Context-Driven Construction Learning

Nancy Chang (nchang@icsi.berkeley.edu)
UC Berkeley, Department of Computer Science and
International Computer Science Institute
1947 Center St., Suite 600, Berkeley, CA 94704

Olya Gurevich (olya@socrates.berkeley.edu)
UC Berkeley, Department of Linguistics
1203 Dwinelle Hall, University of California at Berkeley
Berkeley, CA 94720-2650

Abstract

We present a computational model of how partial comprehension of utterances in context may drive the acquisition of children's earliest grammatical constructions. The model aims to satisfy convergent constraints from cognitive linguistics and crosslinguistic developmental evidence within a statistically driven computational framework. We examine how the tight coupling between contextually grounded language comprehension and learning processes can be exploited to improve the model's ability to search the space of possible constructions. In particular, previously learned constructions may not fully account for all contextually perceived mappings between forms and meanings. In the model, these incomplete analyses directly prompt the formation of new relational mappings that bridge the gap. We describe an experiiment applying the model to the acquisition of English verb island constructions and discuss how the model handles more complex examples involving Russian morphological constructions. Together these demonstrate the viability of the overall approach and representational potential of the model.

Beyond single words

How do children make the leap from single words to complex combinations? The simple act of putting one word in front of another to indicate some relation between their meanings is widely considered the defining characteristic of linguistic competence and the key to unlocking the combinatorial and expressive power of language. A viable account of the acquisition of these combinatorial patterns, or *grammatical constructions*, would thus have significant implications for any theory of language that aspires to cognitive plausibility.

As with most issues impinging on the nature of grammar, linguistic and developmental inquiries into the source of combinatorial constructions have bifurcated along theoretical lines. These reflect divergent assumptions about, among other things, what kind of learning bias children bring to the task, how the target linguistic knowledge should be represented, what kind of data should be considered part of the training input, and how (if at all) language learning interacts with other linguistic and cognitive processes. Theoreticians within the formalist "learnability" paradigm, for example, have generally restricted their attention to the form domain, taking the input for learning to be a set of surface strings (each a sequence of surface forms) and positing relatively abstract structures that govern the combination of linguistic units.

This paper takes as starting point the hypothesis that the learning problem at hand may encompass a broader subset of the child's experience, centrally including meaning as it is communicated in context. We assume along with many theories of language that the basic unit of linguistic knowledge, for both lexical items and larger phrasal and clausal units, is a symbolic pairing of form and meaning, or *construction* (Langacker, 1987; Goldberg, 1995; Fillmore and Kay, 1999). Since the target of learning is rooted in both form and meaning domains, the learner should exploit information from both domains during learning.

Most importantly, we view linguistic constructions as inherently dependent on and supportive of dynamic processes of language *use*, anchored in a communicative context. A crucial but often neglected source of bias in learning constructions must therefore be how much they help the child meet her communicative goals.

This paper presents a computational model of construction learning consistent with these principles, focusing on how language understanding drives language learning. We describe a statistically driven machine learning framework that takes as input a sequence of child-directed utterances paired with their associated situational context, along with the current grammar, or set of constructions; this grammar is initially restricted to lexical items. The utterances are passed to a language understanding system (Bryant, 2003) that produces a partial interpretation, which provides the basis for the learning model to form new constructions. We present results showing how the model acquires simple English "verb island" constructions (Tomasello, 1992), and discuss how the same mechanisms handle the more complex constructions involved in Russian nominal case marking. These studies lend support for the larger program of integrating cognitive and constructional approaches to linguistics, crosslinguistic developmental evidence, and machine learning techniques to address the puzzles of language acquisition.

The Construction Learning model

We briefly describe the construction learning model in terms of (1) the target representation of learning, (2) assumptions about the child language learning scenario, and (3) the computational learning framework; see (Chang, 2004; Chang and Maia, 2001) for more details.

Target representation: embodied constructions

Embodied Construction Grammar (Bergen and Chang, in press; Chang et al., 2002) is a computationally explicit formalism for capturing insights from the construction grammar and cognitive linguistics literature. ECG supports an approach to language understanding based on two linked processes: **analysis** determines what constructions and schematic meanings are present in an utterance, resulting in a *semantic specification* (or *semspec*); the semspec serves to parameterize a **simulation** using active representations (or *embodied schemas*) to produce context-sensitive inferences.

Semantic representations in ECG are richly detailed and cognitively motivated, incorporating image schemas, motor schemas, force-dynamic schemas, and fine-grained representations of event and participant structure. But for ease of exposition, we omit most of this detail in our simple examples below, since it is not crucial for our current focus on the acquisition of the earliest constructions with constituent structure.[1]

We highlight a few aspects of the formalism relevant for the learning model discussion to follow, exemplified by the lexically specific clausal THROW-TRANSITIVE construction shown in Figure 1. The formalism draws from both object-oriented programming languages and constraint-based grammars, including notations for expressing features, inheritance, typing, and unification/coindexation.

construction THROW-TRANSITIVE
  constituents
    t1 : REFERRING-EXPRESSION
    t2 : THROW
    t3 : REFERRING-EXPRESSION
  form
    $t1_f$ `before` $t2_f$
    $t2_f$ `before` $t3_f$
  meaning
    $t2_m$.thrower $\longleftrightarrow$ $t1_m$
    $t2_m$.throwee $\longleftrightarrow$ $t3_m$

Figure 1: Representation of a lexically specific THROW-TRANSITIVE construction, licensing expressions like *You throw the ball*, with separate blocks listing constituent constructions (t1, t2, t3), form constraints (e.g., the word order relation `before`) and meaning constraints (e.g., the identification binding $\longleftrightarrow$).

All constructions have **form** and **meaning** blocks, but the **constituents** block appears only in the complex constructions that are the target of the present learning enterprise. These constituents may be typed as instances of particular constructions, and their form and meaning components (or *poles*) may be referred to (shown with a subscripted f or m) by the constraints listed in the form and meaning blocks. Form constraints are used to capture (partial) word order and other relations between form segments. In the meaning domain, the primary relation is *identification*, or unification, between two meaning entities. In particular, we will focus on role-filler bindings, in which a role (or feature) of one constituent is identified with another constituent. The example construction involves three constituents – two referring expressions and the verb THROW. Their form poles are constrained to come in a specified order, and the meaning poles of the two referring expressions fill the specified roles (thrower and throwee) of the verbal constituent's meaning pole.

[1] These features play a key role in the acquisition of argument structure and grammatical markers; we return to this issue later.

Input: modeling the child learning scenario

Children entering the two-word stage (typically toward the end of the second year) are relatively savvy event participants, having developed a wealth of structured knowledge about the participant roles involved in different events and the kinds of entities likely to fill them (Nelson, 1996; Tomasello, 1992). Their single-word vocabularies typically include names for familiar people and objects, as well as some words for actions. They make use of pragmatic knowledge and joint attention to infer both communicative intentions (Tomasello, 1995) and subtle lexical distinctions (Bloom, 2000), and often respond appropriately to multi-word comments and queries from their parents even in the single-word stage (Bloom, 1973). That is, children can robustly interpret utterances beyond their productive abilities, using (incomplete) lingustic knowledge and relatively sophisticated inference abilities.

These findings suggest that grammar learning may, rather than suffer from the poverty of the stimulus, instead capitalize on the opulence of the substrate. Our learning model thus assumes an ontology of known concepts and an initial lexicon of constructions, represented in ECG. Input data reflects the child's ability to perceive an utterance with a particular intonational contour and segment it into a sequence of word forms, and to pragmatically infer the relevant participants and events in the accompanying situation, as shown in the example input below, where boxed index numbers indicate identification links between participants:

$$
\left[\begin{array}{ll}
\text{Form :} & \left[\begin{array}{l}\text{utterance : throw the ball}\\ \text{intonation : falling}\end{array}\right]\\
\multicolumn{2}{l}{\text{Participants : Mother } \boxed{0}\text{, Naomi } \boxed{1}\text{, Ball } \boxed{2}}\\
\text{Scene :} & \left[\begin{array}{l}\text{Throw}\\ \text{thrower : Naomi } \boxed{1}\\ \text{throwee : Ball } \boxed{2}\end{array}\right]\\
\text{Discourse :} & \left[\begin{array}{l}\text{speaker : Mother } \boxed{0}\\ \text{addressee : Naomi } \boxed{1}\\ \text{speech act : imperative}\\ \text{activity : play}\\ \text{joint attention : Ball } \boxed{2}\end{array}\right]
\end{array}\right]
$$

The example input represents a discourse event in which the mother says "throw the ball" with falling intonation to the child (Naomi). We assume the child can infer (using pragmatic cues) that the corresponding main scene concerns a throwing event to be performed by the child on a particular ball attended to in context. Note here that the action is the inferred intent of the mother, and may or may not be carried out by the child. But the (intended) role-filler structure is assumed in our model to be inferrable in context and thus available to the learning mechanism.

Besides these assumptions, the learning model also draws on findings about the developmental course of construction learning. Early word combinations appear to be lexically specific, with a gradual transition to more general constructions (Tomasello, 1992); crosslinguistically they tend to relate to a small set of basic scenes (Slobin, 1985); and acquisition phenomena are sensitive to a number of usage-based considerations (Tomasello, 2003; Clark, 2003) such as the frequency with which a construction is encountered, the simplicity of its form and meaning, and how easily a particular utterance can be analyzed into its component constructions.

In sum, the model incorporates strong assumptions about the child's conceptual and lexical knowledge and pragmatic abilities, based on developmental evidence. Relatively weak assumptions are made about innate syntactic biases: the ECG formalism allows word order as a possible form constraint. Thus most of the learning bias comes from the meaning domain, and the constructional assumption that forms and meanings are linked.

Computational learning framework

We now describe a computational model of how constructions can be learned from experience. The input is a sequence of utterances paired with their meanings in context, as described in the last section. The learner has access to a language analysis process like that described earlier, which produces a (partial) interpretation of the input utterances based on the current (potentially incomplete) set of constructions. The learning task is then modeled as an incremental search through the space of possible grammars, where the learner adds new constructions on the basis of encountered data. As in the child learning situation, the goal of learning is to converge on an optimal set of constructions, i.e., a grammar that is both general enough to encompass significant novel data and specific enough to accurately predict previously seen data.

A suitable overarching computational framework for guiding the search is provided by the minimum description length (MDL) heuristic (Rissanen, 1978), which is used to find the optimal analysis of data in terms of (a) a compact representation of the data; and (b) a compact means of describing the original data in terms of the compressed representation. The MDL heuristic exploits a tradeoff between competing preferences for smaller grammars (encouraging generalization) and for simpler analyses of the data (encouraging the retention of specific/frequent constructions). This is an approximation of the same tradeoff exploited in previous work applying Bayesian model merging to learning verbs (Bailey, 1997) and context-free grammars (Stolcke, 1994). We extend these approaches to handle the relational structures of the ECG formalism and the process-based assumptions of the model.

Learning strategies. The model may acquire new constructional mappings in two ways:

relational mapping New relational map(s) are formed to account for form-meaning mappings present in the input but unexplained by the current grammar.

reorganization Regularities across known constructions are exploited, either to merge two similar constructions into a more general construction, or to compose two constructions that cooccur frequently into a single construction.

Each construction is also associated with a weight that is incremented as a result of its successful use in analysis.

Algorithms for these operations are given elsewhere (Chang and Maia, 2001; Chang, 2004); relational mapping plays the most crucial role in proposing new relational constraints among constituents and will be illustrated in more detail in the next section.

Evaluating grammar cost. The strategies above provide means for updating the current grammar; the model must then determine which update is optimal at any point in learning, according to some length-based evaluation criterion. We use an approximation of the Bayesian posterior probability of the grammar G given the data D that we call the *cost* of G:

$$\begin{aligned}
\text{cost}(G|D) &= m \cdot \text{size}(G) + n \cdot \text{cost}(D|G) \\
\text{size}(G) &= \sum_{c \in G} \text{size}(c) \\
\text{size}(c) &= n_c + r_c + \sum_{e \in c} \text{length}(e) \\
\text{cost}(D|G) &= \sum_{d \in D} \text{score}(d) \\
\text{score}(d) &= \sum_{x \in d} (\text{weight}_x + p \cdot \sum_{t \in x} |\text{type}_t|) \\
&\quad + \text{height}_d + \text{semfit}_d
\end{aligned}$$

where m and n are learning parameters that control the relative bias toward model simplicity and data compactness. The size(G) is the sum over the size of each construction c in the grammar (n_c is the number of constituents in c, r_c is the number of constraints in c, and each element reference e in c has a length, measured as slot chain length). The cost (complexity) of the data D given G is the sum of the analysis scores of each input token d using G. This score sums over the constructions x in the analysis of d, where weight$_x$ reflects relative (in)frequency, $|\text{type}_t|$ (where t ranges over the constituents of x) denotes the number of ontology items of type t (i.e., the number of alternative fillers for the constituent), summed over all the constituents in the analysis and discounted by parameter p. The score also includes terms for the height of the derivation graph and the semantic fit provided by the analyzer as a measure of semantic coherence.

These criteria favor constructions that are simply described (relative to the available meaning representations and the current set of constructions), frequently useful in analysis, and specific to the data encountered.

Learning from meaning in context

This section describes in greater detail the integration of the learning model with an implemented construction analyzer (Bryant, 2003). We illustrate the analyzer-learner interaction with an example based on the input data shown earlier.

Constructional analysis. On encountering new data, the learner first calls a construction analyzer designed to perform the analysis process described earlier (Bryant, 2003).

The analyzer consists of a set of *construction recognizers* that recognize the input forms of each construction and check whether the relevant semantic constraints are satisfied. The analyzer draws on partial parsing techniques so that utterances not fully covered by known constructions can nevertheless yield partially filled in semantic specifications. Moreover, unknown forms in the input can be skipped, allowing quite simple constructions to provide at least skeletal interpretations of more complex utterances.

In the example, we assume the current grammar includes lexical constructions for *throw* and *ball*, but no word combinations or construction for the article *the*. The utterance "throw the ball" at this stage produces a semspec containing two schemas, corresponding to the meanings of the two recognized constructions, but no associations between them:

```
SCHEMA13 (Ball)
SCHEMA3 (Throw)
    thrower: SCHEMA4 (Human)
    throwee: SCHEMA8 (Physical-Object)
```

Here, `SCHEMA13` corresponds to the meaning pole of the BALL construction, and `SCHEMA3` corresponds to the meaning pole of the THROW construction.

Resolution. We extended the existing analyzer with a resolution procedure that matches the output semspec against the input context. Like other resolution (e.g. reference resolution) procedures, it relies on category/type constraints and (provisional) identification bindings. The resolution procedure attempts to unify each schema and constraint appearing in the semspec with some type-compatible entity or relation in the context. In the example, `SCHEMA13` resolves by this process to the salient Ball in the input, and `SCHEMA3` resolves to the Throw action in context.

Relational mapping. At this point the learner has a partial semspec that through resolution accounts for a subset of the information available in the input context description (namely, the presence of a throwing event and a ball). The learner now searches for a candidate relational mapping present in the input context but not accounted for by the semspec – that is, a form relation that is unused in the current analysis, paired with a meaning relation that is unaccounted for in the semspec. These relations must be structurally isomorphic, that is, their arguments must involve form and meaning poles of the same constituent constructions. In the example, the input includes a number of unexplained meaning relations – for example, the identity of the speaker and addressee, and both Throw schema roles. But only one of these – the binding between the throwee role and the ball – involves meanings that are also accounted for in the input, and for which there is a corresponding form relation over the form poles of the relevant constructions (i.e., an ordering relation between *throw* and *ball*).

The situation is depicted in Figure 2, where the input utterance-context pair are shown as form and meaning schemas and relations on either side of the figure. Constructions found by the analyzer are shown in the center, accounting for the form and meaning schemas drawn with solid heavy lines (i.e., the recognized input and produced semspec). The discovery of structurally isomorphic relations over the form and meaning poles of the two recognized constructions leads to the hypothesis of the new lexically specific THROW-BALL construction shown in the figure (with heavy dotted lines) and formally in Figure 3.

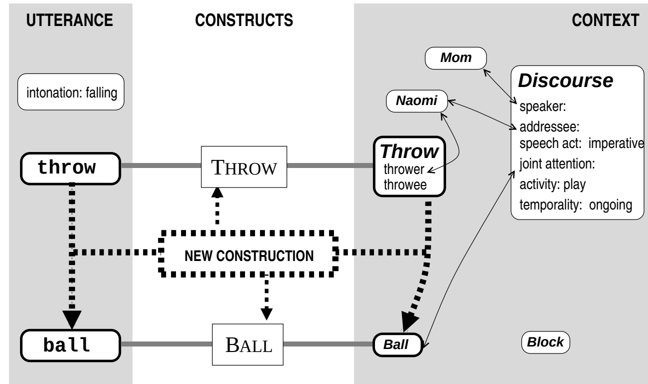

Figure 2: Relational mapping in the learning model for the utterance *throw (the) ball*. Heavy solid lines indicate structures matched during analysis; heavy dotted lines indicate the newly hypothesized mapping.

construction THROW-BALL
  constituents
    t1 : THROW
    t2 : BALL
  form
    $t1_f$ `before` $t2_f$
  meaning
    $t1_m$.throwee $\longleftrightarrow$ $t2_m$

Figure 3: Example learned construction: THROW-BALL learned from the utterance-context pair in Figure 2.

This example illustrates the simplest relational mapping strategy; the requirement of strictly isomorphic form and meaning relations can also be relaxed to allow more complex relational correspondences (expressed using longer constraints). All such mapping strategies are designed to discover how known constructions may fit together in larger structures, thus giving rise to constituent structure.

Once these structured (but lexically specific) constructions are learned, they are subject to reorganization, such that multiple constructions involving *throw* and a specific thrown object may be merged into a generalized *throw*-Object construction (contingent on the MDL learning criteria). We now explore how the model can learn patterns of this kind from a corpus of child-directed utterances.

Experiment: English verb island constructions

The construction learning model was tested in an experiment targeting the acquisition of lexically specific, or itembased, constructions; we focus on patterns centering on specific verbs. This task is of cognitive interest, since "verb island" constructions appear to be learned on independent trajectories (i.e., each verb forms its own "island" of organization (Tomasello, 1992; Tomasello, 2003)).

Input data. The training corpus for the experiment is a subset of the Sachs corpus of the CHILDES database of parent-child transcripts (Sachs, 1983; MacWhinney, 1991) annotated as part of a study of motion utterances (Dan I. Slobin, p.c.). The transcript data consists of parent and child utterances occurring during a joint background activity (e.g., a meal or play). All motion expressions were annotated with descriptions of the inferred speaker meaning and the surrounding discourse and situational context. We used a subset of this corpus containing 829 labeled motion-related child-directed utterances spanning the child's development from 1;3 through 2;6, during which the child makes the transition from the single-word stage. Parental utterances were extracted into input data of the form shown above.

Evaluation criteria. The goal of language learning in our framework is to improve language understanding. We thus defined a quantitative measure intended to gauge how new constructions incrementally improve the model's comprehensive capacity. We defined a grammar G's *coverage* of data D as the percentage of total bindings b in the data (i.e., role-filler bindings relevant to the verb) included in its interpretation (semspec), and measured coverage at each stage of learning. The *throw* subset, for example, contains 45 bindings to the roles of the Throw schema (thrower, throwee, and goal location). At the start of learning, the model has no combinatorial constructions and can account for none of these, but as learning progresses, the model should learn constructions that allow it to cover increasingly more of these bindings.

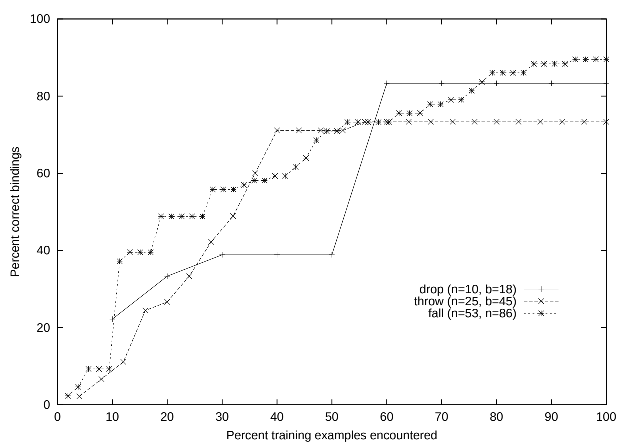

Figure 4: Incremental coverage for three verb islands. (Graphs are scaled relative to subcorpus size.)

Results. Figure 4 shows results for three verb islands: *drop* (n=10 examples), *throw* (n=25), and *fall* (n=50); other verbs followed similar patterns. In all cases coverage gradually improved over the course of learning, as expected, and the model was able to account for a majority of the bindings in the data relatively quickly. But as shown by these examples, the particular learning trajectories were distinct: *throw* constructions show a gradual build-up before plateauing; *fall* has a more fitful climb that seems to converge at an upper bound; and *drop* has an even more jagged rise. A possible explanation for some of these differences may lie in pragmatic differences: *throw* has a much higher percentage of imperative utterances than *fall* (since throwing is pragmatically more likely to be done on command). The relational mapping strategy used in the experiment misses the association of an imperative speech-act with unexpressed agent, which has a more pronounced effect on the learning of *throw* constructions.

Also as expected, the earliest constructions are combinations of specific words (e.g, *throw-ball*, *throw-frisbee*, *you-throw*), giving rise later in learning to more general constructions (e.g., *throw*-Object and Agent-*throw*). Figure 5 shows the number of each type learned.

	lexical	general	total
drop	5	1	6
throw	11	4	15
fall	21	9	30

Figure 5: Number of constructions learned for each verb, including both fully lexically specific constructions and verb island constructions with at least one generalized argument.

Discussion. Despite the small corpus sizes, the results are indicative of the model's ability to acquire useful verb-based constructions. Differences in verb learning lend support to the verb island hypothesis and illustrate how the particular semantic, pragmatic and statistical properties of different verbs can affect their learning course.

Case study: Russian

The verb island experiment demonstrates the model's ability to acquire constituent structure, an essential step in moving beyond lexical items. But the child's learning scenario may be significantly more complicated. We briefly consider some problems that arise for learners of comparable Russian constructions and how the model addresses them.

In Russian, casemarkers suffixed on nouns indicate the participant role played by their associated referents. Word order is thus highly variable: *malchik brosaet devochk-e myach* (boy-NOM throw-3s girl-DAT ball-ACC) and *devochk-e brosaet myach malchik* (girl-DAT throw-3s ball-ACC boy-NOM) have the same participant structure, glossed as 'boy throws ball to girl'. Moreover, the same marker may be ambiguous over multiple class/case combinations (e.g., *-a* indicates either Feminine-I/NOM or Masculine-Animate/ACC).

Flexible word order does not in itself pose an obstacle to the model. Deferring nominal morphology for the moment (see below), the first multi-word constructions learned by the model (via relational mapping) are, like their English equivalents, both verb-specific and fixed-order (e.g., one for each of the examples above). During construction reorganization, the model seeks candidates for merging that are similar in both meaning and form; separate fixed-order constructions involving the same constituents with equivalent participant structures are prime candidates. Generalizing over these constructions leads to a new construction that contains all the shared structure of the original constructions, omitting in this case the order constraints.

Morphological constructions are similar to word combinations in involving constituency, though word-internal. The main difference is that casemarkers do not occur independently of their nominal contexts, and are first learned as part of an unstructured larger unit. Thus the relational mapping strategy for learning constituent structure cannot apply directly. We assume, however, that over time the child is able to segment words into stems and endings, based on general pattern-detection mechanisms (Peters, 1985). Then the model can merge multiple constructions with the same stem and different endings (e.g., merging *devochk-e* (girl-DAT) and *devochk-a* (girl-NOM) yields a stem *devochk-* with no participant role specified). Similarly, a particular casemarker occurring on different stems (but the same verbal context) can be merged to yield a suffix construction whose meaning pole is associated with a specific participant role (or multiple roles, since polysemous markers are allowed). The resulting stem and casemarker constructions may then serve as constituents for larger morphological constructions.

Conclusion

The work described in this paper are best characterized as first steps toward concrete computational validation of our broad research paradigm. The model is intended to offer a detailed picture of the pivotal role meaning in context plays in the acquisition of grammar. It draws on evidence from across the cognitive spectrum arguing for a construction-based grammar formalism, extensive prior knowledge, and a data-driven, incremental learning course.

We have concentrated on the acquisition of constituent structure, as demonstrated by the verb island learning experiment. Note that we have not addressed how the model learns constructions that depend on more general semantic categories; these include both general argument structure constructions corresponding to basic scenes (caused motion, manipulative activity, etc.), and casemarking constructions that generalize across verbs. These categories are not assumed to be universal, but rather must be learned based on the fine-grained semantic structure available in the ECG representation. In ongoing work we are investigating the conditions and assumptions that allow such constructions to emerge. We are also exploring the relative rates of acquisitions of different classes of verbs and continuing to test the robustness of the model to crosslinguistic data.

Acknowledgments

Thanks to Srini Narayanan, Eva Mok, Shweta Narayan, other members of the ICSI Neural Theory of Language group, and anonymous reviewers for useful feedback.

References

Bailey, D. R. (1997). *When Push Comes to Shove: A Computational Model of the Role of Motor Control in the Acquisition of Action Verbs*. PhD thesis, University of California at Berkeley.

Bergen, B. K. and Chang, N. (in press). Simulation-based language understanding in Embodied Construction Grammar. In *Construction Grammar(s): Cognitive and Cross-language dimensions*. John Benjamins.

Bloom, L. (1973). *One word at a time: the use of single word utterances before syntax*. Mouton & Co., The Hague.

Bloom, P. (2000). *How Children Learn the Meanings of Words*. MIT Press, Cambridge, MA.

Bryant, J. (2003). Constructional analysis. Master's thesis, University of California at Berkeley.

Chang, N. (2004). *Constructing Grammar: A computational model of the emergence of early constructions*. PhD thesis, University of California at Berkeley.

Chang, N., Feldman, J., Porzel, R., and Sanders, K. (2002). Scaling cognitive linguistics: Formalisms for language understanding. In *Proc. 1st International Workshop on Scalable Natural Language Understanding*, Heidelberg, Germany.

Chang, N. C. and Maia, T. V. (2001). Learning grammatical constructions. In *Proc. 23rd Cognitive Science Society Conference*, pages 176–181.

Clark, E. V. (2003). *First Language Acquisition*. Cambridge University Press, Cambridge, UK.

Fillmore, C. and Kay, P. (1999). *Construction grammar*. CSLI, Stanford, CA. To appear.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar, Vol. 1*. Stanford University Press.

MacWhinney, B. (1991). *The CHILDES project: Tools for analyzing talk*. Erlbaum, Hillsdale, NJ.

Nelson, K. (1996). *Language in Cognitive Development: The Emergence of the Mediated Mind*. Cambridge University Press, Cambridge, U.K.

Peters, A. M. (1985). *Language Segmentation: Operating Principles for the Perception and Analysis of Language*. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14:465–471.

Sachs, J. (1983). *Talking about the there and then: the emergence of displaced reference in parent-child discourse*, pages 1–28. Lawrence Erlbaum Associates.

Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In Slobin, D. I., editor, *Theoretical Issues*, volume 2 of *The Crosslinguistic Study of Language Acquisition*, chapter 15. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Stolcke, A. (1994). *Bayesian Learning of Probabilistic Language Models*. PhD thesis, Computer Science Division, University of California at Berkeley.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge University Press, Cambridge, UK.

Tomasello, M. (1995). Joint attention as social cognition. In P., C. M., editor, *Joint attention: Its origins and role in development*. Lawrence Erlbaum Associates, Hillsdale, NJ. Educ/Psych BF720.A85.J65 1995.

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