1. An account of basic findings.

Psychological Review, 88, 375-407.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of

featural representations of word meaning. Journal of Experimental

Psychology: General, 126, 99-130.

Mozer, M.C. (1987). Early parallel processing in reading: a connectionist

approach. In M. Coltheart, (Ed.) Attention and Performance, vol. XII: The

Psychology of Reading (pp. 83-104). Hove: Lawrence Earlbaum

Associates

Narayanan, S. (1995). Moving right along: A computational model of

metaphoric reasoning about events. Ph. D Thesis, University of California

at Berkeley.

- Newman, R. L., Connolly J. F., Service, E., McIvor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. Psychophysiology, 2003, 40(4), 640-647.
- Piaget, J. (1952). The origins of intelligence in children. Madison CT: International Universities Press.
- Pinker, S. (1994). On Language. Journal of Cognitive Neuroscience, 6(1), 92-97.
- Pinker, S. (1995). The language Instinct: How the mind creates language. New York: Harper-Collins
- Regier, T. (1996). The human semantic potential: Spatial language and constrained connectionism. Cambridge, Mass.: MIT Press.
- Rogers, T. T. & McClelland, J.L. (2003). The parallel distributed processing approach to semantic cognition, Nature Reviews Neuroscience, Vol 4 (April), 310-322.
- Rohde, D.L.T. & Plaut, D.C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small?, Cognition, 72, 67-109
- Rumelhart, D.E., Hinton, G. E. & Williams, R. J. (1986) Learning internal representations by error propagation. In J. L. McClelland, D. E. Rumelhart and the PDP Research Group, Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations (pp. 318-362). Cambridge, MA: The MIT press.

- Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month old infants. Science, *274*, 1926-1928.
- Seidenberg, M. S., & MacDonald, M. C. (2001). Constraint satisfaction in language acquisition and processing. In Christiansen, M. H. and Chater, N. (eds) Connectionist Psycholinguistics, (pp. 177-211). Westport, Ct.: Ablex Publishing.
- Seidenberg, M.S., & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. Psychological Review, *96*, 523-568.
- Smith, L. B. (1999). Children's noun learning: How general learning processes make specialized learning mechanisms. In MacWhinney, B. (Ed.) (1999). The Emergence of Language. New Jersey: Lawrence Erlbaum Associates.
- Spelke, E. S. (1994). Initial Knowledge: Six suggestions. Cognition, *50*, 443-447.
- Tabor, W. & Tannenhaus, M. K. (2001). Dynamical systems for sentence processing. In Christiansen, M. H. and Chater, N. (eds) Connectionist Psycholinguistics, (pp. 177-211). Westport, Ct.: Ablex Publishing.
- Vihman, M. M. (1996). Phonological Development. Cornwall: TJ Press Ltd.
- Vinson, D., & Vigliocco, G. (2002). A semantic analysis of grammatic class impairments: Semantic representations of object nouns, action nouns and action verbs, Journal of Neurolinguistics, Vol 15(3-5), 317-351.

- Waibel, A., & Hampshire, J. (1989) Building blocks for speech. Byte, August, 235-242.
- Waibel, A., Hanazawa, T., Hinton, G. E., Shikano, K. & Lang, K. J. (1989a). Phoneme recognition using time-delay neural networks. IEEE Transactions of Acoustics, Speech, and Signal Processing, 37, 328-339.
- Weckerly, J., & Elman, J.L. (1992). A PDP approach to processing center-embedded sentences. In Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.
- Wiles, J., & Elman, J. L. (1995). Learning to count without a counter: A case study of dynamics and activation landscapes in recurrent networks. Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society. Cambridge, MA: MIT Press.
- Williams, R. & Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks, Neural Computation, 1(2), 270-280.

Appendix A: Forms and Instructions for Experiment 1, Chapter 2

Attached are the instructions and rating forms used in the feature rating experiment from Experiment 1 in chapter 2. These were originally presented as web pages, since participants completed the experiment on-line via the Web.

Building Sensorimotor Semantic Representations from Human Knowledge - Noun Feature Ratings

On the following pages are a series of various concepts or words, such as "dog", or "kettle". For each of the concepts/words, there is a list of *features*. Please rate *each* concept on *each* feature on a scale of 0 to 10. Try to picture the object or concept mentally as you are making your rating, including its sounds, smells, motions, etc.

In Part I, the feature ratings are between two polar opposites, such as *size: small - large*. In this case a rating of 5 would be average, or size medium; midway between the two extremes. A rating of 1 might be small, a rating of 9, large, while ratings of 0 and 10 would be the *tiniest* and *largest*, respectively, things that you would see in everyday life.

Try to be consistent in your ratings from concept to concept. For example, don't rate "cat" a 7 in intelligence, and then "man" a 4, since cats are clearly not more intelligent than people. However, don't worry if you can't remember exactly your ratings from concept to concept.

In Part II, in contrast, these ratings should be viewed as percentages (0% to 100%). For each concept, rate the percentage of the time that they exhibit this feature. For example, for the concept "apple", the feature "is_gold" might get a rating of 1, indicating that less than 10% of the apples we see might be gold (most are red or green). A 9 indicates the almost definite presence of the feature; for example for the concept "ball", the feature "is_round" might be rated a 9, indicating that at least 90% of the balls that we see are round (some, like a football, are not). A rating of 5 on these features indicates uncertainty as to the presence of the feature, 50% of the time it is there, 50% of the time it is not. A rating of 6 would be 60% present, 40% absent, etc. A rating of 0 means that the concept *never* has the feature, and a rating of 10 means that the concept *always* has the feature. Features may be rated as present even if they are only in one part of a concept, as well. An oilcan might get 9's on both is_conical and is_cylindrical, since its different parts are different shapes (body vs. spout).

Thus in both Part I and Part II you should assign relatively few 0 or 10 ratings on any of the features. To know where to draw the line on these extreme ratings, you should try to limit yourself to the knowledge of the world that an average pre-school child would have. For example, for size, don't compare the concept in question to a microscopic bacteria or to a mountain. You might limit your comparison group to anywhere from the size of a pea (tiny = 0) on up to the size of a house (extremely large=10), for example.

(The concept "is_gold" for apple in the example above is only a 1 and not a 0 because it is remotely conceivable that someone would paint an apple gold, and hence this feature would be true for *that* apple. This is an example of why you should be wary of the extreme 0 and 10 ratings.)

Click Proceed to begin the experiment. Remember, to receive credit you must complete 10 concepts. This will mean 10 passes through the web form, selecting the next concept each time. Each time you submit the data for one concept, you will be linked back again.

[Proceed!](#)

Student ID (For Experimental Credit):

Below is a drop-down box containing all the words in the experiment, in alphabetical order. You must rate only 10 of them, the 10 that were assigned to the phase you signed up for and that you wrote down previously. You will thus have to go through this form 10 times, crossing a word off of your list of 10 each time until you are done (so you don't forget and rate a word twice). Each time you complete a concept, you will go to a completion page, where you can link back to this form if you have not completed all 10 of your words yet.

Please try to imagine the concept that you have selected, and keep that image in mind throughout the ratings. Try to imagine all aspects of the sensory experience of that concept or object. That is, vision, hearing, touch, smell, and if applicable, taste. Do not spend too much time agonizing over the ratings, however, just pick what seems reasonable and keep going.

Part I

Please enter a value between 0 and 10, with 5 being in the middle

of the two opposites.

0-10

Size (small-large)



Weight (light-heavy)



Strength (weak-strong)



Speed (slow-fast)



Temperature (cold-hot)



Cleanliness (dirty-clean)



Tidiness (messy-tidy)



Brightness (dark-light)



Noise (silent-noisy)



Intelligence (stupid-smart)



Goodness (bad-good)



Beauty (ugly-beautiful)



Width/thickness/fatness (thin - thick/fat)



Hardness (soft/pliable - stiff/hard)



Roughness (smooth-rough)



Height (short-tall (Vertical))



- Length (short-long (Horizontal))**
- Scariness (nonthreatening-frightening)**
- Colourfulness (drab-colourful)**

Please enter a value between 0 and 10. A value of 10 means the feature is ALWAYS true or ALWAYS present, a value of 0 means that it is NEVER true or present, and a value of 5 means that it is true about 50% of the time, or for 50% of the instances of that concept. Example: if you think that 60% of the time an apple is red, then rate apples a 6 on is_Red.

Color:

0-10

- is_black**
- is_blue**
- is_brown**
- is_gold**
- is_green**
- is_grey**
- is_orange**
- is_pink**

is_purple

is_red

is_silver

is_white

is_yellow

Shape:

0-10

is_conical

is_crooked

is_curved

is_cylindrical

is_flat

is_liquid

is_rectangular

is_round

is_solid

is_square

is_straight

is_triangular

Surface Texture:

0-10

has_feathers

has_scales

has_fur/hair

is_prickly

is_sharp

is_breakable

Substance:

0-10

made_of_china

made_of_cloth

made_of_leather

made_of_metal

made_of_plastic

made_of_stone

made_of_wood

Locomotion:

0-10

climbs

crawls

flies

leaps

runs

swims

Actions:

0-10

breathes

drinks

eats

makes_animal_noise

sings

talks

Component Parts:

0-10

has_4_legs

has_a_beak

has_a_door

has_a_shell

has_eyes

has_face

has_fins

has_handle

has_leaves

has_legs

has_paws

has_tail

has_teeth

has_wheels

has_whiskers

has_wings

Judgments:

0-10

is_annoying

is_comfortable

is_fun

is_musical

is_scary

is_strong_smelling

is_young

is_old

is_comforting

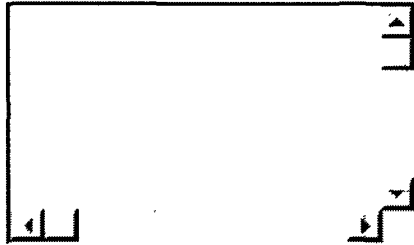
is_lovable

is_edible

is_delicious



Please comment on any features that you found particularly difficult or confusing for this concept (if any)



Indicates Response Required

Submit

Appendix B: Cluster Analysis of 352 Nouns from Chapter 2

For those unfamiliar with reading Hierarchical Cluster Analyses, the following figures are a tree structure. For any concept, follow the branches from the left to the right. The lowest level category including that concept is defined by the first + encountered. E.g. for Puzzle, the concepts Picture and Book share its category. Continuing to the right, Puzzle, Picture and Book join with Present in the next higher-level category, then Necklace joins in at the next level, etc.

Dendrogram using Average Linkage (Between Groups)

* * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * *

		Rescaled Distance Cluster Combine					
C A S E		0	5	10	15	20	25
Label	Num	+-----+-----+-----+-----+-----+					
KLEENEX	163	-+-----+					
TISSUE	310	-+ +--					
NAPKIN	196	-----+ +-----+					
PAPER	210	-----+ +-----+					
FLAG	115	-----+-----+		I			
PICTURE	223	-----+-----+		I			
PUZZLE	247	-----+-----+		I			
BOOK	34	-----+ +--		I			
PRESENT	241	-----+ +--		I			
NECKLACE	197	-----+ +--		I			
MONEY	184	-----+-----+		I			
BIB	28	-----+-----+ +-----+					
DIAPER	89	-----+-----+		I I		I	
BOOTS	35	-----+-----+ +-----+		I I		I	
PURSE	246	-----+-----+ I I I		I		I	
BLANKET	31	-----+-----+ I I I		I		I	
SWEATER	301	-----+ +-----+ I +--		I		I	
SCARF	261	-----+-----+ I I I		I		I	
TOWEL	316	-----+-----+ I I I		I		I	
PILLOW	225	-----+-----+ I I		I		I	
JEANS	156	-----+-----+ I I I		I		I	
PANTS	209	-----+ I I I I		I		I	
BATHROOM	16	-----+ I I +--		I		I	
SLIPPER	277	-----+-----+ +-----+				I	
SOCK	282	-----+ I I I		I		I	
UNDERPANTS	331	-----+ I I I I		I		I	
PAJAMAS	207	-----+ I I I		I		I	
SHORTS	267	-----+ I I I I		I		I	
TIGHTS	309	-----+ I I I		I		I	
HAT	144	-----+ I I I		I		I	
SHIRT	265	-----+ +-- I I		I		I	
SHOE	266	-----+ I I I		I		I	
GLOVES	131	-----+ I I I		I		I	
MITTENS	182	-----+ I I I		I		I	
BELT	26	-----+ +-- I I		I		I	
JACKET	154	-+ +-- I I		I		I	
COAT	74	-----+ I I I		I		I	
DRESS	100	-----+ I I		I		I	
SNOWSUIT	280	-----+ I		I		I	
SNEAKER	278	-----+-----+				I	
BENCH	27	-----+-----+				I	
TABLE	303	-----+ +--		I		I	

* * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * *

C A S E	0	5	10	15	20	25
Label	Num	+-----+-----+-----+-----+-----+				
ROCKING-CH	253	-----+ I	I			
COUCH	80	-----+-----+ +-----+ I				
SOFA	284	-----+ ++	I I			
CHAIR	60	-----+-- I I	I I			
HIGHCHAIR	148	-----+ +-- I	I I			
PLAYPEN	231	-----+ I	I I			
CRIB	85	-----+-----+ I I				
BASKET	14	----+-----+ I I				
PICNIC	222	----+ I	I I			
SANDBOX	258	-----+-----+ I I				
BACKYARD	11	-----+-- I	I I			
YARD	346	-----+ ++ +--	I I			
COUNTRY	81	-----+-- +-- I	+--			
PARK	211	-----+ I I I	I I			
BEACH	18	-----+-----+ I I	I I			
DOWNTOWN	97	-----+-- I I	I I			
FARM	109	-----+ +-- I	I I			
OUTSIDE	203	-----+ I	I I			
PLAYGROUND	230	-----+-- +-- I I				
PORCH	237	-----+ I I I I I				
BASEMENT	13	-----+ I I I I I				
ROOM	255	-----+-- +-- I I I I I				
BEDROOM	23	-----+ I I I I I I I				
LIVINGROOM	174	-----+ I I I I I I I				
KITCHEN	161	-----+-- I I +-- I				
STORE	294	-----+ I +-- I +-----+				
BED	22	----+-----+ I I I I I				
CHURCH	68	----+ I I I I I I				
HOME	149	-----+-- I I I I I				
HOUSE	151	-----+ I I I I I				
SCHOOL	262	-----+-----+ I I I				
POOL	233	-----+-----+ I I				
DISH	90	-----+-----+-- I				
HOSE	150	-----+-----+ I I				
MOVIE	193	-----+-- I I				
TV	329	-----+ +-----+ I I				
RADIO	248	-----+ +-----+ I I				
TAPE	304	-----+-----+ I I I				
PENCIL	215	-----+-----+ I I				
STICK	292	-----+-----+ I I I				
APPLESAUCE	6	-----+-- I I I				
FORK	118	-----+ +-- I I I				
SCISSORS	263	-----+-- I I I I				
ZIPPER	351	-----+ +-- I I I				
KEYS	160	-----+ +-- I I I				

* * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S *

C A S E	0	5	10	15	20	25
Label	Num	+-----+-----+-----+-----+-----+				
NAIL	195	-----+ I	+-+ I I	I I		I
KNIFE	165	-----+ I	I I I I			I
SPOON	287	-----+ I	I I I			I
BATHTUB	17	-----+-----+ I	I I I			I
SHOWER	270	-----+ +--	I I I I			I
SINK	272	-----+ I	I I I I			I
WATCH	338	-----+-+	I I I I			I
WATER	339	-----+ I	I I I I			I
CAMERA	53	-----+-----+ +--	I I			I
GLASSES	130	-----+ I	I I +-+	I		I
CLOCK	70	-----+-+	I I I	I		I
TELEPHONE	307	-----+ +--	I I	I		I
SLIDE	276	-----+ I	I I I I	I		I
WALKER	336	-----+-+	+-+ I	I		I
SPRINKLER	288	-----+ I	I	I I		I
BEADS	19	-----+-----+ I	I	I		I
BUTTON	51	-----+ I	I	I I		I
BOTTLE	36	-----+-+	I	I I		I
POTTY	240	-----+ I	I	I I		I
BOY	40	-----+-+	+-+	I	I	I
JAR	155	-----+ +--	I	I I		I
BOWL	38	-----+ I	I	I I		I
LAMP	168	-----+-+	I	I		I
CUP	86	-----+ I	I	I		I
BRUSH	44	-----+ I	I	I		I
TOOTHBRUSH	315	-----+-----+ I	I	I		I
COMB	77	-----+ +--	I	I		I
BROOM	42	-----+-----+ I	I	I		I
MOP	189	-----+ I	I I	I		I
DOOR	96	-----+-+	I I	I		I
DRAWER	99	-----+ +--	I I	I		I
STAIRS	290	-----+ +--	+-+	I		I
CLOSET	71	-----+ +--	I	I		I
BUCKET	45	-----+-----+ I	I	I		I
ROOF	254	-----+ I	I I I	I		I
HAMMER	142	-----+-----+ I	I	I		I
SHOVEL	269	-----+ I	+-+	I	+-+	
BOX	39	-----+-+	I I	I	I I	
TRAY	321	-----+ +--	I	I	I I	
BLOCK	32	-----+ I	I	I I		I
WINDOW	341	-----+ I	I	I I		I
TOOTH	314	-----+-----+ I		I		I
GARBAGE	124	-----+-----+-----+ I		I		I
TRASH	320	-----+ I	I I I			
ANKLE	3	-----+-----+ I	I I I			

* * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * *

CASE	0	5	10	15	20	25
Label	Num	+-----+-----+-----+-----+-----+				
STREET	298	-+	+-----+	+--+	I	
BELLYBUTTO	25	---+--+	I	I	I I I	
STONE	293	---+ +---+		+---+	I I I	
ROCK	252	-----+		I	+--+	I I
SIDEWALK	271	-----+-----+		I	I I	
PENNY	218	-----+-----+			I I	
SLED	275	-----+-----+			I I	
TRICYCLE	323	-----+-----+		+-----+	I I	
BICYCLE	29	-----+-----+		I	I I I	
MOTORCYCLE	190	-----+	+--+	+--+	I	
STROLLER	299	-----+-----+			I I I	
GAME	122	-----+-----+			I I	
LAWNMOWER	169	-----+-----+			I I	
VACUUM	332	-----+-----+		I	I I	
GARAGE	123	-----+-----+		I	I I	
GAS-STATIO	126	-----+-----+		I I	I I	
DRYER	102	-----+--+		+--+-----+	I I	
OVEN	204	-----+ +---+		I I	I I I	
STOVE	296	-----+ +--+		I	I I I	
WASHING MA	337	-----+-----+		I	I I I	
REFRIGERAT	251	-----+-----+		I	I I I	
CAR	56	-----+-----+		I	I I I	
TRUCK	324	-----+ +--+		I	+--+	I
BUS	48	-----+ I I		I	I	I
TRACTOR	318	-----+--+ +---+			I	I
FIRETRUCK	113	-----+-----+		I	I	I
TRAIN	319	-----+-----+		I	I	I
HELICOPTER	146	-----+-----+			I	I
LADDER	166	-----+-----+				I
GARDEN	125	-----+-----+				I
PLANT	227	-----+ +--+				I
FLOWER	116	-----+ +-----+				I
GRASS	137	-----+-----+			I	I
SAUCE	260	-----+--+			I	I
SOUP	285	-----+ I			I	I
VANILLA	334	-----+-----+			I	I
POP	234	-+---+	I		I	I
SODA	283	-+ I +---+			I	I
DRINK	101	-----+--+ I		I	I	I
JUICE	159	-----+ +--+		I	I	I
ICECREAM	153	---+---+ I		I	I	I
PUDDING	243	---+ I I		+--+	I	I
YOGURT	347	-----+ I		I I	I	I
MILK	181	-----+-----+		I I	I	I
COKE	76	-----+--+		I I	I	I

* * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * *

CASE	0	5	10	15	20	25
Label	Num	+-----+-----+-----+-----+-----+				
COOKIE	78	-----+ +-+ I	I	I	I	
COFFEE	75	-----+ I	I	I	I	
MEDICINE	179	-----+ I	I	I	I	
VITAMINS	335	-----+ +-+ +-+ I	I	I	I	
GLUE	132	-----+ +-+ I	I	I	I	
SALT	257	-----+ I	I	I	I	I
SANDWICH	259	-----+ +-+ I	I	I	I	I
ICE	152	-----+ I	I	I	+-+	I
EGG	105	-----+ I	I	I	I	I
JELLO	157	-----+ I	I	I	I	I
JELLY	158	-----+ +-+ I	I	I	I	I
GUM	139	-----+ I	+-+	I	I	I
PLAYDOUGH	229	-----+ I	I	I	I	I
CEREAL	59	-----+ I	I	I	I	I
POPSICLE	236	-----+ I	I	I	I	I
CAKE	52	-----+ I	I	I	I	I
APPLE	5	-----+ I	I	I	I	I
GRAPES	136	-----+ +-+ I	I	I	I	I
STRAWBERRY	297	-----+ +-+ I	I	I	I	I
FOOD	117	-----+ I	I	I	I	I
MELON	180	-----+ I	I	I	I	I
ORANGE	202	-----+ I	I	I	I	I
PUMPKIN	244	-----+ +-+ I	I	I	I	I
CARROTS	57	-----+ I	I	I	I	I
CRACKER	84	-----+ I	I	I	I	I
TOAST	311	-----+ I	I	I	I	I
CHOCOLATE	67	-----+ I	I	I	I	I
FRENCHFRIE	119	-----+ I	I	I	I	I
NUTS	201	-----+ I	I	+-+	I	I
PRETZEL	242	-----+ I	I	I	I	I
BANANA	12	---+ I	I	I	I	I
BEANS	20	---+ I	I	I	I	I
PEAS	214	---+ +-+ I	I	I	I	I
GREENBEANS	138	-----+ I	I	I	I	+---+
PICKLE	221	-----+ I	+-+	I	I	I
PIZZA	226	-----+ +-+ I	I	I	I	I
PLATE	228	-----+ I	I	I	I	I
HAMBURGER	141	---+ +-+ I	I	I	I	I
PANCAKE	208	---+ I	I	I	I	+-+
DONUT	95	---+ I	I	I	I	I
MUFFIN	194	---+ +-+ I	I	I	I	I
CHEERIOS	62	-----+ I	I	I	I	I
PEANUT BUT	213	-----+ I	I	I	I	I
CORN	79	-----+ I	I	I	I	I
POTATO	238	-----+ I	I	I	I	I

* * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * *

C A S E	0	5	10	15	20	25
Label	Num	+-----+-----+-----+-----+-----+				
NOODLES	198	-----+ I I	I I	I I	I I	
SPAGHETTI	286	-----+ I I I	I I	I I	I I	
CHEESE	63	-+ +--+ I	I I	I I	I I	
CHILD	65	-+-----+ I	I I	I I	I I	
CHICKEN	64	-+ I I	I I	I I	I I	
POPCORN	235	-----+ I	I I	I I	I I	
POTATOCHIP	239	-----+ I	I I	I I	I I	
MEAT	178	-----+ I	I I	I I	I I	
RAISIN	250	-----+ I	I I	I I	I I	
MAN	177	-----+ I	I I	I I	I I	
PARTY	212	-----+ +--+ I	I I	I I	I I	
LIPS	173	-----+ I	I I	I I	I I	
TONGUE	313	-----+ I	+--+ I	I I	I I	
VAGINA	333	-----+ I I I	I I	I I	I I	
MOUTH	192	-----+ I I	I I	I I	I I	
KNEE	164	-----+ +--+ I	I I	I I	I I	
LEG	170	-----+ I I	I I	I I	I I	
STORY	295	-----+ I	I I	I I	I I	
WIND	340	-----+ +--+ I	I I	I I	I I	
BOOBOO	33	-+-----+ I	I I	I I	I I	
OWIE	205	-+ +--+ I	I I	I I	I I	
WORK	345	-----+ +--+ I	I I	I I	I +--+	
RAIN	249	-----+ I I	I I	I I	I I I	
ARM	7	-----+ I I	I I	I I	I I I	
FINGER	111	-----+ +--+ I I	I I	I I	I I I	
HAND	143	-----+ I I I	I I	I I	I I I	
SHOULDER	268	-----+ I I I	I I	I I	I I I	
HAIR	140	-----+ +--+ I I I	I I	I I	I I I	
FEET	110	-----+ I I	I I	I I	I I I	
TOE	312	-----+ I I I	I I	I I	I I I	
PENIS	217	-----+ I I	I I	I I	I I I	
CHEEK	61	-----+ +--+ I	I I	I I	I I I	
TUMMY	325	-----+ I I I	I I	I I	I I I	
BOTTOM	37	-+-----+ I I I	I I	I I	I I I	
BUTTOCKS	50	-+ +--+ I I +--+	I I	I I	I I I	
EAR	104	-----+ I +--+ I	I I	I I	I I I	
CHIN	66	-----+ I I	I I	I I	I I I	
NOSE	199	-----+ I	I I	I I	I I I	
EYE	107	-----+ I I +--+	I I	I I	I I I	
MOON	187	-----+ I I I I	I I	I I	I I I	
STAR	291	-----+ +--+ I I I I	I I	I I	I I I	
SUN	300	-----+ +-----+ I I I I	I I	I I	I I I	
CLOUD	72	-----+ I I I I I	I I	I I	I I I	
SKY	274	-----+ +--+ +--+ I I I	I I	I I	I I I	
LIGHT	171	-----+ I I I I	I I	I I	I I I	

* * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * *

C A S E	0	5	10	15	20	25	
Label	Num	+-----+-----+-----+-----+-----+					
SNOWMAN	279	-----+-----+-----			I I I		
DOLL	93	-----+-----+-----			I I I		
TEDDYBEAR	306	-----+-----+-----			I I I		
BUTTERFLY	49	-----+-----+-----			I I +--+		
GIRL	128	-----+-----+-----			I I I		
FISH	114	-----+-----+-----			I I I		
CAMPING	54	-----+-----+-----			I I I		
CANDY	55	-----+-----+-----			I I I		
TOY	317	-----+-----+-----			I I +-----+		
CIRCUS	69	-----+-----+-----			I I I		I
ZOO	352	-----+-----+-----			I I I		I
LOLLIPOP	175	-----+-----+-----			I I		I
SOAP	281	-----+-----+-----			I I		I
SWING	302	-----+-----+-----			I		I
TREE	322	-----+-----+-----					I
WOODS	344	-----+-----+-----					I
HEN	147	-----+-----+-----					I
ROOSTER	256	-----+-----+-----					I
TURKEY	327	-----+-----+-----					I
DUCK	103	-----+-----+-----					I
GOOSE	133	-----+-----+-----			I		I
PENGUIN	216	-----+-----+-----			I		I
BIRD	30	-----+-----+-----			I		I
OWL	206	-----+-----+-----			+-----+		I
ANT	4	-----+-----+-----			I I		I
BUG	46	-----+-----+-----			I I		I
BEE	24	-----+-----+-----			I I		I
FROG	121	-----+-----+-----			I		I
BAT	15	-----+-----+-----			I		I
ALLIGATOR	1	-----+-----+-----			I		I
TURTLE	328	-----+-----+-----			I I		I
FACE	108	-----+-----+-----			I I		I
HEAD	145	-----+-----+-----			I +-----+		I
FIREMAN	112	-----+-----+-----			I I	I	I
FRIEND	120	-----+-----+-----			I I	I	I
GRANDMA	134	-----+-----+-----			I I	I	I
GRANDPA	135	-----+-----+-----			I I	I	I
LADY	167	-----+-----+-----			I I	I	I
MAILMAN	176	-----+-----+-----			I I	I	I
DOCTOR	91	-----+-----+-----			I I	I	I
DADDY	87	-----+-----+-----			I I	I	I
TEACHER	305	-----+-----+-----			I +--+	I	I
UNCLE	330	-----+-----+-----			I	I	I
PEOPLE	219	-----+-----+-----			I	I	I
SISTER	273	-----+-----+-----			I I	I	I

-

* * * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S *

CASE	0	5	10	15	20	25	
Label	Num	+-----+-----+-----+-----+-----+					
PERSON	220	---+ I I	I	I I	I	I	
MOMMY	183	---+ I I	I	I I	I	I	
NURSE	200	---+ I I	I	I I	I	I	
AUNT	8	---+ I I	I	I I	I	I	
YOUR-OWN-B	348	---++ +---	I	I I	I	I	
WOMAN	343	---+ I I	I I	I I	I	I	
BREAD	41	---++ I	I I	I I	I	I	
GLASS	129	---+ I I	I I	I I	I	I	
BABYSITTER	10	---+ I I	I I	+---	I	I	
BROTHER	43	---++ I	I I	I	I	I	
COWBOY	83	---+ I	+---	I	I	I	
POLICE	232	-----+ I	I	I	I	I	
CLOWN	73	-----+	I	I	I	I	
BABY	9	-----+	I	+-----+			
DRAGON	98	-----+-----+	I	I			
MONSTER	186	-----+	+-----+	I	I		
ELEPHANT	106	-----+	I I	I	I		
MOUSE	191	-----+-----+	I I	I	I		
SQUIRREL	289	-----+	I I I	I	I		
DOG	92	---++	I	+--	I		
PUPPY	245	---+ I	I I	I	I		
CAT	58	---++	I I	I	I		
KITTY	162	---+ I +---	I I	I	I		
YOUR-OWN-P	349	-----+ I I	I I	I	I		
BUNNY	47	-----+	I	+---	I		
LION	172	-----+	I I	I	I		
TIGER	308	---++	I I	I	I		
WOLF	342	---++ ++	I	I	I		
BEAR	21	-----+ ++	I I	I	I		
ANIMAL	2	-----+ I I I	I	I	I		
MONKEY	185	-----+ I I	I	I	I		
PIG	224	---+-----	++	I	I		
SHEEP	264	---+ ++	I	I	I		
COW	82	-----+ ++	I	I	I		
ZEBRA	350	-----+ I I	I	I	I		
DEER	88	---++ ++	I	I	I		
GIRAFFE	127	---++ ++	I	I	I		
DONKEY	94	-----+ ++	I	I	I		
MOOSE	188	-----+	I	I	I		
TUNA	326	-----+					

Appendix C: Forms and Instructions for Experiment 2, Chapter 2

Attached are the instructions and rating forms used in the pilot verb feature generation experiment and the verb feature rating experiment from Experiment 2 in chapter 2. These were originally presented as web pages, since participants completed the experiment on-line via the Web.

Pilot Study – Verb Features Generation

Please list as many features/aspects of the following verbs as you can. There is no rush, please take the time to think about and visualize (if possible) the word in question. Try to focus on physically observable aspects of that verb, rather than on other words, nouns, etc, that it tends to occur with. A “feature” does not have to be a single word, so for “fly” the features might be something like:

Requires wings

Goes fast

Travels from point a to point b

Moves through the air

Etc...

Hit:

_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

Move:

_____	_____
_____	_____

<hr/>	<hr/>
<hr/>	<hr/>
<hr/>	<hr/>

Go:

<hr/>	<hr/>
<hr/>	<hr/>
<hr/>	<hr/>
<hr/>	<hr/>
<hr/>	<hr/>

Put:

<hr/>	<hr/>
<hr/>	<hr/>
<hr/>	<hr/>
<hr/>	<hr/>
<hr/>	<hr/>

Run:

<hr/>	<hr/>
<hr/>	<hr/>
<hr/>	<hr/>
<hr/>	<hr/>
<hr/>	<hr/>

Sleep:

<hr/>	<hr/>
<hr/>	<hr/>
<hr/>	<hr/>

Eat:

Fall:

Carry:

Hold:

Touch:

Walk:

Sit:

Building Sensorimotor Semantic Representations from Human Knowledge - Verb Feature Ratings

On the following pages are a series of various actions or verbs, such as "**hit**", or "**run**". For each of the actions/verbs, there is a list of *features*. Please rate *each* verb on *each* feature on a scale of 0 to 10. Try to picture the object or concept mentally as you are making your rating. Also, rate a given verb in the context of other verbs, for example, the **speed** rating for "**walk**" should be lower than that for "**run**", and the **forcefulness** rating for "**hit**" should be higher than for "**touch**". However, don't worry if you can't remember exactly your ratings from concept to concept.

A **10** rating is the highest level of that feature possible, e.g. "**smash**" might get a **10** on **increases disorder**. A **0** on the other hand indicates the total absence of that feature. In some cases this implies its opposite is true. For example, a rating of **0** for **noisy** would mean total silence. Thus lower ratings (**1-4**) on **noisy** might correspond to relative quiet. Therefore, on some of these dimensions, a rating of **5** can be viewed as the dividing point between the feature and its opposite. In other cases there is no opposite to the feature, merely the absence of it, for example the feature **causes damage**. If this action is not a damaging action at all, then it is appropriate to rate it with a **0**. Otherwise, the larger the number, the more damage this action causes (e.g. up to perhaps a **10** for the verb **destroy**).

Try to avoid a **10** rating unless you are convinced or it is obvious that this verb is the highest possible on that feature. Likewise, avoid a **0** rating unless it's clear that this feature is not present *at all* for this verb, or this verb has the lowest amount possible of that feature. Try using a **9** or a **1** respectively if you are not sure enough to give a **10** or a **0**.

In general for these ratings, you should limit yourself to the experience that a pre-school child might have, that is, very basic physical understandings of themselves and their actions.

Click Proceed to begin the experiment. Remember, to receive credit you must rate 10 verbs. This will mean 10 passes through the web form, selecting the next verb each time. Each time you submit the data for one concept, you will be linked back again.

Proceed!

Verb Ratings Form

Student ID (For Experimental Credit):

Below is a drop-down box containing all the words in the experiment, in alphabetical order. You must rate only 10 of them, the 10 that were assigned to the phase you signed up for and that you wrote down previously. You will thus have to go through this form 10 times, crossing a word off of your list of 10 each time until you are done (so you don't forget and rate a word twice). Each time you complete a concept, you will go to a completion page, where you can link back to this form if you have not completed all 10 of your words yet.

Please try to imagine yourself performing each verb as you proceed, and keep that image in mind while rating the verb on the following featural dimensions. Press TAB to move to the next entry field. Try to ensure that you are consistent across verbs (e.g. don't say that walk is faster than run) but don't spend too much time agonizing over the ratings; just pick what seems reasonable and keep going.

Please enter a value between 0 and 10. A value of 10 means that of all possible verbs, this feature is the most true of this verb, or is of the highest value of all verbs. (E.g., "smash" might get a 10 for "causes damage", or "fly" might get a 10 for speed. A zero means the absence of that feature, e.g. a zero for noise means COMPLETE silence. For normal quiet, try a 1 or a 2 rating.)

Please rate the amount of physical activation of the following body parts involved in this verb (note that limb motion is expressed in terms of joint motion e.g. elbow):

0-10 (0 = no movement or irrelevant to

this action, 5 = average amount)

toes

ankles

knees

hips

torso (e.g. twisting)

shoulders

elbow

wrist

fingers	<input type="checkbox"/>
neck	<input type="checkbox"/>
head	<input type="checkbox"/>
face (General face movement)	<input type="checkbox"/>
eyes	<input type="checkbox"/>
eyebrows	<input type="checkbox"/>
nose	<input type="checkbox"/>
mouth	<input type="checkbox"/>
lips	<input type="checkbox"/>
tongue	<input type="checkbox"/>
requires a specific overall bodily position?	<input type="checkbox"/>
degree of overall body contact involved	<input type="checkbox"/>

Please rate the degree to which the following sensory perceptions/physical observations are present/involved:

0-10

horizontal motion involved	<input type="checkbox"/>
vertical motion involved	<input type="checkbox"/>

- optimum size of actor**
- noisyness (0 = silence)**
- perception - Auditory**
- perception - mental**
- perception - Smell**
- perception - Taste**
- perception - Touch**
- perception - Visual**
- speed (10 = fastest)**
- suddenness (0 - totally expected)**
- tightness (0 = no hold)**
- agitation (physical)**

Physical States: Please rate the degree to which the following are true of a person performing this verb or experiencing its consequences.

0-10

balance (0 = totally unsteady)

- decreases agitation**
- decreases energy**
- decreases hunger**
- decreases thirst**
- decreases tiredness**
- increases agitation**
- increases energy**
- increases hunger**
- increases thirst**
- increases tiredness**
- reactiveness (0 = unreactive to stimuli)**
- tension (0 = completely relaxed)**

Mental State Features: Please rate the degree to which the following mental or cognitive states are typically present in a person performing this verb.

0-10

aggression (0 = complete passivity)

- attention (0 = oblivious to this stimulus)**
- awareness (0 = unaware of anything)**
- control (0 = completely accidental)**
- pleasureable (0 = not at all)**
- painful (0 = not at all)**
- purposeful (0 = completely unintentional)**

Temporal Features: How much of any of the following are involved in the performance of the verb?

0-10

- starts something else**
- ends something else**
- duration of action (0 = instantaneous)**
- periodic action (0 = single action)**
- time pressure involved (e.g. verb "race")**

Physical Requirements/Characteristics: How much are the following features characteristic of or required for performing this verb?

0-10

amount of contact involved between actor and object

involves container/containing

involves supporting something

forcefulness

requires physical object (e.g. target for "hit")

requires a surface

strength involved

involves a trajectory or path from source to goal

Physical Effects: Rate the degree to which performing the verb has the following physical effects, results or consequences.

0-10

interrupts a path or trajectory

causes damage

distance typically travelled (0 = no change in position)

conjoins things

divides things

consumes (e.g. uses up like in "burn")



creates (something new)



destroys



displaces other object (and takes its place)



creates disorder/untidyness



creates order/tidyness



closes/closes down



opens/ opens up



change is involved (0 = totally static e.g. "exist")



transference of something



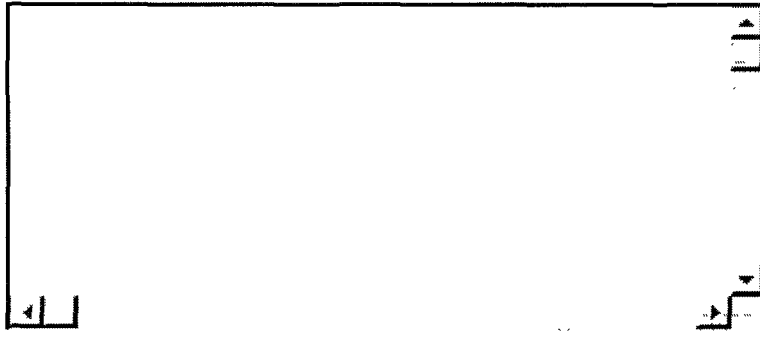
assembles things



disassembles things



Please suggest at least one feature dimension that is not yet listed in the above, and that would have been helpful when rating the above verb. Or, list the LEAST useful feature(s) from the list above for this verb. Also, feel free to make any general comments about the experiment here as well.



Indicates Response Required

Submit

Appendix D: Cluster Analysis of 90 Verbs from Chapter 2

For those unfamiliar with reading Hierarchical Cluster Analyses, the following figures are a tree structure. For any concept, follow the branches from the left to the right. The lowest level category including that concept is defined by the first + encountered.

* * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * *

Dendrogram using Average Linkage (Between Groups)

Rescaled Distance Cluster Combine

C A S E	0	5	10	15	20	25
Label	Num	+-----+-----+-----+-----+-----+				
DRY	21	-+----+				
GRAND	30	-+ I				
TAKE	79	-----+--+				
BRING	4	-----+ I				
PULL	54	----++ +-+				
PUSH	55	----+ +-+ I				
PUT	56	-----+ I +-+				
MOVE	48	-----+ I I				
FIT	26	-----+----+ I				
SHAKE	63	-----+ +-+				
BUY	7	-----+----+ I I				
SHOW	64	-----+ I I I				
BUILD	5	-----+----++ I				
MAKE	47	-----+ I +----+				
OPEN	49	-----+ I I				
RIDE	58	----+-- I I				
THROW	83	---+ +-----+ I +-+				
DRIVE	20	-----+ +-+ I I				
CATCH	9	-----+ I I				
CLEAN	12	-----+ I				
TICKLE	84	-----+----+				
BUMP	6	-----+----+ I I				
HIT	35	-----+ +-----+ I I				
KNOCK	41	-----+ +----+ I				
CUT	16	-----+ I I				
RIP	59	-----+ +----+ I				
SPILL	71	-----+----+ +-----+				
SPLASH	72	-----+ I I				
CHASE	10	-----+-----+ I I				
RUN	60	-----+ +-+ I I				
KICK	39	-----+ I I I				
HIDE	34	-----+----+ I I I				
STAND	73	-----+ +-----+ +-+				
HURRY	37	-----+-- I I I +----+				
SLIDE	68	-----+ +-+ I I I I				
GO	29	-----+-- +-+ I I				
JUMP	38	-----+ I I I I				
STOP	75	-----+ I I I				
SWIM	77	-----+-----+ I I I				
SWING	78	-----+ +----+ I I				
DANCE	17	-----+ I I				
BREAK	3	-----+-----+ I				

C A S E	0	5	10	15	20	25	
Label	Num	+-----+-----+-----+-----+-----+					
SMASH	69	-----+			I		
BLOW	2	-----+-----+			I		
PRETEND	53	-----+ I			I		
SAY	61	-----+-----+-----+			I		
TALK	80	-----+ I	+-----+		+-----+		
LISTEN	44	-----+-----+	I	+-----+	I I		
SING	65	-----+-----+-----+	I	I I	I		
HATE	31	-----+-----+-----+		I I	I		
LICK	42	---+-----+		I I	I		
TASTE	81	---+ +-----+	+-----+	I I	I		
SMELL	70	-----+-----+	I	I I	I		
SIT	66	-----+-----+-----+	I	I I	I		
SLEEP	67	-----+-----+	I I	I I	I		
LOOK	45	---+++	I	+-----+ I I	I		
SEE	62	---+ +-----+	I I	I I I	I		
WATCH	88	-----+ +-----+	I I	I I I	I		
READ	57	-----+-----+ I	I I	I I I	I		
CLAP	11	-----+-----+	I I	I +--+	I		
WAKE	86	-----+-----+ +-----+	+-----+	I I	I		
GIVE	28	-+-----+ I	I I	I I	I		
HAVE	32	-+ +--+ I	I I	I I	I		
COVER	15	-----+ I I	I I	I I	I		
EXIST	23	-----+--+ I	I I	I I	+-----+		
WISH	89	-----+ +--+ I	I I	I I	I I		
STAY	74	-----+-----+ I I	I I	I I	I I		
THINK	82	-----+ I +--+ +--+		I I	I I		
FINISH	25	-----+-----+ I I	I	I I	I I		
LOVE	46	-----+-----+ I I	I	I I	I I		
HEAR	33	-----+-----+ I		I I	I I		
DRAW	18	-+--+	I	I I	I I		
WRITE	91	-+ +-----+	I	+--+	I I		
PAINT	50	---+ +-----+	I	I	I I		
CLOSE	13	-----+--+ +--+ I		I	I I		
FIX	27	-----+-----+ I +--+		I	I I		
SWEEP	76	-----+-----+ I		I	I I		
HUG	36	-----+-----+		I	I I		
LIKE	43	-----+-----+-----+			I I		
PICK	51	-----+-----+-----+			I I		
EAT	22	-----+-----+-----+			I I		
KISS	40	-----+-----+-----+			I I		
DRINK	19	-----+-----+-----+	+-----+		+-----+		
FEED	24	-----+-----+-----+			I I	I	
TOUCH	85	-----+-----+-----+			I I	I	
CARRY	8	-----+-----+-----+			I I	I	
WORK	90	-----+-----+-----+	+-----+		I I	I	
COOK	14	-----+-----+-----+		+-----+	I I	I	
PLAY	52	-----+-----+-----+			I I	I	
WALK	87	-----+-----+-----+			I I	I	
BITE	1	-----+-----+-----+			I I	I	

Appendix E: SRNEngine - A Windows-based neural network simulation tool for the non-programmer

Preface

This appendix is reproduced from Howell and Becker (Submitted c). This paper was first submitted June, 2003, to the journal Behavior Research Methods, Instrumentation, and Computation. It is presently being revised for publication. It describes the self-contained software application that I designed to support the simulation experiments conducted in other chapters of this thesis and in earlier work. While methodological rather than experimental in nature, it is a vital part of my dissertation, since without it I would not likely have been able to complete as many large, complex simulations as I have.

Abstract

SRNEngine is a windows-based application package for training neural networks. The graphical user interface allows the drag-and-drop creation of neural networks

with a variety of architectures, without the need for any programming. At present, these architectures/learning algorithms include Simple Recurrent Networks, Jordan networks, and any kind of feedforward backpropagation network, with up to five each of input, hidden, and output layers (pools of units). A version that adds backpropagation-through-time is in development. The interface is designed to conform to the Microsoft Windows GUI environment that most PC users are already familiar with. SRNEngine includes tools for creating, editing, and manipulating various types of training data, and is especially optimized for working with text/language data, including automatic word-to-input-representation translation at runtime for text corpora. The distributed computing feature allows multiple simulations to be run on a network of workstations, co-ordinated via a central ftp server.

Introduction

Many simulation environments or neural network toolboxes that currently exist, while powerful and flexible, are either designed for those familiar with programming methodologies or are primarily used on a Unix platform. SRNEngine was designed for the Microsoft Windows environment and can be used even by non-programmers. As Macho (2002) points out, one of the options for non-programmers interested in neural networks is to use a spreadsheet program to implement neural network models. However, to implement a neural network within a spreadsheet requires a fair degree of computational

sophistication and effort. SRNEngine was explicitly designed to permit easy creation and training of neural networks. Thus, the SRNEngine application package allows a researcher to design a network by dragging layers graphically onto a screen, and then dragging connections between them to define the activation flow. After specifying a training corpus and output tasks, the network is ready to run. The simplicity of the interface makes the SRNEngine not only an attractive research tool for non-programmers but also a powerful teaching tool. Undergraduates in our lab have found it very easy to use in their research.

Development of a novel neural network application, whether for a research publication or industry, typically requires running numerous simulations to optimize model parameters and to generate a valid sample of model performance statistics. Parallel processing can greatly speed up this process. SRNEngine can be run in distributed-computing mode in order to run multiple simulations in parallel on the idle compute cycles of client PC's. This powerful feature allows the network model to be designed once, and then uploaded to an ftp server. Any number of client installations can then be installed in a special screensaver mode, which uses any idle time on the client computer to download the network specification from the FTP server and run it until complete. When complete, it uploads the results to the server, where they may be retrieved for analysis.

The following sections expand on the interface and distributed computing features of this application. We also describe how it has been optimized for

language/text models (e.g. Howell & Becker, 2000; Howell & Becker, 2001; Howell, Becker, and Jankowicz, 2001; Howell, Trainor, and Sonnadara, 2002) and categorization models (e.g. Howell, Schmidt, Trainor, and Santesso, 2002), the two areas where it has been applied to date¹.

Control Panel

The main screen of SRNEngine is its Control Panel, which houses the main menu and update fields for monitoring the real-time training progress of the network under simulation (See Section 5.4.2). New networks are created by simply dragging layers from a toolbar to the workspace and dropping them in position, then dragging one layer and dropping it onto another to establish the flow of activation.

Learning Algorithms Supported

At present the SRNEngine application supports only the back-propagation of error learning algorithm (Rumelhart, Hinton, & Williams, 1986), although many architectural variants that use this learning algorithm are supported, from arbitrary feedforward architectures to Simple Recurrent Networks (Elman, 1990) and Jordan networks (Jordan, 1986). It is also possible to implement a time-delay neural network (TDNN, Lang, Waibel, & Hinton, 1990) using multiple input 'layers' (pools of input layer units). A back-propagation-through-time (BPTT)

version is currently in development to extend the ability of SRNEngine to handle arbitrarily connected recurrent networks.

Text Corpus Pre-processing Tools

As the name might suggest, the SRNEngine application was originally designed to support the Simple Recurrent Network (SRN) architecture, as it is often used for language modelling. Thus, SRNEngine has an extensive suite of pre-and post-processing tools for language-related models.

SRNEngine is designed to use text-mode human-readable training corpora, and hence translates these human-readable inputs to computable vector representations internally at runtime. This facilitates the use of any textual corpus for training, since the researcher does not have to manually convert the text to vector representations. Text corpora can be converted to any of the following input representations (see Table 5.1): localist word, localist letter-by-letter, letter clusters of arbitrary length, user-defined distributed representations, phoneme-by-phoneme representations, or whole word phonemic representations. The ability to deal with training data and outputs in human-readable form has proved to be a major advantage in our research on language acquisition.

Post-processing, Logging, and Analysis Tools

SRNEngine includes many options for logging performance of a simulation over time. The most basic are error and accuracy logging, but some of the others

available include optional logging of all weights, all deltas, any layer's activations, Euclidean distance to target, per item frequency and accuracy, and more. In addition, prototypes (representations averaged across all exemplars of an input word) may be automatically generated from any hidden layer's activation pattern across the training set, allowing later statistical analysis of networks' hidden layer representations. These log files are easily imported to spreadsheet programs, statistical packages or other analysis tools.

Table 1: Types of Input Representations Supported

Localist Word	e.g. "CAT" = "0 0 0 1"
Localist Letter-by-letter	e.g. "C" = "0 0 0 1", "A" = "0 0 1 0"
Localist Letter Cluster (length <i>n</i>)	e.g. "CA" = "0 0 0 1", "AT" = "0 0 1 0"
User Defined Vector Representations	e.g. "CAT" = "0.4, 0.3, 0.1, 0.5"
Phoneme-by-phoneme Representations	e.g. "ae" = "0 1 0 1", "ng" = "1 0 0 0"
Whole Word Phonemic Representations (automatically generated)	Each word is composed of up to 10 phoneme slots, each containing a 14 bit compressed representation of that phoneme

Operational Details of SRNEngine

The Control Panel displays the current epoch and pattern being trained, the output layer being examined, and the translated output of the network at each pattern. In Figure 5.1, for example, the network is in the process of running 300 epochs of a language acquisition model using a 10,742-word corpus. It is currently running

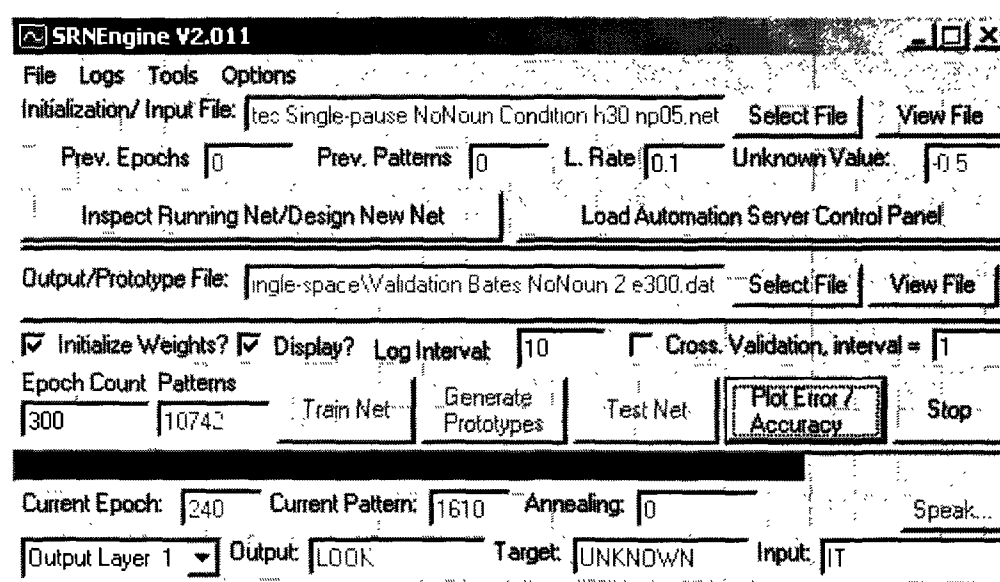


Figure 1 - The control panel of the SRNEngine Simulation Environment.

epoch 240, is on pattern 1610, and the word ‘LOOK’ is being produced at output layer number 1 when no target was specified, and when the input was the word ‘IT’. Furthermore, the error or accuracy curves of the network may be examined at any time during a run. All of this information provides a simple way to check

on the status of the simulation at a glance, although more in-depth visualization on the operation of the network is available from the Inspect Running Net button.

Network Design WorkSpace

This is the other major screen of the application and is where new networks are designed and where currently running networks are examined (See Figure 5.2).

New networks are created by merely dragging layers from a toolbar to the workspace and dropping them in position, then dragging one layer and dropping it onto another to establish the flow of activation. Left-clicking on all layers brings up dialogs that allow the user to specify various parameters for that layer, such as type of activation function, initialization range, etc. Once all parameters have been entered, or the default values accepted, the user may return to the Control Panel and immediately run the network.

On-line Visualization Features

The Workspace screen is also used as the visualization screen, as can also be seen in Figure 5.2. Left-clicking on any layer will display a window containing the current activation values across the nodes of that layer, and the contents will be updated in real-time as the training patterns are processed by the network. Output layers have more options; they can display either activations, mean-squared error per node, target activations, translated output (text or labeled representations), or all of these at once.

Left-clicking on the connections between layers will similarly display the weights between those layers, as seen in the background window in Figure 5.2. These weights can be viewed as Hinton diagrams or as numerical arrays. Right-clicking on a connection will display not the weights themselves, but rather the deltas or changes to those weights that are being made as a result of the current error feedback cycle.

All of these visualization tools have proved to be useful in our work when diagnosing problems in a misbehaving simulation, or when a novice user or student needs to understand exactly how a network is operating time-step by time-step.

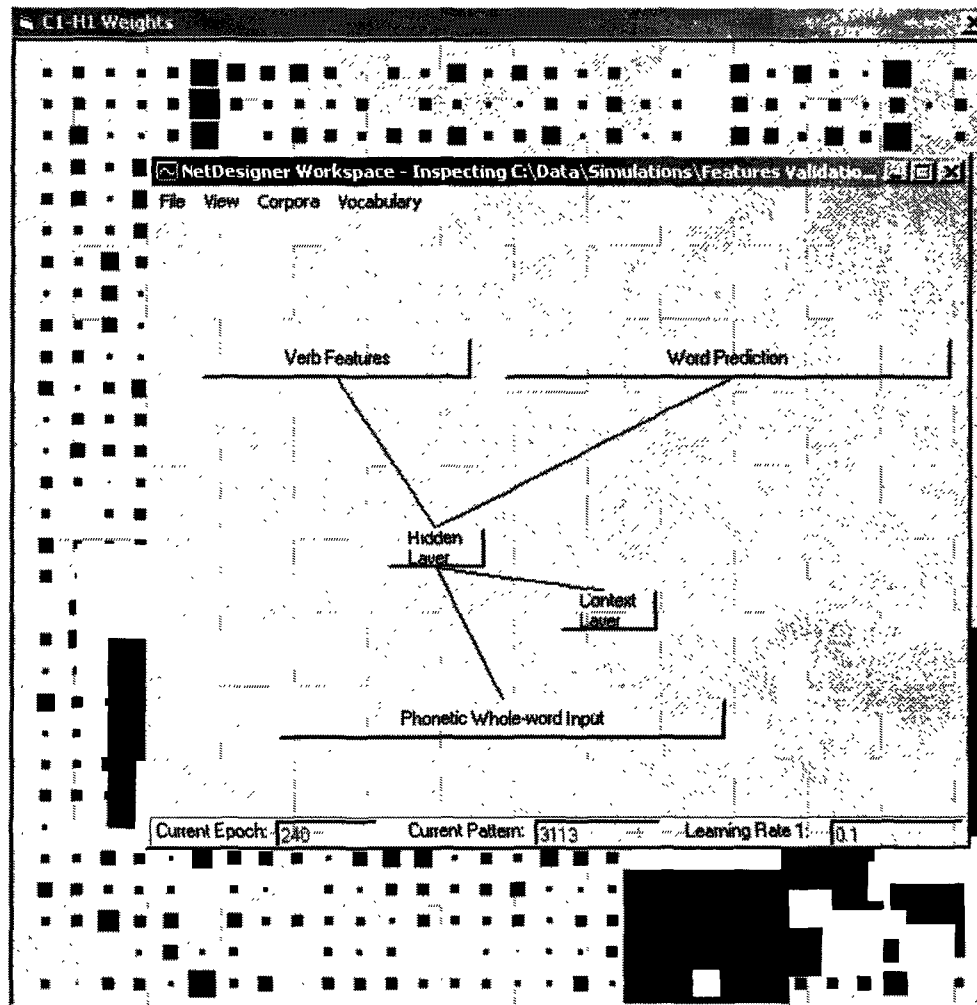


Figure 2: The SRNEngine NetDesigner Workspace screen. This screen is used for both network design and inspecting running networks. The background window is a display of the weights between the context and hidden units. This display is available for any set of weights or layer activations, and can be displayed as Hinton Diagrams (as here) or as numerical arrays. Output activations can also be viewed translated into the correct readable output response, either a word or a categorization response.

Distributed Computing Abilities

One of the most powerful and innovative features of the SRNEngine environment is its support for parallel distributed processing of simulations. In most parallel

computing applications, multiple processors work on several parts of a single problem simultaneously and then recombine the parts to form the solution.

However, most of the neural networks that will be simulated with SRNEngine cannot be broken down into subtasks or partitioned, but must be run in a continuous serial fashion. Another way to parallelize neural networks is by partitioning the training data into non-overlapping sets, training different networks on different data subsets, and periodically averaging the weights.

However, this method is not well suited to applications like language where the data consists of one continuous time-varying stream. Yet another way to parallelize neural networks is across multiple runs of the same network, starting from different random initial weights. This is the form of distributed computing that SRNEngine makes possible, by using unused computer time in a building or department to run many copies of one simulation, or one simulation in each of several conditions, simultaneously.

Much of the early neural network research was conducted at the “existence proof” level, whereby showing that a single network can solve the problem, or can model the data, was the only goal. However, today most researchers realize the need for valid scientific comparisons between alternative network models' performance, or comparisons to control networks lacking one or more critical features. Obviously, in order to be able to compile statistics on network performance, many networks must be run, often on the order of ten or a dozen networks per condition. For example, we have run a three condition

experiment using SRNEngine that investigates the use of semantic features added to a word prediction task (Howell and Becker, 2001). We ran ten networks in each condition, performed a t-test on their performance, and were able to demonstrate that adding features significantly improved word prediction, a finding that has relevance to the literature on the processes of children's word learning.

The SRNEngine application accomplishes this distributed computing by providing the option to be installed as a hidden application on client computers (ScreenSaver mode). When running in ScreenSaver mode, SRNEngine downloads a network to run from a local server (which is simply an account on an FTP server). It then lies dormant while the computer is in use, but as soon as the computer is idle long enough to activate the screensaver, SRNEngine continues from where it last left off. When a simulation is finished, SRNEngine uploads the results to the server and downloads a new simulation to run. The researcher can periodically retrieve the uploaded results via FTP and perform various analyses using a copy of SRNEngine in normal mode. A graphical server management program is also provided that insulates the user even from the need to use an FTP program, by automating the validation and uploading of each network's files to the server. Furthermore, the SRNEngine program checks the server for new versions of its own program each time it downloads a new network to run, enabling auto-upgrading of all client machines.

Conclusion

The SRNEngine neural network simulation engine provides an easy-to-use, graphical neural network design tool as well as a flexible neural network simulator for back-prop and similar models (e.g. SRN's). Full text pre-processing tools are provided for language and phonetic models, and a wide selection of logging and reporting tools are built into the program. The program's distributed-computing ability allows researchers to take advantage of all extra computing cycles in their lab or department, and run many copies of networks or many different networks with easy central administration.

References

- Elman, J. L. (1990). Finding structure in time. Cognitive Science, 14, 179-211.
- Howell, S. R. & Becker, S. (2000). Modelling language acquisition at multiple temporal scales. Proceedings of the 22rd Annual Conference of the Cognitive Science Society, 2000, 1031.
- Howell, S. R., & Becker, S. (2001) Modelling language acquisition: Grammar from the lexicon?, Proceedings of the 23rd Annual Conference of the Cognitive Science Society Conference, 2001, 429-434.
- Howell, S.R., Becker, S., & Jankowicz, D. (2001). Modelling language acquisition: Lexical grounding through perceptual features, Proceedings of the 2001 Developmental and Embodied Cognition Conference, July 31, 2001, Edinburgh.

Howell, S.R., Schmidt, L. A., Trainor, L.J., and Santesso, D.L. (2002). Neural Network Categorization of Infant Emotional States., Presented at the 13th Biennial International Conference on Infant Studies, Toronto, Ontario.

Howell, S.R., Trainor, L.J., and Sonnadara, R. (2002). Neural Network Categorization of Infant-directed and Adult-directed Emotional Speech, Presented at the 13th Biennial International Conference on Infant Studies, Toronto, Ontario.

Jordan, M.I. (1986). Serial order: A parallel distributed processing approach. Institute for Cognitive Science Report 8604. University of California, San Diego.

Lang, K. J., Waibel, A. H., and Hinton, G. E. (1990). A time-delay neural network architecture for isolated word recognition, Neural Networks, vol. 3, no. 1, 33-43.

Macho, S. (2002). Cognitive modeling with spreadsheets. Behavior Research Methods, Instruments, & Computers, vol. 34, no. 1 (February), 19-36.

Rumelhart, D.E., Hinton, G. E. & Williams, R. J. (1986) Learning internal representations by error propagation. In J. L. McClelland, D. E. Rumelhart and the PDP Research Group, Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations, 318-362. Cambridge, MA: The MIT press.