Modeling the Acquisition of Mental State Verbs

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

Abstract

Children acquire mental state verbs (MSVs) much later than other, lower-frequency, words. One factor proposed to contribute to this delay is that children must learn various semantic and syntactic cues that draw attention to the difficult-to-observe mental content of a scene. We develop a novel computational approach that enables us to explore the role of such cues, and show that our model can replicate aspects of the developmental trajectory of MSV acquisition.

1 Introduction

Mental State Verbs (MSVs), such as *think*, *know*, and *want*, are very frequent in child-directed language, yet children use them productively much later than lower-frequency action verbs, such as *fall* and *throw* (Johnson and Wellman, 1980; Shatz et al., 1983). Psycholinguistic theories have suggested that there is a delay in the acquisition of MSVs because they require certain cognitive and/or linguistic skills that are not available during the early stages of language development. For example, MSVs typically occur with a sentential complement (SC) that refers to the propositional content of the mental state, as in *He thinks Mom went home*. Children have to reach a stage of syntactic development that includes some facility with SCs in order to fully acquire MSVs. However, even at 3–5 years old, children are able to process SCs only imperfectly (e.g., Asplin, 2002).

Even when children are able to produce SCs with other verbs (such as verbs of communication, as in *He said Mom went home*), there is a lag before they productively use MSVs referring to actual mental content (Diessel and Tomasello, 2001).[1] Psycholinguists have suggested that young children lack the conceptual ability to conceive that others have mental states separate from their own (Bartsch and Wellman, 1995; Gopnik and Meltzoff, 1997), further delaying the acquisition of MSVs.

Another factor suggested to contribute to the difficulty of acquiring MSVs is their *informational requirements* (Gleitman et al., 2005; Papafragou et al., 2007). Children learn word meanings by figuring out which aspects of an observed scene are referred to by a particular word (Quine, 1960). MSVs often refer to aspects of the world that are not directly observable (i.e., the beliefs and desires of another entity). Thus, in addition to the above-mentioned challenges posed by children's developing linguistic/conceptual abilities, children may simply have difficulty in identifying the relevant mental content necessary to learning MSVs.

In particular, Papafragou et al. (2007) [PCG] have shown that even given adequate conceptual and linguistic abilities (as in adults) the mental events in a scene (the actors' internal states) are not attended to as much as the actions, unless there are cues that heighten the salience of the mental content. PCG further demonstrate that children's sensitivity to such cues lags behind that of adults, suggesting an additional factor in the acquisition of MSVs which

[1] Researchers have noted that children use MSVs in fixed phrases, in a performative use or as a pragmatic marker, well before they use them to refer to actual mental content (e.g., Diessel and Tomasello, 2001; Shatz et al., 1983). Here by "acquisition of MSVs", we are specifically referring to children learning usages that genuinely refer to mental content.

In: R. Levy & D. Reitter (Eds.), *Proceedings of the 3rd Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2012)*, pages 1–10,
Montréal, Canada, June 7, 2012.

is the developmental change in how strongly such cues are associated with the relevant mental content.

We develop a computational model of MSV acquisition (the first, to our knowledge) to further illuminate these issues. We extend an existing model of verb argument structure acquisition (Alishahi and Stevenson, 2008) to enable the representation and processing of mental state semantics and syntax. We simulate the developmental change proposed by PCG through a gradually increasing ability in the model to appropriately attend to the mental content of a scene. In addition, we suggest that even when the learner's *semantic* representation is biased towards the action content, the learner attends to the observed SC *syntax* in an MSV utterance. This is especially important to account for the pattern of errors in child data. Our model thus extends the account of PCG to show that a probabilistic interplay of the semantic and syntactic features of a partial and somewhat erroneous perception of the input, combined with a growing ability to attend to cues indicative of mental content, can help to account for children's developmental trajectory in learning MSVs.

2 Background and Our Approach

To investigate the linguistic and contextual cues that could help in learning MSVs, PCG use a procedure called the Human Simulation Paradigm (originally proposed by Gillette et al., 1999). In this paradigm, subjects are put in situations intended to simulate various word learning conditions of young children. E.g., in one condition, adults watch silent videos of caregivers interacting with children, and are asked to predict the verb uttered by the caregiver. In another condition, subjects hear a sentence containing a nonce verb (e.g., *gorp*) after watching the video, and are asked what *gorp* might mean.

We focus on two factors investigated by PCG in the performance of adults and children in identifying MSVs. The first factor they investigated involved the syntactic frame used when subjects were given a sentence with a nonce verb. PCG hypothesized that an SC frame would be a cue to mental content (and an MSV), since the SC refers to propositional content. The second factor PCG examined was whether the video described a "true belief" or a "false belief" scene: A true belief scene shows an ordinary situation which unfolds as the character in the scene expects — e.g., a little boy takes food to his grandmother, and she is there in the house as expected. The corresponding false belief scene has an unexpected outcome for the character — in this case, another character has replaced the grandmother in her bed. Here the hypothesis was that such false belief scenes would heighten the salience of mental activity in the scene and lead to greater belief verb responses in describing them.

PCG's results showed that both adults and children were sensitive to both the scene and syntax cues, but children's ability to draw on such cues was inferior to that of adults. They thus propose that the difference between children and adults is that children have not yet formed as strong an association as adults between the cues and the mental content of a scene as required to match the performance of adults. Nonetheless, their results suggest that the participating children had the conceptual and linguistic abilities required for MSVs, since they were able to produce them under conditions with sufficiently strong cues.

We simulate PCG's experiments using a novel computational approach. Following PCG, we assume that even when a learner is able to perceive the general semantic and syntactic properties of a belief scene and associated utterance, they may not attend to the mental content in every situation, and that this ability improves over time. We model a developmental change in a learner's attention to mental content: At early stages, corresponding to the state of young children, the learner largely focuses on the action aspects of a belief scene, even in the presence of an utterance using an MSV. Over time, the learner gradually increases in the ability to attend appropriately to the mental aspects of such a scene and utterance, until adult-like competence is achieved in associating the available cues with mental content.

Importantly, our work extends the proposal of PCG by bringing in evidence from other relevant studies on children's ability to process SCs. More specifically, we suggest that when children hear a sentence like *I think Mom went home*, they recognize (and record) the existence of an SC, while *at the same time* they focus on the action semantics as the main (most salient) event. In other words, we assume that children's imperfect syntactic abilities are at least sufficient to recognize the SC usage (Nelson et al., 1989; Asplin, 2002). However, their attention is mostly directed towards the action expressed in the embedded complement, either because mental content is less easily observable than action (Papafragou et al., 2007), or due to the linguistic saliency of the embedded clause (Diessel and Tomasello, 2001; Dehe and Wichmann, 2010). As mentioned above, we model this misrepresentation by considering the possibility of not attending to mental content in a belief scene. Specifically, we assume that (i) the model is very likely to overlook the mental content at earlier stages (corresponding to children's observed behaviour); and (ii) as the model 'ages' (i.e., receives more input), its attentional abilities improve and thus the model is more likely to focus on the mental content as the main proposition. Our results suggest that these changes to the model lead to a match between our model's behaviour and PCG's differential results for children and adults.

3 The Computational Model

A number of computational models have examined the role of interacting syntactic and semantic cues in the acquisition of verb argument structure (e.g., Niyogi, 2002; Buttery, 2006; Alishahi and Stevenson, 2008; Perfors et al., 2010; Parisien and Stevenson, 2011). However, to our knowledge no computational model has addressed the developmental trajectory in the acquisition of MSVs. Here we extend the verb argument structure acquisition model of Alishahi and Stevenson (2008) to enable it to account for MSV acquisition. Specifically, we use their core Bayesian learning algorithm, but modify the input processing component to reflect a developmental change in attention to the mental state content of an MSV usage and its consequent representation, as noted above.

We use this model for the following reasons: (i) it focuses on argument structure learning, and the interplay between syntax and semantics, which are key to MSV acquisition; (ii) it is probabilistic and hence can naturally capture gradient responses to different cues; and (iii) it is incremental, which allows us to investigate changes in behaviour over time. We first give an overview of the original model, and then explain our extensions.

3.1 Model Overview

The input to the model is a sequence of utterances (what the child hears), each paired with a scene (what the child perceives); see Table 1 for an example. First, the *frame extraction component* of the model extracts from the input pair a *frame*—a collection of *features*. We use features that include both semantic properties ('event primitives' and 'event participants') and syntactic properties ('syntactic pattern' and 'verb count'). See Table 2 for examples of two possible frames extracted from the pair in Table 1. Second, the *learning component* of the model incrementally clusters the extracted frames one by one. These clusters correspond to *constructions* that reflect probabilistic associations of semantic and syntactic features across similar usages, such as an agentive intransitive or causative transitive. The model can use these associations to simulate various language tasks as the prediction of a missing feature given others. For example, to simulate the human simulation paradigm setting, we can use the model to predict a missing verb on the basis of the available semantic and syntactic information (as in Alishahi and Pyykkoñen, 2011).

3.2 Algorithm for Learning Constructions

The model clusters the input frames into constructions on the basis of their overall similarity in the values of their features. Importantly, the model learns these constructions incrementally, considering the possibility of creating a new construction for a given frame if the frame is not sufficiently similar to any of the existing constructions. Formally, the model finds the best construction (including a new one) for a given frame F as in:

$$\mathrm{BestConstruction}(F) = \underset{k \in Constructions}{\mathrm{argmax}} P(k|F) \tag{1}$$

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

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

The prior probability of each construction $P(k)$ is estimated as the proportion of observed frames that are in k, assigning a higher prior to constructions

$\text{Think}_{[state, consider, cogitate]}(\text{I}_{[experiencer, preceiver, considerer]}, \text{Go}_{[physical, act, move]}(\text{MOM}_{[agent, actor, change]}, \text{HOME}_{[location, destination]}))$
I think Mom went home.

Table 1: A sample Scene–Utterance input pair.

(a) Interpretation#1 (mental event is attended to)

main predicate	**think**
other predicate	go
event primitives	$\{$ $state, consider, cogitate$ $\}$
event participants	$\{$ $experiencer, perceiver, considerer\}$ $\{$ $preposition, action, perceivable\}$
syntactic pattern	arg1 *verb* arg-S
verb count	2

(b) Interpretation#2 (mental event not attended to)

main predicate	**go**
other predicate	think
event primitives	$\{$ $physical, act, move\}$
event participants	$\{$ $agent, actor, change\}$ $\{$ $location, destination\}$
syntactic pattern	arg1 *verb* arg-S
verb count	2

Table 2: Two frames extracted from the scene–utterance pair in Table 1. The bottom left and right panels of the table describe the two possible interpretations given the input pair. (a) Interpretation#1 assumes that the mental event is the focus of attention. Here **think** is interpreted as the main predicate, which the event primitives and participants refer to. (b) Interpretation#2 assumes that attention is mostly directed to the physical action in the scene, and thus **go** is taken to be the main predicate, which also determines the extracted event primitives and participants. Note that for both interpretations, the learner is assumed to perceive the utterance in full, thus both verbs are heard in the context of the sentential complement syntax (i.e., syntactic pattern with SC and 2 verbs), without fully extracting the syntactic relations between the clauses.

that are more entrenched (i.e., observed more frequently). The likelihood $P(F|k)$ is estimated based on the values of features in F and the frames in k:

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

where i refers to the i^{th} feature of F and j refers to its value. The conditional probability of a feature i to have the value j in construction k, $P_i(j|k)$, is calculated with a smoothed version of:

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

where $\text{count}_i(j,k)$ reflects the number of times feature i has the value j in construction k, and n_k is the number of frames in k. We have two types of features: single-valued and set-valued. The result of the count_i operator for a single-valued feature is based on exact match to the value j, while the result for a set-valued feature is based on the degree of overlap between the compared sets, as in the original model.

3.3 Modeling Developmental Changes in Attending to Mental Content

We extend the model above to account for the increase in the ability to attend to cues associated with MSVs, as observed by PCG. In addition, we propose that children's representation of this situation includes the observed syntax of the MSV. That is, children do not simply ignore the MSV usage, focusing only on the action expressed in its complement — they must also note that this action semantics occurs in the context of an SC usage.

To adapt the model in these ways, we change the frame extraction component to allow two possible interpretations for a mental event input. First, to reflect PCG's proposal, we incorporate a mechanism into the model's frame-extraction process that takes into account the probability of attending to mental content. Specifically, we assume that when presented with an input pair containing an MSV, as in Table 1, a learner attends to the perceptually salient action/state expressed in the complement (here Go) with probability p, and to the non-perceptually salient mental event expressed in the main verb (here Think) with probability $1-p$. This probability p is a function over time, corresponding to the observed developmental progression. At very early stages, p will be high (close to 1), simulating the much greater saliency of physical actions compared to mental events for younger children. With subsequent input, p will decrease, giving more and more attention to the mental content of a scene with a mental event, gradually approaching adult-like abilities.

We adopt the following function for p:

$$p = \frac{1}{\delta \cdot t + 1}, \quad 0 < \delta \ll 1 \tag{5}$$

where t is the current time, expressed as the total number of scene–utterance pairs observed thus far by the model, and the parameter δ is set to a small value to assign a high probability to the physical action interpretation of the scene in the initial stages of learning (when t is small).

We must specify the precise make-up of the frames that correspond to the two possible interpretations considered with probability p and $1 - p$. PCG state only that children and adults differentially attend to the action vs. mental content of the scene. We operationalize this by forming two possible frames in response to an MSV usage. We propose that one of the frames (with probability $1-p$) is the interpretation of the mental content usage, as in Table 2(a). However, we extend the account of PCG by proposing that the other frame considered is not simply a standard representation of an action scene–utterance pair. Rather, we suggest that the interpretation of an MSV scene–utterance pair that focuses on the action semantics does so *within the context of the SC syntax*, given the assumed stage of linguistic abilities of the learner. This leads to the frame (with probability p) as in Table 2(b), which represents the action semantics within a two-verb construction associated with the SC syntax.

4 Experimental Setup

4.1 Input Data

We generate artificial corpora for our simulations, since we do not have access to sufficient data of actual utterances paired with scene representations. In order to create naturalistic data that resembles what children are exposed to, we follow the approach of Alishahi and Stevenson (2008) to build an input-generation lexicon that has the distributional properties of actual child-directed speech (CDS). Their original lexicon contains only high-frequency physical action verbs that appear in limited syntactic patterns. Our expanded lexicon also includes mental state, perception, and communication verbs, all of which can appear with SCs.

We extracted our verbs and their distributional properties from the child-directed speech of 8 children in the CHILDES database (MacWhinney, 2000).[2] We selected 28 verbs from different semantic classes and different frequency ranges: 12 physical action verbs taken from the original model (*come, go, fall, eat, play, get, give, take, make, look, put, sit*), 6 perception and communication verbs (*see, hear, watch, say, tell, ask*), 5 belief verbs (*think, know, guess, bet, believe*), and 5 desire verbs (*want, wish, like, mind, need*). For each verb, we manually analyzed a random sample of 100 CDS usages (or all usages if fewer than 100) to extract distributional information about its argument structures.

We construct the input-generation lexicon by listing each of the 28 verbs (i.e. the 'main predicate'), along with its overall frequency, as well as the frequency with which it appears with each argument structure. Each entry contains values of the syntactic and semantic features (see Table 2 for examples), including 'event primitives', 'event participants', 'syntactic pattern', and 'verb count'. By including these features, we assume that a learner is capable of understanding basic syntactic properties of an utterance, including word syntactic categories (e.g., noun and verb), word order, and the appearance of SCs (e.g., Nelson et al., 1989). We also assume that a learner has the ability to perceive and conceptualize the general semantic properties of events — including mental, perceptual, communicative, and physical actions — as well as those of the event participants. Values for the semantic features (the event primitives and event participants) are taken from Alishahi and Stevenson (2008) for the action verbs, and from several sources including VerbNet (Kipper et al., 2008) and Dowty (1991) for the additional verbs.

For each simulation in our experiments (explained below), we use the input-generation lexicon to automatically generate an input corpus of scene–utterance pairs that reflects the observed frequency distribution in CDS.[3] For an input utterance that contains an MSV, we randomly pick one of the action verbs as the verb appearing within the sentential complement (the 'other predicate').

[2] Corpora of Brown (1973); Suppes (1974); Kuczaj (1977); Bloom et al. (1974); Sachs (1983); Lieven et al. (2009).

[3] The model does not use the input-generation lexicon in learning.

4.2 Setup of Simulations

We perform simulations by training the model on a randomly generated input corpus, and examining changes in its performance over time with periodic tests. Specifically, we perform simulations of the verb identification task in the human simulation paradigm as follows: At each test point, we present the model with a *partial test frame* with missing predicate (verb) values, and different amounts of information for the other features. The tests correspond to the scenarios in the original experiments of PCG, where each scenario is represented by a partial frame as follows:

1. **scene-only scenario**: Corresponds to subjects watching a silent video depicting either an Action or a Belief scene. Our test frame includes values for the semantic features (event primitives and event participants) corresponding to the scene type, but no syntactic features.

2. **syntax-only scenario**: Corresponds to subjects hearing either an SC or a non-SC utterance. The test frame includes the corresponding syntactic pattern and verb count of the utterance type heard, but no semantic features.

3. **syntax & scene scenario**: Corresponds to subjects watching a silent video (with Action or Belief content), and hearing an associated (non-SC or SC) utterance. The test frame includes all the relevant syntactic and semantic features.

We perform 100 simulations, each on 15000 randomly-generated training frames, and examine the type of verbs that the model predicts in response to test frames for the three scenarios. For each scenario and each simulation, we generate a test frame by including the relevant feature values of a randomly-selected physical action or belief verb usage from the input-generation lexicon.

PCG code the individual verb responses of their human subjects into various verb classes. To analogously code our model's response to each test frame, we estimate the likelihood of each of two verb groups, Belief and Action,[4] by summing over the likelihood of all the verbs in that group. In the results below, these likelihood scores are averaged for each test point over the 100 simulations.

[4]The Action verbs include action, communication, and perception verbs, as in PCG. Verbs from the desire group are not considered here, also as in PCG.

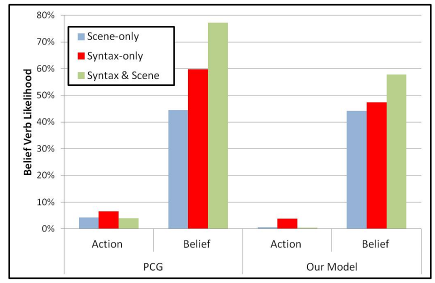


Figure 1: Likelihood of Belief verb prediction given Action or Belief input.

When our model is presented with a test frame containing a Belief scene, we assume that the model (like a language learner) may not attend to the mental content, resulting in one of the two interpretations described in Section 3.3 (see Table 2). We thus calculate the verb class likelihoods using a weighted average of the verbs predicted under the two interpretations. Following PCG, we test our model with two types of Belief scenes: True Belief and False Belief, with the latter having a higher level of belief saliency. We model the difference between these two scene types as a difference in the probabilities of perceiving the two interpretations, with a higher probability for the belief interpretation given a False Belief test frame. In the experiments presented here, we set this probability to 80% for False Belief, and to 60% (just above chance) for True Belief. (Unlike in training, where we assume a change over time in the probability of a belief interpretation, for each presentation of the test frame we use the same probabilities of the two interpretations.)

5 Experimental Results

We present two sets of results: In Section 5.1, we examine the role of syntactic and semantic cues in MSV identification, by comparing the likelihoods of the model's Belief verb predictions across the three scenarios. Here we test the model after processing 15000 input frames, simulating an adult-like behaviour (as in PCG). At this stage, we present the model with an Action test frame (Action scene and/or Transitive syntax), or a Belief test frame

(False Belief scene and/or SC syntax). In Section 5.2, we look into the role of semantic cues that enhance belief saliency, by comparing the likelihoods of Belief vs. Action verb predictions in the syntax & scene scenario. The test frames depict either a True Belief or a False Belief scene, paired with an SC utterance. Here, we test our model periodically to examine the developmental pattern of MSV identification, comparing our results with the difference in the behaviour of children and adults in PCG.

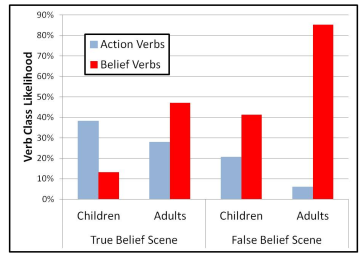
5.1 Linguistic Cues for Belief Verb Prediction

The left side of Figure 1 presents the results of PCG (for adult subjects); the right side shows the likelihood of Belief verb prediction by our model. Similar to the results of PCG, our model's likelihood of Belief verb prediction is extremely low when given an Action test frame (Action scene and/or Transitive syntax), whereas it is much higher when the model is presented with a Belief test frame (False Belief scene and/or SC syntax). Moreover, as in PCG, when the model is tested with Belief content, the lowest likelihood is for the scene-only scenario and the highest is for the syntax & scene scenario.

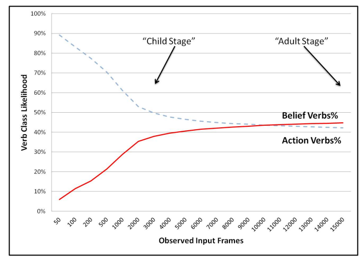
PCG found, somewhat surprisingly, that the syntax-only scenario was more informative for MSV prediction than the scene-only scenario. Our results replicate this finding, which we believe is due to the way our Bayesian clustering groups verb usages together. Non-SC usages of MSVs are often grouped with action verbs that frequently appear with non-SC syntax, and this results in constructions with mixed (action and belief) semantics. When using MSV semantic features to make the verb prediction, the action verbs get a higher likelihood based on such mixed constructions. However, the frequent usage of MSVs with SC results in entrenched constructions of mostly MSVs. Although other verbs, such as *see* and *say*, may also be used with SC syntax, they are grouped with verbs such as *watch* and *tell* into constructions with mixed (SC and non-SC) syntax. When given SC syntax in verb prediction, the more coherent MSV constructions result in a high likelihood of predicting Belief verbs.

5.2 Belief Saliency in Verb Prediction

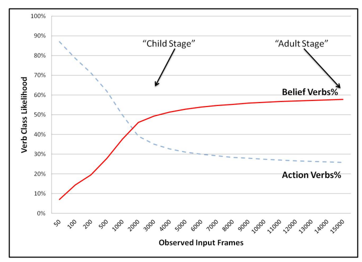
Figure 2(a) shows the PCG results, for children and adults, and for True Belief and False Belief.



(a)



(b)



(c)

Figure 2: Verb class likelihood: (a) PCG results for adults and children (aged 3;7–5;9); (b) Model's results given True Belief; (c) Model's results given False Belief.

Figures 2(b) and (c) present the likelihoods of the model's Belief vs. Action verb prediction, over time, for True and False Belief situations (True/False Belief scene and SC syntax), respectively. We first compare the responses of our model at the final stage of training to those of adults in PCG. At this stage, the model's verb predictions (for both True and False Belief) follow a similar trend to that of adult subjects in PCG. The likelihood of Belief verbs is much higher than the likelihood of Action verbs given a False Belief situation. Moreover, the likelihood of Belief verbs is higher given a False Belief situation, compared to a True Belief situation.

Next, we compare the developmental pattern of Belief/Action verb predictions in the model with the difference in behaviour of children and adults in PCG. We focus on the model's responses after processing about 3000 input pairs, as it corresponds to the trends observed for the children in PCG. At this stage, the likelihood of Belief verbs is lower than that of Action verbs for the True Belief situation, but the pattern is reversed for False Belief; a pattern similar to children's behaviour in PCG (see Figure 2(a)). As in PCG, the likelihood of Belief verb predictions in our model is higher than that of Action verbs for the False Belief situation, in both "child" and "adult" stages, with a larger difference as the model 'ages' (i.e., processes more input). For the True Belief situation also the pattern is similar to that of PCG: Belief verbs are less likely than Action verbs to be predicted at early stages, but as the model receives more input, the likelihood of Belief verbs becomes slightly higher than that of Action verbs.

PCG's hypothesis of greater attention to the action content of a scene implicitly implies that children focus on the action semantics and syntax of the embedded SC of a Belief verb. We have suggested instead that the focus is on the action semantics within the context of the SC syntax of the MSV. To directly evaluate the necessity of our latter assumption, we performed a simulation using both action syntax and semantics to represent the physical interpretation of the belief scene. Specifically, the syntactic features in this representation were non-SC structure with only one verb. Based on these settings, the model predicted high likelihood for the Belief verbs from a very early stage, not showing the same delayed acquisition pattern exhibited by PCG's results. This result suggests that the SC syntax plays an important role in MSV acquisition.

6 Discussion

Various studies have considered why mental state verbs (MSVs) appear relatively late in children's productions (e.g., Shatz et al., 1983; Bartsch and Wellman, 1995). The Human Simulation Paradigm has revealed that adult participants tend to focus on the physical action cues of a scene (Gleitman et al., 2005). PCG's results further show that cues emphasizing mental content lead to a significant increase in MSV responses in such tasks. Moreover, they show that a sentential complement (SC) structure is a stronger cue to an MSV than the semantic cues emphasizing mental content.

In this paper we adapt a computational Bayesian model to analyze such semantic and syntactic cues in the ability of children to identify them. We simulate an attentional mechanism of the growing sensitivity to mental content in a scene into the model. We show that both the ability to observe the obscure mental content and the ability to recognize the use of an SC structure are essential to replicate PCG's observations. Moreover, our results predict the strong association of MSVs to the SC syntax, for the first time (to our knowledge) in a computational model.

Children often use verbs other than MSVs in experimental settings in which MSVs would be the appropriate or correct verb choice (Asplin, 2002; Kidd et al., 2006; Papafragou et al., 2007). Our model presents similar variability in verb choice. One underlying cause of this behaviour in the model is its association of action semantics to SC syntax, due to the tendency to observe the physical cues in a scene associated with an utterance using an MSV with an SC. Preliminary results (not reported here) imply that the association of perception and communication verbs that frequently appear with SC contribute to this pattern of verb choice (see de Villiers, 2005, for theoretical support). Our results require further work to fully understand this behaviour.

Finally, our model will facilitate future work in regards to the *performative usage* of MSVs, in which MSVs do not indicate mental content, but rather direct the conversation. Several studies (e.g., Diessel and Tomasello, 2001; Howard et al., 2008), have referred to the role performative use likely plays in MSV acquisition, since the first MSV usages by children are performative. The semantic properties MSVs take in performative usages is not currently represented in our lexicon. However, the physical interpretation of the mental scene that we have used in our experiments here is similar to the performative usage: i.e., the main perceived action and the observed syntactic structure are the same. At the moment, our results imply that the association of MSVs with their genuine mental meaning is delayed by interpretations of the mental scene which overlook the mental content. In the future, we aim to incorporate the semantic representation of performative usages to better analyze their effect on MSV acquisition.

References

Afra Alishahi and Pirita Pyykkoñen. 2011. The onset of syntactic bootstrapping in word learning: Evidence from a computational study. In *Proceedings of the 33st Annual Conference of the Cognitive Science Society*.

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.

Karen Bartsch and Henry M. Wellman. 1995. Children talk about the mind.

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

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

Paula J. Buttery. 2006. Computational models for first language acquisition. Technical Report UCAM-CL-TR-675, University of Cambridge, Computer Laboratory.

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 University Press.

Nicole Dehe and Anne Wichmann. 2010. Sentence-initial I *think (that)* and *i believe (that)*: Prosodic evidence for use as main clause, comment clause and dicourse marker. *Stuides in Language*, 34(1):36–74.

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

David Dowty. 1991. Thematic Proto-Roles and Argument Selection. *Language*, 67(3):547–619.

Jane Gillette, Lila Gleitman, Henry Gleitman, and Anne Lederer. 1999. Human simulations of lexical acquisition. *Cognition*, 73(2):135–176.

Lila R. Gleitman, Kimberly Cassidy, Rebecca Nappa, Anna Papafragou, and John C. Trueswell. 2005. Hard words. *Language Learning and Development*, 1(1):23–64.

Alison Gopnik and Andrew N. Meltzoff. 1997. Words, thoughts, and theories.

Alice A. Howard, Lara Mayeux, and Letitia R. Naigles. 2008. Conversational correlates of children's acquisition of mental verbs and a theory of mind. *First Language*, 28(4):375.

Carl Nils Johnson and Henry M. Wellman. 1980. Children's developing understanding of mental verbs: Remember, know, and guess. *Child Development*, 51(4):1095–1102.

Evan Kidd, Elena Lieven, and Michael Tomasello. 2006. Examining the role of lexical frequency in the acquisition and processing of sentential complements. *Cognitive Development*, 21(2):93–107.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A large-scale classification of English verbs. *Language Resources and Evaluation*, 42(1):21–40–40.

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.

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

Deborah G. Kemler Nelson, Kathy Hirsh-Pasek, Peter W. Jusczyk, and Kimberly Wright Cassidy. 1989. How the prosodic cues in motherese might assist language learning. *Journal of child Language*, 16(1):55–68.

Sourabh Niyogi. 2002. Bayesian learning at the syntax-semantics interface. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*.

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

Christopher Parisien and Suzanne Stevenson. 2011. Generalizing between form and meaning using learned verb classes. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*.

Amy Perfors, Joshua B. Tenenbaum, and Elizabeth Wonnacott. 2010. Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, 37(03):607–642.

Willard .V.O. Quine. 1960. *Word and object*, volume 4. The MIT Press.

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.