Gradual Acquisition of Mental State Meaning: A Computational Investigation

Libby Barak, **Afsaneh Fazly**, and **Suzanne Stevenson**
Department of Computer Science
University of Toronto
{libbyb,afsaneh,suzanne}@cs.toronto.edu

Abstract

The acquisition of Mental State Verbs (MSVs) has been extensively studied in respect to their common occurrence with sentential complement syntax. However, MSVs also occur in a variety of other syntactic structures. Moreover, other verb classes frequently occur with sentential complements, e.g., Communication and Perception verbs. The similarity in distribution of the various verb classes over syntactic patterns may affect the acquisition of the meaning of MSVs by association. In this study we present a novel computational model to learn verb classes, which allows us to analyze the association of mental verbs to their meaning over a variety of syntactic patterns. Our results point to an important role of the full syntactic preferences of MSVs on top of their occurrences with sentential complements.

Introduction

Mental State Verbs (MSVs), such as *think* and *want*, are initially produced by children much later than verbs from other semantic classes, such as Action (e.g., *throw*), Perception (e.g., *see*), and Communication (e.g., *say*) (Shatz et al., 1983). Even within the class of MSVs different verbs are acquired at different stages. For example, children produce Desire verbs, such as *want* and *wish*, earlier than Belief verbs, such as *think* and *know* (Shatz et al., 1983; Bartsch and Wellman, 1995; Asplin, 2002; de Villiers, 2005; Papafragou et al., 2007).

One group of psycholinguistic studies attributes the observed delay in the acquisition of MSVs in general, and the developmental lag of Belief verbs in particular, to the syntactic requirements of these verbs (de Villiers and Pyers, 2002; de Villiers, 2005; Papafragou et al., 2007). These studies provide evidence from the tendency of MSVs to appear with the Sentential Complement (SC) construction, and argue that since these complex patterns are difficult for children to master (because of the embedded clause), they cause a delay in the acquisition of MSVs. In addition, the later production of Belief verbs compared to Desire verbs is suggested to be due to their association with two different kinds of SC syntax: Desire verbs occur mostly with an infinitival SC (as in *I want (her) to leave*), while Belief verbs occur mostly with a finite SC (a full tensed embedded clause, as in *I think (that) she left*). Notably, infinitivals appear earlier than finite SCs in the speech of young children (Bloom et al., 1984, 1989), and hence are assumed to be easier to acquire.

In contrast, another group of studies argues that the complex syntax of SCs cannot be the sole explanatory factor for the observed delays in the acquisition of MSVs. These studies provide evidence for their argument from several observations: First, children use finite SCs with verbs of Communication (e.g., *say*) and Perception (e.g., *see*) long before they use

them with Belief verbs (Bartsch and Wellman, 1995). The relatively high frequency of these non-MSVs with the SC syntax may thus be a factor that affects the acquisition of MSVs. For instance, de Villiers and Pyers (2002) and Israel (2008) suggest that children first learn to use the complex SC syntax with conceptually simpler verbs that share aspects of the MSV semantics, e.g., Communication and Perception verbs. This stage can be used to break into the full mental meaning of MSVs using the acquired shared meaning with Communication and Perception verbs. Secondly, there is evidence that except for very high-frequency MSVs (such as *think*) many of these verbs frequently appear in constructions other than the SC (Diessel and Tomasello, 2001; Klainerman, 2010). Thus, the non-SC usages of MSVs might also play a special role in their pattern of acquisition.

Together, the above studies suggest that, in order to capture the full developmental trajectory of MSVs, we need to look at the interaction of the following factors: (i) The overall syntactic behaviour of MSVs, including their appearance with the SC and non-SC syntactic patterns; (ii) The syntactic behaviour of other non-MSV verbs, especially Communication and Perception verbs that have a high-frequency of occurrence with the SC syntax; and (iii) The shared semantic properties of MSVs with other semantic classes of verbs.

We examine the interaction of these factors by using a computational model that simultaneously learns argument structure constructions, as well as semantic verb classes (Barak et al., 2014). Specifically, we develop a novel incremental model that forms verb classes by drawing on the similarity of their distribution across these learned constructions. We learn these constructions using the model of Barak et al. (2013) to group usages of verbs into clusters based on the similarity of their linguistic properties. Each learned construction represents an abstraction over similar verb usages, while each verb class captures a higher-level of syntactic and semantic similarity among verb types. The verb classes enable us to expand the analysis in (Barak et al., 2013) that focused on the role of SC syntax in the acquisition of MSVs from the learned constructions. These incrementally-learned verb classes thus allow us to examine the developing interaction of the above three factors, as well as their role in the acquisition of MSVs in their full mental meaning.

Our results suggest how the properties of child-directed speech (CDS) may be guiding the acquisition process of MSVs in the interaction over a variety of verb classes. Moreover, our results point to the role of non-SC syntax in replicating the gradual association of MSVs to SC syntax and mental meaning.
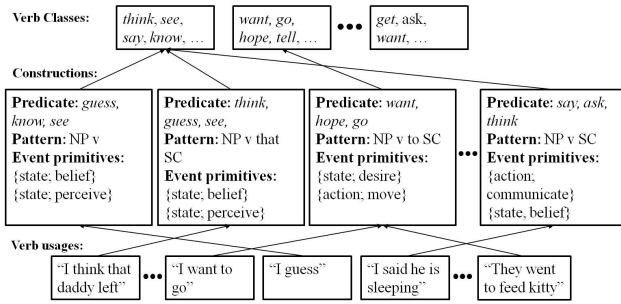
Figure 1: A visual representation of the two layers in the model. The sample constructions present a small subset of the features and semantic properties used in our model. A full frame includes the *predicate* syntactic features (*syntactic pattern*, *argument count*, *preposition*, and *complementizer*) and semantic features (*event primitives* and *event participants*).

The Computational Model

The model we develop here is a significant extension to an existing incremental Bayesian model of verb argument structure acquisition (Alishahi and Stevenson, 2008; Barak et al., 2012). Specifically, our model learns clusters of similar verb usages (*constructions*), as in the model of Barak et al. However, our model adds another layer of abstraction that learns groupings of verbs that exhibit similar distributional patterns of occurrence across the learned constructions. To distinguish between the clusters of the two layers of our new model, we refer to the clusters of verb usages in the first layer as *constructions*, and to the groupings of verbs in the second layer as *verb classes* (see Figure 1).

Overview of the Model

The model learns from a sequence of *frames*, where each frame is a collection of syntactic and semantic *features* representing what the learner might extract from a learning situation. The model incrementally clusters the input frames into *constructions* that reflect probabilistic associations of the syntactic and semantic features across similar verb usages. Each learned cluster is a probabilistic (and possibly noisy) representation of an argument structure construction: e.g., a cluster containing frames corresponding to usages such as *I eat apples*, *She took the ball*, and *He got a book*, etc., represents a Transitive Action construction.[1] Such grouping of similar verb usages allows for some degree of generalization over the observed input, e.g., when encountering a novel verb in a transitive utterance, the model predicts that this verb shares semantic properties with other Action verbs appearing in a similar pattern. Nonetheless, such groupings do not capture the full range of syntactic and semantic behaviour of MSVs and other verbs, which can provide insights into what might be responsible for the observed developmental patterns of MSVs. To capture this, our model learns groupings of verbs that have similar distributions across its learned constructions, forming semantic verb classes that provide a

---

[1]Note that, because the associations are probabilistic, a linguistic construction may be represented by more than one cluster.

higher-level of abstraction over the input.

Figure 1 presents a visualization of the layers of the model. In this example, usages of *think* might be clustered into one construction with *see*, but into another construction with *say*. While each construction represents a different association of semantic and syntactic properties given the usages, the verb classes group all *think*, *see*, and *say* together given their co-occurrence over several constructions. The grouping of these verbs into one verb class associates their occurrence with various finite-SC patterns, i.e., w/o *that*, which is not captured in the constructions. Importantly, the model incrementally and simultaneously learns both of these types of knowledge, allowing us to examine the developmental trajectory of the acquisition of MSVs. In the rest of this section, we first explain how our model learns constructions of verb usages, and then describe how it forms its knowledge of verb classes by drawing on its own learned knowledge of constructions.

The model also includes a component that simulates the difficulty of children in attending to the mental content, e.g., "believing", in a scene that also includes an easier-to-observe physical action, e.g., "walking" (Papafragou et al., 2007). The model simulates this developing attention to mental content as an increasing ability to correctly interpret a scene paired with an SC utterance as having mental semantic properties. At the same time, we assume the child records the observed syntactic properties but erroneously coupling them to the mis-interpreted physical semantic content. For instance, young children may focus on the "making" action in *He thinks Mom made pancakes*, rather than on the "thinking", but record the use of Sentential Complement (SC) pattern (see (Barak et al., 2012) for more details).

Learning Constructions of Verb Usages

The model processes input frames one at a time in a sequence and groups them into clusters on the basis of the overall similarity in the values of their features. Importantly, the model learns these clusters incrementally, and the number and type of clusters is not predetermined. The model considers the creation of a new cluster for a given frame if the frame is not sufficiently similar to any of the existing clusters. Formally, the model finds the best cluster for a given frame F as in:

$$\text{BestCluster}(F) = \underset{k \in Clusters}{\text{argmax}} \, P(k|F) \quad (1)$$

where k ranges over all existing clusters and a new one. Using Bayes rule:

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

The prior probability of a cluster $P(k)$ is estimated as the proportion of frames that are in k out of all observed input frames, thus assigning a higher prior to larger clusters representing more frequent constructions. The likelihood $P(F|k)$ is estimated based on the match of feature values in F and in the frames of k (assuming independence of the features):

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

where i refers to the ith feature of F and j refers to its value, and $P_i(j|k)$ is calculated using a smoothed version of:

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

where $\text{count}_i(j,k)$ is the number of times feature i has the value j in cluster k, and n_k is the number of frames in k.

Learning Verb Classes

In addition to grouping verb usages into constructions, our new model groups verbs into classes on the basis of their distribution across these learned constructions. The algorithm for learning the verb classes is similar to that for learning the constructions, with the exception that the only feature used here is the distribution of the target verb (to be classified) across the learned constructions. Similarly to learning the constructions, the model learns the verb classes incrementally and the number and type of classes is not predetermined. The model considers the creation of a new class for a given verb distribution if the distribution is not sufficiently similar to any of those represented by the existing verb classes.

Formally, after each clustering decision to a construction, the model extracts the current distribution d_v of a verb v over the learned constructions as a smoothed version of the verb's probability of occurrence in each of these clusters:

$$d_v : P(k|v) = \frac{\text{count}(v,k)}{n_v} \quad (5)$$

where $\text{count}(v,k)$ is the number of times the verb v appears in construction k, and n_v is the number of times v has occurred in the input thus far. Each verb class c is represented by a distribution d_c that is a weighted average of the distributions of its members:

$$d_c = \frac{1}{|c|} \sum_{v \in c} \text{count}(v,c) \times d_v \quad (6)$$

where $|c|$ is the size of class c, $\text{count}(v,c)$ is the number of occurrences of v that have been assigned to c, and d_v is the distribution of the verb v (for the instances assigned to the class c).

The model finds the best class for a given verb distribution d_v based on its similarity to the distributions of all existing classes and a new one, as in:

$$\text{BestClass}(d_v) = \underset{c \in Classes}{\text{argmax}} \left(1 - D_{\text{JS}}(d_c \| d_v)\right) \quad (7)$$

where c ranges over all existing classes as well as a new class that is represented as a uniform distribution over the existing constructions. Jensen–Shannon divergence, D_{JS}, is a popular method for measuring the distance between two distributions: It is based on the Kullback–Leibler divergence, but it is symmetric and has a finite value between 0 and 1:

$$D_{\text{JS}}(p\|q) = \frac{1}{2}D_{\text{KL}}(p\|\frac{1}{2}(p+q)) + \frac{1}{2}D_{\text{KL}}(q\|\frac{1}{2}(p+q)) \quad (8)$$

Experimental Setup

Generation of the Input Corpora

We follow the input generation methodology of Alishahi and Stevenson (2008) and Barak et al. (2013) to create naturalistic input that is based on the distributional properties of verbs across a range of syntactic constructions, as observed in CDS. For this, we use an *input-generation lexicon* that contains information about the distributional properties of 71 verbs, covering verbs from different semantic classes and different frequency ranges. Each lexical entry includes the overall frequency of each verb, and its relative frequency with each of a number of observed syntactic constructions. The frequencies are extracted from a manual annotation of a sample of 100 child-directed utterances per verb (or all utterances if less than 100) from a collection of eight corpora from CHILDES (MacWhinney, 2000).[2]

A single input corpus is generated by iteratively selecting a random verb and a one of its argument constructions based on their frequencies according to the lexicon. Since the generation of the input is probabilistic, we conduct 100 simulations for our experiments (each simulation using a different input corpus) to avoid any dependency on specific idiosyncratic properties of a single generated corpus. Importantly, each of the input corpora has the distributional properties observed in CDS, but at the same time the order of presentation of verb usages may vary across the corpora.

Set-up of Simulations

Our goal here is to evaluate how our model interprets a typical usage of an MSV, resembling the *verb identification task* used in many psycholinguistic studies of MSV acquisition (e.g., Asplin, 2002; Papafragou et al., 2007). In this task, participants are asked to predict the meaning of a novel verb (e.g., *gorp*) in a given utterance (e.g., "she gorps that daddy is sleeping") paired with an image/video depicting a mental event. We simulate this task to examine the developmental trajectory of MSV acquisition in our model.

Specifically, we train our model on a randomly generated input corpus of $10,000$ input frames, performing periodic tests: At each test point, we present the model with a *test frame* that represents the psycholinguistic settings of the verb identification task. We then use the model to estimate the likelihood of the *event type* corresponding to the 5 semantic verb classes of Belief, Desire, Perception, Communication, and Action. These likelihoods represent our model's *interpretation* of a test scenario: E.g., the likelihood of the event type Perception given a Belief-fin test frame reflects how likely it is that our model interprets the novel verb in the frame as having a Perception semantics; see the following section for how we estimate the event type likelihoods. We use two kinds of test frames to represent typical usages of MSVs: (i) *Desire-inf* containing a novel verb, the infinitival-SC syntactic properties, and the semantic properties of a randomly-chosen De-

sire verb from our lexicon; and (ii) *Belief-fin* containing a novel verb, the finite-SC syntactic properties, and the semantic properties of a randomly-chosen Belief verb from our lexicon.

Estimating Event Type Likelihoods

Recall that each verb entry in our lexicon is represented as a collection of features, including a set of event primitives — e.g., the set associated with the Belief verb *think* is {*state, cogitate, belief, communicate*}. We estimate each event type likelihood (e.g., Belief likelihood) by averaging over the likelihoods of all *event primitive sets* corresponding to verbs of that class (e.g., all Belief verbs) according to our lexicon.

Formally, we calculate the likelihood of each event primitive set $\mathcal{S}$ given a test frame F_{test}, as in:

$$P(\mathcal{S}|F_{\text{test}}) = \sum_{k \in Clusters} P_{\text{event primitives}}(\mathcal{S}|k)P(k|F_{\text{test}}) \quad (9)$$

where $P(\mathcal{S}|k)$ is the probability of the primitive set $\mathcal{S}$ given construction k, calculated as in Eqn. (4); and $P(k|F_{\text{test}})$ is the probability of assigning the test frame F_{test} to construction k. Note that only the constructions encode the individual semantic and syntactic features (including the event primitives). Hence we need to rely on the model's learned constructions to estimate $P(\mathcal{S}|k)$. However, we can use two ways of estimating $P(k|F_{\text{test}})$: (i) by drawing on the model's learned knowledge as reflected in the constructions; and (ii) by drawing on the model's learned knowledge of verb classes. This would help us understand the role of the model's learned verb classes in the acquisition of MSVs.

For the constructions, we simply calculate $P_{\text{L1}}(k|F_{\text{test}})$ as in Eqn. (2). For the classes, we calculate $P_{\text{L2}}(k|F_{\text{test}})$ as in:

$$P_{\text{L2}}(k|F_{\text{test}}) \approx \sum_{c \in Classes} P(k|c)P(c|F_{\text{test}}) \quad (10)$$

where $P(k|c)$ is the probability of construction k given the distribution d_c of class c over the learned constructions (see Eqn. 6 for how d_c is calculated). To estimate $P(c|F_{\text{test}})$, we compare d_c with $P_{\text{L1}}(k|F_{\text{test}})$ where the latter reflects the probability distribution of F_{test} across the constructions. We use Jensen–Shannon divergence (as in Eqn. (8)) to measure the similarity between the two distributions.

Experimental Results

Psycholinguistic studies have observed that children produce Desire verbs before Belief verbs (Shatz et al., 1983; Bartsch and Wellman, 1995). However, even for Desire verbs, there is still an initial stage when they are produced mostly in non-mental meaning (Bartsch and Wellman, 1995). Most psycholinguistic studies on MSVs, however, have focused on their occurrences with the complex sentential complement (SC) syntax. However, it has been noted that many usages of these verbs are with non-SC patterns, signifying the importance of looking at their full range of syntactic behaviour
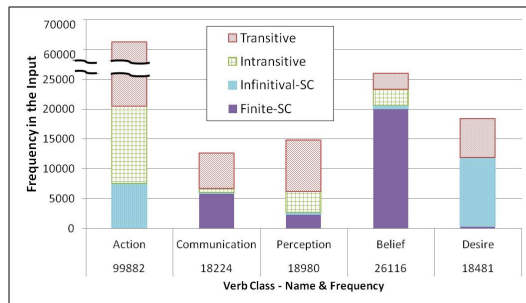

Figure 2: The distribution of our 5 verb classes over the 4 most frequent syntactic constructions for the MSVs. Note that the frequency of Action verbs with the transitive construction is higher than its appearance from the chart. The total frequency of each class, shown on the chart, includes the occurrences of the class with all the syntactic construction in the data.

(Diessel and Tomasello, 2001; Klainerman, 2010). The interaction of Desire and Belief verbs with other (conceptually-simpler) verbs over their shared syntactic and semantic properties might play a role in the acquisition of MSVs (de Villiers and Pyers, 2002; Israel, 2008). We use our model to shed light on the factors that might be responsible for the observed developmental patterns of Desire and Belief verbs.

We test our model's knowledge of MSVs (Belief and Desire verbs) by examining the event type likelihoods (that we estimate as explained above). We say that the model has acquired a solid knowledge of Belief (Desire) verbs if it assigns the highest likelihood to the Belief (Desire) event type when presented with a Belief-fin (Desire-inf) test frame. We also examine the event type likelihoods that our model assigns to the other non-MSV classes (i.e., Perception, Communication, and Action), given each of our two types of test frames. This will help us understand the role of non-MSV verb classes in the acquisition of MSVs. In addition, recall that we include both semantic and syntactic properties in our test frames. Following the settings of the psycholinguistic task that we simulate here, we include the typical syntactic pattern used with MSVs (i.e., the SC syntax) to make the model rely on constructions associated with this pattern to predict the event type of the novel verb. Note that because we include the semantic properties of Belief or Desire in the test frames, it also has the effect of looking into constructions that reflect mental semantics even when associated with non-SC syntactic patterns. This way, we can study the role of both the SC and the non-SC syntax in the acquisition of MSVs.

To evaluate the role of verb classes in the acquisition of MSVs, we compare the developmental patterns in our model arising from each of the two layers. In one case, we estimate event likelihoods given only the knowledge of constructions; in the other, we estimate these likelihoods by using the knowledge of verb classes.

Analysis based on the Learned Constructions

Figure 3(a) presents the likelihood of each event type given the Desire-inf test frame, while Figure 3(b) presents the likelihoods given the Belief-fin test frame. As can be seen, our
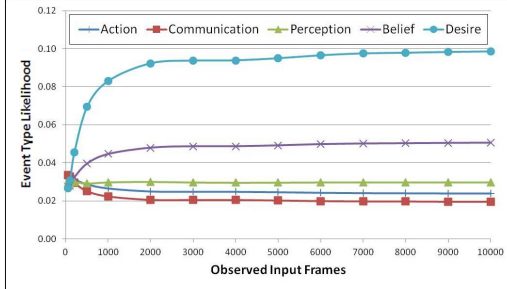
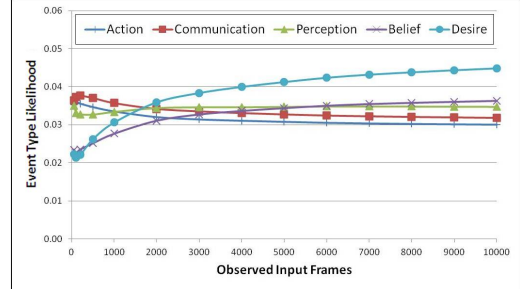(a) Model's likelihoods Desire-inf given test frame
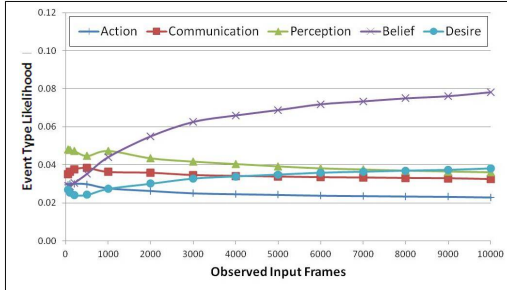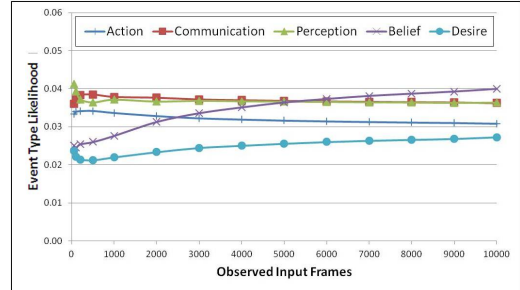

(b) Model's likelihoods given Belief-fin test frame

Figure 3: The likelihoods of each event type given the learned constructions.

(a) Model's likelihoods given Desire-inf test frame

(b) Model's likelihoods given Belief-fin test frame

Figure 4: The likelihoods of each event type given the learned verb classes.

model acquires Desire verbs almost instantly (from very early stages of training), but exhibits a delay in its acquisition of Belief verbs. This pattern is similar to what has been observed in children in that Desires are acquired earlier than Belief verbs; however, it is lacking the observed initial stage of not producing Desire verbs in their mental meanings. Interestingly, the earlier acquisition of Desire verbs can be attributed to its higher frequency (compared to Belief verbs) of appearing with non-SC syntax, e.g., *I want a cookie*, *I like apples*. The different syntactic distributions of Desire and Belief verbs can be seen in Figure 2. Recall that our model incorporates an attentional mechanism whereby when it encounters an SC utterance during the initial stages, it has some difficulty in encoding the mental event due to a competition arising from the action within the SC. Since Belief verbs more frequently appear with the SC syntax, they are more likely to be mis-interpreted at early stages, giving rise to a delay in their acquisition. In contrast, the non-SC usages of Desire verbs are correctly recorded even at the early learning stages. These results point to the importance of looking at both SC and non-SC usages of MSVs.

Looking more closely at Figure 3(b), we can see that early on our model interprets a Belief test frame mostly as having a Perception meaning, and only sometimes as having a Communication meaning. Interpretation of Beliefs as Perception or Communication initially is consistent with the hypotheses of de Villiers and Pyers (2002) and Israel (2008). However, if we look at the distributions of these three verb classes (Belief, Perception, and Communication) in our data (Figure 2), we cannot explain why a Perception interpretation is more

likely than Communication: Compared to Perception verbs, Communication verbs seem to have a distribution closer to that of Belief verbs. Clearly, the constructions do not fully capture the interaction among the different verb classes. In addition, we saw that our model did not show a similar behaviour to children in that it learned Desire verbs too quickly. We attribute this limitation to the fact that the constructions do not capture the interaction between Desire verbs and the other semantic classes (e.g., Action) that could only be captured through generalizations over the full range of syntactic behaviour of all verbs. We now turn into the same analysis using the model's learned verb classes.

Analysis based on the Learned Verb Classes

Figure 4(a) presents the event type likelihoods given the Desire-inf test frame, according to the verb classes.[3] Unlike given the constructions, here we observe a delay in the association of Desire verbs to their mental meaning, as observed in children (Bartsch and Wellman, 1995). The replication of this trend is enabled by capturing the association of Desire verbs to Communication and Action verbs over the use of transitive constructions as well as infinitival-SC: The similarity of the overall syntactic distribution of Desire verbs to Communication and Action verbs can be seen in Figure 2.

Figure 4(b) presents the event type likelihoods for the Belief-fin test frame, which show a delayed acquisition com-

---

[3]Note that the class-based likelihoods are shown on a different scale than those of based on the constructions. We have fewer verb class clusters, and thus we cannot directly compare the likelihood values. Instead we focus on the trends in the relative likelihoods across the 5 event types.

pared to Desire verbs (cf. Figure 4(a)). We replicate this trend when using verb classes, because our incrementally-learned verb classes actually capture the higher similarity of the distribution of Belief verbs across the syntactic patterns to other verb classes, compared with the relatively distinctive distribution of Desire verbs (as can be seen in Figure 2). Moreover, in contrast to the pattern presented from the constructions, initially the model interprets the Belief-fin test frame as either Perception or Communication (with similar likelihoods), which is more in line with what has been suggested in the psycholinguistic literature (de Villiers and Pyers, 2002). Our model's verb classes capture the similar distribution over syntactic properties of Belief and Communication verbs, unlike the results presented for the constructions.

Discussion

We use a computational model of learning verb classes to show the role of a variety of syntactic constructions in learning the meaning of MSVs. While Barak et al. (2013) focused on the role of one syntactic construction, i.e., SCs, our results here point to the importance of looking at the distribution of MSVs over the full range of syntactic constructions. Our results present an initial high likelihood of interpreting usages of MSVs as having non-mental meaning based on the interaction of MSVs with other verb classes based on their syntactic distribution. This can serve as an additional research direction while psycholinguistic studies mostly focused on the cognitive and pragmatic properties of MSVs as a possible cause for the initial production of MSVs in non-mental meaning.

The focus of this work is on the role of the distribution of MSVs over SC and non-SC patterns in their acquisition. We hope to expand this analysis to additional languages that differ in their distributional properties over the syntactic patterns in the future. Moreover, in a preliminary analysis of the formed constructions and verb classes we note that the similarity in semantic properties across verb classes should also play a role in the learning process of MSVs. Notably, we would like to address the possible use of MSVs in nonmental meaning in CDS (Bartsch and Wellman, 1995; Diessel and Tomasello, 2001). We hope to evaluate the role of semantic properties of MSV usages in future work, while carefully assessing the semantic properties of such usages in CDS over time, including possible parental usages of MSVs in non-mental usages in the input children observe.

References

Afra Alishahi and Suzanne Stevenson. 2008. A computational model of early argument structure acquisition. *Cognitive Science*, 32(5):789–834.

Kristen N. Asplin. 2002. *Can complement frames help children learn the meaning of abstract verbs?* Ph.D. thesis, UMass Amherst.

Libby Barak, Afsaneh Fazly, and Suzanne Stevenson. 2012. Modeling the acquisition of mental state verbs.

Libby Barak, Afsaneh Fazly, and Suzanne Stevenson. 2013. Acquisition of desires before beliefs: A computational investigation. *CoNLL-2013*, page 231.

Libby Barak, Afsaneh Fazly, and Suzanne Stevenson. 2014. Learning verb classes in an incremental model.

Karen Bartsch and Henry M. Wellman. 1995. *Children talk about the mind.* New York: Oxford Univ. Press.

Lois Bloom, Lois Hood, and Patsy Lightbown. 1974. Imitation in language development: If, when, and why. *Cognitive Psychology*, 6(3):380–420.

Lois Bloom, Matthew Rispoli, Barbara Gartner, and Jeremie Hafitz. 1989. Acquisition of complementation. *Journal of Child Language*, 16(01):101–120.

Lois Bloom, Jo Tackeff, and Margaret Lahey. 1984. Learning *to* in complement constructions. *Journal of Child Language*, 11(02):391–406.

Roger Brown. 1973. *A first language: The early stages.* Harvard Univ. Press.

Jill G. de Villiers. 2005. Can language acquisition give children a point of view. In *Why Language Matters for Theory of Mind*, pages 199–232. Oxford Univ. Press.

Jill G de Villiers and Jennie E Pyers. 2002. Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development*, 17(1):1037–1060.

Holger Diessel and Michael Tomasello. 2001. The acquisition of finite complement clauses in english: A corpus-based analysis. *Cognitive Linguistics*, 12(2):97–142.

Michael Israel. 2008. Mental spaces and mental verbs in early child English. *Language in the Context of Use: Discourse and Cognitive Approaches to Language*, 37:199.

Lara D Klainerman. 2010. Syntactic collocations patterns and mental state verb acquisition.

A. Kuczaj, Stan. 1977. The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, 16(5):589–600.

Elena Lieven, Dorothé Salomo, and Michael Tomasello. 2009. Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, 20(3):481–507.

B. MacWhinney. 2000. *The CHILDES project: Tools for analyzing talk*, volume 2. Psychology Press.

Anna Papafragou, Kimberly Cassidy, and Lila Gleitman. 2007. When we think about thinking: The acquisition of belief verbs. *Cognition*, 105(1):125–165.

Jacqueline Sachs. 1983. Talking about the There and Then: The emergence of displaced reference in parent–child discourse. *Children's language*, 4.

Marilyn Shatz, Henry M. Wellman, and Sharon Silber. 1983. The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition*, 14(3):301–321.

Patrick Suppes. 1974. The semantics of children's language. *American psychologist*, 29(2):103.