

*A computational model of learning verb subclasses in natural L1 acquisition*

An abstract submitted to PsychoCompLA-2007

Authors:

Misha Becker

Linguistics Department

318 Dey Hall, CB#3155

University of North Carolina

Chapel Hill, NC 27599-3155

[mbecker@email.unc.edu](mailto:mbecker@email.unc.edu)

W. Garrett Mitchener

Department of Mathematics

College of Charleston

Robert Scott Small Building, Room 339

Charleston, SC 29424

[MitchenerG@cofc.edu](mailto:MitchenerG@cofc.edu)

### *A computational model of learning verb subclasses in natural L1 acquisition*

We developed an algorithm to model the learning of three verb classes: raising verbs (e.g., *seem*), control verbs (*try*) and ambiguous verbs that can be used as either (*begin*) (1a-c). These classes of verbs present an interesting learning problem because they are both used with *to*+infinitive complements, yet raising and control verbs have distinct syntactic and semantic properties. Any algorithm that attempts to classify an unknown verb by initially assuming the most restricted class (control) and passing to less restricted classes on the basis of positive evidence (such as use with an expletive subject) cannot distinguish ambiguous verbs from raising verbs. Thus, we developed an algorithm that uses semantic cues in the input.

Previous research (Becker, 2005) pointed to the usefulness of two cues found in sentences containing these verbs: animacy of the sentence subject, and eventivity of the predicate embedded under the main verb (2). Animate subjects are compatible with both raising and control main verbs (though they occur predominantly with control verbs) (2a), but inanimate subjects are compatible only with raising verbs (2b). Embedded stative predicates more commonly occur with raising main verbs (2c), while eventive predicates tend to occur with control main verbs (2d). However, to classify a verb, it is insufficient to only use the proportions at which a verb occurs in each of the four possible semantic frames (animate or inanimate subject plus eventive or stative predicate): Many raising verbs (e.g., *gonna*) occur with animate subjects so often that the crucially informative uses with inanimate subjects are relatively rare, particularly in child-directed speech.

We developed a learning algorithm that maintains numbers representing the verb's preference or aversion to each semantic frame. It automatically discards example sentences that merely reinforce its existing knowledge. This property, which was inspired by linear reward-penalty learning with batch (Yang, 2002), enables it to correctly classify raising verbs like *gonna* that occur frequently with animate subjects. After receiving all input sentences, the algorithm settles to one of twenty rest states. To test the algorithm, we counted the number of occurrences of a few common raising, control and ambiguous verbs with each of the four semantic frames in both the CHILDES database (child-directed speech; MacWhinney 2000) and an annotated version of the Switchboard corpus (adult-directed speech; Taylor et al., 2003, Bresnan et al., 2002). See Tables 1-2. We synthesized input sentences for each verb according to these proportions and fed them to the algorithm. The final states of many runs yield distinct patterns for the three verb classes. In addition, when learning a control verb, each run of the algorithm begins in a neutral state characteristic of raising verbs and drifts toward a state characteristic of control verbs. This effect is harmonious with child grammaticality judgments that vary with age (Becker, 2006): Younger children (age 3) are more likely to accept control verbs where only a raising verb is appropriate, and learn not to make this mistake over several years (by age 5).

- (1) a. John seems to be clever.  
 b. John tried to win the race.  
 c. It began to rain. (raising)  
 c.' John began to write a novel. (control)
- (2) a. Amy seems/tried to be a good waitress. (animate subject)  
 b. The truck seemed/\*tried to roll down the hill. (inanimate subject)  
 c. Gordon seemed to be leaving/?leave. (raising verb with stative/?eventive pred)  
 d. Gordon tried to ?be leaving/leave. (control verb with ?stative/eventive pred)

Table 1. Numbers of Verb Classes with Animate/Inanimate Subject and Eventive/Stative Predicate, CHILDES

Verb subclass	Animate+Eventive	Animate+Stative	Inanimate+Eventive	Inanimate+Stative
Raising	1097	149	37	35
Control	604	110	2	0
Ambiguous	40	4	0	4

Table 2. Numbers of Verb Classes with Animate/Inanimate Subject and Eventive/Stative Predicate, Switchboard

Verb subclass	Animate+Eventive	Animate+Stative	Inanimate+Eventive	Inanimate+stative
Raising	273	241	40	127
Control	716	175	4	0
Ambiguous	673	173	18	32

## References

- Becker, M. (2005) Learning verbs without arguments: The problem of raising verbs, *Journal of Psycholinguistic Research*, 34, 165-191.
- Becker, M. (2006) There Began to be a Learnability Puzzle. *Linguistic Inquiry*, 37, 441-456.
- Bresnan, J., J. Carletta, R. Crouch, M. Nissim, M. Steedman, T. Wasow and A. Zaenen (2002) *Paraphrase analysis for improved generation, LINK project*, Stanford: HCRC Edinburgh-CSLI Stanford.
- MacWhinney, B. (2000) *The Child Language Data Exchange System*, Mahwah: Lawrence Erlbaum Associates.
- Taylor, A., M. Marcus and B. Santorini (2003) The PENN Treebank: An overview. In *Treebanks: The state of the art in syntactically annotated corpora*, ed. by A. Abeille, Dordrecht: Kluwer.
- Yang, C. (2002) *Knowledge and Learning in Natural Language*, New York: Oxford University Press.