A Cognitive Model for the Representation and Acquisition of Verb Selectional Preferences

Afra Alishahi
Department of Computer Science
University of Toronto
afra@cs.toronto.edu

Suzanne Stevenson
Department of Computer Science
University of Toronto
suzanne@cs.toronto.edu

Abstract

We present a cognitive model of inducing verb selectional preferences from individual verb usages. The selectional preferences for each verb argument are represented as a probability distribution over the set of semantic properties that the argument can possess—a *semantic profile*. The semantic profiles yield verb-specific conceptualizations of the arguments associated with a syntactic position. The proposed model can learn appropriate verb profiles from a small set of noisy training data, and can use them in simulating human plausibility judgments and analyzing implicit object alternation.

1 Introduction

Verbs have preferences for the semantic properties of the arguments filling a particular role. For example, the verb *eat* expects that the object receiving its theme role will have the property of being edible, among others. Learning verb selectional preferences is an important aspect of human language acquisition, and the acquired preferences have been shown to guide children's expectations about missing or upcoming arguments in language comprehension (Nation et al., 2003).

Resnik (1996) introduced a statistical approach to learning and use of verb selectional preferences. In this framework, a semantic class hierarchy for words is used, together with statistical tools, to induce a verb's selectional preferences for a particular argument position in the form of a distribution over all the classes that can occur in that position. Resnik's model was proposed as a model of human learning of selectional preferences that made minimal representational assumptions; it showed how such preferences could be acquired from usage data and an existing conceptual hierarchy. However, his and later computational models (see Section 2) have properties that do not match with certain cognitive plausibility criteria for a child language acquisition model. All these models use the training data in "batch mode", and most of them use information theoretic measures that rely on total counts from a corpus. Therefore, it is not clear how the representation of selectional preferences could be updated incrementally in these models as the person receives more data. Moreover, the assumption that children have access to a full hierarchical representation of semantic classes may be too strict. We propose an alternative view in this paper which is more plausible in the context of child language acquisition.

In previous work (Alishahi and Stevenson, 2005), we have proposed a usage-based computational model of early verb learning that uses Bayesian clustering and prediction to model language acquisition and use. Individual verb usages are incrementally grouped to form emergent classes of linguistic constructions that share semantic and syntactic properties. We have shown that our Bayesian model can incrementally acquire a general conception of the semantic roles of predicates based only on exposure to individual verb usages (Alishahi and Stevenson, 2007). The model forms probabilistic associations between the semantic properties of arguments, their syntactic positions, and the semantic primitives

Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition, pages 41–48,
Prague, Czech Republic, June 2007

of verbs. Our previous experiments demonstrated that, initially, this probability distribution for an argument position yields verb-specific conceptualizations of the role associated with that position. As the model is exposed to more input, the verb-based roles gradually transform into more abstract representations that reflect the general properties of arguments across the observed verbs.

A shortcoming of the model was that, because the prediction of the semantic roles was based only on the groupings of verbs, it could not make use of verb-specific knowledge in generating expectations about a particular verb's arguments. That is, once it was exposed to a range of verbs, it no longer had access to the verb-specific information, only to generalizations over clusters of verbs.

In this paper, we propose a new version of our model that, in addition to learning general semantic roles for constructions, can use its verb-specific knowledge to predict intuitive selectional preferences for each verb argument position. We introduce a new notion, a *verb semantic profile*, as a probability distribution over the semantic properties of an argument for each verb. A verb semantic profile is predicted from both the verb-based and the construction-based knowledge that the model has learned through clustering, and reflects the properties of the arguments that are observed for that verb. Our proposed prediction model makes appropriate generalizations over the observed properties, and captures expectations about previously unseen arguments.

As in other work on selectional preferences, the semantic properties that we use in our representation of arguments are drawn from a standard lexical ontology (WordNet; Miller, 1990), but we do not require knowledge of the hierarchical structure of the WordNet concepts. From the computational point of view, this makes use of an available resource, while from the cognitive view, this avoids ad hoc assumptions about the representation of a conceptual hierarchy. However, we do require some properties to be more general (i.e., shared by more words) than others, which eventually enables the model to make appropriate generalizations. Otherwise, the selected semantic properties are not fundamental to the model, and could in the future be replaced with an approach that is deemed more appropriate to child language acquisition. Each argument contributes to the semantic profile of the verb through its (potentially large) set of semantic properties instead of its membership in a single class. As input to our model, we use an automatically parsed corpus, which is very noisy. However, as a result of our novel representation, the model can induce and use selectional preferences using a relatively small set of noisy training data.

2 Related Computational Models

A variety of computational models for verb selectional preferences have been proposed, which use different statistical models to induce the preferences of each verb from corpus data. Most of these models, however, use the same representation for verb selectional preferences: the preference can be thought of as a mapping, with respect to an argument position for a verb, of each class to a real number (Light and Greiff, 2002). The induction of a verb's preferences is, therefore, modeled as using a set of training data to estimate that number.

Resnik (1996) defines the selectional preference strength of a verb as the divergence between two probability distributions: the prior probabilities of the classes, and the posterior probabilities of the classes given that verb. The selectional association of a verb with a class is also defined as the contribution of that class to the total selectional preference strength. Resnik estimates the prior and posterior probabilities based on the frequencies of each verb and its relevant argument in a corpus.

Li and Abe (1998) model selectional preferences of a verb (for an argument position) as a set of nodes in the semantic class hierarchy with a probability distribution over them. They use the Minimum Description Length (MDL) principle to find the best set for each verb and argument based on the usages of that verb in the training data. Clark and Weir (2002) also find an appropriate set of concept nodes to represent the selectional preferences for a verb, but do so using a χ^2 test over corpus frequencies mapped to concepts to determine when to generalize from a node to its parent. Ciaramita and Johnson (2000) use a Bayesian network with the same topology as WordNet to estimate the probability distribution of the relevant set of nodes in the hierarchy. Abney and Light (1999) use a different representational approach: they train a separate hidden Markov model for each verb, and the selectional preference is represented as a probability distribution over words instead of semantic classes.

3 The Bayesian Verb-Learning Model

3.1 Overview of the Model

Our model learns the set of *argument structure frames* for each verb, and their grouping across verbs into *constructions*. An argument structure frame is a set of features of a verb usage that are both syntactic (the number of arguments, the syntactic pattern of the usage) and semantic (the semantic properties of the verb, the semantic properties of each argument). The syntactic pattern indicates the word order of the verb and arguments. A construction is a grouping of individual frames which probabilistically share syntactic and semantic features, and form probabilistic associations across verb semantic properties, argument semantic properties, and syntactic pattern. These groupings typically correspond to general constructions in the language such as transitive, intransitive, and ditransitive.

For each verb, the model associates an argument position with a probability distribution over a set of semantic properties—a semantic profile. In doing so, the model uses the knowledge that it has learned for that verb, as well as the grouping of frames for that verb into constructions.

The semantic properties of words are taken from WordNet (version 2.0) as follows. We extract all the hypernyms (ancestors) for all the senses of the word, and add all the words in the hypernym synsets to the list of the semantic properties. Figure 1 shows an example of the hypernyms for *dinner*, and its resulting set of semantic properties.[1]

The following sections review basic properties of the model from Alishahi and Stevenson (2005, 2007), and introduce extensions that give the model its ability to make verb-based predictions.

3.2 Learning as Bayesian Clustering

Each argument structure frame for an observed verb usage is input to an incremental Bayesian clustering process. This process groups the new frame together with an existing group of frames—a construction—that probabilistically has the most similar semantic and syntactic properties to it. If no construction has sufficiently high probability for the new frame, then a new construction is created for it. We use the probabilistic model of Alishahi and Stevenson (2007) for learning constructions, which is itself an adaptation of a Bayesian model of human categorization proposed by Anderson (1991). It is important to note that the categories (i.e., constructions) are not predefined, but rather are created according to the patterns of similarity over observed frames.

```
Sense 1
dinner
       => meal, repast
           => nutriment, nourishment, nutrition, sustenance,
              aliment, alimentation, victuals
               => food, nutrient
                   => substance, matter
                       => entity
Sense 2
dinner, dinner party
       => party
           => social gathering, social affair
               => gathering, assemblage
                   => social group
                       => group, grouping
```

dinner: {meal, repast, nutriment, nourishment, nutrition, substance, aliment, alimentation, victuals, food, nutrient, substance, matter, entity, party, social gathering, social affair, gathering, assemblage, social group, group, grouping }

Figure 1: Semantic properties for *dinner* from WordNet

Grouping a frame F with other frames participating in construction k is formulated as finding the k with the maximum probability given F:

$$\mathbf{BestConstruction}(F) = \operatorname*{argmax}_k P(k|F) \quad (1)$$

where k ranges over the indices of all constructions, with index 0 representing recognition of a new construction.

Using Bayes rule, and dropping $P(F)$ which is constant for all k:

$$P(k|F) = \frac{P(k)P(F|k)}{P(F)} \propto P(k)P(F|k) \quad (2)$$

The prior probability, $P(k)$, indicates the degree of entrenchment of construction k, and is given by the relative frequency of its frames over all observed frames. The posterior probability of a frame F is expressed in terms of the individual probabilities of its features, which we assume are independent, thus yielding a simple product of feature probabilities:

[1]We do not remove alternate spellings of a term in WordNet; this will be seen in the profiles in the results section.

$$P(F|k) = \prod_{i \in FrameFeatures} P_i(j|k) \quad (3)$$

where j is the value of the i^{th} feature of F, and $P_i(j|k)$ is the probability of displaying value j on feature i within construction k. Given the focus here on semantic profiles, we next focus on the calculation of the probabilities of semantic properties.

3.3 Probabilities of Semantic Properties

The probability in equation (3) of value j for feature i in construction k is estimated using a smoothed version of this maximum likelihood formula:

$$P_i(j|k) = \frac{\mathbf{count}_i^k(j)}{n_k} \quad (4)$$

where n_k is the number of frames participating in construction k, and $\mathbf{count}_i^k(j)$ is the number of those with value j for feature i.

For most features, $\mathbf{count}_i^k(j)$ is calculated by simply counting those members of construction k whose value for feature i exactly matches j. However, for the semantic properties of words, counting only the number of exact matches between the sets is too strict, since even highly similar words very rarely have the exact same set of properties. We instead use the following Jaccard similarity score to measure the overlap between the set of semantic properties, S_F, of a particular argument in the frame to be clustered, and the set of semantic properties, S_k, of the same argument in a member frame of a construction:

$$\mathbf{sem\_score}(S_F, S_k) = \frac{|S_F \cap S_k|}{|S_F \cup S_k|} \quad (5)$$

For example, assume that the new frame F represents a usage of *John ate cake*. In the construction that we are considering for inclusion of F, one of the member frames represents a usage of *Mom got water*. We must compare the semantic properties of the corresponding arguments *cake* and *water*:

cake:	{baked goods,food,solid,substance,matter,entity}
water:	{liquid,fluid,food,nutrient,substance,matter,entity}

The intersection of the two sets is {food, substance, matter, entity}, yielding a $\mathbf{sem\_score}$ of $\frac{4}{9}$.

In general, to calculate the conditional probability for the set of semantic properties, we set $\mathbf{count}_i^k(j)$ in equation (4) to the sum of the $\mathbf{sem\_score}$'s for the new frame and every member of construction k, and normalize the resulting probability over all possible sets of semantic properties in our lexicon.

3.4 Predicting Semantic Profiles for Verbs

We represent the selectional preferences of a verb for an argument position as a semantic profile, which is a probability distribution over all the semantic properties. To predict the profile of a verb v for an argument position arg, we need to estimate the probability of each semantic property j separately:

$$\begin{aligned} P_{arg}(j|v) &= \sum_k P_{arg}(j, k|v) \quad (6) \\ &\propto \sum_k P(k, v) P_{arg}(j|k, v) \end{aligned}$$

Here, j ranges over all the possible semantic properties that an argument can have, and k ranges over all constructions. The prior probability of having verb v in construction k, or $P(k, v)$, takes into account two important factors: the relative entrenchment of the construction k, and the (smoothed) frequency with which v participates in k.

The posterior probability $P_{arg}(j|k, v)$ is calculated analogously to $P_i(j|k)$ in equation (4), but limiting the count of matching features to those frames in k that contain v:

$$P_{arg}(j|k, v) = \frac{\mathbf{verb\_count}_{arg}^k(j, v)}{n_{kv}} \quad (7)$$

where n_{kv} is the number of frames for v participating in construction k, and $\mathbf{verb\_count}_{arg}^k(j, v)$ is the number of those with semantic property j for argument arg. We use a smoothed version of the above formula, where the relative frequency of each property j among all nouns is used as the smoothing factor.

3.5 Verb-Argument Compatibility

In one of our experiments, we need to measure the compatibility of a particular noun n for an argument position arg of some verb v. That is, we need to estimate how much the semantic properties of n conform to the acquired semantic profile of v for arg. We formulate the compatibility as the conditional probability of observing n as an argument arg of v:

$$\mathbf{compatibility}(v, n) = log(P_{arg}(j_n|v)) \quad (8)$$

where j_n is the set of the semantic properties for word n, and $P_{arg}(j_n|v)$ is estimated as in equation (7). However, since j_n here is a set of properties (as opposed to j in equation (7) being a single property), $\mathbf{verb\_count}^k_{arg}$ in equation (7) should be modified as described in Section 3.3: we set $\mathbf{verb\_count}^k_{arg}(j_n, v)$ to the sum of the $\mathbf{sem\_score}$'s (equation (5)) for j_n and every frame of v that participates in construction k.

4 Experimental Results

In the following sections, we first describe the training data for our model. In accordance with other computational models, we focus here on the verb preferences for the direct object position.[2] Next, we provide a qualitative analysis of our model through examination of the semantic profiles for a number of verbs. We then evaluate our model through two tasks of simulating verb-argument plausibility judgment, and analyzing the implicit object alternation, following Resnik (1996).[3]

4.1 The Training Data

In earlier work (Alishahi and Stevenson, 2005, 2007), we used a method to automatically generate training data with the same distributional properties as the input children receive. However, this relies on manually-compiled data about verbs and their argument structure frames from the CHILDES database (MacWhinney, 1995). To evaluate the new version of our model for the task of learning selectional preferences, we need a wide selection of verbs and their arguments that is impractical to compile by hand.

The training data for our experiments here are generated as follows. We use 20,000 sentences randomly selected from the British National Corpus (BNC),[4] automatically parsed using the Collins parser (Collins, 1999), and further processed with TGrep2,[5] and an NP-head extraction software.[6] For each verb usage in a sentence, we construct a frame by recording the verb in root form, the number of the arguments for that verb, and the syntactic pattern of the verb usage (i.e., the word order of the verb and the arguments). We also record in the frame the semantic properties of the verb and each of the argument heads (each noun is also converted to root form); these properties are extracted from WordNet (as discussed in Section 3.1 and illustrated in Figure 1). This process results in 16,300 frames which serve as input data to our learning model.

4.2 Formation of Semantic Profiles for Verbs

After training our model on the above data, we use equation (7) to predict the semantic profile of the direct object position for a range of verbs. Some of these verbs, such as *write* and *sing*, have strong selectional preferences, whereas others, such as *want* and *put*, can take a wide range of nouns as direct object (as confirmed by Resnik's (1996) estimated strength of selectional preference for these verbs). The semantic profiles for *write* and *sing* are displayed in Figure 2, and the profiles for *want* and *put* are displayed in Figure 3. (Due to limited space, we only include the 25 properties that have the highest probability in each profile.)

Because we extract the semantic properties of words from WordNet, which has a hierarchical structure, the properties that come from nodes in the higher levels of the hierarchy (such as *entity* and *abstraction*) appear as the semantic property for a very large set of words, whereas the properties that come from the leaves in the hierarchy are specific to a small set of words. Therefore, the general properties are more likely to be associated with a higher probability in the semantic profiles for most verbs. In fact, a closer look at the semantic profiles for *want* and *put* reveals that the top portion of the semantic profile for these verbs consists solely of such general properties that are shared among a large group of words. However, this is not the case for the more restrictive verbs. The semantic profiles for *write* and *sing* show that the specific properties that these verbs demand from their direct object appear amongst the highest-ranked properties, even though only a small set of words share these properties (e.g., *content,*

[2]To our knowledge, the only work that considers selectional preferences of subjects and prepositional phrases as well as direct objects is Brockmann and Lapata (2003).

[3]Computational models of verb selectional preference have been evaluated through disambiguation tasks (Li and Abe, 1998; Abney and Light, 1999; Ciaramita and Johnson, 2000; Clark and Weir, 2002), but for to evaluate our cognitive model, the experiments from Resnik (1996) are the most interesting.

[4]http://www.natcorp.ox.ac.uk

[5]http://tedlab.mit.edu/~dr/Tgrep2

[6]The software was provided to us by Eric Joanis, and Afsaneh Fazly helped us in using the above-mentioned tools for generating our input corpora.

write	
(0.024)	abstraction
(0.022)	entity
(0.021)	location
(0.020)	substance
(0.019)	destination
(0.018)	relation
(0.015)	communication
(0.015)	social relation
(0.013)	content
(0.011)	message
(0.011)	subject matter
(0.011)	written communication
(0.011)	written language
(0.010)	object
(0.010)	physical object
(0.010)	writing
(0.010)	goal
(0.010)	unit
(0.009)	whole
(0.009)	whole thing
(0.009)	artifact
(0.009)	artefact
(0.009)	state
(0.009)	amount
(0.009)	measure

sing	
(0.020)	abstraction
(0.015)	relation
(0.015)	communication
(0.015)	social relation
(0.013)	act
(0.013)	human action
(0.013)	human activity
(0.013)	auditory communication
(0.012)	music
(0.010)	entity
(0.010)	piece
(0.009)	composition
(0.009)	musical composition
(0.009)	opus
(0.009)	piece of music
(0.009)	psychological feature
(0.008)	cognition
(0.008)	knowledge
(0.008)	noesis
(0.008)	activity
(0.008)	content
(0.008)	grouping
(0.008)	group
(0.008)	amount
(0.008)	measure

Figure 2: Semantic profiles of *write* and *sing* for the direct object position.

message, written communication, written language, ... for *write*, and *auditory communication, music, musical composition, opus, ...* for *sing*).

The examination of the semantic profiles for fairly frequent verbs in the training data shows that our model can use the verb usages to predict an appropriate semantic profile for each verb. When presented with a novel verb (for which no verb-based information is available), equation (7) predicts a semantic profile which reflects the relative frequencies of the semantic properties among all words (due to the smoothing factor added to equation (7)), modulated by the prior probability of each construction. The predicted profile is displayed in Figure 4. It shows similarities with the profiles for *want* and *put* in Figure 3, but the general properties in this profile have an even higher probability. Since the profile for the novel verb is predicted in the absence of any evidence (i.e., verb usage) in the training data, we later use it as the base for estimating other verbs' strength of selectional preference.

want	
(0.016)	entity
(0.015)	object
(0.015)	physical object
(0.014)	abstraction
(0.013)	act
(0.012)	human action
(0.012)	human activity
(0.012)	relation
(0.011)	unit
(0.011)	whole
(0.011)	whole thing
(0.011)	artifact
(0.011)	artefact
(0.008)	communication
(0.008)	social relation
(0.008)	activity
(0.007)	cause
(0.007)	state
(0.007)	instrumentality
(0.007)	instrumentation
(0.007)	event
(0.006)	being
(0.006)	living thing
(0.006)	animate thing
(0.006)	organism

put	
(0.015)	entity
(0.015)	object
(0.013)	physical object
(0.013)	abstraction
(0.011)	unit
(0.011)	whole
(0.011)	whole thing
(0.011)	artifact
(0.011)	artefact
(0.010)	act
(0.009)	relation
(0.008)	human action
(0.008)	human activity
(0.008)	communication
(0.008)	social relation
(0.007)	substance
(0.007)	content
(0.007)	instrumentality
(0.007)	instrumentation
(0.007)	measure
(0.006)	amount
(0.006)	quantity
(0.006)	cause
(0.006)	causal agent
(0.006)	causal agency

Figure 3: Semantic profiles of *want* and *put* for the direct object position.

4.3 Verb-Argument Plausibility Judgments

Holmes et al. (1989) evaluate verb argument plausibility by asking human subjects to rate sentences like *The mechanic warned the driver* and *The mechanic warned the engine*. Resnik (1996) used this data to assess the performance of his model by comparing its judgments of selectional fit against the plausibility ratings elicited from human subjects. He showed that his selectional association measure for a verb and its direct object can be used to select the more plausible verb-noun pair among the two (e.g., <warn,driver> vs. <warn,engine> in the previous example). That is, a higher selectional association between the verb and one of the nouns compared to the other noun indicates that the former is the more plausible pair. Resnik (1996) used the Brown corpus as training data, and showed that his model arrives at the correct ordering of more and less plausible arguments in 11 of the 16 cases.

We repeated this experiment, using the same 16 pairs of verb-noun combinations. For each pair of $<v, n_1>$ and $<v, n_2>$, we calculate the compatibility measure using equation (8); these values are shown in Figure 5. (Note that because these are

A novel verb	
(0.021)	entity
(0.017)	object
(0.017)	physical object
(0.015)	abstraction
(0.010)	act
(0.010)	human action
(0.010)	human activity
(0.010)	unit
(0.009)	whole
(0.009)	whole thing
(0.009)	artifact
(0.009)	artefact
(0.009)	being
(0.009)	living thing
(0.009)	animate thing
(0.009)	organism
(0.008)	cause
(0.008)	causal agent
(0.008)	causal agency
(0.008)	relation
(0.008)	person
(0.008)	individual
(0.008)	someone
(0.008)	somebody
(0.008)	mortal

Figure 4: Semantic profile of a novel verb for the direct object position.

Verb	Plausible		Implausible	
see	friend	-30.50	method	-32.14
read	article	-32.76	fashion	-33.33
find	label	-32.05	fever	-33.30
hear	story	-32.11	issue	-32.40
write	letter	-31.37	market	-32.46
urge	daughter	-36.73	contrast	-35.64
warn	driver	-33.68	engine	-34.42
judge	contest	-39.05	climate	-38.23
teach	language	-45.64	distance	-45.11
show	sample	-31.75	travel	-31.42
expect	visit	-33.88	mouth	-32.87
answer	request	-31.89	tragedy	-33.95
recognize	author	-32.53	pocket	-32.62
repeat	comment	-33.80	journal	-33.97
understand	concept	-32.25	session	-32.93
remember	reply	-33.79	smoke	-34.29

Figure 5: Compatibility scores for plausible vs. implausible verb-noun pairs.

log-probabilities and therefore negative numbers, a lower absolute value of $\mathbf{compatibility}(v, n)$ shows a better compatibility between the verb v and the argument n.) For example, <see,friend> has a higher compatibility score (-30.50) than <see,method> (-32.14). Similar to Resnik, our model detects 11 plausible pairs out of 16. However, these results are reached with a much smaller training corpus (around 500,000 words), compared to the Brown corpus used by Resnik (1996) which contains one million words. Moreover, whereas the Brown corpus is tagged and parsed manually, the portion of the BNC that we use is parsed automatically, and as a result our training data is very noisy. Nonetheless, the model achieves the same level of accuracy in distinguishing plausible verb-argument pairs from implausible ones.

4.4 Implicit Object Alternations

In English, some inherently transitive verbs can appear with or without their direct objects (e.g., *John ate his dinner* as well as *John ate*), but others cannot (e.g., *Mary made a cake* but not **Mary made*). It is argued that implicit object alternations involve a particular relationship between the verb and its argument. In particular, for verbs that participate in the implicit object alternation, the omitted object must be in some sense inferable or *typical* for that verb (Levin, 1993, among others).

Resnik (1996) used his model of selectional preferences to analyze implicit object alternations, and showed a relationship between his measure of selectional preference strength and the notion of typicality of an object. He calculated this measure for two groups of Alternating and Non-alternating verbs, and showed that, on average, the Alternating verbs have a higher strength of selectional preference for the direct object than the Non-alternating verbs. However, there was no threshold separating the two groups of verbs.

To repeat Resnik's experiment, we need a measure of how "strongly constraining" a semantic profile is. We can do this by measuring the similarity between the semantic profile we generate for the object of a particular verb and some "default" notion of the argument for that position across all verbs. We use the semantic profile predicted for the object position of a novel verb, shown earlier in Figure 4, as the default profile for that argument position. Because this profile is predicted in the absence of any evidence in the training data, it makes the minimum assumptions about the properties of the argument and thus serves as a suitable default. We then assume that verbs with weaker selectional preferences have semantic profiles more similar to the default profile

Alternating verbs		Non-alternating verbs	
write	0.61	*hang*	0.56
sing	0.67	*wear*	0.71
drink	0.67	*say*	0.75
eat	0.74	*catch*	0.76
play	0.74	*show*	0.77
pour	0.76	*make*	0.78
watch	0.77	*hit*	0.78
pack	0.78	*open*	0.81
steal	0.80	*take*	0.83
push	0.80	*see*	0.87
call	0.80	*like*	0.87
pull	0.80	*get*	0.87
explain	0.81	*find*	0.87
read	0.82	*give*	0.88
hear	0.87	*bring*	0.89
		want	0.89
		put	0.90
Mean:	0.76	Mean:	0.81

Figure 6: Similarity with the base profile for Alternating and Non-alternating verbs.

than verbs with stronger preferences. We use the cosine measure to estimate the similarity between two profiles p and q:

$$\text{cosine}(p, q) = \frac{p \times q}{||p|| \times ||q||} \tag{9}$$

The similarity values for the Alternating and Non-alternating verbs are shown in Figure 6. The larger values represent more similarity with the base profile, which means a weaker selectional preference. The means for the Alternating and Non-alternating verbs were respectively 0.76 and 0.81, which confirm the hypothesis that verbs participating in implicit object alternations select more strongly for the direct objects than verbs that do not. However, like Resnik (1996), we find that it is not possible to set a threshold that will distinguish the two sets of verbs.

5 Conclusions

We have proposed a cognitively plausible model for learning selectional preferences from instances of verb usage. The model represents verb selectional preferences as a semantic profile, which is a probability distribution over the semantic properties that an argument can take. One of the strengths of our model is the incremental nature of its learning mechanism, in contrast to other approaches which learn selectional preferences in batch mode. Here we have only reported the results for the final stage of learning, but the model allows us to monitor the semantic profiles during the course of learning, and compare it with child data for different age groups, as we do with semantic roles (Alishahi and Stevenson, 2007). We have shown that the model can predict appropriate semantic profiles for a variety of verbs, and use these profiles to simulate human judgments of verb-argument plausibility, using a small and highly noisy set of training data. The model can also use the profiles to measure verb-argument compatibility, which was used in analyzing the implicit object alternation.

References

Abney, S. and Light, M. (1999). Hiding a semantic hierarchy in a Markov model. In *Proc. of the ACL Workshop on Unsupervised Learning in Natural Language Processing*.

Alishahi, A. and Stevenson, S. (2005). A probabilistic model of early argument structure acquisition. In *Proc. of the CogSci 2005*.

Alishahi, A. and Stevenson, S. (2007). A computational usage-based model for learning general properties of semantic roles. In *Proc. of the EuroCogSci 2007*.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98(3):409–429.

Brockmann, C. and Lapata, M. (2003). Evaluating and combining approaches to selectional preference acquisition. In *Proc. of the EACL 2003*.

Ciaramita, M. and Johnson, M. (2000). Explaining away ambiguity: Learning verb selectional preference with Bayesian networks. In *Proc. of the COLING 2000*.

Clark, S. and Weir, D. (2002). Class-based probability estimation using a semantic hierarchy. *Computational Linguistics*, 28(2):187–206.

Collins, M. (1999). *Head-Driven Statistical Models for Natural Language Parsing*. PhD thesis, University of Pennsylvania.

Holmes, V. M., Stowe, L., and Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language*, 28:668–689.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. The University of Chicago Press.

Li, H. and Abe, N. (1998). Generalizing case frames using a thesaurus and the MDL principle. *Computational Linguistics*, 24(2):217–244.

Light, M. and Greiff, W. (2002). Statistical models for the induction and use of selectional preferences. *Cognitive Science*, 26(3):269–281.

MacWhinney, B. (1995). *The CHILDES project: Tools for analyzing talk*. Lawrence Erlbaum.

Miller, G. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography*, 17(3).

Nation, K., Marshall, C. M., and Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *J. of Experimental Child Psych.*, 86:314–329.

Resnik, P. (1996). Selectional constraints: An information-theoretic model and its computational realization. *Cognition*, 61:127–199.