UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

The Empirical Acquisition of Grammatical Relations

Permalink

https://escholarship.org/uc/item/7cz402gb

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 21(0)

Authors

Morris, William C.

Cottrell, Garrison W.

Publication Date

1999

Peer reviewed

The Empirical Acquisition of Grammatical Relations

William C. Morris (wmorris@ucsd.edu)
Department of Computer Science and Engineering, University of California, San Diego
9500 Gilman Drive, Dept. 0114; La Jolla CA 92093-0114

Garrison W. Cottrell (gary@cs.ucsd.edu)
Department of Computer Science and Engineering, University of California, San Diego
9500 Gilman Drive, Dept. 0114; La Jolla CA 92093-0114

Abstract

We propose an account for the acquisition of grammatical relations using the concepts of connectionist learning and a construction-based theory of grammar. The proposal is based on the observation that early production of childhood speech is formulaic and the assumption that the purpose of language is communication. If one assumes that children's comprehension of multiword speech is not globally systematic, but based initially on semi-rote knowledge (so-called "pivot grammars"), a pathway through small-scale systematicity to grammatical relations appropriate to the child's target language can be seen. We propose such a system and demonstrate a portion of the emergence of grammatical relations using a connectionist network.

Introduction

Grammatical relations are frequently a problem for language acquisition systems.[1] In one sense they represent the most abstract aspect of language; subjects transcend all semantic restrictions—virtually any semantic role can be a subject. Where semantics is seen as being related to world-knowledge, syntax is seen as existing on a distinct plane. For this reason there are language theories in which grammatical relations are considered theoretical primitives, the most obvious examples of this are Relational Grammar (Perlmutter, 1982; Perlmutter & Postal, 1983) and Arc-Pair Grammar (Johnson & Postal, 1980).

One approach to learning syntax has been to relegate grammatical relations and their behaviors to the "innate endowment" that each child is born with. There are a number of theories of language acquisition, (e.g., Pinker, 1984, 1989; Hyams, 1986; Borer & Wexler, 1987, 1992) which start with the assumption that syntax is a separate component of language, and that the acquisition of syntax is largely independent of semantic considerations. Accordingly, in these theories there is an innate, skeletal syntactic system present from the very beginning of multiword speech. The acquisition of syntax, then, consists of modifying and elaborating the skeletal system to match the target language.

In order to avoid the need for innate knowledge, we propose a language acquisition system that does not rely on innate syntactic knowledge (Morris, 1998). The proposal is based on Construction Grammar (Goldberg, 1995) and on the learning mechanisms of PDP-style connectionism (Rumelhart & McClelland, 1986). We assume that the purpose of language is communication, and that children learn syntax as part of the mediating mechanism between spoken words and their aggregate meaning (cf. Slobin, 1997: 297). We hypothesize that abstractions such as "subject" emerge through rote learning of particular constructions, followed by the merging of these "mini-grammars". The claim is that in using this sort of a language acquisition system it is possible for a child to learn grammatical relations over time, and in the process accommodate to whatever language-specific behaviors his target language exhibits.

We have made a preliminary study showing that a neural net which is trained with its sole task being the assignment of semantic roles to sentence constituents can acquire grammatical relations. We have demonstrated this in two ways: by showing that this network associates particular subjecthood properties with the appropriate verb arguments, and by showing that the network has gone some distance toward abstracting this nominal away from its semantic content.

Theoretical Proposal

The proposal that we are basing our modeling on involves a three-stage process. The first stage involves rote understanding of speech, reflecting the formulaic speech that children exhibit. The second stage involves progressive abstraction over the formulas, based on both semantic and syntactic similarities. The third stage involves associating the resultant abstractions with specific "subjecthood properties".

---

[1] Grammatical relations are the relationships that noun phrases bear with a clause. These include subjects, objects, and indirect objects.

In the first stage a child learns verb argument structures as separate, individual "mini-grammars". This word is used to emphasize that there are no overarching abstractions that link these individual argument structures to other argument structures. Each argument structure is a separate grammar unto itself.

In the second stage the child develops correspondences between the separate mini-grammars; initially the correspondences are based on both semantic and syntactic similarity,[2] later the correspondences are established on purely syntactic criteria. The transition is gradual, with the role that semantics plays decreasing slowly. The result of the correspondences involves the creation of a larger grammar that includes the constituent mini-grammars. These larger grammars, in turn, will with each other.

For example, the verbs *eat* and *drink* are quite similar to each other, and will "merge" quickly into a larger grammar (while retaining their separate identities within that grammar, however). Similarly, the verbs *hit* and *kick* will merge early, since their semantics and syntax are similar. While all four of these verbs have agents and patients as verb arguments, there are many semantic differences between the verbs of ingestion and the verbs of physical assault, therefore the merge between these two verb groups will occur later in development.

Ultimately, these agent-patient verbs will merge with experiencer-percept verbs (e.g., *like, fear, see, remember*), percept-experiencer verbs (e.g., *please, frighten, surprise*), and others, yielding a prototypical transitive construction, with an extremely abstract argument structure. The verb-arguments in these abstract argument structures can be identified as "A", the transitive actor, and "O", transitive patient (or "object"). In addition there is prototypical intransitive argument structure with a single argument, "S", the intransitive "subject". (This schematic description was first put forward by Dixon, 1979.)

In the third stage, the child begins to associate the abstract arguments of the abstract transitive and intransitive constructions with the "bridging constructions" that instantiate the properties of, for example, clause coordination, control structures, and reflexivization. So, for example, an intransitive-to-transitive bridging construction will associate the S of an intransitive first clause to the deleted co-referent A of a transitive second clause. This will enable the understanding of a sentence like *Max arrived and hugged everyone*. Similarly, a transitive-to-intransitive bridging construction will map the A of an initial transitive clause to the S of a following intransitive clause; this will enable the understanding of a sentence like *Max hugged Annie and left*.

From beginning to end this is a usage-based acquisition system. It starts with rote-acquisition of verb-argument structures, and by finding commonalities, it slowly builds levels of abstraction. Through this bottom-up process, it accommodates to the target language. (For other accounts of usage based systems, see also Schlesinger, 1982,1988; Bybee, 1985; Langacker, 1987, 1991a, 1991b; St. John & McClelland, 1990; Tomasello, 1992; Elman et al., 1996.)

Psycholinguistic Evidence

The above proposal is based on the notion that children start with rote behavior, progress through a period in which small-scale systematic behaviors emerge in many very limited arenas, and finally come to a period in which the numerous small-scale systematic behaviors draw together into a small number of large-scale systematic behaviors. There is considerable evidence for this notion.

Child language specialists have noted for years that the earliest multiword utterances that children produce are formulaic. Many of the formulas appear to be "frozen phrases", or unanalyzed sequences that children can treat as individual lexical items (Peters, 1983; Barton & Tomasello, 1994). Other formulas, frequently referred to as "pivot grammars", consist of word combinations in which one element is fixed and the other "open"; examples include *more* + X, as in *more juice, more banana*, etc., and X + *allgone*, as in *juice allgone, milk allgone*, etc. (Braine, 1963, 1976; Bloom, 1973; Brown, 1973; Horgan, 1978). Pine, Lieven, & Rowland (forthcoming) have collected samples of 12 children's speech over six months starting when the children acquired vocabularies of approximately 100 words. They found that each child's five most common pivot formulas could account for between 65 and 90 percent of the child's subject + verb combinations.

Pivot grammars point toward the isolation of early systematic behavior. It is evident that syntactic sophistication likewise appears in isolated small-scale "puddles" of systematic behavior. Tomasello (1992) has shown that when children first begin to vary the types of syntax that they use with a verb, the novel use does *not* extend to other verbs. For a considerable period, children learn syntactic variation on a verb-by-verb basis. Tomasello refers to this analysis as the "Verb-Island Hypothesis".

There appear to be discernable patterns to the ways that children's grammars coalesce. Studies by Bloom, Lifter, & Hafitz (1980) and Clark (1996) examined the semantics behind sets of verbs that shared morphology, i.e., the verbs that appeared with *-ed, -s,* or *-ing.*. Interestingly, they found that these sets of verbs did tend to share semantic features. Pine, Lieven, & Rowland (forthcoming) analyzed the overlap of verb sets with these same verb endings, and found that, by and large, the sets defined by common morphology were distinct from one another. That is, morphological generalization was limited to small groups of semantically similar verbs.

---

[2] For our purposes, "syntactic similarity" refers to similarity of constructions. In English this primarily refers to word order, in other languages case marking performs the same function. In a connectionist network that is sensitive to word order, the instantiation of syntactic similarity is in similarity of trajectories through activation space.
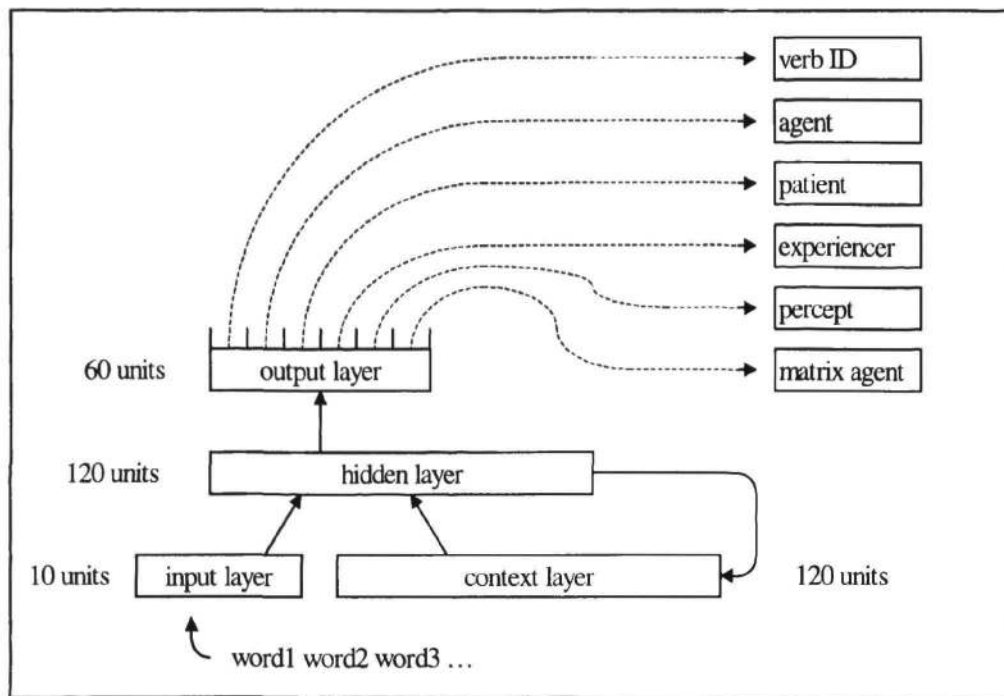
Figure 1: The Network

Finally, we note that studies by Maratsos, Kuczaj, Fox, & Chalkley (1979) and Maratsos, Fox, Becker, & Chalkley (1985), among others, show a surprising differential in children's rate of comprehension of the passive voice depending on which verbs are used. Passive sentences with verbs having agent and patient arguments (as in *Larry was hit by Marvin*) are understood at a higher rate than passive sentences with verbs having experiencer and percept verb arguments (as in *Larry was seen by Marvin*). The difference was dramatic at age 4 (85% comprehension for agent-patient verbs vs. 34% comprehension for experiencer-percept verbs), and did not disappear until the age of ten or eleven (Maratsos et al., 1985). This study demonstrates that even such a thoroughly abstract operation as passivization is learned not as an abstract operation defined in terms of grammatical relations, but as a semantically restricted operation which later "grows" into a greater abstraction. Upon investigation, Maratsos showed that there was a corresponding gap in the parental input to children; in child-directed speech there appear to be few, if any, passive sentences with experiencer-percept verbs.[3]

A Connectionist Simulation

In this section we present a connectionist simulation to test whether a network could build abstract relationships corresponding to "subjects" and "objects" given an English-like language with a variety of grammatical constructions.

This was done in such a way that there is no "innate" knowledge of language in the network. In particular, there are no architectural features that correspond to "syntactic elements", i.e., no grammatical relations, no features that facilitate word displacement, and so forth.

The motivation behind the network is the notion that merely the drive to map input words to output semantics is sufficient to induce the necessary internal abstractions to facilitate the mapping. This is an instantiation of the notion that the sole purpose of language is communication, and syntax is emergent. We were preceded in this approach by St. John & McClelland (1990). Our network and task is a simplified version of the one used by them. We differ from them in that our noun meanings are extremely simplified, while our syntactic constructions are more complex.

Our model uses a Simple Recurrent Network (Elman, 1990) implemented using the Stuttgart Neural Network Simulator (SNNS). The network is shown in Figure 1.

The input layer is ten units wide; each pattern represents one of 56 words or one of 2 punctuations. Each of these is represented by a unique pattern of 5 zeroes and 5 ones. The input consists of sentences drawn from the vocabulary of 56 words and two punctuations. Of these 56 words, 25 are verbs, 25 are nouns, and remaining 6 are a variety of function words. All of the nouns are proper names. Of the verbs, five are unergative (intransitive, with agents as the sole arguments, e.g., *run, sing*), five are unaccusative (intransitive, with patient arguments, e.g., *fall, roll*), ten are "action" transitives (with agent & patient arguments, e.g., *hit, kick, tickle*), and five are "experiential" transitives (with experiencer & percept arguments, e.g., *see, like, remember*).

---

[3] There are, however, a number of percept-experiencer verbs in passive sentences, e.g., *I was surprised by...*, or *He was frightened by ....*

440

In addition there was a "matrix verb", *persuade,* which was used for embedded sentence structures. The five remaining words were *who, was, by, and,* and *self.* The two punctuations were "period" and "reset". The network was trained to hold the output values during a "period" and to reset the output values at a "reset".

The output layer is 60 units wide. These are divided into 6 fields that are 10 units wide. The first field is the verb identifier, the second through the fifth are the identifiers for the agent, the patient, the experiencer, and the percept. (Note that at most only two of these four fields should be asserted at a single time.) The sixth field is the "matrix agent" field, which will be explained below. The internal identifiers of the nouns are different from (and unrelated to) the external identifiers—the internal identifiers each have only two units asserted.

Using the back-propagation learning procedure (Rumelhart, Hinton, & Williams, 1986) the network was taught to assign the proper noun identifier(s) to the appropriate role(s) for any of a number of sentence structures. Thus for the sentence, *Sandy persuaded Kim to kiss Larry,* the matrix agent role is filled by *Sandy,* the agent role is filled by *Kim,* and the patient role is filled by *Larry.* In the sentence, *Who did Larry see,* the experiencer role is filled by *Larry* and the percept role is filled by *who.*

Training was conducted with 50 epochs, with 10,000 sentences in each epoch. The learning rate was 0.2, initial weights set within a range of ± 1.0. There was no momentum. The learning function was online back-propagation using the mean-square error criterion. The squashing function was the standard 0-1 logistic function.

Examples of the types of sentences in the training set are shown in examples 1-6. The numbers in parentheses indicate the percentage of the total training corpus represented by each type of sentence. Semantic roles (i.e., agent, patient, experiencer, percept, or matrix agent) present in each example sentence are indicated in parentheses after the example.

1. Simple declarative intransitives (18%)
 e.g., *Sandy jumped* (agent)
  Sandy fell (patient)
2. Simple declarative transitives (26%)
 e.g., *Sandy kissed Kim* (agent & patient)
  Sandy saw Kim (exper. & percept)
3. Simple declarative passives (6%)
 e.g. *Sandy was kissed.* (patient)
4. Questions (20%)
 e.g., *Who did Sandy kiss?* (agent & patient)
  Who kissed Sandy? (agent & patient)
  Who did Sandy see? (exper. & percept)
  Who saw Sandy? (exper. & percept)
5. Control (equi-NP) sentences (25%)
 e.g., *Sandy persuaded Kim to run.*
   (matrix agent & agent)
  Sandy persuaded Kim to fall.

   (matrix agent & patient)
  Sandy persuaded Kim to kiss Max.
   (matrix agent, agent, & patient)
  Sandy persuaded Kim to see Max.
   (matrix agent, exper. & percept)
6. Control (equi-NP) sentences with questions (5%)
 e.g., *Who did Sandy persuade to run/fall?*
   (questioning embedded subject, whether agent or patient, of an intransitive verb)
  Who persuaded Sandy to run/fall?
   (questioning matrix agent; note embedded intransitive verb)
  Who persuaded Sandy to kiss/see Max?
   (questioning matrix agent; note embedded transitive verb)
  Who did Sandy persuade to kiss Max?
   (questioning embedded agent)

The distribution of verb types (agent-only, patient-only, agent-patient, and experiencer-percept types) is intended to be as realistic as possible for such a small vocabulary. The distribution of constructions is intended to provide some syntactic richness and to lay the groundwork for our generalization tests. The richness of the syntax here is out of proportion to the vocabulary size. In future work we hope to reduce the imbalance between vocabulary and syntax.

The test There were two systematic gaps in the data presented to the network; both involved experiential verbs: passive sentences with experiential verbs, e.g., *Sandy was seen by Max,* and questioning embedded subjects in transitive clauses with experiential verbs, e.g., *Who did Sandy persuade to see Max?*

Neither of these sentence types occurred with experiencer-percept verbs; all of the training involving these constructions used only agent-patient verbs. The test involved probing these gaps.

The network was not expected to generalize over these two systematic gaps in the same way. The questioning-of-embedded-subject-sentences gap is part of an interlocking group of constructions which "conspire" to compensate for the gap. The "members of the conspiracy" are the transitive sentences (group 2 above), the questions (group 4), and the control sentences (group 5). These sentences are related to each other, and they should cause the network to treat the agents of action verbs and the experiencers of experiential verbs similarly in the context of embedded clauses. The passive gap has no such compensating group of constructions. Only the transitive sentences (group 2) provides support for the passive generalization; as we shall see, these were insufficient to bridge the gap.

The sentences in groups 5 & 6 involve the "subjecthood property" of control structures (or equi-NP deletion). Part of the point of including these was to show that a network

Table 1: Sentence comprehension using Euclidean distance decisions

Sentence description	Percent correct	
Simple active clauses, action verbs	97.6%	
Simple active clauses, experiential verbs	97.6%	
Simple passive clauses, action verbs	91.8%	
Simple passive clauses, experiential verbs	6.2%	⇦ Failure to generalize
Control (equi-NP) structures	83.6%	
Questioning embedded subjects, action verbs	91.4%	
Questioning embedded subjects, experiential verbs	67.4%	⇦ Successful generalization

can learn to associate subjecthood properties with the appropriate nominal.

The Results

In Table 1 we show the result of testing a variety of constructions, some forms of which were trained, and two were not. Five hundred sentences of each listed type were tested. The results were computed using Euclidean distance decisions—each field in the output vector was compared with all possible field values (including the all-zeroes vector), and the fields assigned the nearest possible correct value. For a sentence to be "correct" all of the output values had to be correct. The two salient lines are for simple passive clauses with experiential verbs, which had a 6.2% success rate, and questioning embedded subjects with experiential verbs, which had a 67.4% success rate. (These lines are italicized in the table below.)

The relationship between these rates is as expected. The near complete failure of generalization for simple passive clauses with experiential verbs shows that the nonappearance of experiential verbs in the passive voice in the training set causes the network to learn the passive voice as a semantically narrow alternation. This is similar to an undergeneralization found by Maratsos et al. (1979; 1985), in which 4- and 5-year old children were shown to not comprehend passive sentences containing experiencer-percept verbs at an age when they could readily understand passive sentences containing agent-patient verbs. This gap has been shown by Maratsos et al. (1985) to be one that actually exists in parental input to children.

On the other hand, the questioning of embedded subjects with experiential verbs, which likewise did not appear in the training set, showed much greater generalization, in all likelihood because there is a "conspiracy of syntactic constructions" surrounding this gap. As a result we are seeing a level of abstraction, with the network able to "define", in some sense, the gap in terms of the embedded subject rather than merely an embedded agent.

This second gap, that of the questioning of embedded subjects of experiential verbs, is an unnatural omission, unlike the passive voice gap with experiential verbs. There does not appear to be a corresponding gap in English. This omission was introduced in the training set precisely because it is unnatural, and the network went a considerable distance in overcoming its effects, thus demonstrating the emergence in the network of the subject, i.e., a syntactic constituent abstracted away from semantics. This was demonstrated by the fact that the network had begun to define the gap found in the questioning of embedded agents of action verbs in more semantically abstract terms, thus allowing the network to correctly interpret two-thirds of the sentences involving the questioning of embedded experiencers.

Discussion and Conclusions

This simulation and the planned extensions of it are intended to demonstrate that the most abstract syntactic aspects of language are learnable. There are two broad areas in which this is explored: control of "subjecthood" properties and demonstration of abstraction across semantic roles.

In the area of control of properties, this simulation demonstrated that the network was capable of learning equi-NP deletion (also known as "control constructions"). This is shown in the ability of the network to correctly process sentences such as *Sandy persuaded Kim to run* (these are shown in groups 5 & 6, in section 4 above). As was seen above, the network was able to correctly understand these sentences at a rate of 84%.

The network's ability to abstract from semantics was shown in the ability of the network to partially bridge the artificial gap in the training set, that of the questioned embedded subject of experiential verbs. The network was able to define the position in that syntactic construction in terms of a semantically-abstract entity, that is, a subject rather than an agent.

We are currently examining the representation of "subject" in the network's hidden layer. In the examination we are seeing the way that the network has developed a partially-abstract representation of the subject. We are also seeing the limitations of abstraction; the network's representation of the subject of a given sentence is also partially specified in semantically loaded units. And, as we have seen in the Maratsos (1985) study, this appears to be appropriate to the way that humans learn language. (See also Goldberg, 1995, for a theoretical analysis that predicts this semantically-limited scope to certain syntactic constructions.)

We believe that we have shown that syntax as a separate entity from semantic processing is an unnecessary assumption. Rather, what we see in our network is that "syntax", in the usual understanding of that term, is part and

parcel of the *processing* required to map from a sequence of input words to a set of semantic roles.

Acknowledgments

References

Bloom, Lois. (1973). *One Word at a Time: The Use of Single Word Utterances Before Syntax.* The Hague: Mouton.

Bloom, L., K. Lifter, & J. Hafitz. (1980). Semantics of verbs and the development of verb inflection in child language. *Language 56,* 386-412.

Brown, Roger. (1973). *A First Language/The Early Stages.* Cambridge MA: Harvard University Press.

Bybee, Joan. (1985). *Morphology: A Study of the Relation between Meaning and Form.* Amsterdam: John Benjamins.

Clark, Eve V. (1996). Early verbs, event-types, and inflections. In C. E. Johnson & J. H. V. Gilbert (eds.), *Children's Language,* Volume 9. Mahwah NJ: Lawrence Erlbaum Associates.

Dixon, Robert M. W. (1979). Ergativity. In *Language,* 55. 59-138.

Elman, Jeffrey L. (1990). Finding Structure in Time. In *Cognitive Science* 14: 179-211.

Elman, Jeffrey L., Elizabeth A. Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi, & Kim Plunkett. (1996). *Rethinking Innateness: A Connectionist Perspective on Development.* Cambridge MA: MIT Press.

Goldberg, Adele E. (1995). *A Construction Grammar Approach to Argument Structure.* Chicago: University of Chicago Press.

Horgan, D. (1978). How to answer questions when you've got nothing to say. *Journal of Child Language, 5.* 159-165.

Johnson, David E., & Paul M. Postal. 1980. *Arc pair grammar.* Princeton NJ: Princeton University Press

Keenan, E. L. (1976). Towards a Universal Definition of "Subject." In Li, C. (ed.) *Subject and Topic.* New York: Academic Press.

Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar,* Vol. 1, *Theoretical Prerequisites.* Stanford: Stanford University Press.

Langacker, Ronald W. (1991a). *Foundations of Cognitive Grammar,* Vol. 2, *Descriptive Application.* Stanford: Stanford University Press.

Langacker, Ronald W. (1991b). *Concept, Image, and Symbol: The Cognitive Basis of Grammar.* Berlin: Mouton de Gruyter.

Marantz, Alec P. (1984). *On the nature of grammatical relations.* Cambridge MA: MIT Press.

Maratsos, Michael, Stanley A. Kuczaj II, Dana E. C. Fox, & Mary Anne Chalkley. (1979). Some Empirical Studies in the Acquisition of Transformational Relations: Passives, Negatives, and the Past Tense. In W. A. Collins (ed.) *Children's Language and Communication: The Minnesota Symposia on Child Psychology,* volume 12. Hillsdale NJ: Lawrence Erlbaum Associates.

Maratsos, Michael, Dana E. C. Fox, Judith A. Becker, & Mary Anne Chalkley. (1985). Semantic restrictions on children's passives. *Cognition,* 19. 167-191.

Morris, William C. (1998). *Emergent Grammatical Relations: An Inductive Learning System.* Ph.D. dissertation, University of California, San Diego.

Perlmutter, David. 1982. Syntactic representation, syntactic levels, and the notion of subject, in P. Jacobson and G. Pullum (eds.) *The Nature of Syntactic Representation,* Dordrecht: Reidel.

Perlmutter, David, & Paul Postal. 1983. Toward a universal definition of the passive, in D. Perlmutter (ed.) *Studies in Relational Grammar I.* Chicago: The University of Chicago Press.

Pine, Julian M., Elena V. M. Lieven, & Caroline F. Rowland. (Forthcoming). Comparing different models of the development of the English verb category. MS.

Pinker, Steven. (1984). *Language Learnability and Language Development.* Cambridge MA: Harvard University Press.

Pinker, Steven. (1989). *Learnability and Cognition: The Acquisition of Argument Structure.* Cambridge MA: MIT Press.

Rumelhart, David E., Geoffrey E. Hinton, & Ronald J. Williams. (1986). Learning Internal Representations by Error Propagation. In D. E. Rumelhart & J. L. McClelland. (eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Volume 1: *Foundations.* Cambridge MA: The MIT Press.

Rumelhart, David E., & James L. McClelland. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Volume 1: *Foundations.* Cambridge MA: The MIT Press.

Schlesinger, Itzchak M. 1982. *Steps to language: Toward a theory of language acquisition.* Hillsdale NJ: Lawrence Erlbaum Associates.

Schlesinger, Itzchak M. 1988. The Origin of Relational Categories. In Y. Levy, I. M. Schlesinger, and M. D. S. Braine (eds.) *Categories and Processes in Language Acquisition.* Hillsdale NJ: Lawrence Erlbaum Associates.

Slobin, Dan I. (1997). The Origins of Grammaticizable Notions: Beyond the Individual Mind. In D.I. Slobin (ed.) *The Crosslinguistic Study of Language Acquisition* Volume 5: *Expanding the Contexts.* Mahwah NJ: Lawrence Erlbaum Associates.

St. John, Mark F., & James F. McClelland. (1990). Learning and Applying Contextual Constraints in sentence Comprehension. *Artificial Intelligence, 46.* 217-257.

Tomasello, Michael. (1992). *First verbs: A case study of early grammatical development.* Cambridge UK: Cambridge University Press.