A computational model of learning semantic roles from child-directed language

Afra Alishahi [a]; Suzanne Stevenson [b]

[a] Department of Computational Linguistics and Phonetics, Saarland University, Saarbrücken, Germany
[b] Department of Computer Science, University of Toronto, Toronto, Canada

First published on: 26 May 2009

PLEASE SCROLL DOWN FOR ARTICLE

A computational model of learning semantic roles from child-directed language

Afra Alishahi

Department of Computational Linguistics and Phonetics, Saarland University, Saarbrücken, Germany

Suzanne Stevenson

Department of Computer Science, University of Toronto, Toronto, Canada

Semantic roles are a critical aspect of linguistic knowledge because they indicate the relations of the participants in an event to the main predicate. Experimental studies on children and adults show that both groups use associations between general semantic roles such as Agent and Theme, and grammatical positions such as Subject and Object, even in the absence of familiar verbs. Other studies suggest that semantic roles evolve over time, and might best be viewed as a collection of verb-based or general semantic properties. A usage-based account of language acquisition suggests that general roles and their association with grammatical positions can be learned from the data children are exposed to, through a process of generalisation and categorisation.

In this paper, we propose a probabilistic usage-based model of semantic role learning. Our model can acquire associations between the semantic properties of the arguments of an event, and the syntactic positions that the arguments appear in. These probabilistic associations enable the model to learn general conceptions of roles, based only on exposure to individual verb usages, and without requiring explicit labelling of the roles in the input. The acquired role properties are a good intuitive match to the expected properties of various roles, and are useful in guiding comprehension in the model to the most likely

interpretation in the face of ambiguity. The learned roles can also be used to select the correct meaning of a novel verb in an ambiguous situation.

1. INTRODUCTION

Semantic roles, such as Agent, Theme, and Recipient in (1) and (2) below, are a critical aspect of linguistic knowledge because they indicate the relations of the participants in an event to the main predicate.

(1) Mom$_{Agent}$ gave this$_{Theme}$ to her$_{Recipient}$.
(2) Mom$_{Agent}$ gave her$_{Recipient}$ this$_{Theme}$.

Moreover, it is known that people use the associations between roles and their syntactic positions to help guide on-line interpretation (e.g., Carlson & Tanenhaus, 1988; Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006; McRae, Ferretti, & Amyote, 1997; Trueswell, Tanenhaus, & Garnsey, 1994). For example, upon hearing the partial utterance *Mom gave her . . .* , most hearers would guess that *her* refers to the Recipient of the giving event, whereas in *Mom gave this . . .* they would assume that *this* is the Theme of the event. How children acquire this kind of complex relational knowledge, which links predicate-argument structure to syntactic expression, is still not well understood. Fundamental questions remain concerning what the nature of semantic roles is, how they are learned, and how associations are established between roles and the grammatical positions the role-bearing arguments appear in.

Early theories suggested that roles are drawn from a pre-defined, universal inventory of semantic symbols or relations, and that innate 'linking rules' that map roles to sentence structure enable children to infer associations between role properties and syntactic positions (e.g., Pinker, 1989). However, numerous questions have been raised concerning the plausibility of innate linking rules for language acquisition (e.g., Bowerman, 1990; Fisher, 2000; Kako, 2006).

An alternative, usage-based view is that children acquire roles gradually from the input they receive, by generalising over individually learned verb usages (e.g., Lieven, Pine, & Baldwin, 1997; Tomasello, 2000). For instance, Tomasello (2000) claims that, initially, there are no general labels such as Agent and Theme, but rather verb-specific concepts such as 'hitter' and 'hittee', or 'sitter' and 'thing sat upon'. Recent experimental evidence confirms that access to general notions like Agent and Theme is age-dependent (Shayan & Gershkoff-Stowe, 2007). It remains unexplained, though, precisely how verb-specific roles metamorphose to general semantic

roles. Moreover, experiments with children have revealed the use of verb-specific biases in argument interpretation (Nation, Marshall, & Altmann, 2003), as well as of strong associations between general roles and syntactic positions (Fisher, 1994, 1996, 2002). However, specific computational models of such processes have been lacking.

We have proposed a usage-based computational model of early verb learning that uses Bayesian clustering and prediction to model language acquisition and use. Our previous experiments demonstrated that the model learns basic syntactic constructions such as the transitive and intransitive, and exhibits patterns of errors and recovery in their use similar to those of children (Alishahi & Stevenson, 2008). A shortcoming of the model was that roles were explicit labels, such as Agent, which were assumed to be 'perceptible' to the child from the scene. In this paper, we have extended our model to directly address the learning and use of semantic roles.

Our Bayesian model associates each argument position of a predicate with a probability distribution over a set of semantic properties – a *semantic profile*. We show that initially the semantic profiles of an argument position yield verb-specific conceptualisations of the role associated with that position. As the model is exposed to more input, these verb-based roles gradually transform into more abstract representations that reflect the general properties of arguments across the observed verbs. We further establish that such representations can be useful in guiding the argument interpretation of ambiguous input, as well as in aiding word learning in unclear contexts. Our focus is on developing a computational-level model (Marr, 1982), characterising the functional capacities of human language acquisition rather than specific psychological processes that implement those functions.[1]

2. THE MULTIDISCIPLINARY STUDY OF SEMANTIC ROLE LEARNING

2.1 Linguistic theories

The notion of thematic roles was first introduced by semanticists as the relationship between a predicate and its arguments (Fillmore, 1968; Jackendoff, 1972). However, this notion was extensively used by syntacticians as a theoretical device to explain argument indexing (i.e., linking grammatical relations to semantic structure) and grammatical generalisation (Chomsky, 1981; Pinker, 1984). In many theories of syntax such as Government and

---

[1] Bayesian models are shown to be successful in modelling a variety of cognitive processes, and are becoming increasingly popular in cognitive modelling. For a broad overview of recent Bayesian models of cognition, see Griffiths, Kemp, and Tenenbaum (2008).

Binding Theory (Chomsky, 1981) and Lexical–Functional Grammar (Bresnan, 2001; Falk, 2001), thematic roles are believed to be discrete, limited in number, and universal. The mapping between roles and the sentence structure is defined through a set of universal 'linking rules'. These rules are argued to be innate, and to help children in learning the syntax of their language. A strong version of these rules suggests that the mapping of a thematic role to a particular grammatical function is rigid (Baker, 1988; Pinker, 1984). A weaker position proposes that thematic roles and syntactic positions are matched by means of a hierarchy, such that the highest-ranked thematic role occupies the highest-ranked syntactic position (i.e., the Subject), continuing down the two hierarchies in parallel until one runs out of arguments (Grimshaw, 1990; Jackendoff, 1990; VanValin & LaPolla, 1997).

Researchers have proposed various lists of thematic roles, mostly different in size and granularity. But there is little consensus about the 'correct' set of thematic roles. That is mainly due to the fact that, in order for the universal linking rules to be useful, it should be possible to assign each argument of every verb in the language to one and only one thematic role. That is, what a verb semantically entails about each of its arguments must permit us to assign the argument, clearly and definitely, to some role or other, and what the meaning of the verb entails about every argument must always be distinct enough that two arguments clearly do not fall under the same role definition. However, it seems that there is no cross-linguistically consistent definition of thematic roles that satisfies these criteria.

Dowty (1991) proposes a different theory of thematic roles: the Proto-Role Hypothesis. According to his theory, thematic roles draw from a pool of more basic semantic properties such as sentience, volition, and movement. No single thematic role necessarily has all of these properties, and some have more than others. The mapping of an argument to a grammatical position is decided based on the number of proto-role properties that the argument has in a particular event; for example, in a causal action event, the argument that demonstrates more proto-agent properties is assigned to the Subject position. The proposed proto-roles are based on analysis of linguistic data. However, Dowty (1991) does not give any explicit account on whether these proto-roles are innate or learned from experience.

2.2 Psycholinguistic studies

Although several experimental studies have been performed to explore the role of thematic roles in language processing (see, for example, Kuperberg et al., 2006; Trueswell et al., 1994), there is little agreement on what the nature of thematic roles is. McRae et al.'s (1997) experiments on human subjects' ranking of role/filler featural similarity for Agent and Patient roles suggests that thematic roles might best be viewed as verb-specific concepts.

On the other hand, there is evidence that in the absence of any verb-specific information, humans have some conception of the general semantic roles in familiar constructions. Kako (2006) shows that human subjects assign proto-role properties to grammatical roles such as Subject and Object, even for a novel verb (or a familiar verb in an unusual construction). Based on such results, proponents of the usage-based framework for language acquisition and use have suggested that children do not have access to a pre-defined set of thematic roles or proto-role properties. Instead, children learn thematic roles gradually from the input they receive, through a process of categorisation and generalisation (e.g., Lieven et al., 1997; Tomasello, 2000). However, a detailed account of the transformation of verb-specific to general semantic roles, and their association with grammatical positions, is yet to be proposed.

There are few experiments on how children learn general semantic roles. Shayan and Gershkoff-Stowe (2007) show that children indeed demonstrate a pattern of gradually learning thematic roles, and that both age and access to linguistic cues affect the learning process. In their experiments, 3- and 5-year-old children's knowledge of Agent and Patient roles is tested in simple transitive events with two participants. It is shown that older children have a better understanding of the relational similarity across Agents and Patients in animated scenes presented side-by-side on a computer screen. It is also shown that providing linguistic cues (i.e., using a novel verb in a transitive frame to describe the scenes) makes it easier for both age groups to grasp relational similarity across semantic roles. Moreover, experiments with children have revealed the use of verb-specific biases in argument interpretation (Nation et al., 2003), as well as of strong associations between general roles and syntactic positions (Fisher, 1994, 1996, 2002). However, the how and when of the emergence and learning of thematic roles, and their mapping to syntactic positions, is yet to be explored through further psycholinguistic experiments and computational modelling.

2.3 Computational models

The study of the learnability of general roles can benefit from computational modelling, by exploring explicit mechanisms for a psychologically plausible usage-based learning strategy for semantic roles. Computational experiments must show the feasibility of learning a general conception of semantic roles from individual verb usages, as well as of establishing an association between general roles and syntactic positions, without built-in knowledge of linking rules.

Learning relationships between entities or concepts has been studied through a number of computational models. Doumas and Hummel (2005) and Doumas, Hummel, and Sandhofer (2008) present DORA, a connectionist model that can discover relational concepts from unstructured

examples by comparing instances of relations and extracting their common features. The model can also learn 'meta relations', i.e., the relations between other relations (e.g., learning the *same_colour* relationship based on the relations *red*(x), *red*(y), *blue*(x), and *blue*(y)). In a similar vein, Kemp, Tenenbaum, Griffiths, Yamada and Ueda (2006) propose a Bayesian model of discovering systems of related concepts which, given various sets of entities, clusters these entities into different 'kinds', and chooses the most likely relations between kinds. These models, however, are presented for and evaluated in non-linguistic domains, and are not concerned about the specific problems that language learners face. Specifically, they do not attempt to learn the association between relationship arguments and syntactic positions, or to generalise the argument profiles across different relations.

When attempting to learn the thematic relations from instances of language use, the predicate terms of the language such as verbs and adjectives can be considered as the relational concepts, and nouns as the arguments. However, learning the relations in the language domain is especially complicated due to linguistic ambiguity, and the variation in the argument structure of most verbs. For example, the verb *break* can be used with one argument (*the window broke*), two arguments (*he broke the window*), or three arguments (*he broke the window with a hammer*). On the other hand, even the same set of verb arguments can be expressed in more than one way (e.g., *she gave me the book* and *she gave the book to me*). Therefore, the existing models of learning non-linguistic relations cannot be directly applied to the language domain.

Within models of language acquisition, a number of computational models learn verb-specific roles that are not generalised. For example, Chang (2004) presents a model that learns associations between form relations (typically word order) and meaning relations (typically role-filler bindings) from input data, and uses them in language comprehension. However, the acquired associations are not generalised beyond the scope of the individual verbs. Verb-specific roles are, in nature, very similar to the selectional preferences of a verb (i.e., the properties that a verb demands from its arguments), and many computational models have been proposed to learn them. Most of these models use WordNet (Miller, 1990) as their underlying ontology: they first initialise WordNet concepts with their frequency of use as the particular argument of a verb, and then find the appropriate level in the WordNet hierarchy for capturing the verb's restrictions on that argument (e.g., Clark & Weir, 2002; Resnik, 1996). However, none of these models generalise their acquired verb-based knowledge to a higher level, which would yield constraints on the arguments of general constructions such as the transitive or intransitive.

In a contrasting approach, many computational systems model human learning of the assignment of general pre-defined roles to sentence constituents. McClelland and Kawamoto (1986), Allen (1997) and Morris,

Cottrell, and Elman (2000) present connectionist models for assigning roles to constituents of sentences, by using as input a multi-feature representation of the semantic properties of arguments and the surface structure of the sentence. The output of these models is the assignment of a limited number of fixed thematic roles such as Agent and Instrument to the arguments of a verb. These models can also guess certain properties (semantic features) of a missing argument in a sentence, but the roles themselves are not learned. Allen's (1997) model treats the representation of thematic roles differently in that each role is further elaborated by additional proto-role units. However, the explicit role-labelling of the arguments in the training data is critical to these models, and it has not been demonstrated that they can learn general roles based only on the semantic properties of the arguments and the set of proto-role properties specified in the training data. Bates and MacWhinney (1989) and Matessa and Anderson (2000) take a different approach to role assignment, as learning the relative importance of specific cues such as word order, noun animacy, and case inflection. Both models learn a cue dominance hierarchy of the language from input data, but the roles themselves are assumed to be pre-defined.

As in our previous work on argument structure acquisition, all of these computational models require explicit labelling of the arguments that receive the same role in order to learn the association of the roles to semantic properties and/or syntactic positions. In this paper, we show that our extended model can learn general semantic profiles of arguments, as well as their association with grammatical positions, without the need for role-annotated training data, or even a list of known roles to draw from.

3. A COMPUTATIONAL MODEL FOR LEARNING SEMANTIC ROLES

We present an implemented computational model of the learning of general semantic roles. We take a statistical usage-based approach, in which the model learns semantic conceptions of general roles given examples of verb usages paired with a semantic representation of the event. Following Alishahi and Stevenson's (2008) proposal, generalisation in our model is achieved through an unsupervised Bayesian algorithm that groups similar verb usages: the general constructions of language, such as transitive and intransitive constructions in English, can be modelled as a cluster of individual verb usages that share similar syntactic and semantic features.

We propose a new view of semantic roles as a distributed representation of semantic properties that each verb argument can take on. In this view, general semantic roles are not pre-defined universal labels, or sets of fixed and innately specified proto-role properties. Instead, each role is a

generalisation of the semantic properties of the arguments appearing in a particular syntactic position in input data. More specifically, we represent each general semantic role in the form of a *semantic profile*, a probability distribution over the semantic properties of the arguments. Moreover, our model forms probabilistic associations between syntactic positions of arguments, their semantic properties, and the semantic primitives of the verb. These associations are generalised (through the constructions) to form more abstract notions of role semantics. We show through computational simulation that the model can use these associations to guide language comprehension and acquisition in the face of ambiguity.

A preliminary version of this model was presented in Alishahi and Stevenson (2007b). In this paper we describe an improved version of the model, present additional experiments, and discuss the results and their import for modelling the acquisition of general semantic roles. Our assumptions regarding the input to the model and the properties of a verb usage are described in detail in Sections 3.1 and 3.2, respectively. The process of learning general constructions and the formation of semantic profiles for argument positions are described in Sections 3.3 and 3.4.

3.1 The input to our model

We assume that, upon observing a simple event, the child can infer certain semantic properties of both the event itself and the arguments that participate in it. Moreover, if the child hears a sentence while watching the scene, they can establish a link between the linguistic description of the event and the relevant semantics inferred through observation. We use such pairings of utterance and semantic elements as the basic input to our model.

In real situations, it is not always easy to find the appropriate semantics for the utterance from the full representation of the perceived aspects of the scene, a well-known problem referred to as *referential uncertainty* (e.g., Pinker, 1989; Quine, 1960). Learning the correct meaning for both verbs and nouns has been suggested to be based on cross-situational observation (Fisher, Hall, Rakowitz, & Gleitman, 1994; Pinker, 1989); that is, frequent usage of a word in the presence of a concept or an event guides the child to establish a link between that word and its corresponding meaning. A number of computational models have dealt with the problem of picking the right meaning for an utterance, as well as learning the meaning of individual words (e.g., Fazly, Alishahi, & Stevenson, 2008; Yu, Ballard, & Aslin, 2005; Siskind, 1996). To simplify the learning problem that we focus on in this paper, we assume here that the (non-trivial) task of picking out the appropriate semantics for the utterance from the full representation of the perceived aspects of the scene has been performed, and that for each word, the child knows the corresponding meaning.

The cross-situational account of word learning is confirmed by child experimental data: Forbes and Farrar (1995) show that training with variable events makes children and adults more likely to generalise verbs to label modified test events. However, many verbs describe a particular perspective on events that cannot be inferred merely by cross-situational analysis. For example, 'buying' and 'selling' almost always happen at the same time. The knowledge of thematic roles and their association with grammatical positions has been suggested to guide children in acquiring the meaning of verbs in such cases. We will return to this issue in Section 6.2.2, and show a similar trend in the behaviour of our model. However, as input data to our model, we use a small set of the most frequent verbs in a corpus of child-directed conversations, assuming that their meaning can be inferred through successive usage of each verb in the presence of the corresponding event in an unambiguous context. For example, the verb *eat* is mostly used in a context where no counterpart *feeding* event is happening, and therefore the issue of finding the right perspective does not cause a problem for the word learner. We then show that the general conceptions that the model acquires based on this limited set of verbs can be successfully used to guide the acquisition of other verbs in ambiguous contexts.

3.2 Verb usages and argument properties

We mentioned in the previous section that the input to our model consists of pairings of a natural language utterance and the semantic elements of an event that the utterance describes. It has been shown that children are sensitive to certain properties of the event, such as causation and movement, from an early age (Cohen & Oakes, 1993; Gentner, 1978; Fisher, 1996; Naigles & Kako, 1993). Also, children show sensitivity to the semantic properties of the arguments, such as animacy and being human (Braine, Brody, Fisch, Weisberger, & Blum, 1990; Fisher, 1996; Gropen, Pinker, Hollander, & Goldberg, 1991). Some of these properties are intrinsic to the actual participants in an event; for example, in an eating scene, the observer may notice the eater's gender and age, as well as more general properties such as being human. Others are exhibited by the participants in virtue of the roles they take on in the event, such as 'moving in a rhythmic manner' in a dancing scene. We capture each of these types of properties in the semantic representation of an event. We make no claims regarding the psychological reality of the specific properties associated with the participants in an event. Rather, our goal is to show that given sets of such properties, our model learns roles based on the shared properties of the arguments that are mapped to certain grammatical positions in similar usages, and the shared properties of the events that are expressed in similar syntactic patterns.

Lexical properties: {meal, repast, nutriment, nourishment, sustenance, ...}
Event-based properties: {non-independently exist, affected, change, ...}

Sara eat lunch

Semantic primitives: {act, consume}

Lexical properties: {woman, adult female, female, person, individual, somebody, human, ...}
Event-based properties: {volitional, affecting, animate, independently exist, consuming, ...}
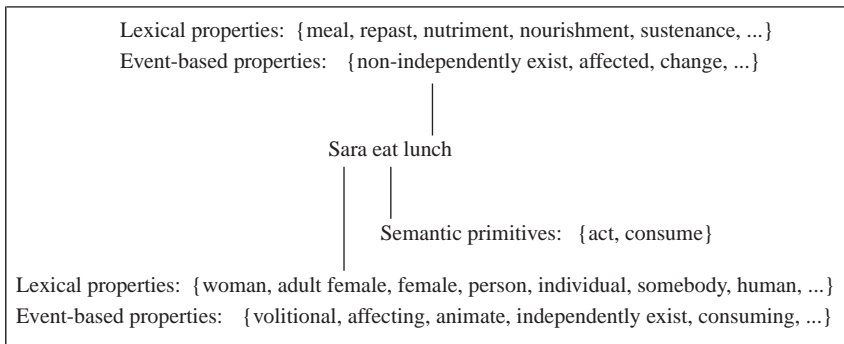
Figure 1. A sample verb usage: an utterance paired with the inferred semantic information.

Figure 1 shows a sample verb usage, consisting of a natural language utterance paired with the semantic information that is inferred through observing the corresponding event given to our model as a sequence of words in root form.

The meaning of the utterance is represented as three sets of semantic features:

- *Semantic primitives of the verb*: the basic characteristics of the predicate are described as semantic primitives (e.g., {cause, become, rotating}). Some of the primitives are general and shared by many verbs (e.g., 'movement' or 'act'), whereas others are verb-specific (e.g., 'consume' or 'play').
- *Lexical properties of each argument*: the inherent properties of the argument (e.g., {woman, adult, person, ... }). These lexical semantic properties are independent of the event that the argument participates in.
- *Event-based properties of each argument*: the properties that the argument takes on in virtue of how it participates in the event. Some of these properties are similar to the proto-role properties proposed by Dowty (1991) (e.g., 'cause' or 'affected') but others are verb-specific (e.g., 'eating' or 'falling').

We explain later how we choose the properties for events and arguments in our experiments.

3.3 General constructions as groups of verb usages

A construction in our model is a group of verb usages that are 'similar enough', according to the probabilities over their features, to be grouped together. The notion of 'similar enough' is described in detail in the next

section of the paper; here we focus on the general properties of constructions, and their representation in the lexicon.

Verb usages are described by both syntactic and semantic features. The values for some features, such as the syntactic pattern and the number of arguments, come from a limited set, due to the regular nature of natural languages. Therefore, there is often a dominant syntactic pattern and argument number among usages in each construction. However, due to the probabilistic nature of our model, a few 'outliers' might be found in each construction as well. Features with more varied values (such as semantic properties of the arguments and semantic primitives of the verbs) are less uniform, but typically overlap in value in a construction. Therefore, the primary property of constructions in our model is that they determine a probabilistic association between syntactic and semantic features. For example, usages such as *He made dinner* and *She ate soup* may be grouped into the same (transitive) construction. While the usages share the verb semantic primitive *act*, they differ in others (*create* for the former, and *consume* for the latter). If this observation holds across a number of usages that exhibit this form, then we would find a higher probability for the primitive *act* given this construction than for the other semantic primitives. In this way, constructions probabilistically generalise the properties of a set of usages. Figure 2 shows a portion of the acquired lexicon, showing the lexical entries of verbs containing frames, and their links to constructions. A formal account of the acquisition of constructions in our model is presented in Section 4.1.

3.4 Semantic profiles as probability distributions over argument properties

Psycholinguistic experiments have shown that children are sensitive at an early age to the association between grammatical positions, such as Subject
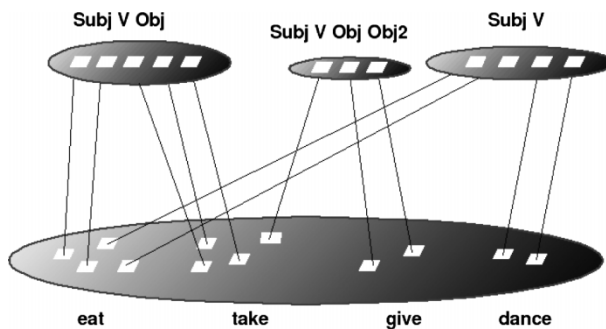

Figure 2. A portion of the lexicon showing three constructions. The usages of each verb are shown as white squares, and linked to the appropriate constructions. For example, *eat* has been seen twice in a transitive construction and twice in an intransitive construction.

and Object, and the properties of the roles that are typically associated with those positions (e.g., Fisher, 1996). Our model learns similar associations between the semantic properties of arguments, their syntactic positions, and the semantic primitives of verbs. Such associations lead to the formation of a conception of the general semantic role for each argument position in a syntactic pattern, similar to the conceptions that humans have of these roles.

More specifically, by knowing the general semantic properties of an event (such as causal action or directed motion), the number of the arguments that participate in the event, and the syntactic pattern that the event is described in (such as the transitive pattern), we want to induce probabilities for both lexical and event-based properties that each of the arguments in the event can take. For example, by knowing that 'blick' refers to some novel causal action where one argument turns another argument around, and by hearing the utterance *duppy is blicking the luff*, we want to guess both the lexical properties of *duppy* and *luff* (i.e., what they are) and their event-based properties (i.e., what properties they exhibit in that particular event). To formalise our guess, we define a semantic profile as two probability distributions, one over all the lexical properties that a word in a language can have, and one over all the event-based properties that an argument can take on in a scene. We give a formal definition of a semantic profile in Section 4.2.

4. DETAILS OF THE ROLE LEARNING MODEL

Our model incrementally learns from the processing of each verb usage that is input to it. The system represents each verb usage in the form of an *argument structure frame* (or simply *frame*), which is a set of form and meaning features. Figure 3 shows the extracted frame from the verb usage shown in Figure 1. Each frame records the head verb, the number of arguments, the semantic primitives of the verb, and the lexical and event-based properties of each of the arguments. The frame also records the syntactic pattern of the utterance; currently, this syntactic pattern encodes

Head verb	*eat*
Number of arguments	2
Syntactic pattern	arg1 *verb* arg2
Semantic primitives of verb	{act, consume}
Lexical properties of argument 1	{woman, adult female, female, person, individual, ...}
Event-based properties of argument 1	{volitional, affecting, animate, independently exist, ...}
Lexical properties of argument 2	{meal, repast, nutriment, nourishment, sustenance, ...}
Event-based properties of argument 2	{non-independently exist, affected, change, ...}

Figure 3. The argument structure frame extracted from the verb usage *Sara ate lunch* in Figure 1.

the relative positions of the verb, each of its arguments, and the function words (e.g., prepositions) used in the utterance. The extracted frame is stored in the lexical entry of the head verb.

Each frame is presented to an unsupervised Bayesian clustering process, which groups it together with an existing group of frames that probabilistically has the most similar properties to the new frame. Our Bayesian clustering approach is detailed in Section 4.1. The acquired constructions are used to predict a semantic profile for each role; this prediction process is described in Section 4.2.

4.1 Learning as Bayesian clustering

Each extracted frame is input to an incremental Bayesian clustering process that groups the new frame together with an existing group of frames – a construction – that probabilistically has the most similar properties to it. If none of the existing constructions has sufficiently high probability for the new frame, then a new construction is created, containing only that frame. We use an extended version of Alishahi and Stevenson's (2008) probabilistic model, which is itself an adaptation of a Bayesian model of human categorisation proposed by Anderson (1991).[2] 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.

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

$$\textbf{BestConstruction}(F) = \underset{k}{\arg\max} \, P(k|F) \qquad (1)$$

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:

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

where $P(k)$ is the prior probability of construction k, and $P(F|k)$ is the posterior probability of frame F for a construction k.

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.

$$P(k) = \frac{n_k}{n+1} \qquad (3)$$

---

[2] Anderson's (1991) model is shown to be directly equivalent to the Dirichlet process mixture model (Antoniak, 1974; Neal, 2000) if the parameters of the two models are set appropriately.

where n is the total number of observed frames, and n_k is the number of frames participating in construction k. The normalising factor is set to $n + 1$ rather than n, to accommodate the probability mass of a potential new construction ($k = 0$). The prior probability of the new construction is calculated as

$$P(0) = \frac{1}{n + 1} \tag{4}$$

Thus, the prior probability of an existing construction is proportional to its size (Equation 3), following the intuition that it is more probable for a newly observed frame to come from a more entrenched construction. Moreover, the prior probability of a new construction is inversely proportional to the number of observed frames overall (Equation 4), capturing the intuition that the more exposure the child has to the language, the less likely it is that a brand new construction will be observed. This approach is referred to as the Chinese Restaurant process (Antoniak, 1974; Pitman, 2006), and has been used in other Bayesian models of cognitive processes such as categorisation (Anderson, 1991), and learning relational concepts in non-linguistic domains (Kemp et al., 2006).

The conditional probability of a frame F given that it belongs to construction k is expressed in terms of the individual likelihood probabilities of its features, which we assume are independent, thus yielding a simple product of feature probabilities:

$$P(F|k) = \prod_{i \in FrameFeatures} P_i(j|k) \tag{5}$$

where j is the value of the ith feature of F, and $P_i(j|k)$ is the probability of displaying value j on feature i within construction k. This probability is estimated using a smoothed version of the following maximum likelihood formulation, reflecting the emphasis on usage statistics in child language acquisition:

$$P_i(j|k) = \frac{\mathbf{count}_i^k(j)}{n_k} \tag{6}$$

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 features with a single value such as the number of arguments and the syntactic pattern, $\mathbf{count}_i^k(j)$ is calculated by simply counting those members of construction k whose value for feature i exactly matches j. However, we do not treat certain features such as the lexical or event-based properties of the arguments as individual binary features: because we calculate the likelihood of a frame as a product of the likelihood probabilities of its features, and because the number of the semantic properties of the arguments is potentially large, treating them as individual

features will result in them dominating the single-valued features such as the syntactic pattern or the number of the arguments. Instead, we represent these features as a set, and use a simple set similarity score to compare them with each other. We describe the details of the estimation of the likelihood probabilities for each of the features in Appendix A.

4.2 Representation of semantic profiles

In Section 3.4, we defined a semantic profile as two probability distributions, one over all the lexical properties that a word in a language can have, and one over all the event-based properties that an argument can exhibit in a scene. These probability distributions are useful in many language tasks, such as sentence processing and word learning (we discuss using the profiles in these tasks in Sections 6.2.1 and 6.2.2). The semantic profile for a grammatical position is predicted based on some combination of the syntactic pattern of an utterance and its number of arguments, and the semantic properties of the predicate. For example, if the hearer observes an unknown noun in the Subject position of a transitive sentence, she would want to guess the lexical and event-based properties that the argument can take on. In addition, knowing the properties of the event results in more accurate profiles. For example, the Object of a directed motion verb might have different properties than that of a causal action verb, even if both usages have the same syntactic pattern.

Unlike learning a frame, where we compared the semantic properties as sets of values, in forming a semantic profile we want to know how likely it is for an individual property to be assigned to a grammatical position in a construction. Therefore, predicting a semantic profile requires looking at the probability of all the individual properties, j_p, for an unobserved feature i, based on the observed features in a frame F and the learned constructions. In our previous work (Alishahi & Stevenson, 2008), we showed how Bayesian prediction can be used to simulate various types of language use. Here we use the same mechanism to estimate the property probabilities in a semantic profile:

$$
\begin{aligned}
P_i(j_p|F) \quad &= \quad \sum_k P_i(j_p|k)P(k|F) \\
&\propto \quad \sum_k P_i(j_p|k)P(k)P(F|k)
\end{aligned}
\tag{7}
$$

Here, k ranges over all constructions, and the probabilities $P(k)$ and $P(F|k)$ are determined as in our learning module.[3] To estimate the likelihood $P_i(j_p|k)$, we use a modified version of Equation (6) in which $\mathbf{count}_i^k(j_p)$ is the number of frames in construction k that include property j_p in the ith feature of the frame, which is a set of properties:

---

[3] Because our goal is to look at the general and not verb-specific profiles, we do not include the head verb of a usage as a feature in F.

$$p_i(j_p|k) = \frac{\text{count}_i^k(j_p)}{n_k} \tag{8}$$

A vector of the resulting probabilities $P_i(j_p|F)$ over all j_p forms the semantic profile of that argument. Note that the Bayesian formulation used above is not the absolute probability of an argument having a property j_p. Instead, $P(j_p|F)$ reflects the *relative* probability of the property j_p for a particular argument in the frame F, compared with all the possible properties that an argument can have. That is, $\Sigma_{j_p} P(j_p|F) = 1$. The absolute probabilities can be easily calculated by multiplying the relative probabilities by the total number of distinct values that a feature i can take. Using relative probabilities instead of the absolute ones allows us to represent a semantic profile as a probability distribution, and integrate it within our Bayesian model. That way, we can easily compare two different candidate profiles for a particular argument position, a method we use in Section 6.2 in order to resolve ambiguity in sentence comprehension and language acquisition.

A semantic profile contains all the properties ever observed in an argument position. As learning proceeds, a profile may include a large number of properties with very low probability. In order to display the profiles we obtain in our results section below, we create truncated profiles which list the properties with the highest probabilities, in decreasing order of probability value. To avoid an arbitrary threshold, we cut the ordered list of properties at the widest gap between two consecutive probabilities across the entire list.

5. EXPERIMENTAL SET-UP

Through a number of computational experiments, we demonstrate that our model can learn intuitive profiles for general semantic roles, and can use the learned profiles in language processing tasks. The creation of the input corpora for our simulations is described in Section 5.1. Section 5.2 describes how noise is added to the input corpora.

5.1 The input corpora

As described in Section 3.1, we want our model to learn its general knowledge of semantic roles from the usages of a small set of the most frequent verbs in child-directed data. For this purpose, we used a portion of the CHILDES database (MacWhinney, 1995). We extracted the 20 most frequent verbs in mother's speech to each of Adam, Eve, and Sarah, and selected 13 verbs from those in common across these three lists. We constructed an input-generation lexicon based on these 13 verbs, including their total frequency among the three children. We also assigned each verb a set of possible argument structure frames and associated frequencies, which were manually compiled by
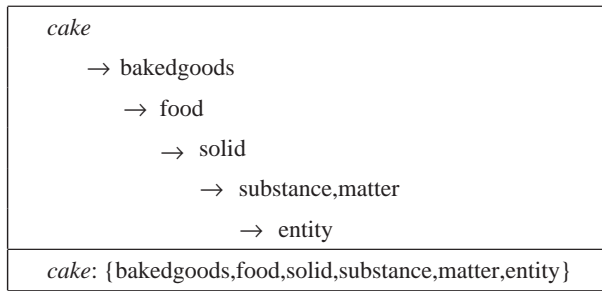
```
cake
    → bakedgoods
        → food
            → solid
                → substance,matter
                    → entity
```

cake: {bakedgoods,food,solid,substance,matter,entity}

Figure 4.   Semantic properties for *cake* from WordNet.

examination of 100 randomly sampled uses of a verb from all conversations of the same three children. Finally, from the sample verb usages, we extracted a list of head nouns and prepositions that appeared in each argument position of each frame, and added these to the lexicon.

For each noun in the lexicon, we extracted a set of lexical properties from WordNet (Miller, 1990), as follows. We hand-picked the intended sense of the word, extracted all the hypernyms (ancestors) for that sense, and added all the words in the hypernym synsets to the list of the semantic properties. Figure 4 shows an example of the hypernyms for *cake*, and its resulting set of semantic properties.[4] These properties are later used to induce a semantic profile for each argument position in a construction. Note that the behaviour of our model does not depend on the actual properties extracted from WordNet, and these properties can later be replaced by another resource deemed more appropriate in the context of child language acquisition. However, the WordNet-extracted properties have the desirable property that some of them are more general than others, and are shared by a number of nouns in our lexicon.

For each verb frame, we manually compiled a set of semantic primitives for the event as well as a set of event-based properties for each of the arguments. We chose these properties from what we assumed to be known to the child at the stage of learning being modelled, drawing on linguistic proposals concerning fundamental event properties (e.g., Jackendoff, 1990; Dowty, 1991; Rappaport Hovav & Levin, 1998). The verb primitives and event-based argument properties describe the coarse-level semantics of an event, as well as the finer-grained (verb-based) meaning distinctions among our experimental verbs (e.g., *cause* for a wide range of events, as opposed to *playfully* for a 'playing' event). Examples of these properties can be seen in the frames and profiles reported in the results section.

---

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

It is important to note that the input-generation lexicon is not used by the model in language learning and use, but only in producing the experimental corpora. For each simulation in our set of experiments, an input corpus of verb usages is automatically randomly generated using the frequencies in the input-generation lexicon to determine the probability of selecting a particular verb and argument structure. Arguments of verbs are also probabilistically generated based on the word usage information for the selected frame of the verb. The resulting corpora are further processed to simulate noise and incomplete data, as described next.

5.2 Adding noise

We assume that, at the point of learning modelled by our system, the child is able to recognise the syntactic pattern of the utterance and the semantic properties of the event. However, in reality, the input to children is often noisy or incomplete. For example, the child might mishear the utterance, or might not be able to extract the correct syntactic pattern from it. Similarly for the semantic information, the child might not be able to recognise the semantic properties that the participants have in a particular event, or the semantic properties of the event itself. More problematic for the process of learning, children might sometimes misinterpret the perceived scene or utterance by filling the gaps based on their own (imperfect) linguistic knowledge, which also leads to noisy data. We simulate two types of noise in our input corpora: incomplete data (where some pieces of information are missing) and misinterpreted data (where the child replaces the missing data with her own inference). Other types of noise, such as ungrammatical or incomplete sentences, are not currently modelled in the input.

During the input-generation process, two generated input items out of every five have one of their features randomly removed. One of these modified input items is used to simulate incomplete data. This modified input pair is left as is – i.e., with one feature missing – in the generated corpus. The other modified input pair is used to simulate noise. During a simulation, the missing feature of this input pair is replaced with the most probable value predicted for it at that point in learning; the completed input pair is then used in the learning process. This corresponds to a child using her own inferred knowledge to fill in information missing from an observed scene/utterance pair. The resulting input pair is noisy, especially in the initial stages of learning.[5]

---

[5] In predicting a syntactic pattern for an incomplete frame, it is possible that the pattern will have place-holders for more arguments than are present in the scene representation. In cases such as these, when creating the corresponding utterance, the excess argument slots in the predicted pattern are simply left blank.

6. EXPERIMENTAL RESULTS

In this section, we report experimental results from a number of simulations of our model. The experiments are divided into two groups: In Section 6.1, we examine the process of learning semantic profiles for general roles from usage data. In Section 6.1.1, we look at some of the semantic profiles that our model acquires for argument positions in various constructions, which are traditionally associated with general thematic roles such as Agent, Theme, and Destination. We show that the generated profiles match the intuitive expectations for these roles. In Section 6.1.2 we show that our model can learn multiple semantic roles for a single grammatical position such as Subject or Object, depending on the semantic primitives of the verb that participates in the syntactic pattern. In Section 6.1.3, we demonstrate the process of role generalisation, where the model starts from learning verb-specific profiles for a particular argument position in a construction, and later moves to a more general profile for the same position as a result of exposure to more input.

The second group of experiments, reported in Section 6.2, focuses on how the acquired profiles can be used in various language tasks. In Section 6.2.1, we examine how the semantic profiles can be used in sentence comprehension and ambiguity resolution. In Section 6.2.2, we look at how these profiles are used in verb learning, especially in cases where the context is ambiguous and the meaning of a novel verb cannot be inferred only from observation.

In the following sections, we represent each semantic profile as an ordered list of properties and their probability values in that profile, as calculated by Equation (8). In most experiments, we test our model on 200 input items, since receiving additional input after 200 items ceases to make any substantial difference in the output. We discuss this issue in more detail in Section 6.1.3.

6.1 The acquisition of profiles from input

6.1.1 Formation of semantic profiles for roles

Psycholinguistic experiments have shown that humans have a conception of general semantic roles. Specifically, Kako (2006) shows that human subjects who hear a transitive sentence assign more proto-agent properties (such as *cause* and *motion*) to the Subject of the sentence, and more protopatient properties (such as *changed* and *created*) to the Object of the sentence. Even in the absence of a known verb and known arguments (e.g., *The grack mecked the zarg*), the results are the same. Our model shows the same behaviour; that is, it associates a relatively high probability to the relevant properties in the semantic profile it creates for each grammatical position in a familiar construction.

We train our model on 200 randomly generated input items, and then present it with a test input item containing a novel verb *gorp* appearing in a familiar construction, with unknown nouns appearing as its arguments. The test input can be represented as a partial frame, where the properties of the verb and arguments are left out. As an example, a test pair for a novel verb appearing in a transitive construction looks as follows:

<div align="center"><i>x gorp y</i></div>

number of arguments	2
syntactic pattern	arg1 *verb* arg2

We then produce a semantic profile for each of the unknown arguments to reveal what the model has learned about the likely semantic properties for that position in the corresponding construction. We average the obtained probabilities over five simulations on different random input corpora. In each of the reported profiles below, a semantic property with a high relative probability (compared with other properties in the list) is more likely to be associated with the argument.

Our model learns semantic profiles for argument positions in various constructions. Figure 5 shows the top portion of the predicted semantic profiles for the arguments in the Subject and Object positions of a transitive construction (corresponding to *x* and *y* in the *gorp* test input above). The emerging semantic profile for each argument position demonstrates the intuitive properties that the argument receiving that role should possess. For example, the lexical portion of the semantic profile for an argument that appears in the Subject position in a transitive construction (the left box of Figure 5) demonstrates the properties of an animate entity, most likely a human. In contrast, the lexical portion of the semantic profile for an argument in the Object position (the right box of Figure 5) most likely corresponds to a physical entity. Moreover, the event-based portion of the profile for the Subject position shows a tendency towards more Agent-like properties such as 'independently existing', 'sentience', 'volition', etc., as opposed to more Theme-like properties associated with the Object position, such as 'undergoing change'. In both profiles, the general properties receive a higher probability than the verb-specific properties such as 'making' and 'eating' for Agent and 'falling' and 'being made' for Theme.

Similarly, Figure 6 demonstrates the predicted semantic profiles for the argument positions in a directed motion construction, traditionally considered as Agent and Destination (as in *Joe*_{Agent} *went to school*_{Destination}). Again, the profiles show intuitively plausible properties for each role. Interestingly, the lexical portion of the Agent profile shown in Figure 6 is

PARTIAL FRAME: TRANSITIVE	
Number of arguments	2
Syntactic pattern	arg1 *verb* arg2

ARGUMENT 1 (AGENT)		ARGUMENT 2 (THEME)	
Probability	Event-based property	Probability	Event-based property
0.048	independently exist	0.086	state
0.048	sentient	0.031	independently exist
0.035	animate	0.031	change
0.035	change	0.031	change possession
0.035	affected		
0.035	change emotional		
0.035	becoming		
0.013	volitional		
0.013	possessing		
0.013	getting		

Probability	Lexical property	Probability	Lexical property
0.054	entity	0.056	entity
0.040	object	0.037	object
0.040	physical object	0.037	physical object
0.026	being	0.023	unit
0.026	organism	0.023	artifact
0.026	living thing	0.023	artefact
0.026	animate thing	0.023	whole
0.015	person	0.023	whole thing
0.015	individual	0.018	abstraction
0.015	someone	0.014	being
0.015	somebody	0.014	organism
0.015	mortal	0.014	living thing
0.015	human	0.014	animate thing
0.015	soul	0.014	person
0.015	causal agent	0.014	individual
0.015	cause	0.014	someone
0.015	causal agency	0.014	somebody
0.014	unit	0.014	mortal
0.014	artifact	0.014	human
⋮	⋮	⋮	⋮

Figure 5. Semantic profiles of argument positions Agent and Theme in a transitive construction.

very similar to the lexical portion of the Agent profile shown in Figure 5, since the agents of both constructions are usually either humans or animals. However, the event-based portion of the two Agent profiles are different after the first few properties. For the Object position, the predicted profiles

PARTIAL FRAME: DIRECTED MOTION	
Number of arguments	2
Syntactic pattern	arg1 *verb to* arg2

ARGUMENT 1 (AGENT)	
Probability	Event-based property
0.127	independently exist
0.124	sentient
0.124	volitional
0.122	animate
0.121	change location
0.119	motion
0.119	direction
0.093	going
0.026	coming

Probability	Lexical property
0.056	entity
0.056	object
0.056	physical object
0.055	being
0.055	organism
0.055	living thing
0.055	animate thing
0.055	person
0.055	individual
0.055	someone
0.055	somebody
0.055	mortal
0.055	human
0.055	soul
0.055	causal agent
0.055	cause
0.055	causal agency
0.014	female
0.014	female person

ARGUMENT 2 (DESTINATION)	
Probability	Event-based property
0.340	location
0.324	destination
0.321	path
0.015	source

Probability	Lexical property
0.200	location
0.127	entity
0.087	relation
0.087	spacial relationship
0.087	preposition
0.038	part
0.037	region
0.021	abstraction
0.015	abode
0.015	residence
0.015	address
0.015	geographic point
0.015	geographical point
0.015	point
0.012	attribute
0.012	opening
0.012	gap
0.012	space
0.012	amorphous shape
0.012	shape
⋮	⋮

Figure 6. Semantic profiles of argument positions Agent and Destination in a directed motion construction.

for these two events are completely different: whereas the Theme profile in Figure 5 demonstrates the properties of a physical object, the profile in Figure 6 demonstrates the properties of a location.

6.1.2 Multiple possible roles for a position

One grammatical position in a syntactic pattern can be assigned to different semantic roles in different usages. For example, two usages *he saw her* and *he got worse* both have the same syntactic pattern ('arg1 *verb* arg2'), but the Subject and Object of the first usage are usually considered as Experiencer and Stimulus, whereas the Subject and Object of the second usage are considered as Theme and State. It is vital for a model of semantic role learning to be able to distinguish the arguments of different types of verbs when those arguments occur in the same syntactic position.

To test this, we examine the semantic profiles for the Subject and Object positions in two transitive usages of novel verbs, where the only difference between the two usages is the semantic primitives of the event. We use the primitives associated with the verbs *see* and *get* in the above examples. Figure 7 shows the partial frames for these usages, and the predicted semantic profiles for the Subject and Object positions of each usage.

This experiment is crucial in showing that the model does not simply associate a single semantic profile with a particular argument position. The model forms a complex association among a syntactic pattern, an argument position, and the semantic primitives of the verb, allowing it to make a distinction between different roles assigned to the same position in the same syntactic pattern.

6.1.3 Verb-based vs. general semantic profiles

We have shown that our model can learn general conceptions of semantic roles. However, because the model learns these general profiles from instances of verb usage, we expect these profiles to go through a gradual generalisation process, where they initially reflect the properties of specific verb arguments, and become more general over time. We tracked this generalisation process for the acquired semantic profiles. For example, Figure 8 shows the semantic profile for the argument in the Object position in a transitive usage. The left box shows the semantic profile right after the first transitive usage. In this particular simulation, the first transitive verb in the corpus is *eat*, and its second argument in that usage is *sandwich*. The semantic profile thus reflects the properties of a sandwich, and not the general properties of that argument position. The profile becomes more general after processing 50 input items, shown in the right box of Figure 8. Since we do not include any semantic primitives for the main event in the partial frame, the profile also reflects properties from usages such as *she went home* or *it came out*, which have the

PARTIAL FRAME: PERCEPTION	
Number of arguments	2
Syn. pattern	arg1 *verb* arg2
Verb primitives	{perceive}

PARTIAL FRAME: CHANGE OF STATE	
Number of arguments	2
Syn. pattern	arg1 *verb* arg2
Verb primitives	{change state}

ARGUMENT 1 (EXPERIENCER)	
Probability	Event-based property
0.201	independently exist
0.188	sentient
0.188	animate
0.181	visual
0.174	seeing

Probability	Lexical property
0.059	entity
0.059	object
0.059	physical object
0.055	being
0.055	organism
0.055	living thing
0.055	animate thing
0.055	person
0.055	individual

ARGUMENT 1 (THEME)	
Probability	Event-based property
0.162	independently exist
0.140	sentient
0.113	animate
0.083	change
0.083	affected
0.083	change emotionally

Probability	Lexical property
0.065	entity
0.065	object
0.065	physical object
0.050	being
0.050	organism
0.050	living thing
0.050	animate thing
0.049	person
0.049	individual

ARGUMENT 2 (STIMULUS)	
Probability	Event-based property
0.445	independently exist
0.444	perceivable

Probability	Lexical property
0.058	entity
0.054	object
0.054	physical object
0.028	being
0.028	organism
0.028	living thing
0.028	animate thing
0.028	unit
0.028	artifact

ARGUMENT 2 (STATE)	
Probability	Event-based property
0.383	state

Probability	Lexical property
0.131	abstraction
0.113	attribute
0.102	bad
0.087	quality
0.074	badness
0.037	entity
0.035	property
0.028	negative
0.024	relation

Figure 7. Semantic profiles of syntactic positions Subject and Object in two different transitive usages: Experiencer and Stimulus in a perception event are presented on the left side, and Theme and State in a change of state event are presented on the right side.

PARTIAL FRAME: TRANSITIVE	
Number of arguments	2
Syntactic pattern	arg1 *verb* arg2

ARGUMENT 2 AFTER 5 ITEMS	
Probability	Event-based property
0.124	independently exist
0.124	change
0.124	affected
0.124	stationary
0.124	change matter
0.124	eaten
0.124	vanished

Probability	Lexical property
0.062	entity
0.057	substance
0.057	matter
0.056	food
0.056	nutrient
0.056	nutriment
0.056	nourishment
0.056	nutrition
0.056	sustenance
0.056	aliment
0.056	alimentation
0.056	victuals
0.055	group
0.055	grouping
0.055	dish
0.055	snack food

ARGUMENT 2 AFTER 50 ITEMS	
Probability	Event-based property
0.164	location
0.164	destination
0.137	path
0.106	change
0.093	independently exist
0.066	change possession
0.040	affected
0.040	stationary
0.040	change matter
⋮	⋮

Probability	Lexical property
0.122	entity
0.084	location
0.044	object
0.044	physical object
0.044	unit
0.044	artifact
0.044	artefact
0.044	whole
0.044	whole thing
0.035	relation
0.028	part
0.028	spacial relationship
0.028	preposition
0.028	region
0.022	instrumentality
0.022	instrumentation
0.022	substance
0.022	matter
0.022	implement
0.015	food
0.015	nutrient
⋮	⋮

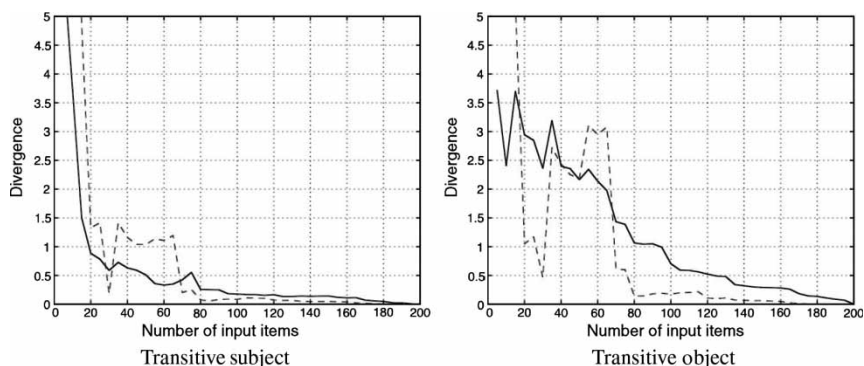**Figure 8.**   The evolution of the Transitive Object role.

Figure 9. Learning curves for semantic profiles. The x-axis is time (number of inputs), and the y-axis is divergence from the profile that the model eventually converges to. Solid and dashed lines show the divergence between the lexical and event-based portions of the profiles, respectively.

same syntactic pattern and number of arguments. (As shown in Section 6.1.2, adding the appropriate semantic primitives that indicate the semantic type of the main verb results in more specific semantic profiles for a particular grammatical position.) Nevertheless, the general transitive profiles in Figure 8 illustrate the process of evolving from a verb-specific profile (e.g., 'eaten' and 'vanished' in the earlier profile) to a more general one, as a result of processing more input.

To observe the trend of moving from a more specific to a more general semantic profile for each argument position, we need to compare the semantic profile for an argument position at a given point in learning, and the profile for that position that the model eventually converges to at the end of each simulation. More technically, we need to measure the divergence between the two probability distributions represented by these semantic profiles. We use a standard divergence measure, Relative Entropy, for this purpose.[6] This measure shows how different the two semantic profiles are, with a value of zero indicating two identical profiles. Figure 9 shows the profile divergence for Subject and Object positions of a transitive construction after every 5 input items over a total of 200 items, averaged over 5 simulations. The divergence between the lexical portion of the profiles is shown by solid lines, and the divergence between the event-based portion of the profiles is shown by dashed lines. Figure 9 shows that the profile for the Subject position (i.e., the Agent) is learned faster than the profile for the Object position (i.e., the Theme), which is a much less constrained role. The

---

[6] $RelativeEntropy\ (P \| Q)\ =\ \Sigma_i P(i) \log \frac{P(i)}{Q(i)}$, where P and Q are probability distributions.

curves show that the model stabilises on the final profiles between 150 and 200 input items, when receiving more inputs ceases to make any substantial difference in the profiles.

6.2 Using the acquired profiles in language tasks

6.2.1 Who's blicking? Using semantic profiles in comprehension

Semantic roles are helpful in on-line ambiguity resolution, by guiding adults to the interpretation that best matches the role expectations of a given verb for a particular position (e.g., Carlson & Tanenhaus, 1988; Trueswell et al., 1994). For example, hearing an animate noun as the direct object of a verb such as *give* prompts the hearer to interpret it as the Recipient of the event, whereas an inanimate noun heard as the direct object of the same verb leads the hearer to interpret it as the Theme of the event. Nation et al. (2003) have shown that young children also draw on the expectations of the arguments imposed by the main verb in on-line argument interpretations.

Children's use of the associations between argument properties and syntactic positions is most evident in the work of Fisher and colleagues (e.g., Fisher, 1996). For example, in Fisher's (1996) Experiment 1, 3- and 5-year-olds were taught novel transitive and intransitive verbs for unfamiliar Agent–Patient events. For example, one girl rolls another on a wheeled dolly by pulling with a crow bar, and the experimenter says *Look, she's blicking her over there*! or *Look, she's blicking over there*! The identities of the Subject and Object were obscured by using ambiguous pronouns, yielding sentences which differed only in their syntactic pattern. Children's interpretation of a novel verb in its sentence context was assessed by asking them to choose the participant in each event that appears in the Subject position (*Which one is blicking her over there*? vs. *Which one is blicking over there*?). The children interpreted the verbs differently depending on the sentence structure, though neither sentence explicitly identified one participant in the event as the Subject. Both 3- and 5-year-old children picked the causal agent (as opposed to the other participant in the event) as the Subject of a transitive sentence almost all the time, while they picked the causal agent as the Subject of an intransitive sentence only about half the time (i.e., at chance levels).

Here we show that our model, in using its acquired semantic profiles to predict the best interpretation of an ambiguous input, exhibits probabilistic preferences that are compatible with children's behaviour in Fisher's experiment. We set up the computational experiment as one where we compare the probability of the two interpretations of the same scene, based on the properties associated with the Subject position. For example, in the case of a transitive usage, as a response to *Which one is blicking her over there*?, the child might point to the Agent of the event. This behaviour shows that the child has associated the first argument (i.e., Subject) with Agent-like

properties. Alternatively, by pointing to the Theme of the event, the child shows that she has associated the Subject of the transitive sentence with Theme-like properties. Each of these interpretations can be represented as one frame. The two interpretations for the transitive context are shown in Figure 10.

We give each frame F to the model, and have it calculate:

$$\textbf{match\_score}(F) \;=\; \max_{k} \log(P(k)P(F|k)) \tag{9}$$

where $P(k)$ and $P(F|k)$ are from the learning model (Equations (3) and Equation (5), respectively). The measure **match_score** corresponds to how well the given input matches the model's acquired knowledge of language: for a 'preferred' usage, there is a high chance that the model has learned a well-entrenched construction (i.e., with high prior probability $P(k)$) that has a high compatibility with the current frame F (i.e., the posterior probability $P(F|k)$). Therefore, having a high **match_score** for a frame shows that the model considers that frame as compatible with its previously acquired knowledge. When comparing two alternative interpretations, a higher **match_score** for the frame corresponding to one of the interpretations indicates that the model has a preference for that interpretation. We show this preference as

$$\textbf{pref}(A, B) \;=\; \textbf{match\_score}\,(F_A) - \textbf{match\_score}\,(F_B) \tag{10}$$

where A and B are two alternative interpretations, and F_A and F_B are the frames representing those interpretations. A positive value for **pref**(A,B) shows a preference for interpretation A. In the context of the *blicking*

SUBJECT AS AGENT

Number of arguments	2
Syntactic pattern	arg1*verb* arg2
Verb primitives	{cause, move}
arg1 lexical properties	{woman, adult female, female, person, individual, ...}
arg1 event-based properties	{volitional, affecting, animate, ind. exist, cause, cause movement}
arg2 lexical properties	{woman, adult female, female, person, individual, ...}
arg2 event-based properties	{ind. exist, change location, affected, motion, manner}

SUBJECT AS THEME

Number of arguments	2
Syntactic pattern	arg1 *verb* arg2
Verb primitives	{cause, move}
arg1 lexical properties	{woman, adult female, female, person, individual, ...}
arg1 event-based properties	{ind. exist, change location, affected, motion, manner}
arg2 lexical properties	{woman, adult female, female, person, individual, ...}
arg2 event-based properties	{volitional, affecting, animate, ind. exist, cause, cause movement}

Figure 10. Transitive condition: *She blick her.*

SUBJECT AS AGENT

Number of arguments	1
Syntactic pattern	arg1 *verb*
Verb primitives	{cause, move}
arg1 lexical properties	{woman, adult female, female, person, individual, ...}
arg1 event-based properties	{volitional, affecting, animate, ind. exist, cause, cause movement}

SUBJECT AS THEME

Number of arguments	1
Syntactic pattern	arg1 *verb*
Verb primitives	{cause, move}
arg1 lexical properties	{woman, adult female, female, person, individual, ...}
arg1 event-based properties	{ind. exist, change location, affected, motion, manner}

Figure 11.   Intransitive condition: *She blick*.

experiment explained above, having a positive value for **pref**(Agent, Theme) means that the model has recognised the first argument *she* as the Agent, while a negative value means that the model has interpreted *she* as the Theme. We also compared analogous frames using an intransitive utterance instead of a transitive one to describe the scene. These frames are shown in Figure 11.

Figure 12 shows the results after processing 10 and 100 input usages, averaged over 10 simulations. As noted above, Fisher (1996) finds that both younger and older children have a very strong tendency to interpret the
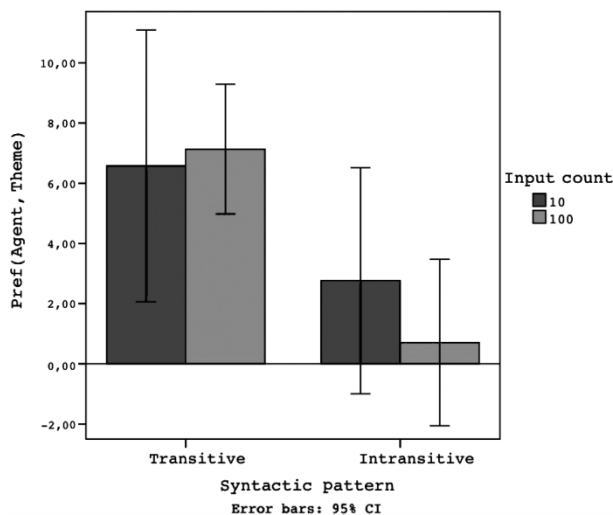

Figure 12.   The preference of the model towards interpreting the Subject as Agent (as opposed to Theme) for the Transitive/Intransitive conditions.

Subject of a transitive utterance as the Agent, but show no significant preference towards choosing the Agent or the Theme of the event as the Subject of an intransitive sentence. Our results mimic this trend. For the Transitive condition, there is a very strong (and similar) preference to interpret the Subject as the Agent after both 10 and 100 inputs. For the Intransitive condition, there is a much smaller and highly inconsistent preference to interpret the Subject as Agent very early, but no preference towards one interpretation over another by 100 inputs. As Fisher notes, children's behaviour in the face of intransitives is not incorrect since there are various intransitive uses; indeed, for our model, some of the verbs in our lexicon can have an intransitive usage where the Subject is the Agent (e.g., *she came* or *he ate*), whereas others map the Subject to the Theme of the event (e.g., *window broke*).

The same experiment in Fisher (1996) also includes a condition with sentence contexts which differ in their use of the preposition *to* or *from*. For example, a novel manner of delivery and of receipt is represented, and described as *She is pilking the ball over to her* or *She is pilking the ball away from her*. Again, the child is asked to select who is pilking the ball. If the child points to the Agent of the event, it means that she has associated the Subject of the utterance with Agent-like properties. Otherwise, she has associated the Subject with Recipient-like properties. Fisher finds a tendency for children, especially the 5-year-olds, to choose an Agent interpretation more often when the preposition is *to* than when it is *from* (with the latter showing essentially random choice).

Similarly to the previous simulations, we set up the experiment as one where we compare the probability of the two interpretations of the scene, one which associates the Subject of the sentence with the properties of an Agent, and one which associates it with the properties of the Recipient. The two frames that represent these interpretations for the sentence with *to* are shown in Figure 13. Analogous frames, using the preposition *from* are shown in Figure 14. (Note that the features for these frames are identical to those of the frames in Figure 13, except for the syntactic pattern feature.) We calculate **pref**(Agent, Recipient) for each condition separately. Again, having a positive value means that the model has recognised the first argument *she* as the Agent, while a negative value means that the model has recognised *she* as the Recipient.

Figure 15 shows the results after processing 10 and 100 input usages, averaged over 10 simulations. Similar to Fisher's results, our model does not show a strong, consistent preference nor a large difference between responses with *to* and *from* early on (at 10 inputs). However, as with children, with additional input, we do find a much stronger preference for interpreting the Subject as Agent when the preposition is *to*. Also compatible with Fisher's results, the preference for interpreting the Subject as Agent is much stronger

SUBJECT AS AGENT

Number of arguments	3
Syntactic pattern	arg1 *verb* arg2 *to* arg3
Verb primitives	{cause, possess}
arg1 lexical properties	{woman, adult female, female, person, individual, ...}
arg1 event-based properties	{volitional, sentient, animate, ind. exist, affecting, cause, giving, ...}
arg2 lexical properties	{toy, artifact, artefact, object, physical object, entity, whole, ...}
arg2 event-based properties	{affected, change, change possession, ind. exist}
arg3 lexical properties	{woman, adult female, female, person, individual, ...}
arg3 event-based properties	{animate, ind. exist, affected, possessing, given, beneficiary}

SUBJECT AS RECIPIENT

Number of arguments	3
Syntactic pattern	arg1 *verb* arg2 *to* arg3
Verb primitives	{cause, possess}
arg1 lexical properties	{woman, adult female, female, person, individual, ...}
arg1 event-based properties	{animate, ind. exist, affected, possessing, given, beneficiary}
arg2 lexical properties	{toy, artifact, artefact, object, physical object, entity, whole, ...}
arg2 event-based properties	{affected, change, change possession, ind. exist}
arg3 lexical properties	{woman, adult female, female, person, individual, ...}
arg3 event-based properties	{volitional, sentient, animate, ind. exist, affecting, cause, giving, ...}

Figure 13.   To condition: *She pilk doll to her.*

SUBJECT AS AGENT

Number of arguments	3
Syntactic pattern	arg1 *verb* arg2 *from* arg3
Verb primitives	{cause, possess}
arg1 lexical properties	{woman, adult female, female, person, individual, ...}
arg1 event-based properties	{volitional, sentient, animate, ind. exist, affecting, cause, giving, ...}
arg2 lexical properties	{toy, artifact, artefact, object, physical object, entity, whole, ...}
arg2 event-based properties	{affected, change, change possession, ind. exist }
arg3 lexical properties	{woman, adult female, female, person, individual, ...}
arg3 event-based properties	{animate, ind. exist, affected, possessing, given, beneficiary}

SUBJECT AS RECIPIENT

Number of arguments	3
Syntactic pattern	arg1 *verb* arg2 *from* arg3
Verb primitives	{cause, possess}
arg1 lexical properties	{woman, adult female, female, person, individual, ...}
arg1 event-based properties	{animate, ind. exist, affected, possessing, given, beneficiary}
arg2 lexical properties	{toy, artifact, artefact, object, physical object, entity, whole, ...}
arg2 event-based properties	{affected, change, change possession, ind. exist }
arg3 lexical properties	{woman, adult female, female, person, individual, ...}
arg3 event-based properties	{volitional, sentient, animate, ind. exist, affecting, cause, giving, ...}

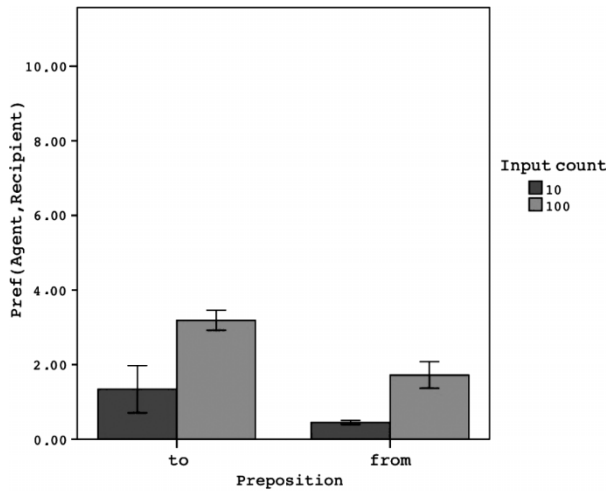**Figure 14.**   From condition: *She pilk doll from her.*

Figure 15.   The preference of the model towards interpreting the Subject as Agent (as opposed to Recipient) for the To/From conditions.

for *to* than *from* sentences at the later inputs. However, with *from*, we do find a small preference for the Agent interpretation (where Fisher's results indicate roughly chance behaviour); because of the low frequency of *from* in our input data, our model's behaviour is more influenced by other syntactic patterns than children's responses indicate. Note that, also in accord with Fisher's results, the preferences of the model for the Subject–Agent association in utterances with both *to* and *from* are lower than for the same association in a transitive utterance (see Figure 12); in our model, this is a direct consequence of the high frequency of the transitive construction in the data.

The results reported in this section confirm that the model is able to use its learned associations between semantic properties and argument positions to appropriately guide interpretation of an ambiguity. These results predict that very early on, children (like our model) would experience some difficulty in this type of task, when drawing on the knowledge of a less commonly observed construction.

6.2.2 Syntactic bootstrapping: using semantic profiles in verb learning

As mentioned earlier, we assume that our model already knows the meaning of a small set of verbs at the point it starts to learn about general semantic roles. Learning the correct meaning for both verbs and nouns has been suggested to be based on cross-situational observation (Pinker, 1989;

Fisher et al., 1994). However, it has been argued that for certain verbs or situations, using cross-situational learning is not enough. For example, Gleitman (1990) argues that paired verbs such as *chase* and *flee* can always be used to describe the same event. Therefore, just by watching a chasing scene, it is impossible for the child to decide whether the verb describing the scene means 'chase' or 'flee'. It is suggested that in such situations, children rely on the syntactic structure of the sentence that describes the event to induce the correct meaning of the verb, a hypothesis known as syntactic bootstrapping (Gleitman, 1990; Landau & Gleitman, 1985). For example, consider the case of *chase/flee*, where the child hears *Look! The fox is gorping the rabbit*. Based only on watching the event, and without using the word order information of the accompanying sentence, the child cannot decide whether *gorp* means chase or flee. However, because of her previous experience with other transitive verbs, the child knows that the argument that appears in the Subject position is more likely to be the one that causes the described event. In this example, that argument is *fox*, and therefore the event referred to by *gorp* is more likely to be initiated by the fox, hence *gorp* must mean chase.

Here we show that our model can use its acquired knowledge of language in guiding verb meaning in such ambiguous situations. We simulate the

CHASE INTERPRETATION

Number of arguments	2
Syntactic pattern	arg1 *verb* arg2
Verb primitives	{cause, move}
arg1 lexical properties	{entity, physical object, object, animate thing, living thing, ...}
arg1 event-based properties	{volitional, sentient, animate, ind. exist, cause, affecting, change location, motion, cause motion}
arg2 lexical properties	{entity, physical object, object, animate thing, living thing, ...}
arg2 event-based properties	{volitional, sentient, animate, ind. exist, affected, change location, motion}

FLEE INTERPRETATION

Number of arguments	2
Syntactic pattern	arg1 *verb* arg2
Verb primitives	{cause, move}
arg1 lexical properties	{entity, physical object, object, animate thing, living thing, ...}
arg1 event-based properties	{volitional, sentient, animate, ind. exist, affected, change location, motion}
arg2 lexical properties	{entity, physical object, object, animate thing, living thing, ...}
arg2 event-based properties	{volitional, sentient, animate, ind. exist, cause, affecting, change location, motion, cause motion}

Figure 16. Different interpretations of *gorp* in *fax gorp rabbit*, where *gorp* can mean either chase or flee.

chase/flee example as the problem of choosing the correct interpretation from the two possible ones upon hearing the sentence *fox gorp rabbit*, as shown in Figure 16. The frames that represent the two possible interpretations are identical in all features except for the event-based properties of the arguments. In the 'chase' interpretation of *gorp* in *fox gorp rabbit*, the properties 'cause, affecting, cause motion' are associated with the first argument, i.e., the fox, and the property 'affected' is associated with the second argument, the rabbit. These associations are reversed in the 'flee' interpretation of *gorp* in *fox gorp rabbit*, where the rabbit is the cause of the event.

As in the previous section, we compare two interpretations using the measure **pref**(Chase, Flee). The results are shown in Figure 17. As can be seen, the model shows a preference towards interpreting *gorp* as chase from early on (i.e., after processing 10 input items), and this preference becomes more pronounced after processing more input (i.e., after processing 100 input items). Note that the compared frames in this experiment (shown in Figure 16) have much more in common than the frames being compared with each other in the experiments of the previous section. Not only do the participants have similar lexical properties, they have similar event-based properties as well (e.g., both the fox and the rabbit are volitional, animate, in motion, changing location, etc.). What distinguishes the two interpretations from each other here is a very small set of event-based properties (i.e., 'cause, affecting, cause motion' for one argument, and 'affected' for the other). Therefore, we do not see as strong a preference in this case as we saw in the
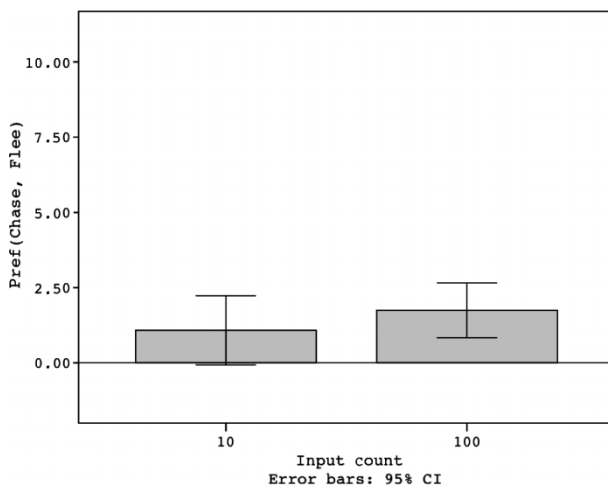

Figure 17. The preference of the model towards interpreting the main verb *gorp* as chase (as opposed to flee), in *fox gorp rabbit*.

simulations reported in the previous section. However, the model does exhibit preferences even based on such small differences in the properties that the arguments in each interpretation can have.

7. GENERAL DISCUSSION

The traditional view of semantic roles as innately specified and universal has been questioned by recent experimental studies. For example, McRae et al. (1997) suggest that semantic roles are verb-specific concepts; specifically, they view semantic roles as a property-based and event-specific representation of verb selectional restrictions. We present a similar account in Alishahi & Stevenson (2007a), an extension of the current model in which the semantic profiles yield verb-specific conceptualisations of the arguments associated with a syntactic position, and verb-based profiles are used in simulating human plausibility judgements on selectional preferences. However, the purely verb-based view of semantic roles cannot explain the experimental findings that demonstrate children's and adults' access to a general conception of semantic roles even in the absence of familiar verbs (Fisher, 1994, 1996, 2002; Kako, 2006). The goal of the current paper is to show that general semantic roles can be acquired from verb-specific ones through a process of categorisation and generalisation. Our Bayesian model of role learning is an elaboration of the usage-based view of role learning proposed by Lieven et al. (1997) and Tomasello (2000), and provides a detailed account of how verb-specific roles transform into general semantic profiles for arguments in various linguistic constructions.

Our model establishes a link between the syntactic pattern of an utterance and the semantic properties of the scene. Over time, common properties of the arguments that appear in the same syntactic position in a construction are generalised to form a profile that represents the role associated with that position. In our experimental results, we show that our model can form intuitive profiles for general roles such as Agent, Theme and Destination in various constructions, even when the verb appearing in the construction is not known to the model. The profiles predicted for the arguments of a novel verb are a result of generalising the actual arguments that appear in similar usages. We further show that the acquired profiles can guide the model in language comprehension when faced with ambiguous input, an effect that has been studied in human language processing (e.g., Carlson & Tanenhaus, 1988; Fisher, 1996; Trueswell et al., 1994).

The proposed model also makes a number of predictions about the process of learning general semantic roles and associating them with grammatical positions. First, the model learns a profile for a number of semantic roles such as Agent, Theme or Instrument, but it demonstrates a

different time-course for learning a stabilised profile for these roles. For example, as discussed in Section 6.1.3, the semantic profile that the model maintains for the Transitive Subject (or Agent) converges much faster than the one learned for the Transitive Object (or Theme). Also, at the beginning, the event-based profiles for both roles undergo much more change, whereas the lexical profiles change more smoothly. The model makes another interesting prediction with respect to using semantic profiles in word learning. In Section 6.2.2, we showed that the model can choose the correct interpretation of a novel verb using the semantic profiles that it has associated to the arguments of the construction. Figure 17 shows that, at the early stages of learning, this ability improves as the model receives more input, but at the later stages of learning the performance of the model does not change considerably upon receiving more input. Further empirical research is needed to investigate these predictions.

In the context of modelling child language acquisition, one of the strengths of the proposed model is its incremental nature. The model processes one input item at a time, and the constructions emerge and grow as a response to the input. However, this strategy results in a formation of constructions that is very sensitive to the order of the input items: the similarity between the early verb usages determines the number of the new constructions that are created. It also means that, at some point in learning, the acquired constructions may not be perfect. In other words, the model may not reach the globally optimum solution. It often happens that one (intuitive) construction is divided between two clusters, or one single cluster contains usages that we would recognise as two different constructions. The Bayesian formulation of the induction of semantic profiles allows for contribution from all of the constructions, and therefore reduces the unwanted effect of an imperfect construction structure on the behaviour of the model. Nevertheless, the structure of the constructions does affect the output, and in a few simulations it is not possible for the model to overcome the early mistakes it has made even when provided with more input (see Alishahi & Stevenson, 2008 for a detailed discussion). Although it is not cognitively plausible to revise all of the previously formed constructions after processing each new usage, it can be assumed that some constructions are revised and reorganised at certain intervals (see Parisien, 2008 for a variation of this idea). In fact, it has been suggested that children consolidate and reorganise what they learn during their sleep (Dumay et al., 2004; Orban et al., 2006; Walker, 2005). Sensitivity to the order of input can also be reduced by considering more than one partitioning of the observed frames into clusters at each point of time, and using repeated sampling (or Monte Carlo method) to approximate the complete Bayesian formulation (for example, see Sanborn, Griffiths, & Navarro, 2006 for a cognitively plausible sampling proposal).

As for the nature of semantic roles, our model follows Dowty's (1991) Proto-Role hypothesis and Kako's (2006) experimental findings that semantic roles are not single-valued and fixed labels, but a collection of semantic properties, in our case associated with various probabilities. Our semantic profiles (i.e., probability distributions over the argument properties) that represent each role emerge from a generalisation process over the semantic properties of the arguments that have been observed in the input. This hypothesis is based on the assumption that children are sensitive to certain lexical and event-based properties of the event and its arguments (Braine et al., 1990; Cohen & Oakes, 1993; Fisher, 1996; Gentner, 1978; Naigles & Kako, 1993). Note, however, that we do not make any claims regarding the actual properties that take part in a semantic profile. In fact, a number of cross-linguistic studies have suggested that the speakers of different languages show sensitivity to different properties of the events and their arguments (Allen, Özyürek, Kita, Brown, Furman, Ishizuka, & Fujii 2007; Choi & Bowerman, 1991). The flexibility of our usage-based account of role learning allows for language-specific profiles to be acquired as a response to the properties of the input. However, a shortcoming of the current model is that it assumes that children have access to the full set of properties of an argument from the very beginning. Much research has explored how children acquire a conceptual hierarchy and learn conceptual relations (e.g., Doumas et al., 2008; Xu & Tenenbaum, 2007). In the future, we plan to look at a more realistic setting for our experiments where the conceptual distinctions and relations are developed over time. Expanding on the vision expounded in Gentner and Namy (2006), it will be important to show that concept- and relation-learning can occur in an integrated fashion with the acquisition of verb argument structure in order to adequately model the full picture of child cognitive development.

A similar issue arises with respect to our assumptions about the knowledge of relevant properties of verbs. The acquisition of general semantic roles in our model is based on detecting and generalising the properties of a verb and its arguments in the input usage data. In other words, knowing (at least part of) the meaning of the verbs in the input usages is a prerequisite for the generalisation process and the formation of the semantic profiles. However, researchers have debated whether the meaning of verbs can be learned without using any knowledge of the syntactic structure of the language, and the mapping of the semantic roles to the syntactic positions in a sentence (Fisher, 2000; Landau & Gleitman, 1985; Gleitman, 1990). Specifically, 'paired' verbs that can describe various perspectives on the same event (e.g., *chase/flee, buy/sell, eat/feed*) pose a challenge for the cross-situational and observation-based account of verb learning. Drawing

on work in structural analogy for concept and relation learning, Fisher (2000) proposes that structural alignment can account for verb learning in a manner that requires little language-specific syntactic knowledge to give the child an initial sentence-structural guide for interpreting verbs. According to this hypothesis, using only partial or pre-syntactic descriptions of sentences, children obtain information about verb meaning essentially by structural analogy, aligning a sentence representation with a conceptual representation that shares its skeletal structural properties.

Our view on the acquisition of verb meaning and semantic roles is compatible with this account. The frame features that we extract from a verb usage are similar to the cues used for Fisher's structural alignment (the number of arguments, the word order in a sentence, and the semantic properties of the arguments and the event). In our view, verb learning is an iterative process: the model can learn a sketch of the general roles based only on a very small set of basic frequent verbs, whose meanings can be inferred from successive usage and observation of the corresponding event/activity. Moreover, we assume that knowing some of the general and specific properties of these verbs and their arguments is enough to boost the process of acquiring initial semantic profiles. These profiles can be used in the acquisition of the more ambiguous verbs, which in turn (after being learned) can contribute to the richness of the previously acquired profiles, or to the formation of new profiles. In this paper, we have investigated the first steps of using pre-syntactic knowledge of word meaning in forming a general conception of the semantic roles, and of applying these conceptions in guiding verb learning in ambiguous situations. Given the compatibility of the approach with structural alignment as assumed by Fisher (2000), future work will need to explore whether our framework can be combined, as suggested above, with elements of an analogical approach to concept and relation learning, to achieve the integration of language and concept learning as proposed in Gentner and Namy (2006). Our Bayesian model has the potential for drawing these pieces together in an iterative and integrated account of learning verb meaning and semantic roles.

In conclusion, we have shown that our Bayesian model of early verb learning, extended to include sets of semantic properties for arguments, can acquire associations between those properties and the syntactic positions of the arguments. These probabilistic associations enable the model to learn general conceptions of roles, based only on exposure to individual verb usages, and without requiring explicit labelling of the roles in the input. Because of the model's Bayesian formulation, the roles naturally metamorphose from verb-specific to highly general properties. The acquired role properties are a good intuitive match to the expected properties of various

roles, and are useful in guiding comprehension in the model to the most likely interpretation in the face of ambiguity. The learned roles can also be used to select the correct meaning of a novel verb in an ambiguous situation.

REFERENCES

Alishahi, A., & Stevenson, S. (2007a). A cognitive model for the representation and acquisition of verb selectional preferences. *Proceedings of the ACL Workshop on Cognitive Aspects of Computational Language Acquisition*, (pp. 41–48).

Alishahi, A., & Stevenson, S. (2007b). A computational usage-based model for learning general properties of semantic roles. In *Proceedings of the 2nd European Cognitive Science Conference*, (pp. 425–430).

Alishahi, A., & Stevenson, S. (2008). A computational model for early argument structure acquisition. *Cognitive Science*, *32*(5), 1–46.

Allen, J. (1997). Probabilistic constraints in acquisition. In A. Sorace, C. Heycock, & R. Shillcock (Eds.), *Proceedings of the GALA97 Conference on Language Acquisition, Edinburgh*, (pp. 300–305).

Allen, S., Özyürek, A., Kita, S., Brown, A., Furman, R., Ishizuka, T., & Fujii, M. (2007). Language-specific and universal influences in children's syntactic packaging of Manner and Path: A comparison of English, Japanese, and Turkish. *Cognition*, *102*(1), 16–48.

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

Antoniak, C. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics*, *2*(6), 1152–1174.

Baker, M. C. (1988). *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3–76). New York: Cambridge University Press.

Bowerman, M. (1990). Mapping thematic roles onto syntactic functions: Are children helped by innate linking rules? *Linguistics*, *28*, 1253–1289.

Braine, M., Brody, R., Fisch, S., Weisberger, M., & Blum, M. (1990). Can children use a verb without exposure to its argument structure? *Journal of Child Language*, *17*(2), 313–342.

Bresnan, J. (2001). *Lexical functional syntax*. Oxford, UK: Blackwell Publishers.

Carlson, G., & Tanenhaus, M. (1988). Thematic roles and language comprehension. In W. Wilkins (Ed.), *Syntax and semantics: Vol. 21. Thematic Relations* (pp. 263–291). San Diego, CA: Academic Press.

Chang, N. (2004). Putting meaning into grammar learning. In *Proceedings of the First Workshop on Psycho-computational Models of Human Language Acquisition* (pp. 17–24).

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, *41*(1–3), 83–121.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, the Netherlands: Foris Publications.

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

Cohen, L., & Oakes, L. (1993). How infants perceive a simple causal event. *Developmental Psychology*, *29*(3), 421–433.

Doumas, L. & Hummel, J. (2005). A symbolic-connectionist model of relation discovery. In *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 606–611).

Doumas, L., Hummel, J., & Sandhofer, C. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review*, *115*(1), 1–43.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547–619.

Dumay, N., Gaskell, G., & Feng, X. (2004). A day in the life of a spoken word. In *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 339–344). Mahwah, NJ: Lawrence Erlbaum Associated, Inc.

Falk, Y. N. (2001). *Lexical-functional grammar: An introduction to parallel constraint-based syntax*. CSLI Publications.

Fazly, A., Alishahi, A., & Stevenson, S. (2008). A probabilistic incremental model of word learning in the presence of referential uncertainty. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society.*

Fillmore, C. (1968). The case for case. In E. Back & R.J. Harms (Eds.), *Universals in linguistic theory* (pp. 1–88). London: Holt, Rinehart and Winston.

Fisher, C. (1994). Structure and meaning in the verb lexicon: Input for a syntax-aided verb learning procedure. *Language and Cognitive Processes*, *9*(4), 473–517.

Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, *31*(1), 41–81.

Fisher, C. (2000). From form to meaning: A role for structural analogy in the acquisition of language. *Advances in Child Development and Behavior*, *27*, 1–53.

Fisher, C. (2002). Structural limits on verb mapping: the role of abstract structure in 2.5-year-olds' interpretations of novel verbs. *Developmental Science*, *5*(1), 55–64.

Fisher, C., Hall, G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: syntactic and conceptual constraints on vocabulary growth. *Lingua*, *92*(1), 333–375.

Forbes, J., & Farrar, M. (1995). Learning to represent word meaning: What initial training events reveal about children's developing action verb concepts. *Cognitive Development*, *10*(1), 1–20.

Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development*, *49*(4), 988–98.

Gentner, D., & Namy, L. (2006). Analogical processes in language learning. *Current Directions in Psychological Science*, *15*(6), 297–301.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 135–176.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (in press). Bayesian models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational cognitive modeling*. Cambridge: Cambridge University Press.

Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.

Gropen, J., Pinker, S., Hollander, M., & Goldberg, R. (1991). Affectedness and direct objects: The role of lexical semantics in the acquisition of verb argument structure. *Cognition*, *41*(1–3), 153–195.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.

Kako, E. (2006). Thematic role properties of subjects and objects. *Cognition*, *101*, 1–42.

Kemp, C., Tenenbaum, J., Griffiths, T., Yamada, T., & Ueda, N. (2006). Learning systems of concepts with an infinite relational model. In *Proceedings of the National Conference on Artificial Intelligence* (pp. 381–388).

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes*, *21*(5), 489–530.

Landau, B., & Gleitman, L. R. (1985). *Language and experience: evidence from the blind child*. Cambridge, MA: Harvard University Press.

Lieven, E. V. M., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, *24*, 187–219.

MacWhinney, B. (1995). *The CHILDES project: tools for analyzing talk* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Manning, C. D., & Schütze, H. (2001). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.

Matessa, M., & Anderson, J. R. (2000). Modelling focused learning in role assignment. *Language and Cognitive Processes*, *15*(3), 263–292.

McClelland, J. L., & Kawamoto, A. H. (1986). *Mechanisms of sentence processing: Assigning roles to constituents of sentences* (pp. 272–325). Cambridge, MA: MIT Press.

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, *12*(2/3), 137–176.

Miller, G. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography*, *17*(3), 235–244.

Morris, W. C., Cottrell, G. W., & Elman, J. L. (2000). A connectionist simulation of the empirical acquisition of grammatical relations. In S. Wermter & R. Sun (Eds.), *Hybrid neural symbolic integration*. Berlin: Springer Verlag.

Naigles, L., & Kako, E. (1993). First contact in verb acquisition: defining a role for syntax. *Child Development*, *64*(6), 1665–1687.

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

Neal, R. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, *9*(2), 249–265.

Orban, P., Rauchs, G., Balteau, E., Degueldre, C., Luxen, A., Maquet, P., & Peigneux, P. (2006). Sleep after spatial learning promotes covert reorganization of brain activity. *Proceedings of the National Academy of Sciences of the USA, 103*(18), 7124–7129.

Parisien, C. (2008). *An incremental Bayesian model for learning syntactic categories*. Master's thesis, Department of Computer Science, University of Toronto.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1989). *Learnability and cognition: the acquisition of argument structure*. Cambridge, MA: MIT Press.

Pitman, J. (2006). *Combinatorial stochastic processes*. Berlin: Springer.

Quine, W. (1960). *Word and object*. Cambridge, MA: MIT Press.

Rappaport Hovav, M., & Levin, B. (1998). Building verb meanings. In *The projection of arguments: lexical and computational factors* (pp. 97–134). Stanford, CA: CSLI Publications.

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

Sanborn, A., Griffiths, T., & Navarro, D. (2006). A more rational model of categorization. *Proceedings of the Twenty-eighth Annual Conference of the Cognitive Science Society*.

Shayan, S. & Gershkoff-Stowe, L. (2007). How do children learn Agent and Patient roles? *Poster presentation at SRCD Biennial Meeting Boston, Massachusetts, USA*.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*, 39–91.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, *74*, 209–253.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*(3), 285–318.

VanValin, R. D., & LaPolla, R. J. (1997). *Syntax: structure, meaning and function*. Cambridge, UK: Cambridge University Press.

Walker, M. P. (2005). A refined model of sleep and the time course of memory formation. *Behavioural and Brain Sciences, 28*, 51–104.

Xu, F., & Tenenbaum, J. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272.

Yu, C., Ballard, D. H., & Aslin, R. N. (2005). The role of embodied intention in early lexical acquisition. *Cognitive Science*, *29*, 961–1005.

APPENDIX A: LIKELIHOOD PROBABILITIES
OF FRAME FEATURES

As described in Section 4.1, the calculation of the prior probabilities of the constructions in our model follows the intuition of the Chinese Restaurant process (Antoniak, 1974; Pitman, 2006), which implies that as the model processes more inputs, it becomes less likely that a new construction is created. However, ideally we want our model to detect an 'unusual' frame which does not conform to any of the existing constructions, and place it in a new construction, even after processing a large number of input items. That is, if a feature value j in a new frame is unseen for a construction k, we want the likelihood $Pi(j|k)$ to be low enough to lead to the creation of a new construction. Therefore, we use Laplace's law (Manning & Schütze, 2001) to estimate the maximum likelihood of a feature value:

$$P_i(j|k) = \frac{\mathbf{count}_i^k(j) + \lambda}{n_k + \lambda \alpha_i} \tag{11}$$

where n_k is the number of frames participating in construction k, $\mathbf{count}_i^k(j)$ is the number of those with value j for feature i, and λ and α_i are smoothing factors. The parameter λ implicitly determines the importance of a distinct feature value in creating a new construction: it is set to a small constant so that the constructions which have no members with value j for feature i have a low (but non-zero) probability. The parameter α_i is set to (an estimate of) the number of possible values that feature i can take on. Note that, for a new construction, because $\mathbf{count}_i^0(j)$ and n_0 both have the value 0, the estimated conditional probability is $1/\alpha_i$ – that is, for a new construction, all feature values are equally likely. We present a theoretical method for estimating the smoothing factors in Appendix B. However, note that the smoothing factors contribute more to the likelihood of a small construction, but as more frames are added to each construction, the effect of smoothing becomes weaker, reflecting the higher confidence of the model in its captured knowledge.

As mentioned in Section 4.1, for features with a single value such as the syntactic pattern, $\mathbf{count}_i^k(j)$ is calculated by simply counting those members of construction k whose value for feature i exactly matches j. However, for those features whose value is represented as a set, such as the semantic properties of the arguments, 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 compare the set value of a particular feature in the observed frame, S_1, and the set value of the corresponding feature in a member frame of a construction, S_2, using the Jaccard similarity score:[7]

$$\mathbf{sem\_score}(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} \tag{12}$$

This measure captures the overlap between two sets of properties. Due to the hierarchical nature of the semantic property features, more similar arguments have more overlap in their corresponding feature sets. For example, assume that the new frame represents the verb usage *John ate cake*, and one of the members of the construction that we are considering represents *Mom got water*. We must compare the lexical 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}, and the union is {baked goods, liquid, fluid, food, solid, nutrient, substance, matter, entity}, therefore the **sem_score** for these sets is $\frac{4}{9}$.

---

[7] The selected semantic properties and the corresponding similarity score 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.

In general, to calculate the likelihood probability for the set of properties, we set **count**$_i^k(j)$ in Equation (11) to the sum of the **sem_score**'s for the new frame and every member of construction k, and normalise the resulting probability over all possible sets of properties in our lexicon:

$$P_{properties}(s_j|k) = \frac{\sum_{f \in k} \textbf{sem\_score}(S_j, S_f) + \lambda|S_j|}{\sum_{l \in \text{Lexicon}} \sum_{f \in k} \textbf{sem\_score}(S_l, S_f) + \lambda\alpha_{prop}} \tag{13}$$

where S_j is a feature value (i.e., a set of properties) in the current frame, S_f represents the properties of the corresponding feature in a frame f that belongs to construction k, S_l represents the properties of a lexeme l, and λ and $|S_j|/\alpha_{prop}$ are the smoothing factors (where $|S_j|$ is the size of the set S_j). For a new construction, the default probability of a set S_j will be equal to $|S_j|/\alpha_{prop}$, giving a higher default probability to a larger set. Intuitively, the sum in the numerator represents how much the meaning of the new set S_j resembles that of frames grouped in construction k. The double sum in the denominator, on the other hand, represents how distinct the semantic properties of the frames in construction k are from all the existing properties in the language. For example, if we are looking at the lexical properties of an argument, each word in the lexicon is compared against all the words in the construction. The latter factor is computationally expensive, and we use a pre-calculated similarity matrix to estimate it. However, we can assume that humans have an idea of how general or unique the meanings associated with a lexeme or a construction are, and do not have to perform such extensive calculations every time they process a new input.

APPENDIX B: SETTING THE PARAMETERS OF THE MODEL

For the purpose of implementing the model and running simulations, we need to assign values to parameter α_i for each feature i in Equation (11). This parameter determines the default likelihood probability $1/\alpha_i$ for feature i in a new construction. That is, for a new construction for which n_k and $count_i^k(j)$ are zero, the probability of having any value j for feature i is $1/\alpha_i$. Assuming that all feature values are equally likely, α_i is an estimate of the number of possible values that feature i can take on.

In our model, each frame has six different features: the head verb, the number of arguments, the syntactic pattern, the semantic primitives of the head verb, the lexical properties of the arguments, and the event-specific properties of the arguments. We assign $\alpha_{head} = 15$ and $\alpha_{patt} = 15$ as an upper bound on the number of the predicates and the distinct syntactic patterns that can be observed in the input. We assume that each verb usage can have 0 to 5 arguments, therefore $\alpha_{argnum} = 6$. The estimated α_i value for other features is determined by the number of semantic primitives, lexical properties, and event-based properties, as defined in Section 5.1: $\alpha_{prim} = 12$, $\alpha_{lexical} = 493$, and $\alpha_{event\_specific} = 50$.

We also need to set the parameter λ in Equation (11). This parameter determines the default probability of a value j for a feature i in an existing construction k, where none of the frames in k have value j for their feature i. For all features i in frame F, we need this probability (i.e., λ) to be small enough that the resulting $P_i(j|k)$, where j differs from the value in k, penalises construction k in the competition in equation (1). If we wanted each construction k to contain only frames that have the same values on all of the features, we would want the posterior probability of k to be less than that of a new construction, or $P(k|F) < P(0|F)$. The extreme case would be when only one construction exists which contains all n observed frames, every one of which has the same values as F for all features, except for feature i. By simplifying the inequality

$P(k|F) < P(0|F)$ for this case, we obtain an upper bound of $\lambda < \prod_i \frac{1}{\alpha_i}$. This gives us an upper-bound of 10^{-7} for the value of λ. We set λ to a moderate value, 10^{-5}, to allow for constructions which have frames with similar but not identical values for some features.