TITLE: Computational Models of Learning the Raising-Control Distinction

For the special issue on Computational Models of First Language Acquisition

ABSTRACT:

We discuss several algorithms to model the learning of three verb classes: raising verbs (e.g., *seem*), control verbs (e.g., *try*) and verbs that are ambiguous between being raising or control (e.g., *begin*). These classes of verbs present an interesting learning problem because they occur in sentences with similar surface forms, yet raising and control verbs have distinct syntactic and semantic properties. Previous research pointed to the usefulness of two cues found in sentences containing these verbs: animacy of the sentence subject, and the eventivity of the predicate embedded under the main verb. We discuss three algorithms that attempt to classify a verb as raising, control, or ambiguous given sample sentences with information about subject animacy and predicate eventivity based on proportions found in two corpora of naturalistic speech (one adult-directed and the other child-directed). Simple proportions of the semantic frames are insufficient to classify the verbs. A perceptron trained through a boosting-like minimization process gives better results, but is overly sensitive to the sample size. We develop an on-line accumulator algorithm that can successfully learn to distinguish the three classes of verbs through a gradual process, similar to that observed in children.

KEYWORDS: boosting, child language acquisition, learning algorithm, perceptron, raising-control, syntax

1 Introduction

One of the fundamental problems in language learning, in addition to segmenting the speech stream and mapping lexical meanings onto individual lexical items, is that of determining the hierarchical structure that underlies the string of words that form a given sentence. The learner receives help toward this end from various sources. Basic word order is learned very early, such that infants at 14 to 17 months of age, who are not yet speaking in sentences, are able to distinguish correct from incorrect word order (SVO vs. *SOV for English) in comprehension (Hirsh-Pasek and Golinkoff, 1991; Fernald, 1992). Thus, once a child has figured out the basic word order of his language, he can assume that the subject will precede the predicate and, within the predicate, the verb will precede the object (for English). Children also make assumptions about the grammatical relations between adjacent words. For example, children assume that the noun phrases (NPs) adjacent to a verb are thematic *arguments* of the verb. In other words, if a sentence has the string form NP V NP, one NP is the subject and one NP is the object (Fisher et al., 1991).

The source of these supporting assumptions is subject to debate. According to some researchers, children rely on distributional information in the input in order to discover word order patterns (Tomasello, 2000). According to others, these assumptions are made available as part of the child's Universal Grammar apparatus (Crain, 1991). We hold the view that although learners surely exploit distributional patterns in the input in learning their particular language (something we rely upon in our demonstration below), learners must also make certain a priori assumptions about language. In particular, learners must assume that predicate argument relations can be *but need not be* local. In some cases, as we will see, argument relations extend across clause boundaries. Learners must also assume that not only can there be hidden structure (e.g. silent arguments) within sentences, but that pairs of sentence strings that appear to be the same on the surface can in fact be associated with different underlying structures.

The question for the language learner is this: how can you determine the structural properties of predicates that distribute similarly, yet ultimately have different structural properties? In this paper, we discuss the case of distinguishing raising verbs (e.g. *seem*) from control verbs (e.g. *try*). Raising and control verbs overlap in one of the syntactic environments they occur in. Although the verb classes are distinguishable in certain other environments, the presence of a class of ambiguous verbs (verbs that can be either raising or control) gives rise to a subset conundrum for the learner: it turns out that the learner cannot simply adopt a default assumption as to the category of a novel verb, because there is no true subset relation between the raising and control classes (cite author's paper).

In the following sections we develop learning models that use probabilistic tendencies in the input to approximate the way in which actual language learners (i.e., children) might acquire the raising and control verb classes over time. We begin with an overview of the syntactic and semantic properties of raising and control verbs, and the reason why a subset-based solution is not viable. We then present the results of searches of corpora of spoken adult language which gives a picture of what the input to children potentially looks like. We use the Switchboard corpus of adult-directed speech and the CHILDES corpus of child-directed speech. Finally, we describe our learning models and discuss the results of our implementation.

To give a brief preview, we investigate three learning models that attempt to classify a verb as raising, control, or ambiguous, from sample sentences essentially of the surface form

(1) John likes to run
    SUBJECT MAIN-VERB *to* PREDICATE

Since there is no surface syntactic information in such sentences to determine which class the main verb belongs to, the data includes basic semantic information about the subject (whether it is animate or inanimate) and predicate (whether it is eventive or stative). All three classification procedures show significant distinctions between data from child-directed speech and adult-directed speech.

The first learning algorithm is to look at the proportions with which a verb occurs in different semantic frames. This gives an elementary means of classifying verbs, however, there are no thresholds on the proportions that correctly classify all the data from the corpora.

The second is a perceptron trained jointly with a set of threshold functions. This algorithm can correctly classify the adult-directed speech on which it was trained, but it is sensitive to the size of the input corpus and misclassifies several verbs from the child-directed speech. Furthermore, it must be given the entire set of training data at once.

We develop a third algorithm that accumulates each verb's preferences for each of four different semantic frames on-line, and these preferences classify the verb as raising, control, or ambiguous. This algorithm is less sensitive to the corpus size, and with some tuning, it works well on both child- and adult-directed speech.

2 Raising and Control

Both raising and control verbs can occur in the string in (2).

(2) Scott ______ to paint with oils.

a. Scott$_i$ tends [t_i to paint with oils] (raising)

b. Scott$_i$ likes [PRO$_i$ to paint with oils] (control)

The primary difference between the two constructions is that in the control sentence there is a thematic (semantic, selectional) relationship between the main verb and the subject, while in the raising sentence there is no such relation: the subject of the sentence is thematically related only to the lower predicate (*paint with oils*). Structurally speaking, according to movement-based syntactic frameworks the subject in (2a) is said to *raise* up to its surface position, leaving behind a trace (with which it is coindexed), while in (2b) the subject is generated in the main clause and *controls* the reference of PRO, the silent subject of the embedded clause. Thus, the string in (2) represents a case where, until the learner has acquired the syntactic and semantic properties of the main verb, the learner cannot simply take a string of input and immediately deduce the correct underlying structure.

There are other types of sentence frames that do distinguish these classes of verbs. For instance, control verbs cannot occur with an expletive (semantically empty) subject (e.g. *it, there*), while raising verbs can. This is because control verbs assign a θ-role to their subject, and expletives cannot bear a θ-role (Chomsky, 1981).

(3) There tend to be arguments at poker games.

(4) * There like to be arguments at poker games.

Additionally, there are sentence frames that allow some control verbs but no raising verbs, such as transitive or intransitive frames.[1,2]

(5) John likes bananas.

(6) * John tends bananas.

Taking a naïve view of the learning procedure one might hypothesize that, since only control verbs are banned from sentences like (3), a learner should assume, given a sentence

[1]The raising verbs *tend* and *happen* have homophonous forms that are (in)transitive, e.g. *John tends sheep*, or *Interesting things happened yesterday*. The general problem of homophony is significant for learning but is beyond the scope of this paper.

[2]Verbs in this frame could also be non-control transitive or intransitive verbs, like *eat*; of interest here are verbs that could occur in both a transitive/intransitive frame and a frame with an infinitive complement. We ignore here the further problem that transitive or intransitive verbs can occur with an adjunct infinitive clause, as in *John runs to stay in shape*.

like (2), that the novel verb is a control verb. If the learner has guessed incorrectly, she will eventually encounter an input sentence like (3), and this datum would provide the triggering evidence to change her grammar. Furthermore, if the learner has guessed correctly, she would have data such as (5) to confirm her categorization of the verb.

However, we contend that such a strategy is insufficient, and that learners need to draw on multiple sources of information in order to determine the category of a novel verb encountered in the syntactic environment (2) in the input. This is because of the existence of verbs that are ambiguous between having a raising or a control interpretation. The ambiguous verbs include *begin*, *start*, *continue* and *need*. As discussed by Perlmutter (1970) these verbs can occur in raising contexts (e.g., with an expletive subject), but they can also have a control interpretation when they occur with an animate subject.

(7) It began to rain. (raising)

(8) Rodney began to talk to Zoe. (control)

Moreover, ambiguous verbs also occur in single clause frames as transitive or intransitive verbs.

(9) The game continued all afternoon.

(10) The referee started the match.

Since ambiguous verbs can occur in all of the environments that both raising and control verbs can, their existence raises a challenge for language learners. *Begin* will be heard with an expletive subject, as in (7), where it will be analyzed as a raising verb, and it will be heard with an animate subject as in (8), where it should be analyzed as a control verb. But *tend*, which is unambiguously raising, will also be heard with expletive subjects (3) and animate subjects (*Scott tends to paint with oils*). In the absence of explicit negative evidence (Chomsky, 1959; Marcus, 1993) how will a learner determine that *tend* is not ambiguous and therefore functions as a control verb when it occurs with an animate subject? The upshot is that learners cannot simply assume a novel verb heard in (2) is a control verb, or even that a verb in (7) is an unambiguous raising verb. Instead, learners need to use input cues in a probabilistic manner in order to distinguish all three categories of verbs (cite author's paper).

Previous research has shown that two types of cues from within the ambiguous string (such as (2)) provide information about whether a given ambiguous string is likely to be a raising or a control sentence. These cues come from whether the subject is animate or inanimate (see (11a–b)) and whether the predicate inside the infinitive clause is stative or eventive (see (12a–b)).

(11) a. *Samantha* likes to be tall. (*animate*)

b. *The tower* seems to be tall. (*inanimate*)

(12) a. Samantha hates to *mow the lawn.* (*eventive*)

b. Samantha seems to *be happy.* (*stative*)

The evidence comes from a psycholinguistic experiment in which adults were asked to fill in the main verb in an incomplete sentence (cite author's paper). The properties of subject animacy and predicate eventivity were systematically manipulated. Participants gave significantly more control verbs when the sentence had an animate subject or an eventive lower predicate, and they gave significantly more raising verbs when the sentence had an inanimate subject or a stative lower predicate. While these cues indicate tendencies for these verb classes (not definitive restrictions; cf. *Samantha hates to be tall*), the psycholinguistic data show that they are very strong tendencies.

Thus, there is information in the input that can lead a learner toward distinguishing the classes of raising and control verbs. There is empirical evidence that young children make use of some of these cues. For example, when the subject of the sentence is inanimate, 3-

and 4-year-olds interpret a verb in the frame in (2) as if it were a raising verb, even if it is actually a control verb in the adult grammar (cite author's paper). Also, children analyze verbs occurring with an expletive subject as raising verbs (cite author's paper).

As we discuss below, corpus data from both adult-directed and child-directed speech (from Switchboard and CHILDES) show that the preferences adults show for placing control verbs with animate subjects and raising verbs with inanimate or expletive subjects, for example, are also reflected in patterns of usage in naturalistic speech. We hypothesize that learners can draw on these patterns in the input, synthesizing them together with assumptions based on UG principles about the mapping between surface strings and structure in language (two superficially similar strings need not have the same structure) and about the nature of thematic relations, in order to learn the structural properties of raising, control and ambiguous predicates.

3 Description of Data

We have searched two sources of naturalistic spoken language. The Switchboard corpus (Taylor et al., 2003) contains naturalistic adult-to-adult speech recorded in phone conversations. The CHILDES database (MacWhinney, 2000) contains dozens of corpora of speech to children, much of it recorded in spontaneous conversations between parents or researchers and young children.

3.1 CHILDES

Due to the need for handcoding of the CHILDES data, our dataset of child-directed speech is somewhat limited in size. We analyzed the mothers' speech (the *MOT tier) in all of the Adam, Eve and Sarah files within the Brown (1973) corpus. The total number of *MOT utterances in the Brown corpus is over 59,500.

We began the search by searching for all occurrences of specific raising, control and ambiguous verbs followed by the word *to*, using the Clan program (search string syntax: `combo +t*mot +s‘‘(seem+seems+seemed)^ * ^ to’’ adam*.cha`). Each utterance in the output was then coded by hand (by the first listed author) for whether the subject was animate or inanimate, and whether the predicate inside the infinitive phrase was eventive or stative. Cases in which the subject was null were counted if it was clear from the context what the referent of the subject was. For example, Adam's mother's question "Want to give this to Ursula?" was clearly directed at Adam ("(Do you) want to ...") and so it was counted as having an (implied) animate subject. Utterances in which the lower predicate was elided or unclear were not counted. Subject animacy was judged according to whether the referent was living or nonliving (also whether it would be replaced with a *he/she* pronoun or *it* pronoun, with the exception that insects would likely be referred to with *it* but are alive). Predicate eventivity was judged according to whether the predicate typically occurs in the present progressive where it has an on-going meaning (these are eventive: e.g., *John is walking*) or whether it occurs in the plain present tense without a habitual meaning (these are stative: e.g., *John knows French*). The results, summing across the three children's mothers, are given in Tables 1–3. All verb occurrences here are those with an infinitival complement.

Table 1: Mothers' Distribution of Raising Verbs

Verb	Animate+Eventive	Animate+Stative	Inanimate+Eventive	Inanimate+Stative
seem	0	4	4	5
used	32	13	2	3
going	1065	132	31	27
Total	1097	149	37	35
	88% eventive		51% eventive	

All of the verb classes are heavily skewed towards having an animate subject and an

Table 2: Mothers' Distribution of Control Verbs

Verb	Animate+Eventive	Animate+Stative	Inanimate+Eventive	Inanimate+Stative
want	354	53	2	0
like	156	54	0	0
try	86	0	0	0
love	7	3	0	0
hate	1	0	0	0
Total	604	110	2	0
	85% eventive		100% eventive	

Table 3: Mothers' Distribution of Ambiguous Verbs

Verb	Animate+Eventive	Animate+Stative	Inanimate+Eventive	Inanimate+Stative
start	5	0	0	0
begin	1	0	0	0
need	34	4	0	4
Total	40	4	0	4
	91% eventive		0% eventive	

eventive predicate. The main difference between the classes is that the raising verbs also have non-zero occurrences with inanimate subjects and stative predicates. With the exception of the verb *need*, an ambiguous verb, and possibly *want* (which is traditionally categorized as purely control, but according to some dialects it can occur with an expletive subject and therefore may also be ambiguous) the other verbs do not occur with inanimate subjects and rarely with stative predicates.

3.2 Switchboard

The Switchboard corpus (Taylor et al., 2003) contains over 100,000 utterances of adult-to-adult spontaneous speech recorded in telephone conversations. The corpus is parsed, and a portion of it has been annotated to indicate the animacy of each NP (Bresnan et al., 2002). We searched through the annotated corpus using the program Tgrep2 (Rohde, 2005), which searches for hierarchical structures, for all occurrences of specific raising, control and ambiguous verbs followed by an infinitive complement. The numbers of occurrences with animate versus inanimate subjects were then tallied. Animate subjects were those annotated as being human, animal or organizations. Inanimate subjects were those tagged as a place, time, machine, vehicle, concrete or nonconcrete.

Subsequent to this first search, all output utterances were then coded by hand for whether the infinitive predicate was eventive or stative. This was carried out by entering all of the output of the first search into spreadsheets and having three different research assistants code the predicates according to the same criteria used for the CHILDES data (if it occurs in present progressive with an on-going meaning it is eventive; if it occurs in plain present with a non-habitual meaning it is stative). The degree of coder agreement varied among the verb classes, with the least agreement with raising verbs (78% agreement, based on *seem*) to the most agreement with ambiguous verbs (93% agreement, based on *need*; they had 88% agreement with control verbs, based on *want*). Disagreements were resolved by going with the majority result (2 out of 3 coders' judgments) except in cases where there was a 3-way split (1 coder judged stative, 1 judged eventive and 1 judged unclear) or in the very few cases where 2 coders appeared to have made errors (e.g., judging *understand* to be eventive) in which case the first author made the judgment call. Such cases amounted to less than 1% of the data. The results are given in Tables 4–6.

The numbers from the Switchboard search are larger than those from CHILDES due to the much larger amount of data, and perhaps in part to differences in child-directed versus adult-directed speech (with the exception of *going-to/gonna*, which appears to be relatively rare in the Switchboard data). The main trend in the Switchboard data is that while the raising verbs are evenly split between having an eventive or a stative predicate when the subject is animate, and there are many occurrences of these verbs with inanimate subjects,

Table 4: Distribution of Raising Verbs in (Annotated) Switchboard

Verb	Animate+Eventive	Animate+Stative	Inanimate+Eventive	Inanimate+Stative
seem	24	57	23	71
used	156	96	2	35
going	44	11	3	6
tend	36	37	10	9
happen	13	20	2	6
Total	273	241	40	127
	55% eventive		24% eventive	

Table 5: Distribution of Control Verbs in (Annotated) Switchboard

Verb	Animate+Eventive	Animate+Stative	Inanimate+Eventive	Inanimate+Stative
want	342	123	3	0
try	149	12	1	0
like	181	33	0	0
love	18	0	0	0
hate	20	7	0	0
choose	6	0	0	0
Total	716	175	4	0
	80% eventive		100% eventive	

control verbs are overwhelmingly biased towards having an eventive predicate and almost never occur with inanimate subjects. Ambiguous verbs, as a group, are in between the raising and control classes on both counts: like the raising verbs they have nonzero numbers of occurrences with both inanimate subjects and stative predicates, but like control verbs they show a bias for eventive predicates when the subject is animate.

4 Learning Models

Research over the past several years has shown that children, even prelinguistic infants, are very good at noticing statistical patterns in the world around them, and it has been suggested

Table 6: Distribution of Ambiguous Verbs in (Annotated) Switchboard

Verb	Animate+Eventive	Animate+Stative	Inanimate+Eventive	Inanimate+Stative
need	208	62	3	23
have	442	105	5	6
start	14	1	7	2
begin	0	4	2	1
continue	9	1	1	0
Total	673	173	18	32
	80% eventive		36% eventive	

that children make use of these regularities and patterns in acquiring language. Various models of input-based language learning have been proposed over the years for learning different aspects of language: past tense morphology (Rumelhart and McClelland, 1986), constituent order (Saffran et al., 1996), grammatical structure (Gomez and Gerken, 1997; Hudson-Kam and Newport, 2005), and verb argument structure (Alishahi and Stevenson, 2005a,b). Some approaches make hybrid use of both input patterns and UG principles, as in Yang's account of parameter setting using the Variational Learning paradigm (Yang, 2002), while others rely more or less wholly on input for learning. All of these proposals incorporate the fact that many patterns in language are of a probabilistic nature. For example, a given verb can occur in various syntactic frames, but it may be more likely to occur in some than in others (Lederer et al., 1995).

Alishahi and Stevenson (2005a,b; henceforth A&S), although having a somewhat different goal from our work, likewise focus on the learning of verb-argument structure. They adopt a Bayesian framework to model the phenomenon of children's overgeneralization errors in using intransitive verbs in a transitive frame with a causative meaning (causative meaning is compatible with a transitive syntactic frame but not typically with an intransitive syntactic frame). For example, children sometimes say "Adam fall toy" to mean Adam makes the toy fall (Bowerman, 1982). In A&S's model, similar syntactic frames (e.g., transitive, intransitive, ditransitive, etc.) are grouped together according to their shared semantic properties, where semantic properties are understood as (combinations of) primitive features such as CAUSE or MOVE. Syntactic frames, which include the verb, are associated with semantic properties with a certain probability. The more frequently a given semantic feature appears in general, the higher its probability of being associated with a given individual syntactic frame. A&S show that after running their learning simulation on 800 input utterance-meaning pairs using the most common verbs found in the mothers' speech in the Brown (1973) corpus, their learner managed to learn these verbs with the expected U-shaped learning curve. Moreover, the learner made some of the same overgeneralizations in sentence frame use in the production portion of the test that actual children make (e.g. inserting an intransitive verb in a transitive frame with causative meaning).

Crucially for A&S, the learner is able to deduce both the syntactic frame of the sentence they are perceiving, and also the meaning of the utterance (based upon perception of the nonlinguistic scene that is cooccurring with the utterance). The problem we are interested in is significantly more difficult than the one tackled by A&S (and, therefore, these assumptions do not hold), for two reasons. One is that syntactic ambiguity is involved in parsing the string, such that we do not assume that the learner can immediately deduce the structure upon hearing the string. Secondly, given the abstractness of the verb meanings we are interested in (cf. *seem*, *want*), we do not assume that the child can immediately determine the meanings of these verbs based on observation of the environment.[3] In fact, as A&S correctly point out, the syntactic frame of a verb and its lexical meaning are closely tied together, such that if children could immediately determine the meaning of *want* or *seem* upon hearing it in a sentence and observing a scene, knowledge of the syntactic properties of the sentence would follow. But for the reasons just cited neither of these pieces of information (the structure of the sentence, the meanings of verbs) are available a priori to learners for the types of sentences we are interested in.

[3]The idea that children could deduce the meaning of any verb, even concrete ones, based on observation of the world is challenged in the syntactic bootstrapping literature; see Gleitman (1990).

We assume, instead, that the learner, around age 3 years, is able to detect clausal and phrasal boundaries (known to be achieved by age 9–10 months; Hirsh-Pasek et al. 1987; Jusczyk et al. 1992), identify common nouns and verbs (e.g. *cat*, *dog*, *book*, *be happy*, *run*, *be small*), know which nouns denote animate things (*cat*) and which denote inanimate things (*book*), and we assume children at this age also know the basic word order of their language (see above for empirical support).

Our approach is to develop a model of how children might make use of distributional patterns in the input to distinguish raising verbs and control verbs, and additionally to distinguish the ambiguous verbs from either nonambiguous set. We focus on only a small subset of the verbs to be acquired, representative of each of the three classes (raising, control, ambiguous).

5 A comparison of learning algorithms

The learning problem at hand is: Determine whether a particular verb may be used in raising or control syntax (or both) given a set of sentences using that main verb, along with information about whether the subject is animate or inanimate, and whether the predicate in the embedded clause is eventive or stative, giving four possible semantic frames: animate+eventive, animate+stative, inanimate+eventive, inanimate+stative. These will be abbreviated as the symbols AE, AS, IE, and IS. In this section, we examine three learning strategies: simple proportions of semantic frames, a perceptron, and a special accumulator algorithm.

CHILDES		
try	C	1.000
want	C	0.866
going	R	0.849
need	A	0.810
seem	R	0.000

Switchboard		
try	C	0.920
want	C	0.731
going-st	R	0.706
need	A	0.703
going	R	0.688
tend	R	0.391
seem	R	0.137

Table 7: Table of verbs in decreasing order of the proportion of sentences using the verb that have an animate subject and an eventive predicate. The line *going-st* in the Switchboard table is a synthetic data point which is used as a sensitivity test. It comes from adding 4 animate+eventive sentences to the *going* data in Table 4. The center column of each table indicates whether the verb is raising (R), control (C), or ambiguous (A).

5.1 Ordering by proportions

Given the data as in Section 3, a reasonable place to start is to compute for each verb the fraction of sentences of each type in which it occurs. As seen in Table 7, sorting the verbs according to this simple calculation comes close to separating them into the three classes. However, in each corpus, the verb *going* (a raising verb) is problematic. In the CHILDES data, it occurs above *need* (ambiguous). In the Switchboard data, it occurs below *need,* however, this ordering is not robust. As a sensitivity test, we create a synthetic data point *going-st* by adding four animate+eventive sentences to the counts from *going*, which could easily come from one additional conversation. Note that this minor change is just enough to move *going* out of order to the line *going-st.*

The key indicator that a verb is of the raising class is its use with inanimate subjects (particularly expletives) and/or stative predicates. However, these types of sentences are fairly rare in comparison to sentences with animate subjects and eventive predicates (see section 3 above). The raising predicate *going to/gonna* is particularly troublesome because it is common in the child-directed speech we observed and the vast majority of its uses are with an animate subject and an eventive predicate. Therefore, a learning algorithm hearing a verb in many animate+eventive sentences must determine whether the other three types of sentences are grammatical but uncommon, indicating that the verb is raising, or that the other three types of sentences are vanishingly rare, indicating that the verb is control. Consequently, it must ignore most but not all of the animate+eventive sentences.

Thus, simply looking at the proportion of animate+eventive sentences fails. Instead, an appropriate algorithm should have some sort of threshold: It must count the number of occurrences of each verb in each sentence type, but once enough data has arrived to confirm that the verb can definitely be used in a certain sentence type, the decision is made and further instances of the verb in that sentence type give no additional information.

The rest of this section describes in detail two more elaborate learning algorithms we developed for this verb classification problem. The first is a modified perceptron trained through a process related to boosting. The second is an accumulator algorithm similar to linear reward-penalty with batch (LRPB) (Yang, 2002) except that it is constantly pulled toward a discrete set of rest states, so its long term behavior is characterized by occasional shifts from one rest state to another.

5.2 The perceptron algorithm

5.2.1 Mathematical description

The perceptron is a simple classifier that takes as input a vector of numbers between -1 and 1. The sign of a linear combination of these inputs categorizes the input into one of two classes, one for positive output and one for negative output.

We would like to train a perceptron to classify verbs as raising, control, or ambiguous based on how often they occur in the four sentence types. Specifically, for each semantic frame $s \in \{\mathrm{AE}, \mathrm{AS}, \mathrm{IE}, \mathrm{IS}\}$, there is an input function

(13) $$\phi_s(x) = \tanh\left(a_s(x - c_s)\right).$$

These functions translate sentence counts into inputs that fall between -1 to 1 and pass through a center point $(c_s, 0)$ with steepness controlled by the constraint $\phi_s'(c_j) = a_s$. Each verb will be classified as ±raising and ±control, thus we need two linear combinations of the inputs with weights $\lambda_{t,s}$ where t is the feature R for raising or C for control, and s is the sentence type. Given a vector of counts,

$$\mathbf{n} = \begin{pmatrix} n_{\mathrm{AE}} \\ n_{\mathrm{AS}} \\ n_{\mathrm{IE}} \\ n_{\mathrm{IS}} \end{pmatrix}$$

the classifier function

$$\begin{aligned} P_{\mathrm{R}}(\mathbf{n}) &= \lambda_{\mathrm{R}} \cdot \phi(\mathbf{n}) \\ &= \lambda_{\mathrm{R,AE}}\phi_{\mathrm{AE}}(n_{\mathrm{AE}}) + \lambda_{\mathrm{R,AS}}\phi_{\mathrm{AS}}(n_{\mathrm{AS}}) + \lambda_{\mathrm{R,IE}}\phi_{\mathrm{IE}}(n_{\mathrm{IE}}) + \lambda_{\mathrm{R,IS}}\phi_{\mathrm{IS}}(n_{\mathrm{IS}}) \end{aligned}$$

predicts whether the verb can be used with raising syntax. A positive output indicates that the perceptron thinks the count vector $\mathbf{n}$ is from a raising verb and a negative number if not. Since the verbs in question have a second binary feature, we need a second function

$$\begin{aligned} P_{\mathrm{C}}(\mathbf{n}) &= \lambda_{\mathrm{C}} \cdot \phi(\mathbf{n}) \\ &= \lambda_{\mathrm{C,AE}}\phi_{\mathrm{AE}}(n_{\mathrm{AE}}) + \lambda_{\mathrm{C,AS}}\phi_{\mathrm{AS}}(n_{\mathrm{AS}}) + \lambda_{\mathrm{C,IE}}\phi_{\mathrm{IE}}(n_{\mathrm{IE}}) + \lambda_{\mathrm{C,IS}}\phi_{\mathrm{IS}}(n_{\mathrm{IS}}) \end{aligned}$$

which gives a positive number if the perceptron thinks the count vector $\mathbf{n}$ is from a control verb, and a negative number if not. An ambiguous verb should generate positive output for both P_{R} and P_{C}.

Perceptrons are traditionally trained with a feedback algorithm that iteratively improves the $\lambda_{t,s}$, but for this problem the steepness constants a_s and centers c_s used in computing the inputs must also be selected. We therefore define a joint training process in terms of an optimization problem. For each verb v, let $\mathbf{n}_v$ be its count vector, let y_{R} be 1 if the verb is raising and -1 if not, and let y_{C} be 1 if the verb is control and -1 if not. Define a penalty called *loss* by

$$L_v = \exp\left(-y_{\mathrm{R},v} P_{\mathrm{R}}(\mathbf{n}_v)\right) + \exp\left(-y_{\mathrm{C},v} P_{\mathrm{C}}(\mathbf{n}_v)\right)$$

If P correctly classifies $\mathbf{n}_v$, then the signs of $y_{\mathrm{R},v} P_{\mathrm{R}}(\mathbf{n}_v)$ and $y_{\mathrm{C},v} P_{\mathrm{C}}(\mathbf{n}_v)$ are both positive, which yields a small loss. Conversely, misclassification results in large loss. The net penalty or *risk* is defined to be the sum of the losses for each verb in a training set T,

$$R(T) = \sum_{v \in T} L_v = \sum_{v \in T} \exp\left(-y_{\mathrm{R},v} P_{\mathrm{R}}(\mathbf{n}_v)\right) + \exp\left(-y_{\mathrm{C},v} P_{\mathrm{C}}(\mathbf{n}_v)\right) \tag{14}$$

The learning algorithm is to find the values of a_s, c_s, and $\lambda_{t,s}$ that minimize R. If there happen to be any parameter settings that correctly classify all the training data, then R has no absolute minimum because those values of $\lambda_{t,s}$ can be scaled without changing the output of the classifier. We therefore impose the constraints

$$\sum_{s \in \{\mathrm{AE,AS,IE,IS}\}} \lambda_{\mathrm{R},s} = 1 \text{ and } \sum_{s \in \{\mathrm{AE,AS,IE,IS}\}} \lambda_{\mathrm{C},s} = 1$$

The risk function (14) is the same as the one used in AdaBoost, so the learning algorithm here may also be understood as a variant on boosting (Schapire., 2003).

5.2.2 Results and interpretation

The learning algorithm was programmed into Mathematica, using its built-in simulated annealing algorithm to minimize R. The training data was the count vectors for all verbs in the Switchboard corpus. The steepness, center, and weight constants so generated are shown in Table 8. This classifier correctly classifies all 16 verbs in the training set. It also correctly classifies the synthetic point *gonna-st* as raising but not control. It also correctly

a_{AE}	$= 0.0125776$	c_{AE}	$= 2.83354$	$\lambda_{\text{R,AE}}$	$= -120.642$	$\lambda_{\text{C,AE}}$	$= 65.6656$
a_{AS}	$= 0.0425202$	c_{AS}	$= -5.80528$	$\lambda_{\text{R,AS}}$	$= 47.935$	$\lambda_{\text{C,AS}}$	$= -9.03134$
a_{IE}	$= 442.997$	c_{IE}	$= 0.00437486$	$\lambda_{\text{R,IE}}$	$= 53.7703$	$\lambda_{\text{C,IE}}$	$= -11.5359$
a_{IS}	$= 3.3847$	c_{IS}	$= 5.25297$	$\lambda_{\text{R,IS}}$	$= 19.9371$	$\lambda_{\text{C,IS}}$	$= -44.0984$

Table 8: Perceptron classifier generated from the Switchboard data.

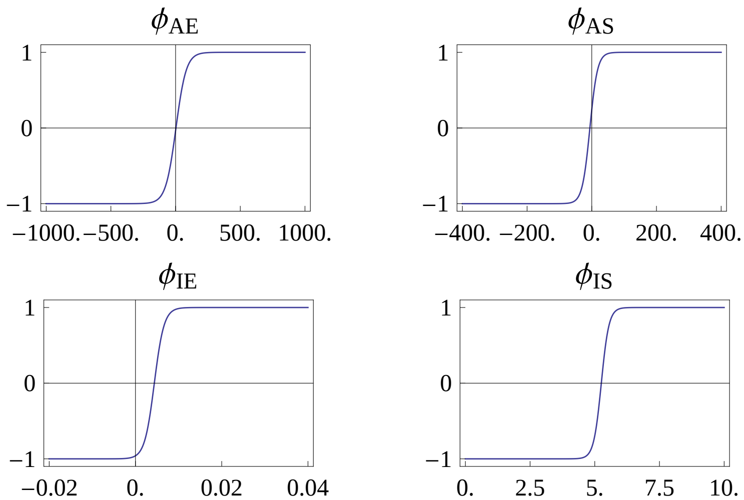


Figure 1: Perceptron input functions

classifies *try* and *want* from the CHILDES data. However, it misclassifies *need* as control when it should be ambiguous, *gonna* as ambiguous when it should be raising, and *used* as ambiguous when it should be raising.

To explain how this classifier works, the functions ϕ_s are graphed in Figure 1. The AE and AS functions have very shallow slopes, so they contribute very little information unless there are many sentences of those types. The IE function has a steep slope and is centered just to the right of zero, so $\phi_{\text{IE}}(x)$ is essentially 1 if $x > 0$ and -1 if $x = 0$. Likewise, $\phi_{\text{IS}}(x)$ is essentially 1 if $x > 5$ and -1 if $x \leq 5$. The weights λ have very intuitive signs: $\lambda_{\text{R,AS}} < 0$ so AS sentences are a counter-indicator for raising verbs, and the other three sentence types are favorable indicators. Weights for control verbs have the opposite signs, so AS sentences are favorable indicators for control verbs and the others are counter-indicators.

The optimization problem has 16 unknowns and 2 constraints for a total of 14 degrees of freedom. Given that there are only 16 training points, there is some chance that it has overfitted the data. However, boosting algorithms typically do not suffer from the overfitting problems seen in some other learning algorithms, and the resulting classifier is sensible. So, it seems that the results of this computational experiment are meaningful.

Unfortunately, this is an off-line algorithm, that is, it requires a complete sample of sentences to run, and it is awkward to formulate the algorithm in a way that can incorporate new data one sentence at a time. The classifier does not give the correct output for several verbs in the CHILDES data. Some of the misclassified verbs are fairly close to the decision boundary: If two more inanimate+stative sentences are added to the *need* data, then the ϕ_{IS} input changes, and the resulting synthetic point is classified properly.

Another aspect of its failure is that the counts of sentence type are implicitly assumed to come from a specific sample size. For example, if the counts of all the training data are doubled, as if twice as many sentences were collected to form the corpus, the center parameter c_{IS} nearly doubles to 9.1 to compensate. The CHILDES data includes a huge number of *gonna* sentences, so for this verb, the overall sample size huge compared to the sample from Switchboard. In contrast, the sample size for other verbs is smaller for CHILDES. It is therefore not surprising that this classifier has trouble with the CHILDES data. Simply scaling the sentence counts by the sample size does not solve this problem, as discussed in Section 5.1.

5.3 The accumulator algorithm

Given the limitations of the perceptron, we now develop an on-line learning algorithm that is more independent of the corpus size.

5.3.1 Informal description

Let us assume that for each relevant verb, the brain maintains a memory register that can be in one of several discrete states, indicating various mixtures of preference for raising and control constructions. If a sample sentence comes along that reinforces the current state of the register, then the state should remain nearly the same. If an isolated sample sentence comes along that contradicts the current state of the register, then it might be noise and should be ignored. However, if several sample sentences are given that contradict the current state, then it should change.

The accumulator algorithm works by examining a set of sentences with a common main verb one by one. For each of the four possible semantic frames, the algorithm maintains a number that measures how readily the verb may be used in a sentence of that type. To give a physical analogy, one could represent these numbers by cutting egg cartons so as to make four linear strips of five cups each, and placing a marble in the central cup of each strip. (See Figure 2 for a single marble in its strip, and Figure 3 for a picture of all four strips.) Each strip represents one of the sentence types. When the algorithm receives a sentence, it gives the marble corresponding to that type a small push to the right, and it gives the other marbles a small push to the left. Eventually the pushes may build up enough to push a marble into the next cup. The farther to the right a marble is, the more readily the verb may be used in that sentence type. So, for a strict control verb like *try*, the animate+eventive marble typically ends up in the rightmost cup, and the other marbles end up one or two cups to the left, indicating that *try* is used infrequently with the other types. For a raising verb like *seem* that is used frequently in all four types of sentences, all four marbles stay close to the center.

The pushes to the right are weaker if the marble is already to the right of center. This means that no marble ever gets pushed off the end of its strip, and that once the algorithm establishes that the verb can be used in a particular semantic frame, it attaches less importance to new sample sentences of that type. Also, each marble moves with a certain stickiness as if it were covered in honey. This means that once pushed, it rolls down slowly enough that the small pushes have a chance to accumulate and push it into the next cup.

Once all the sample sentences have been seen, each marble comes to rest at the bottom of one of the five cups in its strip, as in Figure 7, and the state of the algorithm can be represented by drawing a row of squares for each of the strips and coloring each square corresponding to the cup where one of the marbles settled. In an initial numerical experiment, sentences based on proportions from the CHILDES corpus were fed to the algorithm. Figures 5 through 6 show the positions of the four marbles as functions of time.

As a more thorough numerical experiment, the algorithm was run 200 times for each of several different verbs. The sample sentences were constructed based on the proportions of sentence types found in the two corpora (CHILDES and Switchboard) with 100 runs for each corpus, and each run of the algorithm was given 100 sample sentences. The histograms in Figures 8 through 12 show what fraction of the runs yielded various final states for the different verbs.

Section 5.3.2 gives the mathematical details of the algorithm. Three numerical parameters are introduced that control the algorithm's behavior: The overall strength of the small pushes is represented by the number γ and the duration of the pushes is controlled by σ. The stickiness of the marbles is controlled by β. The results of the numerical experiments and reasonable values for γ, σ, and β are presented in Section 5.4.

5.3.2 Mathematical details

For each verb, the algorithm maintains four numbers between -1 and 1, one for each semantic frame. A positive number indicates a preference for that frame and a negative number indicates an aversion. Each number represents the location of a particle in a force field

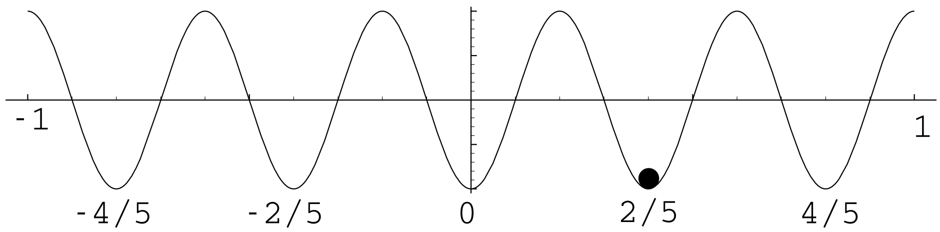


Figure 2: A particle in the potential $v(x) = \cos 5\pi x$, resting at the bottom of one of the wells. The potential exerts a force on the particle proportional to $-dv/dx$, so the particle tends to move downhill and settle at the bottom of a well. Other forces due to the influence of sample sentences can push the particle uphill, and possible into another well.

whose potential is

$$v(x) = -\cos(5\pi x) \tag{15}$$

The potential has 5 wells between -1 and 1. A particle at location x experiences a force equal to $-v'(x)$ due to the potential. Particles can move freely but tend to settle at the bottoms of the wells at $-4/5$, $-2/5$, 0, $2/5$ and $4/5$, thereby discretizing the state space. Each verb is represented by four particles, whose locations at time t are denoted $x_{\mathrm{AE}}(t)$, $x_{\mathrm{AS}}(t)$, $x_{\mathrm{IE}}(t)$, and $x_{\mathrm{IS}}(t)$. For convenience, we introduce the vector notation

$$\mathbf{x}(t) = \begin{pmatrix} x_{\mathrm{AE}}(t) \\ x_{\mathrm{AS}}(t) \\ x_{\mathrm{IE}}(t) \\ x_{\mathrm{IS}}(t) \end{pmatrix}, \quad V(\mathbf{x}) = v(x_{\mathrm{AE}}) + v(x_{\mathrm{AS}}) + v(x_{\mathrm{IE}}) + v(x_{\mathrm{IS}}). \tag{16}$$

Now the force on the whole particle system due to the force field at time t is easily expressed as $-\operatorname{grad} V(\mathbf{x}(t))$.

In addition to the force field, each sample sentence exerts a force on each particle as follows. Each semantic frame is associated with a particular pattern vector, the entries of which represent the scale of such a sentence's force on the four particles:

$$(17) \qquad \mathbf{p}_{\mathrm{AE}} = \begin{pmatrix} 1 \\ -1/4 \\ -1/4 \\ -1/2 \end{pmatrix}, \quad \mathbf{p}_{\mathrm{AS}} = \begin{pmatrix} -1/4 \\ 1 \\ -1/2 \\ -1/4 \end{pmatrix}, \quad \mathbf{p}_{\mathrm{IE}} = \begin{pmatrix} -1/4 \\ -1/2 \\ 1 \\ -1/4 \end{pmatrix}, \quad \mathbf{p}_{\mathrm{IS}} = \begin{pmatrix} -1/2 \\ -1/4 \\ -1/4 \\ 1 \end{pmatrix}.$$

The pattern $\mathbf{p}_{\mathrm{AE}}$ for animate eventive sentences comes from setting the AE entry (first entry) to 1, setting the entries for frames that differ in one aspect (IE and AS) to $-1/4$, and setting the entry for the frame that differs in both aspects (IS) to $-1/2$. The total of the pattern is 0. The AE particle is given a push toward 1, and the other particles are given a push toward -1. The other three pattern vectors are constructed similarly. If a sentence of type T arrives at time t_0, then it creates a force on the particles with potential

$$(18) \qquad L(\mathbf{x}, t, T, t_0) = \begin{cases} 0 & \text{if } t < t_0, \\ \frac{1}{2} e^{-(t-t_0)/\sigma} \left\|\mathbf{x} - \mathbf{p}_T\right\|^2 & \text{if } t \geq t_0. \end{cases}$$

This sentence's contribution to the potential pushes the particle system toward the pattern $\mathbf{p}_T$. The exponential part causes the force to weaken over time, and σ is a decay rate constant to be determined. The force at time $t \geq t_0$ is given by $-\operatorname{grad} L(\mathbf{x}(t), t, T, t_0)$ where the gradient is taken with respect to the entries of $\mathbf{x}$. See Figure 3 for an example.

Finally, to ensure that the particles settle down, we add friction terms, so that each particle experiences a damping force proportional to its velocity. In vector notation, these forces are $-d\mathbf{x}/dt$.

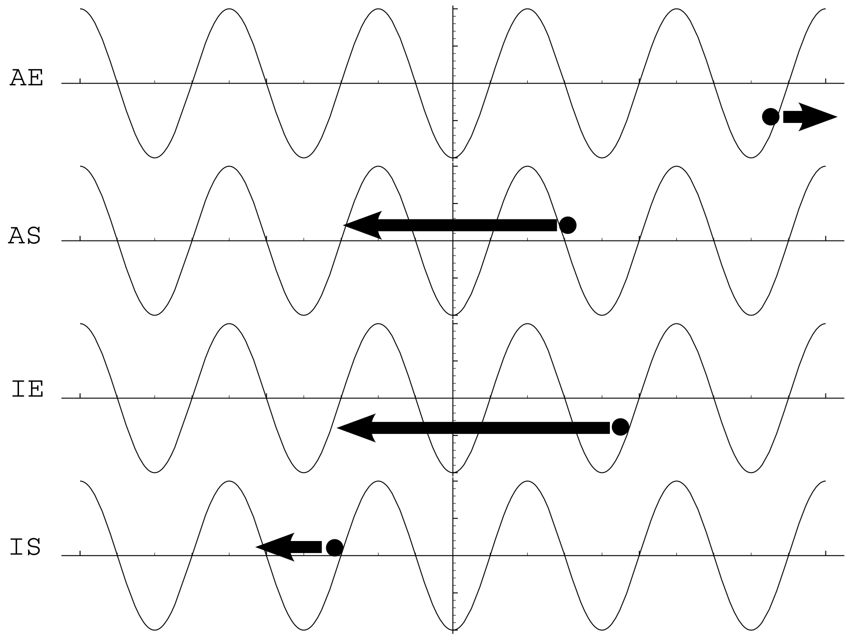


Figure 3: Four particles in potentials, representing the state of the algorithm after it has seen some data. The arrows indicate the relative strengths of the forces exerted on the particles by a sentence of type AE (animate subject, eventive predicate). The size of the force on each particle is proportional to the distance between that particle's location and the corresponding entry of $\mathbf{p}_{\mathrm{AE}}$.

The overall behavior of the particles is governed by the differential equation

(19)
$$\begin{aligned}\frac{d^2\mathbf{x}(t)}{dt^2} &= \mathbf{F}_{\mathrm{field}} + \sum_{\mathrm{inputs}} \mathbf{F}_{\mathrm{input}} + \mathbf{F}_{\mathrm{friction}} \\ &= -\operatorname{grad} V(\mathbf{x}(t)) - \gamma \left(\sum_i \operatorname{grad} L(\mathbf{x}(t), t, T_i, t_i) \right) - \beta \frac{d\mathbf{x}(t)}{dt}\end{aligned}$$

The constants γ and β control the relative magnitudes of the forces from input sentences and friction. The sum is over all inputs, where the i-th input is a sentence of type T_i that arrives at time t_i.

5.4 Results

The accumulator learning algorithm includes three unknown constants, σ, γ, and β. In addition, the rate of arrival of input sentences is unknown. However, some experimenting

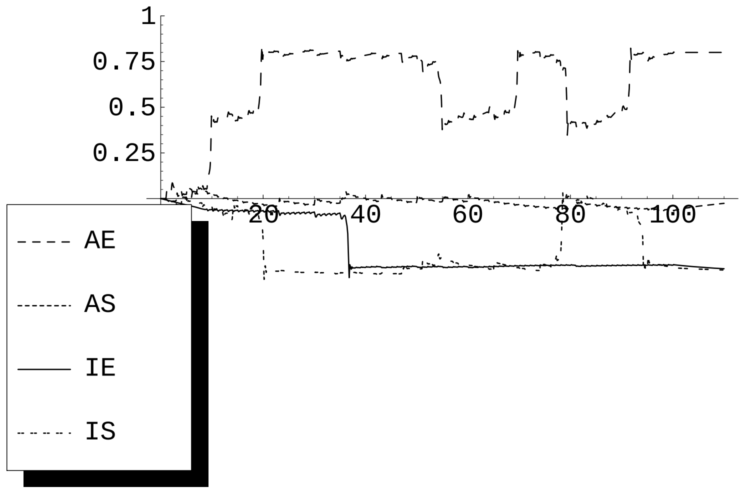


Figure 4: Particle dynamics tests for *need* using proportions matching the CHILDES corpus.

shows that if sample sentence t_i arrives at $t = i$ (a rate of one input per time unit), then the following parameter values give reasonable results:

(20)
$$\sigma = 10$$
$$\gamma = 8$$
$$\beta = 8$$

With these values, the differential equation (19) may be solved by standard numerical methods.

As a first test of the algorithm, we pick a verb and feed 100 randomly created sentences to the algorithm. Since the algorithm only cares about two pieces of semantic information from each input, there is no need to generate actual sentences, so each simulated input need only indicate a semantic. The semantic frames are generated in proportions matching that verb's occurrence with animate/inanimate subjects and with eventive/stative predicates in the CHILDES data. The particle dynamics run out to time $t = 110$ to give the system 10 time units to settle after the last input arrives. See Figures 4–6 for time traces of the particle positions for each verb.

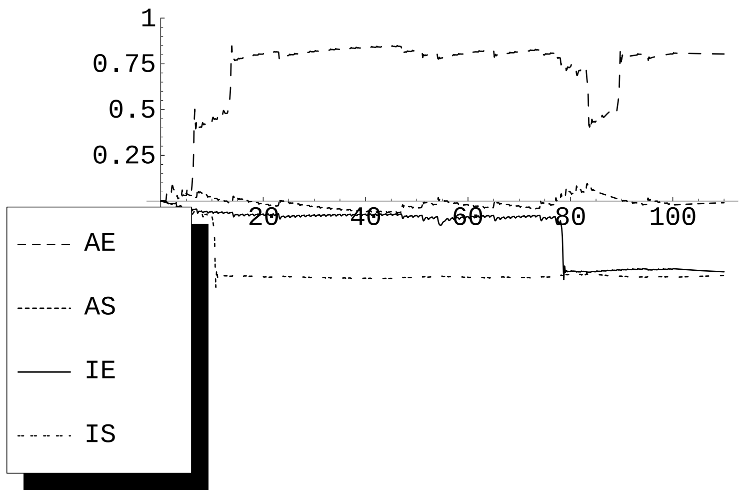


Figure 5: Particle dynamics tests for *want* using proportions matching the CHILDES corpus.

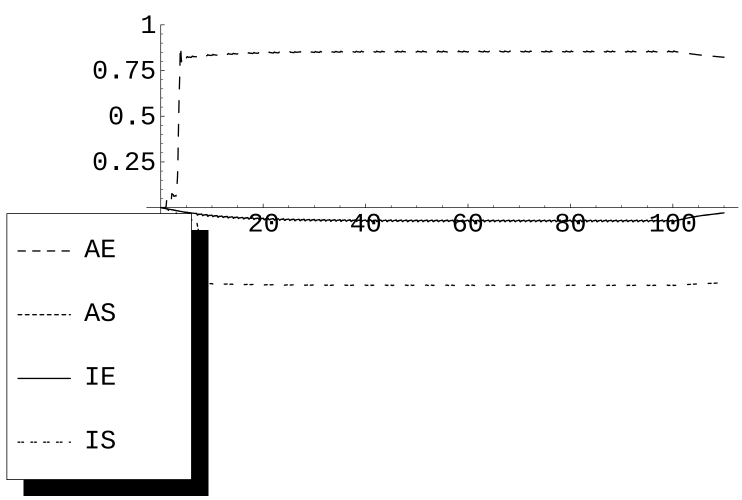


Figure 6: Particle dynamics tests for *try* using proportions matching the CHILDES corpus.

The verb *try* is quite distinct, showing a very strong preference for animate subjects and eventive predicates. The verbs *need* and *want* show an intermediate pattern with some preference for animate+eventive. Typically *want* is considered a control verb, but there are several non-control idiomatic uses that some speakers accept, such as "It looks like it wants to rain." Some of these uses are present in the CHILDES data, as in "This one just doesn't want to go right," (Sarah, file 135). So, the fact that the algorithm displays similar behavior for *want* and *need* is to be expected.

Rather than look at just one run of the algorithm, it is helpful to examine its behavior for many randomly generated input samples. The fact that the algorithm distinguishes raising, control, and ambiguous verbs is made clearest by running the algorithm many times on different sets of randomly generated sentences for each verb, and building a histogram of its final states. See Figures 7 through 13 for histograms showing the proportion of times the algorithm came to rest in various final states in 100 random trials. Since there is so little CHILDES data for *seem*, its Switchboard histogram is displayed next to a histogram generated from an even mixture of the four sentence types. There is even less CHILDES data for *tend,* so its Switchboard histogram is displayed alone. Note that *try, seem,* and *tend* are particularly distinct, which agrees with the fact that *try* is very strongly control, and *seem* and *tend* are very strongly raising.

There are three patterns (labeled A, B, and C) that occur many times when learning the verbs *try, want, need,* and *going to/gonna*, as shown in Figure 14. For each combination of corpus and verb, these patterns occur at characteristic frequencies. The strongly raising verbs *seem* and *tend* do not exhibit these patterns. This observation suggests that the fraction of runs of the algorithm that end in one of these three patterns might make a suitable index for classifying verbs. The new index is a sum of three terms,

(21) $$H = A + B + \max\{A, B, C\}$$

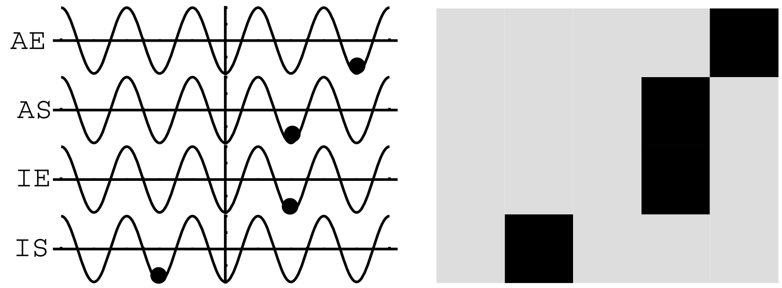


Figure 7: Histogram key: The final state of the algorithm will be represented by a 4 row by 5 column array of colored squares. The algorithm maintains one particle for each of four sentence types (AE, AS, IE, IS). At the end of the learning process, each particle settles to the bottom of one of five wells in the basic potential (left). Each particle is represented by a row of five squares, and the one corresponding to its rest position is colored black (right).

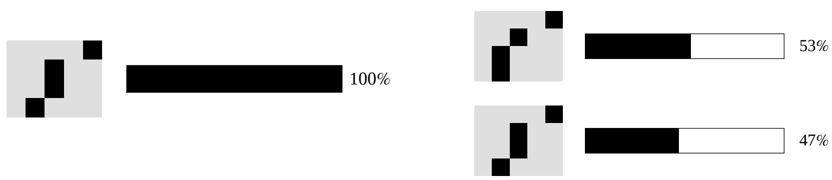


Figure 8: Histograms of the learning algorithm for *try* using the sentence type proportions from CHILDES (left) and Switchboard (right).

where A is the percentage of times the verb causes the algorithm to end up in pattern A, and likewise for B and C. The $A + B$ term is included because it is large for control verbs. The maximum of A, B, and C is included because it is rather low for *going* and other control verbs: They can be used in a greater variety of patterns which leads to a greater variety of final states. The result of ordering most of the verbs listed in Section 3 by H is shown in Figure 15. The verbs used to design the algorithm (*try*, *want*, *need*, *going*, *tend*, *seem* and the synthetic points *seem-eq* (in which all four frames occur at equal rates) and *going-st*) occur in the correct order, and with ample margin between verb types. Most of the other verbs also occur in the correct order. The exceptions are *continue*, *start*, and *begin*, which are rare in the corpus, and *have*, which has so many uses that it poses a learnability problem all by itself.

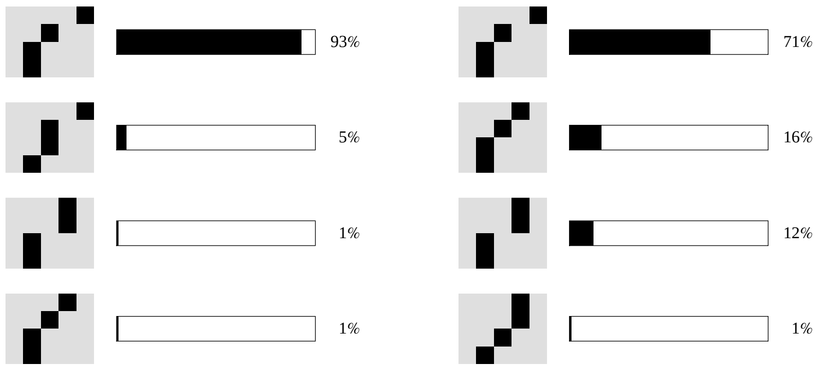


Figure 9: Histograms of the learning algorithm for *want* using the sentence type proportions from CHILDES (left) and Switchboard (right).

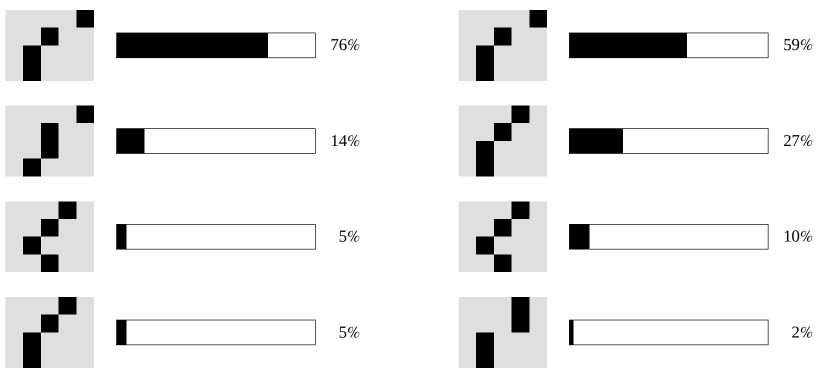


Figure 10: Histograms of the learning algorithm for *need* using the sentence type proportions from CHILDES (left) and Switchboard (right). Only the most common patterns are shown.

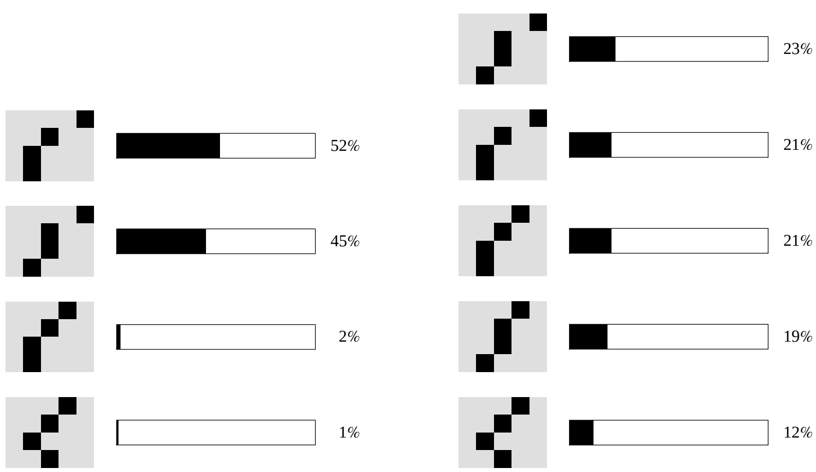


Figure 11: Histograms of the learning algorithm for *going to/gonna* using the sentence type proportions from CHILDES (left) and Switchboard (right). Only the most common patterns are shown.

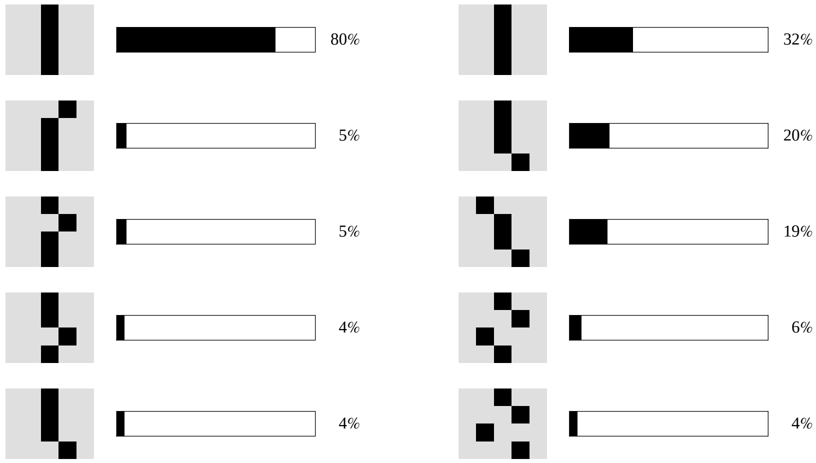


Figure 12: Histograms of the learning algorithm for *seem-eq* using equal sentence type proportions (left) and the sentence type proportions for *seem* from Switchboard (right). Only the five most common patterns are shown.

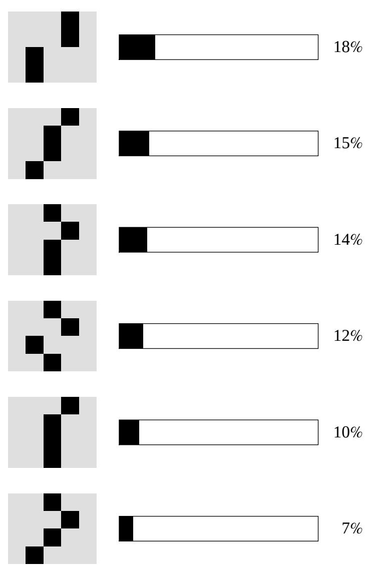
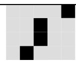
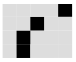
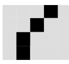


Figure 13: Histograms of the learning algorithm for *tend* using the sentence type proportions from Switchboard. Only the six most common patterns are shown. There were too few occurrences of *tend* in CHILDES to justify feeding any data for *tend* to the algorithm.

CHILDES		try	want	need	going
A		100	5	14	45
B		0	93	76	52
C		0	1	5	2
	other	0	1	5	1

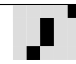
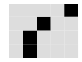
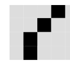
Switchboard		try	want	need	going
A		47	0	0	23
B		53	71	59	21
C		0	16	27	21
	other	0	13	14	35

Figure 14: The three most frequently occurring patterns (labeled A, B, and C) and the percentage of trial runs in which they occur.

CHILDES			
try		C	200
want		C	191
⇓ *raising*		*178.5*	
need		A	166
⇑ *control*		*157.5*	
going		R	149
used	+	R	69
seem-eq		R	0

Switchboard			
choose	+*	C	200
love	+*	C	200
like	+	C	196
have	+!	A	186
continue	+*!	A	180
try		C	153
hate	+*	C	148
want		C	142
⇓ *raising*			*130*
need		A	118
⇑ *control*			*105*
going-st		R	92
going		R	67
used	+	R	34
tend		R	2
start	+*!	A	2
seem		R	0
happen	+	R	0
begin	+*!	A	0

Figure 15: Verbs sorted by the index $H = A + B + \max\{A, B, C\}$. Thresholds are indicated halfway between verbs of different types. The verbs not used in designing the algorithm are marked with a plus (+). The verbs out of order are marked with an exclamation point (!). The verbs occurring in fewer than 30 sentences in the corpus are marked with a star (*).

6 Discussion

We began with the broad question of how language learners determine the underlying structure of a given string, given that even with knowledge of basic word order of a language, many sentence strings are potentially compatible with multiple underlying structures. Focusing on the case of a string that could host either a raising or a control verb in the matrix verb position, we showed that the learner cannot simply resolve the ambiguity of this string with a subset-type strategy. That is, the learner cannot assume that a novel verb in this string is a control verb, on the supposition that an incorrect assumption will be defeated by hearing the verb with an expletive subject.

The reason is that the class of ambiguous verbs (*begin, start, need*, etc.) occur in all of the sentential environments that both raising and control verbs do. Thus, there is no proper subset relation between raising and control verbs. Although evidence of a verb's occurrence with expletive subjects provides useful information to learners (and learners are certainly expected to use this information), we have argued that learners must additionally rely on semantic cues within the ambiguous string in a probabilistic manner, in order to distinguish the classes of raising, control and ambiguous verbs. Based on experimental evidence from adult speakers, we identified four relevant semantic features: animate vs. inanimate subjects, and eventive vs. stative embedded predicates.

The simple classification strategy of looking at the proportions of a verb's use in four sentence types fails. Some verbs are used overwhelmingly with animate subjects and eventive predicates, and these drown out the fact that the verb can be used in distinctly raising contexts.

A perceptron, trained jointly with functions that translate sentence counts into numbers between -1 and 1, does better. It learns thresholds such that if a verb is seen at least some minimum number of times in a sentence type, then the verb may be definitely be used in that context. It can be trained using by solving a minimization problem similar to the one used by AdaBoost. This learning algorithm can correctly classify all of the verbs as seen in the Switchboard corpus. However, it is an off-line algorithm, and it is sensitive to the size of the input corpus. Furthermore, it has many parameters.

We therefore developed an on-line accumulator algorithm with graduated output and the ability to ignore inputs that simply confirm what it already knows. Tests of this algorithm reveal that it is capable of distinguishing different classes of verbs from the frequencies of their use in basic semantic frames. Running the algorithm once on a set of sample sentences with a particular main verb yields a pattern that describes the verb's preference for animate and inanimate subjects, and eventive and stative predicates. Control verbs can be identified by their stronger preference for animate subjects and eventive predicates. Ambiguous verbs are harder to identify by a single run of the algorithm, but repeating the algorithm with more data yields stronger distinctions. An index H based on the fraction of runs of the algorithm that end in each of three states sorts verbs from control to raising that correctly orders the most common such verbs (except for *have*) in the CHILDES and Switchboard corpora, with reasonably large margins between the verb classes. This algorithm has fewer parameters than the perceptron, and it gives better results for CHILDES after being tuned for the Switchboard data. However, the thresholds between verb classes are clearly different between the two corpora, so it is not yet possible to claim that it correctly classifies all the CHILDES verbs. Other disadvantages of this algorithm are that it is complex, and the selection of parameters is somewhat ad-hoc. Furthermore, it is quite different from the well-studied algorithms of statistical learning theory, and further research should be done on it.

The accumulator algorithm is designed to mimic the gradual learning process observed by Author (cite author's paper). When learning a verb, the initial state of the particles is neutral, allowing all types of sentences, and only with a significant amount of data do some types become strongly preferred or dispreferred. This parallels the tendency of young children to accept control verbs in syntactic contexts that are appropriate only for raising syntax, and gradually learn the proper usage as they acquire the adult grammar. The algorithm learns without any need for a subset relationship between the sets of possible raising and control constructions.

The histograms of final states and the index H reveal that the accumulator algorithm behaves quite differently when given data matching the proportions of the two different corpora. In particular, the Switchboard proportions for *going to/gonna* yield a much wider variety of patterns than the CHILDES proportions. It would seem that child-directed speech is even more heavily biased toward animate subjects with eventive predicates than typical adult conversation. Such sentence strings could be associated with either raising or control syntax and in that sense offer less information than the other three kinds of sentences concerning the class of the main verb. If children are discarding such sentences as uninformative, then the bias in child-directed speech in favor of them might contribute to the tendency of young children to accept control verbs in raising constructions: Initially, the child-directed speech they hear contains insufficient information and they misclassify many verbs. As they age, they hear more adult conversation, which contains more informative sentence types and should eventually lead them to learn the proper class for each verb.

Importantly, while we incorporate the frequency of occurrence of particular syntactic and semantic frames into the accumulator learning model, we assume that the learner makes certain assumptions about language structure prior to experience. For instance, the learner must assume that similar strings can be associated with divergent underlying structures, and that semantic relationships need not be local (i.e., the subject of the sentence might be semantically related only to a predicate in a lower clause, not the immediately string-adjacent verb). In addition, to derive the semantic properties of these verbs, learners must be biased to assume that inanimate or expletive subjects are unlikely to be agents (along the lines of Dowty (1991) or Keenan (1976)), and therefore that verbs that occur with these subjects are unlikely to assign them a thematic role. These assumptions are necessitated by the particular learning problem at hand: the classes of raising, control and ambiguous verbs could not be distinguished without these assumptions (for instance, on a purely input-based learning model). However, if these assumptions are in place for learning this particular set of verb classes, they should in principle be available for learning other classes of verbs. We hold the view, then, that learning brings these assumptions about language to bear on the language learning task in general.

Sorting by proportion of AE sentences, the perceptron, and the accumulator all exhibit divergence between the Switchboard and CHILDES data: The thresholds between the different classes of verbs are unequal, and a classifier trained on one corpus does not work well on the other. There is therefore evidence that either these particular corpora are unrepresentative of actual speech, or more likely, that child-directed speech uses generally different proportions of animate and eventive predicates than adult directed speech. In future work, the statistical differences between adult and child directed speech should be studied, including the extent to which oddities in child language acquisition may be attributed to these differences versus features of the underlying learning algorithm.

Acknowledgments

This material is based upon work supported by the National Science Foundation.

References

Alishahi, Afra, and Suzanne Stevenson. 2005a. The acquisition and use of argument structure constructions: A Bayesian model. In *Proceedings of the ACL 2005 Workshop on Psychocomputational Models of Human Language Acquisition*.

Alishahi, Afra, and Suzanne Stevenson. 2005b. A probabilistic model of early argument structure acquisition. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*.

Bowerman, Melissa. 1982. Evaluating competing linguistic models with language acquisition data: Implications of developmental errors with causative verbs. *Quaderni di Semantica* 3:5–66.

Bresnan, Joan, Jean Carletta, Richard Crouch, Malvina Nissim, Mark Steedman, Tom Wasow, and Annie Zaenen. 2002. *Paraphrase analysis for improved generation, link project*. Stanford, CA: HRCR Edinburgh-CLSI Stanford.

Brown, Roger. 1973. *A first language*. Cambridge, MA: Harvard University Press.

Chomsky, Noam. 1959. Review of *verbal behavior*. *Language* 35:26–58.

Chomsky, Noam. 1981. *Lectures on government and binding: The Pisa lectures*. New York: Mouton de Gruyter.

Crain, Stephen. 1991. Language acquisition in the absence of experience. *Behavioral and Brain Sciences* 14.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67:547–619.

Fernald, Ann. 1992. Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In *The adapted mind: Evolutionary psychology and the generation of culture*, ed. et al. Barkow. Oxford: Oxford University Press.

Fisher, Cynthia, Henry Gleitman, and Lila R. Gleitman. 1991. On the semantic content of subcategorization frames. *Cognitive Psychology* 23:331–392.

Gleitman, Lila. 1990. The structural sources of verb meanings. *Language Acquisition* 1:3–55.

Gomez, Rebecca, and LouAnn Gerken. 1997. Artificial grammar learning in one-year-olds: Evidence for generalization to new structure. In *Proceedings of BUCLD 21*, ed. Elizabeth Hughes, Mary Hughes, and Annabel Greenhill, 194–204.

Hirsh-Pasek, Kathy, and Roberta Golinkoff. 1991. Language comprehension: A new look at some old themes. In *Biological and behavioral aspects of language acquisition*, ed. N. Krasnegor, D. Rumbaugh, M. Studdert-Kennedy, and R. Shiefelbusch, 301–320. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hirsh-Pasek, Kathy, D.G. Kemler Nelson, P.W. Jusczyk, K.W. Cassidy, B. Druss, and L. Kennedy. 1987. Clauses are perceptual units for young infants. *Cognition* 26:269–286.

Hudson-Kam, Carla, and Elissa Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1:151–196.

Jusczyk, Peter, Kathy Hirsh-Pasek, K.G. Kemler Nelson, L.J. Kennedy, A. Woodward, and J. Piwoz. 1992. Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology* 24:252–293.

Keenan, Edward. 1976. Toward a universal definition of subjects. In *Subject and topic*, ed. Charles Li. New York: Academic Press.

Lederer, Anne, Henry Gleitman, and Lila Gleitman. 1995. Verbs of a feather flock together: Semantic information in the structure of maternal speech. In *Beyond names for things: Young children's acquisition of verbs*, ed. Michael Tomasello and William E. Merriman, 277–297. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

MacWhinney, Brian. 2000. *The child language data exchange system*. Mahwah, NJ: Lawrence Erlbaum Associates.

Marcus, Gary. 1993. Negative evidence in language acquisition. *Cognition* 46:53–85.

Perlmutter, David M. 1970. The two verbs *begin.* In *Readings in English transformational grammar*, ed. Roderick A. Jacobs and Peter S. Rosenbaum, 107–119. Waltham, Mass.: Ginn.

Rohde, Douglas L. T. 2005. Tgrep2 user manual. Manuscript.

Rumelhart, David, and J.L. McClelland. 1986. On learning the past tenses of English verbs. In *Parallel distributed processing, volume 2*. Cambridge, MA: MIT Press.

Saffran, Jenny, Richard Aslin, and Elissa Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274:1926–1928.

Schapire., Robert E. 2003. The boosting approach to machine learning: An overview. In *Nonlinear estimation and classification*, ed. D. D. Denison, M. H. Hansen, C. Holmes, B. Mallick, and B. Yu. New York: Springer.

Taylor, Ann, Mitchell Marcus, and Beatrice Santorini. 2003. The penn treebank: An overview. In *Treebanks: The state of the art on syntactically annotated corpora*, ed. Anne Abeillé. Dordrecht: Kluwer.

Tomasello, Michael. 2000. Do young children have adult syntactic competence? *Cognition* 74:209–253.

Yang, Charles. 2002. *Knowledge and learning in natural language*. New York: Oxford University Press.