A Structured Context Model for Grammar Learning

Nancy Chang and Eva Mok

Abstract—We present a structured model of context that supports an integrated approach to language acquisition and use. The model extends an existing formal notation, Embodied Construction Grammar (ECG), with representations for tracking both entities and events in discourse and situational context. The notation employs an intermediate level of granularity between low-level sensorimotor representations (such as that suitable for dynamic models of action and events for grounded language learning) and the more schematic representations needed for learning and using grammar. The resulting model allows existing systems for simulation-based language understanding and comprehension-driven grammar learning to represent, interpret and acquire a variety of contextually grounded construction.

I. INTRODUCTION

Language acquisition is inherently context-dependent: every utterance a child hears is rooted within an ongoing stream of activity involving multiple participants interacting in structured ways. From the earliest communicative gestures and first words through word combinations and grammatical constructions, children learn not just the sound patterns (or gestures) of their input language(s) but also the ways in which those patterns are used to achieve communicative goals, all within specific situational contexts.

Most theories of language acquisition acknowledge that by the time children acquire their first words, they have amassed considerable sensorimotor and social-interactional expertise, which are deployed to infer, express and achieve goals. Nonetheless, aspects of situational and pragmatic context play a relatively limited role in most models of the acquisition of syntax. In some cases, this omission is theoretically motivated: models based on syntactico-centric theories (including Generative Grammar and its successors) restrict their attention to formal syntax, with detailed semantic and pragmatic information relegated to either the lexicon or general inference processes. From a more practical perspective, situational context poses far greater challenges for both representation and data collection than the simple surface strings assumed as input in more traditional approaches.

Two current streams in the literature take a more inclusive approach to linguistic representation. Construction-based theories of grammar [1-3], along with work in the broader cognitive linguistics community, take the basic unit of language to be a *construction*, or form-meaning mapping. In particular, syntactic patterns are, in this framework, inherently paired with aspects of meaning, where meaning encompasses any aspect of semantics or pragmatics, including the context of use. On this view, syntactic patterns and grammatical markers may differ from lexical items in size of level of abstraction, but they can still be represented, used and learned in similar ways. In particular, usage-based theories of construction learning [4, 5] have proposed that children's earliest constructions are lexically specific and motivated by individual instances of use, only gradually giving rise to the more abstract patterns of form and meaning associated with grammar.

The other relevant stream comes from computational models of grounded language learning. These models take a bottom-up approach that emphasizes the situated nature of language learning, exposing robotic and simulated agents to sensorimotor input accompanied by linguistic input in a dynamic environment [6-8]. While work in this area has focused on lexical acquisition and concrete physical domains, the background assumptions provide a more realistic approximation of the child's ability to exploit situational context in language learning.

The current work takes an approach that is consistent with both of these streams in addressing the need for incorporating context into models of grammar learning. We build on previous work on Embodied Construction Grammar (ECG) [9], a computational formalism designed to support models of language acquisition and use. The formalism captures many insights from cognitive linguistics and construction-based theories of language, and provides a suitable target of learning for all stages of language learning. Its meaning representations are also, however, *embodied*, in that they provide an interface to dynamic structures that model aspects of action and perception. It thus provides an intermediate level of representation between abstract linguistic theory and sensorimotor grounding.

In this paper we present a structured, dynamic context representation that addresses the need for context in language understanding. We first review some linguistic phenomena that shed light on the role of context in early child language, drawing examples from typologically diverse languages (Section II) before summarizing the theoretical framework within which our current work is situated (Sections III & IV). We then describe extensions to the ECG framework that allow constructions to refer to structured representations of both situational and discourse context and express diverse contextual constraints (Section

Both authors are with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, and the International Computer Science Institute, 1947 Center Street Ste. 600, Berkeley, CA 94704. (phone: 510-666-2889; email: {nchang, emok}@ icsi.berkeley.edu).

V). Finally, we show how interleaved processes of constructional analysis, reference resolution and grammar learning can exploit the resulting integrated context model to express, understand and learn a variety of contextually grounded constructions (Section VI).

II. THE PROBLEM OF CONTEXT

The contextual fluidity of language use has been extensively documented in the literature: speakers use language to accomplish communicative goals rooted in specific contexts, and many utterances make sense only relative to those contexts. This context-dependence is especially pronounced in parent-child interactions, which tend to focus on objects and events in the immediate environment. Crosslinguistically, many of the earliest words and expressions are tied to specific goals and social interactions. English-speaking children as young as 14 to 18 months use expressions like *hi* and *bye-bye* as part of set social routines or *there* and *uh-oh* to indicate goal achievement or failure [10-13]. Such expressions require children to learn contextual constraints on their use.

More generally, most parent-child interactions refer in some way to ongoing activities and their participants, as seen in the following dialogue [14]:

```
*MOT: are they clean yet ?
*CHI:  they clean . ...
*MOT: they're pretty clean .
*CHI:  let me wash them . ...
*MOT: just wash a little bit right in there .
*MOT: that's it .
```

Example 1. An English dialogue between Eve (age 2) and her mother, with no expressed antecedent for the pronoun *they*.

The pronoun *they* in the mother's first utterance has no antecedent in the discourse, but the child has no difficulty determining its referent in the context of the washing scenario in question (her hands, as indicated by corpus annotation). Likewise, the mother uses the expression *in there* to indicate a location that must be inferred in context. Such situation-based antecedents for pronouns and other *indexicals* are typical of child-parent interactions, and more common than in adult interactions.

Situational context plays an even more important role when referents are omitted entirely. Such omissions occur occasionally even in languages like English, but they are commonplace in many languages, especially "pro-drop" languages in which pronouns are typically omitted. In pro-drop languages with rich inflectional morphology (such as Spanish or Turkish), verb inflections provide cues about the gender and number of omitted referents. In pro-drop languages with minimal morphology (such as Mandarin Chinese), the lack of morphological cues forces hearers to rely more heavily on context and knowledge about typical events to infer the intended referents, as illustrated in the dialogue below [15] (literal translation in parentheses):

```
*CHI:   zang1.           (dirty)
*MOT:   zang1 le .       (dirty ASPECT)
*CHI:   ei xi3+xi3 .     (INTERJ wash+wash)
*MOT:   en xi3 xi3 .     (INTERJ wash+wash)
*MOT:   xing2 .          (alright)
```

Example 2. A Mandarin dialogue between HaoYu (age 2) and his mother; subject and object of the verb *xi3* ('wash') are unexpressed.

Neither these utterances nor the preceding context mention the item to be washed, and the verb *xi3* ('wash') is never expressed with either a subject or a direct object, even pronominally. To make sense of these utterances, the child must integrate what he hears with information from the scene. Such omission is a general feature of Mandarin adult-to-adult conversation, where 45.6% of subjects and 40.1% of objects are omitted [16].

These examples suggest that models of language understanding and language learning must be able to represent and refer to aspects of the surrounding situational and discourse context. Specific lexical items as well as more general constructions might apply only when an appropriate referent (e.g., a plural, cleanable object; or a location in focus) is present or some other contextual condition holds. Relevant information includes not just stimuli currently perceptible in the environment but also some history of the preceding sequence of utterances and events.

Constructions should be able to express such contextual constraints and thus explicitly direct processes of language understanding to seek appropriate specific referents. The reference resolution process, in turn, must translate such constraints into a search of the current context. Finally, learning processes must associate such contextual constraints with constructions based on instances of use.

III. EMBODIED LANGUAGE UNDERSTANDING

The current work is situated within a larger effort to build models of cognition and language that satisfy convergent constraints from biology, psychology, linguistics and computation [17]. The overarching goal of the Neural Theory of Language (NTL) project is to bridge the gap between behavior and the brain through successive levels of computational modeling. While some levels hew closely to neural computation (low-level computational biology and structured connectionist models) [18], we focus here on the computational level, which provides an intermediate level of description between cognitive and linguistic theories and more detailed and biologically inspired structures.

A central hypothesis explored by the NTL project is that language understanding exploits many of the same structures used for action, perception, imagination, memory and other neurally grounded processes, which forms the embodied basis of meaning (for linguistic and psychological evidence see [19-22]). Crucially, these embodied representations are parameterized, and language serves to supply the necessary parameters for *simulations* that lead to deep understanding.

The model of language understanding substantiating this

hypothesis involves several interacting structures and processes (Figure 1). The *analysis* process determines which constructions are instantiated by a given utterance, drawing on linguistic knowledge, conceptual knowledge (entity and event types), and the current communicative context. Along with a structural analysis analogous to a syntactic parse tree, the analysis process yields an interpretation of the sentence: a graph of embodied semantic schemas, called the *semantic specification* (or *semspec*). The *reference resolution* process links schemas in the semspec with contextually available referents. This *resolved semspec* provides parameters for a *simulation* process that activates embodied conceptual structures to produce new inferences and update the context.
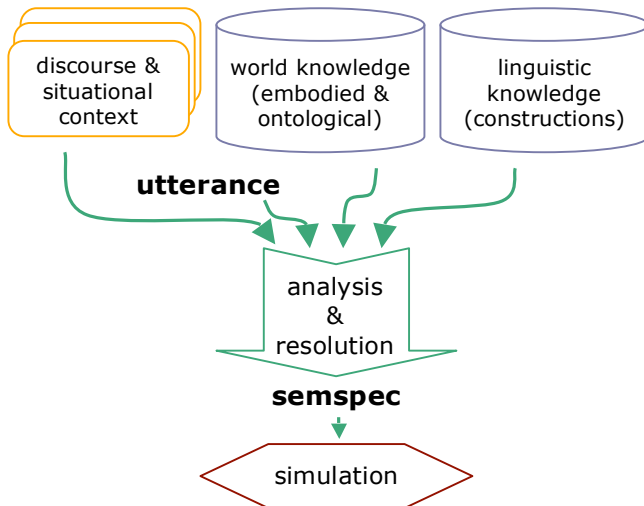

Fig. 1. Simulation-based language understanding: An utterance is passed along with the current context to an analysis process that draws on general knowledge and linguistic knowledge (expressed using Embodied Construction Grammar). The result of analysis and reference resolution is a semantic specification (semspec) that parameterizes a dynamic simulation using embodied representations.

Previous systems within the NTL project have implemented basic versions of both the analysis and simulation processes. Simulation is modeled using both Bayesian networks and Petri-net-based representations (called *x-schemas*) inspired by motor control and perception [23, 24]. The analysis process is most relevant for current purposes. The current analyzer system extends partial parsing and constraint-based parsing techniques to include semantic information [28]. To date, the analyzer has been applied to interpret a variety of early constructions in English, Mandarin Chinese and German.

A key limitation of the existing analyzer is that it is not explicitly connected to context: its input is a sentence in isolation, without its situational or discourse context. The lack of access to the evolving discourse and situational context is problematic in light of the phenomena discussed in Section II. The goal of the current work can thus be stated more concretely as incorporating such information in a model of context that can serve both the language understanding model we have described in this section — in particular, as the key data structure needed to support

reference resolution — and the larger model of language learning of which it is a subcomponent (discussed further in Section V). Our approach is to extend the formal tools of ECG, previously used for representing both embodied meanings and linguistic knowledge, to include contextually available entities and events. We describe ECG in more detail in the next section.

IV. EMBODIED CONSTRUCTION GRAMMAR

The crucial interface between language and simulation in this framework is supplied by Embodied Construction Grammar (ECG) [9, 25], a unification-based computational formalism for representing linguistic knowledge. As in other construction-based grammars [1, 3, 26], *constructions* express generalizations linking the domains of form and meaning. Constructions vary in size (from morphemes and lexical items to phrasal and clausal units) and specificity (from idioms to more abstract grammatical constructions); and they encompass information that crosscuts traditional levels of linguistic analysis (e.g., phonological, morphological, syntactic, semantic and pragmatic).

In ECG, meaning is represented using *embodied schemas* that specify parameters for simulation; constructions link forms to these embodied schemas. The ECG schema and construction formalisms include mechanisms for expressing type constraints, ordering constraints, *identification* (unification) constraints, self-reference, constituency and dependency relations. Computationally, both constructions and schemas are implemented using typed feature structures with unification constraints, organized in an inheritance hierarchy. (See [9] for more formalism details.)

Fig. 2. A Wash-Action schema, including roles for the washer and washee of specified ontological types; and an X-wash-Y construction with three constituents (two referring expressions and the verb wash) that are ordered and unified as shown.

Figure 2 shows three example ECG structures: (1) Washing actions are represented using the Wash-Action schema, which summarizes the main participant roles in a washing scene. (2) The lexical Wash construction (*i.e.,* the verb *wash*) links the Wash-Action schema with a form (simplified here as an orthographic string). (Other lexical constructions, not shown, link names like *Mom* or *Baby* to schemas representing the relevant individuals.) (3) The more complex X-wash-Y construction has three constructional *constituents* (named x, w and y) and expresses several

relational constraints on the form and meaning components of its constituents (indicated by subscripted f and m), including ordering and identification constraints.

These structures, along with appropriate referential constructions, supply the linguistic knowledge underlying sentences like "Mom wash Baby" or "Baby wash hands": the analysis and resolution processes produce a semspec instantiating the Wash-Action schema with its roles appropriately bound; these roles also specify parameters for a simulation based on a washing x-schema (not shown). It is this dynamic structure that captures the sensorimotor details involved in executing a washing action (*e.g.*, the energy expended by the washer and the iterative nature of the washing process, or the relative state of cleanliness of the washee before and after washing). That is, the semspec provides a narrow interface between linguistically specified relations and full-blown inference, allowing constructions to remain relatively schematic, and thus more simply expressed, learned and extended to novel situations.

V. REPRESENTING CONTEXT

In this section we introduce a structured representation of discourse and situational context. This model captures information needed in the reference resolution process and the constructional constraints that parameterize that process.

The context model has four main components, each represented as a structured ECG schema: the **Discourse** schema, the **ContextElement** schema, the **Event** schema, and the **DiscourseSegment** schema. The central organizing structure of the model is the Discourse schema (Figure 3), a structured representation of a sequence of communicative acts involving participants and objects. Past communicative acts, both non-linguistic and linguistic, are accessible as the situational-history and the discourse-history, respectively.

Each role in the Discourse schema is filled by a list of items, all subcases of the ContextElement schema, the main high-level schema in our ontology for representing potential discourse referents. Such referents include not only physically present participants and objects, such as humans, cups, or shirts, but also events in the situation and previous utterances (represented using the Event and DiscourseSegment schemas, respectively).

```
schema Discourse
  roles
    participants: ContextElement
    objects: ContextElement
    situational-history: Event
    discourse-history: DiscourseSegment
```
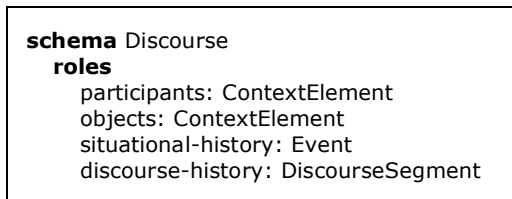
Fig. 3. The Discourse schema is the central bookkeeping structure for a discourse, including its relevant participants, objects, and a running history of both linguistic and non-linguistic context.

In Figure 4, the schemas listed above are used to represent the context after the first utterance in Example 1. The model is partitioned into sections corresponding to different schema types. Note that the utterance "are they clean yet?" is included in the discourse history as DS01, a **DiscourseSegment**. This schema contains meta-level information about the communicative act, including its speaker, addressee(s), shared attentional-focus, speech-act, and an analysis of the utterance (not shown).

```
Discourse:
  Discourse01
    participants: Eve (1) , Mother (2)
    objects: Hands (3)
    situational-history: Wash-Action (4)
    discourse-history: DS01 (5)

Participants:
  Eve (1)                  Mother (2)
    category: child          category: parent
    gender: female          gender: female
    name: Eve                name: Eve
    age: 2                   age: 33

Objects:
  Hands (3)
    category: BodyPart
    part-of: Eve (1)
    number: plural
    accessibility: accessible

Situational History:
  Wash-Action (4)
    washer: Eve (1)
    washee: Hands (3)

Discourse History:
  DS01 (5)
    speaker: Mother (2)
    addressee: Eve (1)
    attentional-focus: Hands (3)
    content: {"are they clean yet?"}
    speech-act: question
```
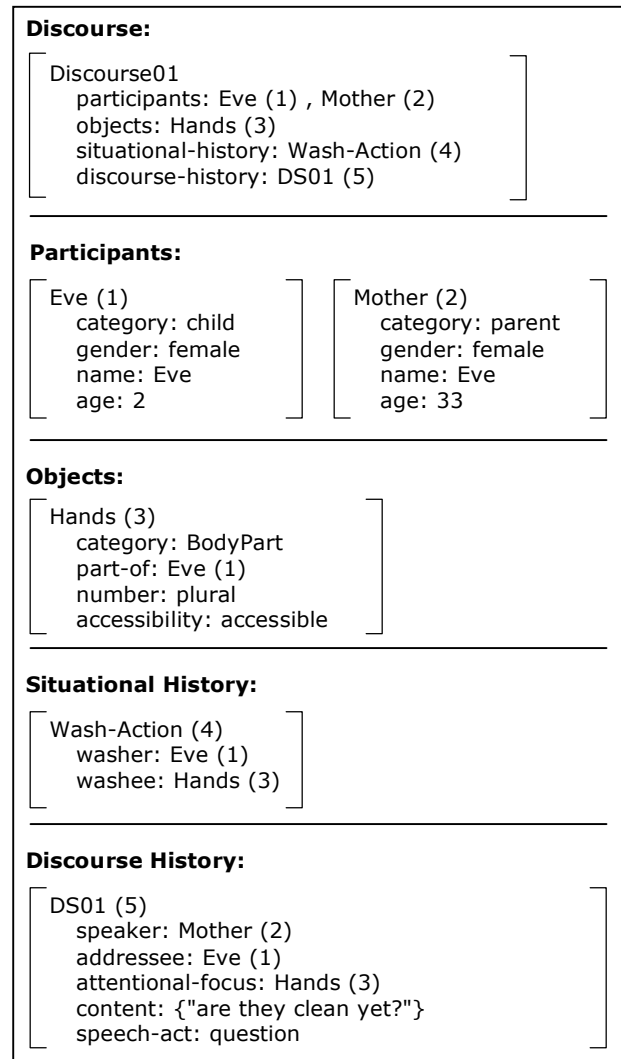
Fig. 4. A feature structure representing the context after the first utterance from Example 1. Numbers in parentheses indicate coindexed structures or slots. The central washing event of the discourse is captured by the Wash-Action schema.

Together these schemas serve as a snapshot of a specific point in the discourse. Both situational and discourse elements are represented using ECG schemas. It is thus straightforward for constructions to make explicit reference to them and for the analysis and reference resolution processes to treat them uniformly.

The evolving discourse is updated by adding new events and utterances to its situational and discourse history lists. These new context items might be detected based on separate input-monitoring procedures, for example by event recognition and speech detection systems based on low-level real-time sensory input. The current implementation uses transcripts of parent-child interactions that have been

manually annotated to simulate the unfolding sequence of utterances and events in the discourse.

VI. EXPLOITING CONTEXT

The schematic structures introduced above provide a basis for capturing the evolving discourse and situational context. This section describes how the existing ECG formalism and its associated processes of language understanding and learning exploit the structured context model.

A. Expressing contextual constraints in ECG

Since all components of the context model are represented using ECG schemas, constraints on these components are easily expressed using existing formal mechanisms of ECG constructions. In particular, constructions associated with referring expressions (such as pronouns, proper nouns and noun phrases) typically assert conditions that must hold of their intended referents. These conditions are represented in a special schema, called a ReferentDescriptor, which serves as input to the reference resolution process that finds an appropriate referent in the current context (i.e., locates the ContextElement that best fits the constraints in the ReferentDescriptor).

The ReferentDescriptor schema, shown in Figure 5, is intended to capture all linguistically specified information that may be relevant to identifying a referent. This information includes: (i) a restricted set of ontological features, such as number, gender, (ontological) category and case; (ii) a more open-ended set of possible restrictions and modifiers that apply to the referent; and (iii) specifically contextual restrictions, such as the referent's accessibility [27] (specifying whether it is *accessible* in the current discourse context, newly introduced, etc.). The resolved-ref slot plays a special role in linking language and context: this slot is coindexed with the particular ContextElement found by the reference resolution process.

```
schema ReferentDescriptor
    roles
        number
        grammatical-gender
        ontological-category
        case
        accessibility
        modifiers
        resolved-ref: ContextElement
```
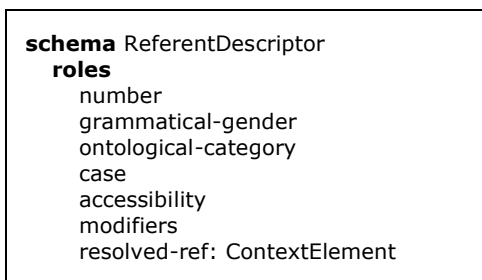
Fig. 5. The ReferentDescriptor schema captures constraints on the referent through a set of features (some omitted for brevity). Resolution is the processing of finding the ContextElement that best fits these constraints, with the result stored in the resolved-ref role.

We illustrate the ReferentDescriptor schema in constructions for the pronouns *them* and *you*, represented in ECG as shown in Figure 6. Each of these constructions evokes an instance of the ReferentDescriptor schema (locally referred to as rd) and binds its meaning (self$_m$) to its rd.resolved-ref role. The remaining constraints assert other relevant restrictions on the referent. Note that the You construction also binds its meaning directly to the current addressee (where DS denotes the current discourse segment).

```
construction Them
    form
        self_f.orth ← "them"
    meaning: ContextElement
        evokes ReferentDescriptor as rd
        self_m ↔ rd.resolved-ref
        rd.number ← plural
        rd.accessibility ← accessible
        rd.case ← object
```

```
construction You
    form
        self_f.orth ← "you"
    meaning: ContextElement
        evokes ReferentDescriptor as rd
        self_m ↔ rd.resolved-ref
        self_m ↔ DS.addressee
        rd.accessibility ← accessible
```
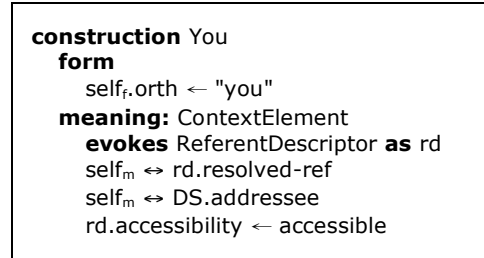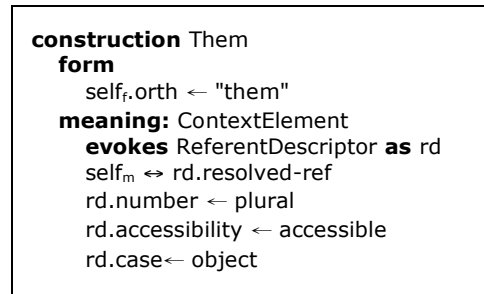
Fig. 6. Both the them and you constructions link an orthographic form with a ContextElement that is bound to an accessible resolved referent of an evoked ReferentDescriptor. While *them* asserts several constraints on the features of its referent, *you* is bound directly to the addressee in the current discourse segment.
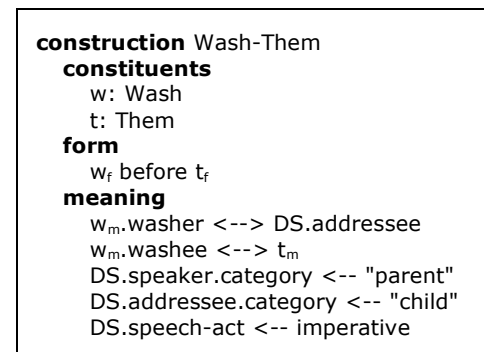
```
construction Wash-Them
    constituents
        w: Wash
        t: Them
    form
        w_f before t_f
    meaning
        w_m.washer <--> DS.addressee
        w_m.washee <--> t_m
        DS.speaker.category <-- "parent"
        DS.addressee.category <-- "child"
        DS.speech-act <-- imperative
```

Fig. 7. A possible two-word imperative construction "Wash them" constrained to be spoken by a parent to a child. Discourse-based constraints are expressed using the speaker and addressee roles in the DiscourseSegment (DS) schema.

Referring expressions are not the only class of constructions that can impose contextual constraints. Many other constructions may assert situational restrictions: they may require that a specific type of event is accessible in context, as described earlier for utterances like "bye-bye", or they may require a particular social status of one of the discourse participants. As an example, the expression "wash them" may at some point in a child's development be learned as a lexically specific word combination that is used exclusively by a parent toward the child as an imperative. Figure 7 depicts these contextual constraints on the associated hypothetical construction, again using the variable DS to refer to the current discourse segment and thus access its associated roles appropriately.

As these examples illustrate, the context model we have introduced, paired with the ReferentDescriptor schema, provide special structured schemas for organizing all information related to the context that may be relevant to constructional analysis or reference resolution.

B. Incorporating context into language analysis

The context-based extensions to ECG introduced in the last section employ the same underlying representations (feature structures) as used in the current analyzer implementation [28]. They thus do not alter the basic mechanics of constructional analysis: the search for potential constructions whose constraints are satisfied, and the unification of the relevant meanings into the semspec.

The availability of contextual information does, however, allow a more principled approach to reference resolution than previously available. Besides being able to interpret references to roles in the current context (*e.g.*, addressee), the analyzer can also invoke reference resolution to seek the most likely ContextElement matching a ReferentDescriptor instantiated with the relevant constructional constraints. This interleaving of analysis and resolution can impose new context-based constraints that restrict the search space of analyses beyond those in the constructions alone, thus potentially reducing ambiguity and speeding processing.

We briefly illustrate how the context model shown in Figure 4 is updated based on the hypothetical utterance "Wash them" (Figure 8). Many schemas remain the same, such as those for the participants and objects present. DS02 is the DiscourseSegment containing the current utterance, whose semspec is shown at the bottom of the figure.

At this point the constructional analysis process has determined that the constraints of the lexical Wash and Them constructions, as well as the larger Wash-Them construction, have been satisfied. The semspec shown in the figure is the result of unifying the various constraints of these constructions. The main schema of interest is the ReferentDescriptor rd01, which is evoked and constrained by the Them construction. The Wash-Them construction binds its resolved referent to the washee slot in the Wash-Action schema. Neither construction specifies the actual referent here, however; it is the reference resolution process that determines that the schema's resolved-ref slot is bound to the Hands described in the Discourse.

In its simplest form, the resolution process finds the element (or set of elements) in context that unifies with this ReferentDescriptor. Here the resolution appears trivial, since Hands is the only object in context that satisfies the constraint of plurality. Note, however, that different approaches to resolution may vary in both how wide a net is cast in the search for referents (for example, whether ontological knowledge would lead it to include other plural objects, such as the child's ears or the parent's hands) and how the resulting set of potential referents is ordered (for example, by salience, recency in discourse, or measures such as attentional focus). The particular resolution algorithm is not at issue for current purposes, though it remains the subject of ongoing work in the NTL project [29].
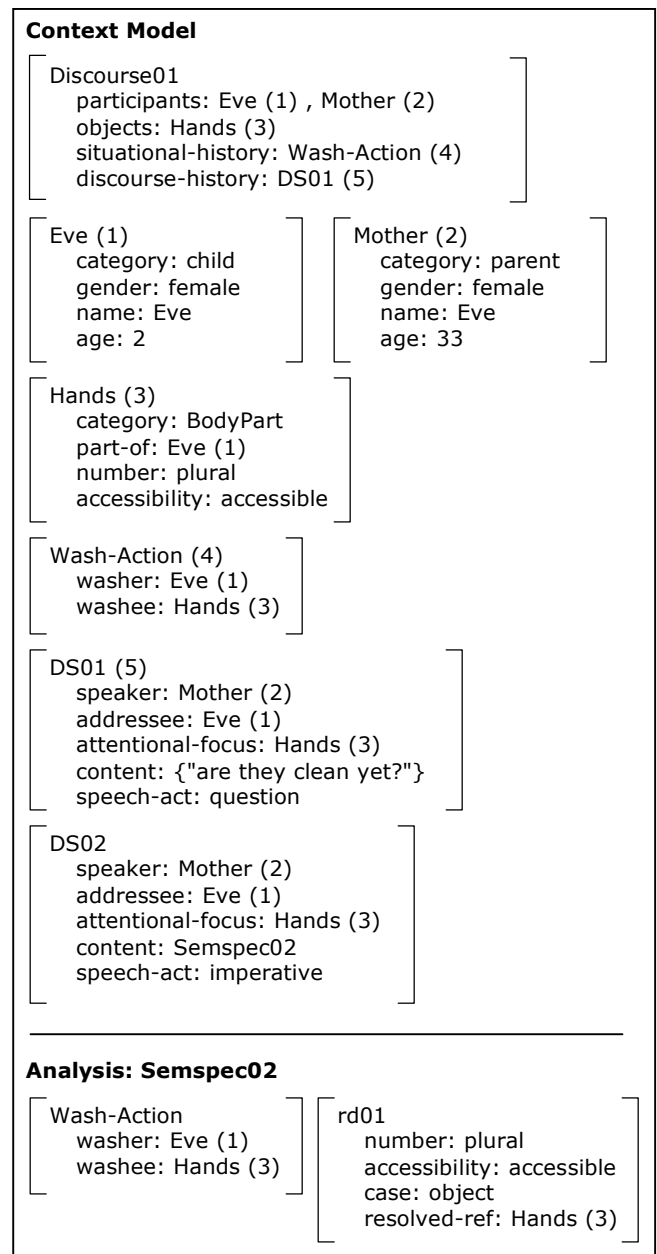
Context Model

Discourse01
    participants: Eve (1) , Mother (2)
    objects: Hands (3)
    situational-history: Wash-Action (4)
    discourse-history: DS01 (5)

Eve (1)
    category: child
    gender: female
    name: Eve
    age: 2

Mother (2)
    category: parent
    gender: female
    name: Eve
    age: 33

Hands (3)
    category: BodyPart
    part-of: Eve (1)
    number: plural
    accessibility: accessible

Wash-Action (4)
    washer: Eve (1)
    washee: Hands (3)

DS01 (5)
    speaker: Mother (2)
    addressee: Eve (1)
    attentional-focus: Hands (3)
    content: {"are they clean yet?"}
    speech-act: question

DS02
    speaker: Mother (2)
    addressee: Eve (1)
    attentional-focus: Hands (3)
    content: Semspec02
    speech-act: imperative

Analysis: Semspec02

Wash-Action
    washer: Eve (1)
    washee: Hands (3)

rd01
    number: plural
    accessibility: accessible
    case: object
    resolved-ref: Hands (3)

Fig. 8. The context model updated after analyzing the utterance "Wash them." The referent descriptor rd01 is bound to the contextually available Hands schema.

The resolution procedure is also mediated by linguistically specified features of the referent. Unlike pronouns, referring expressions in general may evoke referents not previously in context. Narratives typically construct contexts in which to situate the action, and indefinite articles (such as English *a/an*) introduce new (or *inaccessible*) entities into the discourse; a presumed referent (*e.g.* lunch) may also be introduced by a sentence like "Have you eaten yet?". In these cases, or when resolution fails to find a suitable ContextElement in the context model, its function is to create a new referent based on the constraints given by the ReferentDescriptor and make it accessible as a new entity.

C. Leveraging context in language learning

As noted earlier, the ECG formalism is designed to serve not only as the basis for simulation-based language understanding but also as the target of a usage-based model of the acquisition of early multiword constructions [5]. This learning model takes input of sentences paired with their communicative contexts and learns the optimal grammar (a set of constructions represented using ECG) according to a minimum description length (MDL) heuristic. This heuristic incorporates costs for both the size of the grammar and size of the data according to the grammar.

The model focuses on argument structure constructions (*e.g.*, the caused-motion construction containing an agent, a patient and a path, as exemplified by the sentence "I threw the ball into the basket"). Its behavior and assumptions are consistent with Tomasello's [5] hypothesis that children's earliest constructions are lexically specific (e.g., X's *hitting* Y, or *I'm* Y-*ing it)*. Children later generalize these item-based constructions into more general constructions. A key assumption is that learning is *usage-based*, both in the sense that more frequent, useful constructions are learned first, and in the sense that language learning depends on language use – in particular, the learning model is tightly integrated with the language understanding model described earlier.

The overall understanding-learning loop is detailed in [30, 31] and summarized in Figure 9. Essentially, the model invokes the analysis and resolution processes described earlier to produce a *partial analysis*, rather than the usual full semspec suitable for simulation. The difference between this partial analysis and the relatively richer understanding of the situation available in the current context can prompt the formation (or *hypothesis*) of new form-meaning mappings (*i.e.*, constructions) to bridge the gap. Constructions can also be learned based on the *reorganization* of the existing set of constructions: constructions that are structurally similar or compatible (*e.g.*, those that have shared constituents or shared meanings) can be merged into more general constructions with variable arguments, split into smaller constructions or joined into larger constructions. These learning operations serve as the means for searching the space of grammars, and are evaluated according to the MDL-based cost.

The context model we have presented is compatible with all the mechanisms in the existing learning model, and is easily incorporated to extend the model's ability to learn contextual constraints. Any feature of the context model can be hypothesized as a constraint on a new construction, as long as it is present in the input. For example, a Wash-Them construction might be hypothesized to capture the mapping between the word order relationship in the utterance "wash them" and the role binding between the associated washing schema and the context-determined referent of *them*, as well as the remaining constraints available in context. Subsequent learning input might lead to a similar Wash-It construction, and eventually a generalized Wash-X construction. Crucially, existing learning operations can accommodate contextual constraints as easily as other constraints.
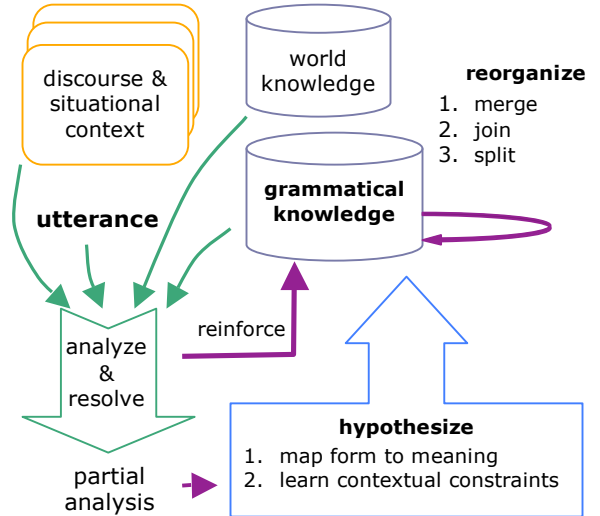

Fig. 9. Construction learning model: Each input utterance is analyzed using the current grammar, yielding a partial analysis, which is then resolved against the current discourse and situational context to produce a coherent meaning. New constructions are hypothesized to explain new information obtained through resolution, and existing constructions are reorganized via generalization over constructions based on shared constituents or constraints.

The current implementation of the learning model does not address potential interactions between concept and language learning, and it assumes a relatively stable initial lexicon. It nonetheless captures many aspects of the learning task faced by the child, in contrast to grammar induction systems that focus almost exclusively on syntax [32, 33]. Pilot tests of the learning model on child-directed utterances drawn from the CHILDES corpus [34] have demonstrated successful learning of simple English constructions, where the analyzer's performance improves as more training data is encountered, and partially abstract constructions like the Wash-X construction allow successful interpretations of previously unseen sentences. Larger-scale testing of the system on larger datasets and on multilingual input are currently under way.

VII. CONCLUSION AND FUTURE WORK

In this paper we have examined the role of context in language acquisition and use. We have focused on showing how a structured but relatively high-level description of discourse and situational context can be incorporated into an existing simulation-based model of language understanding and its associated usage-based model of language learning.

The model we have presented is intended primarily to explore the linguistic and cognitive viability of the overall NTL framework by capturing aspects of the environment encountered by children learning language, as well as prior knowledge they bring to the task. This includes not just embodied sensorimotor routines and categorical knowledge of entities and events but also the kinds of social reasoning and pragmatic inference acquired in the first year of life.

While the inclusion of such rich input is more complex and difficult to model than other possible learning input, it also offers a more realistic basis for addressing the basic conundrums of language acquisition without requiring recourse to unmotivated assumptions about syntactic biases.

The approach taken here should also, however, prove useful for more applied efforts to automatically acquire linguistic patterns beyond lexical items based on naturalistic input. In particular, any attempt to learn multiword grammatical constructions must face the kinds of representational issues addressed by our model.

We have focused here on simple cases of pronominal reference in child-directed language, but we believe the model will scale to accommodate the representational needs of more exotic phenomena in both child-centered and adult language crosslinguistically. In addition, we expect that the basic learning model will prove robust to incorporating more realistic representations of the surrounding context (as input) and the inferential abilities of the child, as well as learning operations motivated by the complementary processes of language production.

One challenge that is the subject of continuing research concerns the handling of omitted arguments, which as noted earlier is a common phenomenon in both child language and in pro-drop languages (and illustrated for Mandarin Chinese in Example 2). All of the processes described here, from constructional analysis and reference resolution to grammar learning, must be expanded to flexibly allow constituents to be omitted while still allowing implicit reference to the context model. To this end, probabilistic versions of the analyzer and the learning model are both ongoing work within the NTL project [29, 35].

References

[1] W. Croft, *Radical Construction Grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press, 2001.

[2] C. Fillmore and P. Kay, *Construction grammar*. Stanford, CA: CSLI, 1999.

[3] A. E. Goldberg, *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press, 1995.

[4] J. Bybee and J. Scheibman., "The effect of usage on degrees of constituency: the reduction of don't in English.," *Linguistics*, vol. 37, pp. 575-596, 1999.

[5] M. Tomasello, *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press, 2003.

[6] T. Regier, *The Human Semantic Potential*. Cambridge, MA: MIT Press, 1996.

[7] D. Roy, "Learning Words and Syntax for a Visual Description Task," *Computer Speech and Language*, vol. 16, 2002.

[8] L. Steels, "Evolving grounded communication for robots," *Trends in Cognitive Science*, vol. 7, pp. 308-312, 2003.

[9] B. K. Bergen and N. Chang, "Embodied Construction Grammar in simulation-based language understanding," in *Construction Grammars: Cognitive Groundings and Theoretical Extensions*, J.-O. Östman and M. Fried, Eds. Philadelphia, PA: John Benjamins, 2005, pp. 147-190.

[10] L. Bloom, E. Tinker, and C. Margulis, "The words children learn: Evidence against a noun bias in early vocabularies," *Cognitive Development*, vol. 8, pp. 431–450, 1993.

[11] P. S. Dale and L. Fenson, "Lexical development norms for young children," *Behavior Research Methods, Instruments & Computers*, vol. 28, pp. 125-127, 1996.

[12] K. Nelson, J. Hampson, and L. K. Shaw, "Nouns in early lexicons: Evidence, explanations and implications.," *Journal of Child Language*, vol. 20, pp. 61-84, 1993.

[13] A. Ninio, "Expression of communicative intents in the single-word period and the vocabulary spurt.," in *Children's language*, vol. 8, K. E. Nelson and Z. Reger, Eds. Hillsdale, NJ: Erlbaum, 1995.

[14] R. Brown, *A first language: The early stages.* Cambridge, MA: Harvard University Press, 1973.

[15] T. Tardif, "Adult-to-child speech and language acquisition in Mandarin Chinese," vol. Unpublished doctoral dissertation: Yale University., 1993.

[16] Y. J. Kim, "Subject/object drop in the acquisition of Korean: A cross-linguistic comparison," *Journal of East Asian Linguistics*, vol. 9, pp. 325-351, 2000.

[17] N. Chang, J. Feldman, and S. Narayanan, "Structured Connectionist Models of Language, Cognition and Action," presented at Ninth Neural Computation and Psychology Workshop, Plymouth, UK, 2004.

[18] J. A. Feldman and D. H. Ballard, "Connectionist models and their properties," *Cognitive Science*, vol. 6, pp. 205-254, 1982.

[19] L. W. Barsalou, "Perceptual symbol systems," *Behavioral and Brain Sciences*, vol. 22, pp. 577-660, 1999.

[20] A. M. Glenberg and D. A. Robertson, "Indexical understanding of instructions," *Discourse Processes*, vol. 28, pp. 1-26, 1999.

[21] G. Lakoff, *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press, 1987.

[22] R. W. Langacker, *Foundations of Cognitive Grammar, Vol. 1*: Stanford University Press, 1987.

[23] D. R. Bailey, J. A. Feldman, S. Narayanan, and G. Lakoff, "Modeling Embodied Lexical Development," presented at Proc. 19th Cognitive Science Society Conference, 1997.

[24] S. Narayanan, "KARMA: Knowledge-based Action Representations for Metaphor and Aspect," vol. Ph.D. dissertation: University of California Berkeley., 1997.

[25] B. Bergen, N. Chang, and S. Narayan, "Simulated Action in an Embodied Construction Grammar," presented at Proc. 26th Cognitive Science Society Conference, 2004.

[26] P. Kay and C. Fillmore, "Grammatical constructions and linguistic generalizations: the What's X doing Y? construction," *Language*, vol. 75, pp. 1-33, 1999.

[27] K. Lambrecht, *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge, UK: Cambridge University Press, 1994.

[28] J. Bryant, "Recovering coherent interpretations using semantic integration of partial parses," presented at 3rd workshop on Robust Methods in Analysis of Natural language Data (ROMAND-04), Geneva, Switzerland, 2004.

[29] J. Bryant, "A Psychologically Plausible Model of Constructional Analysis," in *Department of EECS*, vol. Dissertation Proposal: University of California, Berkeley, 2005.

[30] N. Chang and O. Gurevich, "Context-driven construction learning," presented at 26th Annual Meeting of the Cognitive Science Society, 2004.

[31] N. Chang, S. Narayanan, and M. Petruck, "Putting Frames in Perspective," presented at 2002 International Conference on Computational Linguistics, 2002.

[32] A. Clark, "Unsupervised Language Acquisition: Theory and Practice," vol. PhD thesis: University of Sussex., 2001.

[33] D. Klein, "The Unsupervised Learning of Natural Language Structure," vol. Ph.D. Thesis: Stanford University., 2005.

[34] B. MacWhinney, *The CHILDES project: Tools for analyzing talk.* Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.

[35] E. Mok, "Learning Contextual Constructions," in *Department of EECS*, vol. Dissertation Proposal: University of California, Berkeley, 2005.