# A Connectionist Model of Verb Subcategorization

**Hinrich Schütze**
Center for the Study of Language and Information
Stanford University
Stanford, California 94305-4115
schuetze@csli.stanford.edu

**Abstract**

Much of the debate on rule-based vs. connectionist models in language acquisition has focussed on the English past tense. This paper investigates a new area, the acquisition of *verb subcategorization*. Verbs differ in how they express their arguments or *subcategorize* for them. For example, "She gave him a book." is good, but "She donated him a book." sounds odd. The paper describes a connectionist model for the acquisition of verb subcategorization and how it accounts for overgeneralization and learning in the absence of explicit negative evidence. It is argued that the model presents a better explanation for the transition from the initial rule-less state to final rule-like behavior for some verb classes than the symbolic account proposed by Pinker (1989).

## Introduction

The debate as to whether connectionist or symbolic models provide a better account for linguistic knowledge and how their accounts differ has centered on the English past tense (e.g. (Rumelhart and McClelland, 1986),(Pinker and Prince, 1988),(MacWhinney and Leinbach, 1991),(Marcus et al., 1992),(Daugherty and Seidenberg, 1993), (Daugherty and Hare, 1994)). This paper investigates a new area, the acquisition of *verb subcategorization*.

(1) a. She gave a book to the student.
  b. She gave the student a book.
  c. She donated a book to the church.
  d. ∗ She donated the church a book.

Verbs differ in how they express their arguments or *subcategorize* for them. For example, both (1a) and (1b) are grammatically correct, but only (1c) is an acceptable way of expressing the arguments of "donate". The alternation exhibited by "give" in this example is called the dative alternation. The verb "give" is said to undergo the alternation, whereas "donate" does not allow it.

A detailed account for the acquisition of verb subcategorization has been proposed by Pinker (1989) and Gropen et al. (1989). One of the major insights of their theory is that the meaning of verbs is an important determinant of subcategorization. For example, verbs of continuous imparting of force such as "pull" only allow the prepositional construction (as in (1a)), verbs of instantaneous imparting of force such as "throw" allow both the prepositional (1a) and the dative construction (1b). Rote learning cannot explain successful acquisition since subcategorization frames are used productively for novel verbs that are semantically compatible with a given subcategorization frame ("He emailed me the paper."), while no such generalization is observed for semantically incompatible verbs

In section 2, this paper describes a connectionist model for the acquisition of the dative alternation that incorporates the importance of semantic factors for subcategorization. The model is shown to account for overgeneralization and learning in the absence of explicit negative evidence. Section 3 discusses how symbolic and connectionist models of subcategorization differ and argues that the connectionist model presents a better explanation for the transition from the initial rule-less state to final rule-like behavior for some verb classes than the symbolic account proposed by Pinker (1989).

## The Connectionist Model

This section presents a model of the acquisition of subcategorization that has the crucial properties of the learning situation of the human child. First, no explicit correction of errors occurs. Second, the initial analysis of the occurring subcategorization frames strongly suggests generalizations that are in fact ungrammatical. This leads to overgeneralizations from which the model recovers in subsequent learning.

### Architecture

The basic architecture of the model is shown in Figure 1. The input consists of ten semantic microfeatures representing semantic properties of verbs important for subcategorization and one localist representation for each of the 56 verbs, i.e. each of the verbs is represented by a separate node which is on if the verb is presented to the net and off otherwise. The localist representations feed into a block of two hidden units which is fully connected with the main hidden layer consisting of ten hidden units. The microfeature block is also fully connected with the ten hidden units. Finally, the hidden units propagate their activation to two output units in the block "selections". One of the units is for the frame "VERB + THEME + *to* + GOAL", abbreviated as PREP. An example of this frame is (1a). The other output unit is for the frame "VERB + GOAL + THEME", abbreviated as DITR (example (1b)). The net is trained using the backpropagation algorithm (Rumelhart et al., 1986). Training proceeds by presenting one of the verbs, propagating activation forward through the network to the selection units, computing the error between the selected frames and the actual occurring frame, and backpropagating the error. An example for input and output is given below.

One reason for connecting the 56 verb nodes and the 10-unit block by a "bottleneck" of only two hidden units is that the high connectivity of a network that would fully connect the hidden units and the block of 56 verb nodes would be inefficient and implausible cognitively. More importantly, this architecture reflects the intuition that verb subcategorization is mainly determined by semantic properties of verbs (represented by the ten microfeatures) and that the child's learning process is heavily biased against using idiosyncracies for selecting subcategorization frames. However, there is the possiblity of using idiosyncratic features for subcat selection if no other reliable predictor can be found. The bottleneck architecture thus implements the insight from (Pinker, 1989) that the selection of subcategorization frames is largely dependent on semantic properties and only moderately influenced by idiosyncracies.

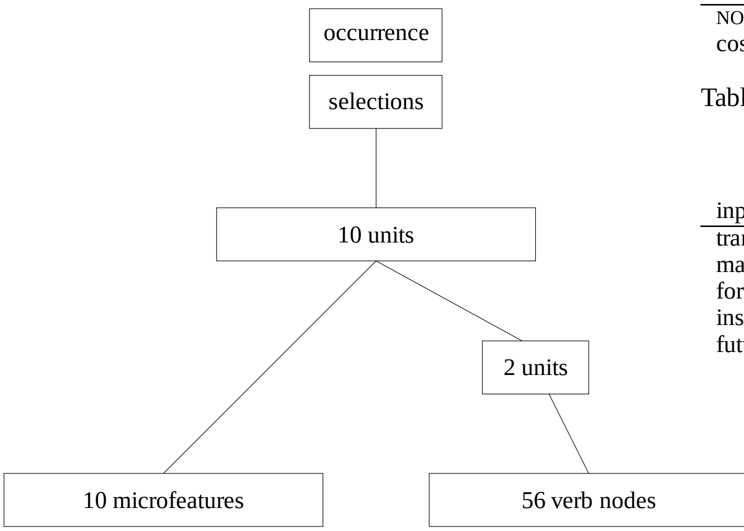


Figure 1: The architecture of the subcat frame selector.

The rationale for the training regime is that each occurrence of a verb with a subcategorization frame is a learning experience for the child. If his or her current grammar specifies the incorrect frame, for instance DITR for *donate*, then each occurrence of *donate* with the frame PREP will push the grammar towards the correct subcategorization specification.

The 56 verbs used in the experiment are mainly verbs of change of possession taken from Pinker (1989:110ff.). Their categorization into subclasses in Table 1 is also similar to Pinker's. The main difference is that *allocate*, *allot*, *assign*, *bequeath*, *grant*, *award*, *advance*, *forward* were categorized as simple verbs of giving, rather than verbs of future having, since they do not seem to imply a more immediate transfer than for instance *sell* or *send*. *donate* was added as an idiosyncratic word: semantically it is part of the GIVING class, but it does not exhibit the dative alternation.

The microfeatures used for encoding the verbs are shown in Table 2. They represent semantic distinctions identified as important by Pinker.

Table 3 shows how the verbs in Table 1 are encoded using the features. Again, I follow Pinker's proposal.

Table 4 gives the possible subcategorization frames of the verb classes covered in the experiment.

The network was trained in 800 epochs, where each epoch consisted in presenting all verbs to the network. Table 5 dis-

| |
|---|
| GIVING: transfer from giver to recipient<br>give pass hand sell pay trade lend loan serve feed send allocate allot assign bequeath grant award advance forward |
| DONATE<br>donate |
| TRANSPORT: manner of transport specified<br>ship mail |
| INSTANTANEOUS: instantaneous imparting of force causing motion<br>throw toss flip slap kick poke fling blast |
| CONTINUOUS: continuous imparting of force causing motion<br>carry pull push schlepp lift haul lower |
| DIRECTION: direction of transfer specified<br>bring take |
| FUTURE: verbs of future having<br>offer promise leave guarantee pledge reserve |
| NON-TRANSFER: verbs of future not having<br>cost spare envy begrudge bet refuse ask save charge fine forgive |

Table 1: Verbs of change of possession covered by the model.

| input feature | definition |
|---|---|
| transfer | transfer takes place or will take place |
| manner | specified manner |
| force | imparting of force |
| instantaneous | instantaneous vs. continuous imparting of force |
| future | (possible) transfer in the future |

Table 2: Semantic microfeatures used for encoding.

| class | transfer | manner | force | instantaneous | future |
|---|---|---|---|---|---|
| GIVING | + | – | | | |
| DONATE | + | – | | | |
| TRANSPORT | + | + | | | |
| INSTANTANEOUS | + | + | + | + | |
| CONTINUOUS | + | + | + | – | |
| DIRECTION | + | – | + | – | |
| FUTURE | + | | | | + |
| NON-TRANSFER | – | – | | | |

Table 3: Microfeature encoding of the classes in 1.

| class | DITR | PREP |
|---|---|---|
| GIVING | + | + |
| DONATE | – | + |
| TRANSPORT | + | + |
| INSTANTANEOUS | + | + |
| CONTINUOUS | – | + |
| DIRECTION | + | + |
| FUTURE | + | + |
| NON-TRANSFER | + | – |

Table 4: Possible subcategorization frames for the classes in 1.

| input pattern | example output | target 1 | target 2 |
|---|---|---|---|
| +transfer:1.0 | PREP:0.6 | PREP:1.0 | PREP:0.0 |
| –transfer:0.0 | DITR:0.4 | DITR:0.0 | DITR:1.0 |
| +manner:0.0 | | | |
| –manner:1.0 | | | |
| +force:1.0 | | | |
| –force:0.0 | | | |
| +instantaneous:0.0 | | | |
| –instantaneous:1.0 | | | |
| +future:0.0 | | | |
| –future:0.0 | | | |

Table 5: Example of input/target patterns for *bring*.

plays an example input pattern for *bring*, that can be directly derived from the encoding given in Table 3. During early training, the subcat selector will select incorrect preferences for the two frames, for example the 0.6/0.4 example output in Table 5. The model is trained with the two correct subcategorization frames for *bring*, PREP and DITR. Note that in any given occurrence of *bring*, only one subcategorization frame is used. Consequently, only one of the target patterns is used in any given training step. For those verbs that do not exhibit the dative alternation, only one of these frames will occur in training. This setup guarantees that only positive evidence is available during training.

### Analysis of Results

| class | DITR | PREP |
|---|---|---|
| GIVING | 0.5 | 0.5 |
| DONATE | 0.0 | 1.0 |
| TRANSPORT | 0.5 | 0.5 |
| INSTANTANEOUS | 0.5 | 0.5 |
| CONTINUOUS | 0.0 | 1.0 |
| DIRECTION | 0.5 | 0.5 |
| FUTURE | 0.5 | 0.5 |
| NON-TRANSFER | 1.0 | 0.0 |

Table 6: Preferred subcat frames after training.

Table 6 presents subcategorization frames output by the network after training. For each class, the activation level of the two output units is shown when the network was presented with one of the class members as input. (There were no intra-class differences for the level of precision shown in Table 6.) All verbs are categorized correctly as exhibiting the dative alternation or allowing only one frame. However, the crucial test is the performance for verbs not seen during training. The correct categorization in Table 6 could be the result of rote learning. Generalization to new words would fail in this case.

Table 7 shows results for unseen verbs from the 8 classes, demonstrating that the underlying generalizations were captured by the model. Unseen verbs were encoded by presenting their semantic microfeatures to the net, but with zero activation for the 56 localist nodes that represent the verbs seen during training. Unseen verbs from the classes GIVING, INSTANTANEOUS, FUTURE, NON-TRANSFER, and to a lesser degree TRANSPORT are productive in that new instances have the same or very similar subcat preference compared to the verbs in the training set. For DIRECTION, there is interference from the large number of verbs in CONTINUOUS that share the

| class | DITR | PREP |
|---|---|---|
| GIVING | 0.5 | 0.5 |
| DONATE | 0.5 | 0.5 |
| TRANSPORT | 0.4 | 0.6 |
| INSTANTANEOUS | 0.5 | 0.5 |
| CONTINUOUS | 0.1 | 0.9 |
| DIRECTION | 0.3 | 0.7 |
| FUTURE | 0.5 | 0.5 |
| NON-TRANSFER | 1.0 | 0.0 |

Table 7: Preferred subcat frames for unseen verbs.

fact that they describe a transfer involving non-instantaneous imparting of force. As one would expect, generalization for a verb like *donate* that is an exception fails completely. Since it has the same properties as a verb from the GIVING class, it is classified as a verb that allows both subcat frames.

While the model was trained on the 56 verbs overgeneralization occurred initially for *donate* (i.e. the ditransitive construction "he donated her a book" was judged to be acceptable), since it does not fit very well into the generalizations that determine subcategorization. Figure 2 shows how the model recovered from the initial overgeneralization in the later phase of training. Subcategorization selection for verbs other than *donate* is correct after about 100 epochs and doesn't change in later epochs.

## Discussion

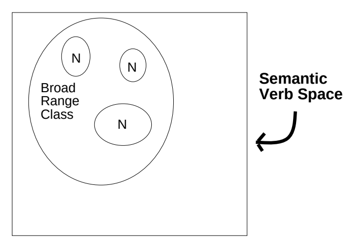


Figure 3: Narrow-range classes ("N") and broad-range classes in Pinker's model of subcategorization.

### The rule-based account.

The proposal put forth in (Pinker, 1989) can be schematized as shown in Figure 3. The square symbolizes the semantic space of all verbs. Verbs are governed by broad-range rules (or classes) and narrow-range rules (or classes). The large circle ("Broad Range Class") is the broad-range class of verbs that is semantically compatible with the subcat frame in question (e.g. dative construction). A verb outside of the circle yields a completely ungrammatical sentence if used with the frame. Any verb in the broad-range class can be understood when used with the subcat frame, but the construction is marginal for many verbs. The small circles marked N are narrow-range classes. For their members, the subcat frame is completely grammatical. If a new verb is created that falls semantically into one of these classes, then it will fully participate in the construction. Verbs inside the B circle but outside of the N circles are overgeneralizations in children's language or creative use in adult language.

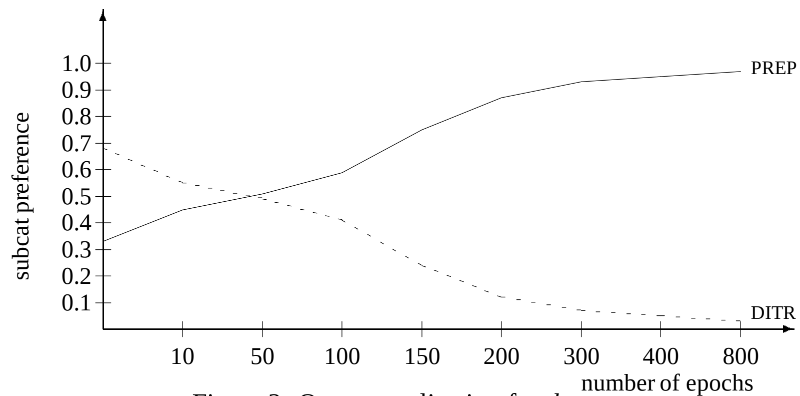


Figure 2: Overgeneralization for *donate*.

## Overgeneralization

Two cases of overgeneralization are distinguished in (Pinker, 1989): extragrammatically conditioned overapplication of broad-range rules and incorrect semantic representations for individual verbs (page 292). Children recover from the second type of error when they learn the correct semantics of a verb. The child recovers from overgeneralization of the first type when he or she learns that a broad-range rule is only a necessary condition for, say, the dative alternation and that the sufficient condition of the applicability of a narrow-range rule must also be satisfied. So this account relies crucially on the formation of narrow-range rules. In errors of the second type, recovery from overgeneralization amounts to acquisition of a narrow-range rule. Narrow-range class formation is based on semantic structure: "Upon noticing that a pair of individual verbs are morphologically and semantically related in a way captured by a nonaffixing broad-range rule, the learner would create a rule whose semantic operations mapped the narrow conflation class specification for one onto the narrow conflation class specification of the other. In other words, the generalization that the learner would make would be: if verb X alternates, other verbs with the same grammatically relevant semantic structure alternate, too." (Pinker 1989:274)

The seed for the formation of a narrow-range rule is hence the comparison of (the semantics of) individual verbs. Unfortunately, the simplest scheme for comparison would not model acquisition correctly as noted by Pinker (p. 278): transfer of subcat frames does not only occur in the case of identical semantic representations (that is the new and old verb's semantics differ only in grammatically irrelevant idiosyncracies). For example, the definition of "manner of speaking" words also covers "manner of communication" verbs (p. 215) which wrongly predicts that the ungrammaticality of the ditransitive construction for manner-of-speaking verbs ("$*$ He shouted her the name.") extends to communication verbs ("John faxed her the name.") which is not true, since the latter sentence is acceptable. There may be ways of loosening comparisons between lexicosemantic structures that give correct predictions for verb subcategorization (cf. Pinker 1989:279). For example, one could define a match to be successful even if one or more values of the features of the verbs' semantic structures are not the same thus relaxing strict identity requirements. However, this would not make the right predictions, for example "throw" and "pull" differ in only one feature, instantaneous vs. continuous imparting of force. At this point it is not clear whether the symbolic model could be extended to give a fully explicit account of semantic similarity, narrow-rule formation and thus overgeneralization.

In an alternative explanation for overgeneralization in the past tense, Clark (1993:103) proposes that an overregularized form such as "goed" may be used in the same period as a correct form such as "went" because the child has not yet recognized that "went" is a form of "to go". Evidence for this position is that one also finds "wents" and "wenting" in children's speech. However, since it must be obvious to the child that the same verb is used in both dative and prepositional constructions, a similar account for overgeneralization does not seem to be possible in the case of verb subcategorization.

In contrast to the symbolic account, the connectionist model exhibits overgeneralization and recovery from it in a natural way. The property of backpropagation crucial here is that it implements gradient descent, i.e. parameters are modified so that a maximal decrease in error is achieved at each step in learning. As a result, the main generalizations are learned in the first phase, since acquiring them covers most cases and increases the error most rapidly. Idiosyncracies, for which the main generalizations make wrong predictions, are then learned in the second phase of learning. No additional assumptions are thus necessary to account for overgeneralization in a connectionist framework. The phenomenon is naturally explained by the properties of gradient descent learning as it is implemented by backpropagation.

## Negative Evidence

Verb subcategorization presents a puzzle to child language research because children learn restrictions on how certain verbs express there arguments although they don't have access to any explicit negative information that a verb does not take a certain subcategorization frame. If the child learns to transform "Mary gave the book to him." into "Mary gave him the book.", then why isn't "Mary donated the book to him" transformed in the same way, if negative evidence is not available? This dilemma is called Baker's paradox. (Marcus, 1993) argues against recent research that claims that children can in fact draw on negative evidence for learning. He demonstrates that the amount of explicit negative evidence actually present is not sufficient to escape from Baker's paradox.

However, the model presented in section 2 shows that **implicit** negative evidence can solve Baker's paradox. To illustrate, consider the verbs of non-transfer. Although the model is never told that these verbs disallow the prepositional construction, it correctly generalizes to the impossibility of this frame for known and novel verbs (Table 7) from the implicit observation that this incorrect subcat frame never occurs in the input. This suggests that implicit negative evidence is sufficient to overcome Baker's paradox, even if explicit negative evidence can be shown to be insufficient.

### Transition from rule-less to rule-governed state.

One of the virtues of a rule-based account lies in the fact that many regularities are most elegantly described by rules. An example would be the "instantaneous" class in Table 1: There doesn't seem to be an exception to the narrow-class rule posited for this class in (Pinker, 1989). If a verb satisfies the description of the "instantaneous" class, then it can be used with both subcat frames.

However, if rules are posited, one needs to explain where they come from. Narrow range classes are apparently not innate since they are specific to English. So in the beginning of learning, the child does not possess any of the narrow-range rules. Again, we are faced with the above-mentioned problem of how narrow-range rules are learned. For the case of the past-tense, Marcus et al. (1992:6) write that the "-ed" rule is created

> . . . presumably by abstracting the regular pattern from a set of regular forms accumulated over time from parental speech and juxtaposed as past and stem forms of the same verb . . .

The problem is that there are several candidates for rules in the first stages of learning: it is not clear what criterion the child would apply to his or her early vocabulary to determine that "-ed" will become a rule and, say, the fit/fit/fit, quit/quit/quit type of regularity will not.

This problem is no less serious in the case of subcategorization learning (unless we are willing to believe that narrow-range rules are innate). Suppose the problem of similarity comparisons mentioned above could be solved somehow. Presumably, the transition from non-rule to rule would correspond to the realization by the child that a large cluster of verbs with similar lexico-semantic grids is really one of the narrow-range rules in the language. Instead of looking up the closest neighbor of a new verb, subcategorization behavior would then be determined by the rule. However, the decision whether a group of verbs is a "large cluster of verbs with similar lexico-semantic grids" is necessarily quantitative. There is always noise in the input, either in the form of incorrect subcategorization frames (Gropen et al., 1989:251), or in the form of exceptions. For example "present" and "bestow" belong to the "giving" class semantically, but using them in the ditransitive construction results in marginal and completely incorrect sentences, respectively. So one cannot demand that a transition from non-rule to rule can only occur if there isn't a single exception or it would never apply. If quantitative factors play an important role in the transition from non-rule to rule, it is hard to see how a purely symbolic learning procedure could succeed in acquiring verb subcategorization correctly.

In the connectionist model in section 2, the problem of transition doesn't arise because a unified framework for rules, analogy, and exceptions is provided. A rule simply emerges if no evidence to the contrary (or very little) is present in training, so that the regularity gets ever stronger in training to the point of exhibiting rule-like behavior. See (Plunkett and Marchman, 1993) for an extensive discussion of the transition from "rote-like" to rule-like behavior in connectionist models.

## Conclusion

This paper compares the explanatory adequacy of a rule-based and a connectionist model for the acquisition of verb subcategorization. The comparison suggests that, while rules are succinct descriptions of the adult linguistic system, a connectionist account is more successful at accounting for the time course of learning, including the phenomena of overgeneralization, learning in the absence of explicit negative evidence, and the transition from the initial rule-less state to final rule-like behavior for some verb classes.

## References


Clark, E. V. (1993). *The Lexicon in Acquisition*. Cambridge University Press, Cambridge.

Daugherty, K. and Hare, M. (1994). What's in a rule? the past tense by some other name might be called a connectionist net. In Mozer, M. C., Smolensky, P., Touretzky, D. S., and Weigend, A. S., editors, *Proceedings of the 1993 Connectionist Models Summer School*. L. Erlbaum Associates, Hillsdale NJ.

Daugherty, K. and Seidenberg, M. (1993). Why no mere mortal has ever flown out to center field but people often say they do. In *Proc. of the Fifteenth Annual Conference of the Cognitive Science Society*, pages 383–388, Boulder CO.

MacWhinney, B. and Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb model. *Cognition*, 40:121–157.

Marcus, G. F. (1993). Negative evidence in language acquisition. *Cognition*, 46:53–85.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., and Xu, F. (1992). *Overregularization in Language Acquisition*, volume 57:4 of *Monographs of the Society for Research in Child Development*. The University of Chicago Press.

Pinker, S. (1989). *Learnability and Cognition*. The MIT Press, Cambridge MA.

Pinker, S. and Prince, A. (1988). On language and connectionism. *Cognition*, 28(1):73–194.

Plunkett, K. and Marchman, V. (1993). From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition*, 48:21–69.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In Rumelhart, D. E., McClelland, J. L., and the PDP Research Group, editors, *Parallel Distributed Processing. Volume 1* . The MIT Press, Cambridge MA.

Rumelhart, D. E. and McClelland, J. L. (1986). On learning the past tenses of English verbs. In McClelland, J. L., Rumelhart, D. E., and the PDP Research Group, editors, *Parallel Distributed Processing. Volume 2*. The MIT Press, Cambridge MA.