Title
Learning Grammatical Constructions

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 23(23)

ISSN
1069-7977

Authors
Chang, Nancy C.
Maia, Tiago V.

Publication Date
2001

Peer reviewed

Learning Grammatical Constructions

Nancy C. Chang (nchang@icsi.berkeley.edu)
International Computer Science Institute
1947 Center Street, Suite 600, Berkeley, CA 94704 USA

Tiago V. Maia (tmaia@cse.buffalo.edu)
State University of New York at Buffalo
226 Bell Hall, Buffalo, NY 14260-2000 USA

Abstract

We describe a computational model of the acquisition of early grammatical constructions that exploits two essential features of the human grammar learner: significant prior conceptual and lexical knowledge, and sensitivity to the statistical properties of the input data. Such principles are shown to be useful and necessary for learning the structured mappings between form and meaning needed to represent phrasal and clausal constructions. We describe an algorithm based on Bayesian model merging that can induce a set of grammatical constructions based on simpler previously learned constructions (in the base case, lexical constructions) in combination with new utterance-situation pairs. The resulting model shows how cognitive and computational considerations can intersect to produce a course of learning consistent with data from studies of child language acquisition.

Introduction

This paper describes a model of grammar learning in which linguistic representations are grounded both in the conceptual world of the learner and in the statistical properties of the input. Precocity on both fronts has previously been exploited in models of lexical acquisition; we focus here on the shift from single words to word combinations and investigate the extent to which larger phrasal and clausal constructions can be learned using principles similar to those employed in word learning. Our model makes strong assumptions about prior knowledge – both ontological and linguistic – on the part of the learner, taking as both inspiration and constraint the course of development observed in crosslinguistic studies of child language acquisition.

After describing our assumptions, we address the representational complexities associated with larger grammatical constructions. In the framework of Construction Grammar (Goldberg, 1995), these constructions can, like single-word constructions, be viewed as mappings between the two domains of *form* and *meaning*, where form typically refers to the speech or text stream and meaning refers to a rich conceptual ontology. In particular, they also involve relations among multiple entities in both form (e.g., multiple words and/or phonological units) and meaning (multiple participants in a scene), as well as mappings across relations in these two domains. We introduce a simple formalism capable of representing such relational constraints.

The remainder of the paper casts the learning problem in terms of two interacting processes, construction hypothesis and construction reorganization, and presents an algorithm based on Bayesian model merging (Stolcke, 1994) that attempts to induce the set of constructions that best fits previously seen data and generalizes to new data. We conclude by discussing some of the broader implications of the model for language learning and use.

Conceptual and lexical prerequisites

Children learning their earliest word combinations bring considerable prior knowledge to the task. Our model of grammar learning makes several assumptions intended to capture this knowledge, falling into two broad categories: representational requirements for ontological knowledge; and the ability to acquire lexical mappings.

Infants inhabit a dynamic world of continuous percepts, and how they process and represent these fluid sensations remains poorly understood. By the time they are learning grammar, however, they have amassed a substantial repertoire of concepts corresponding to people, objects, settings and actions (Bloom, 1973; Bloom, 2000). They are also competent event participants who have acquired richly structured knowledge about how different entites can interact (Tomasello, 1992; Slobin, 1985), as well as sophisticated pragmatic skills that allow them to determine referential intent (Bloom, 2000).

Few computational models of word learning have addressed the general problem of how such sensorimotor and social-cultural savvy is acquired. Several models, however, have tackled the simpler problem of how labels (either speech or text) become statistically associated with concepts in restricted semantic domains, such as spatial relations (Regier, 1996), objects and attributes (Roy and Pentland, 1998), and actions (Bailey et al., 1997; Siskind, 2000). Such models assume either explicitly or implicitly that lexical items can be represented as *maps* (i.e., bidirectional associations) between representations of *form* and *meaning* that are acquired on the basis of input associations.[1] Most of these also produce word senses whose meanings exhibit category and similarity

[1]Typically, supervised or unsupervised training is used to induce word categories from sensorimotor input, which is described using continuous or discrete features; models vary in the degree of inductive bias present in the input feature space.

effects like those known to be pervasive in human cognition (Lakoff, 1987): concepts cluster into categories with prototype structure and graded category membership.

For our current spotlight on the acquisition of grammatical structures, we will make a similar set of simplifying assumptions. We do not attempt to model the complex reasoning and inference processes needed to infer the appropriate intended meaning of an utterance in context; rather, we take as input a representation of the inferred meaning in a given situational context. We also assume that lexical maps like those produced by the word-learning models described above are available as input to the grammar-learning process.

For present purposes, the precise variant of word learning is not at issue, as long as several representational requirements are met. Lexical maps should facilitate the identification of similar concepts and provide some basis for generalization. They must also be able to capture the kinds of event-based knowledge mentioned above: the meanings of many early words and constructions involve multiple entities interacting within the context of some unified event (Bloom, 1973) or basic scene (Slobin, 1985). Fortunately, these representational demands have long been recognized in the context of adult constructions, and semantic descriptions based on *frames* relating various participant *roles* have been developed by, e.g., the Berkeley FrameNet project (Baker et al., 1998). Frame-based representations can capture the relational structure of many concepts, including not only early sensorimotor knowledge but also aspects of the surrounding social and cultural context.

It will be convenient to represent frames in terms of individual role bindings: Throw.thrower:*Human* and Throw.throwee:*Object* together bind a Throw frame with a *Human* thrower acting on an *Object* throwee. Note that although this representation highlights relational structure and obscures lower-level features of the underlying concepts, both aspects of conceptual knowledge will be crucial to our approach to language learning.

In the current model, ontological knowledge is represented with an inheritance hierarchy in which frames are represented as feature structures (i.e., attribute-value matrices) and role bindings are handled by unification. Our initial set of constructions contains a number of lexical form-meaning maps, where for simplicity we further constrain these to be mappings from orthographic forms to feature-structure meanings, as in Bailey (1997).

We now turn to the representationally more complex case of grammatical constructions, before addressing how such constructions are learned.

Grammatical Constructions

We base our representations of grammatical knowledge on ideas from Construction Grammar (Goldberg, 1995) and Cognitive Grammar (Langacker, 1987). In these approaches, larger phrasal and clausal units are, like lexical constructions, pairings of form and meaning. A key observation in the Construction Grammar tradition is that the meaning of a sentence may not be strictly predictable from the meaning of its parts; the syntactic pattern itself may also contribute a particular conceptual framing. For example, the CAUSED-MOTION construction underlying *Pat sneezed the napkin off the table* imposes a causative reading on the typically non-causative verb *sneeze*, and the need for an agentive recipient in the DITRANSITIVE construction renders *Harry kicked the door the ball* somewhat anomalous.

On this account, syntactic patterns are inextricably linked with meaning, and grammaticality judgments are rightly influenced by semantic and pragmatic factors. The interpretation and acceptability of an utterance thus depends not only on well-formedness conditions but also on the structure of the language user's conceptual ontology and on the situational and discourse context.

The main representational complexity introduced with these multiword constructions is the possibility of structure in the form pole. As mentioned above, although individual lexical items can evoke complex frames with multiple participant roles (e.g., *bye-bye*, *baseball*), the actual mapping between the form and meaning pole is necessarily straightforward. With multiple form units available, however, additional structures arise, both within the form pole itself and, more significantly, in the *relational correlations* between the form and meaning poles.[2] That is, a multiword construction may involve a more complex, *structured map* between its form and meaning poles, with maps between form and meaning relations whose arguments are also mapped.

In addition to the sound patterns of individual words, the form pole includes intonational contours, morphological inflections and word order. As with single words, the meaning pole encompasses the much larger set of frame-based conceptual knowledge. The constructional mapping between the two domains typically consists of a set of form relations (such as word order) corresponding to a set of meaning relations (such as role-filler bindings).

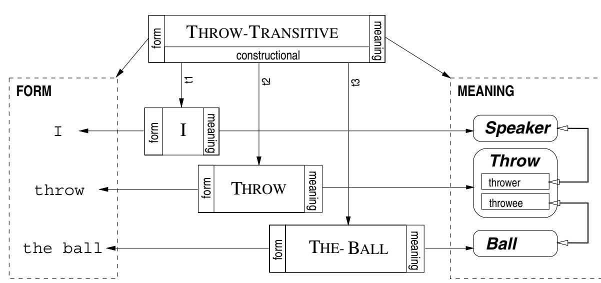

Figure 1: A constructional analysis of the sentence, *I throw the ball*, with form elements at left, meaning elements at right and some constituent constructions linking the two domains in the center.

As an example, Figure 1 gives an iconic representation of some of the possible constructions involved in an analy-

[2]See Gasser and Colunga (2000) for arguments that the ability to represent relational correlations underlies infants' reputed aptitude for statistically driven learning of concrete and abstract patterns.

sis of *I throw the ball*. The lexical constructions for I, THROW and THE-BALL[3] all have simple poles of both form and meaning. But besides the individual words and concepts involved in the utterance, we have several word order relationships (not explicitly shown in the diagram) that can be detected in the form domain, and bindings between the roles associated with Throw and other semantic entities (as denoted by the double-headed arrows within the meaning domain). Finally, the larger clausal construction (in this case, a verb-specific one) has constituent constructions, each of which is filled by a different lexical construction.[4] Crucially, the clausal construction serves to associate the specified form relations with the specified meaning relations, where the arguments of these relations are already linked by existing (lexical) maps. For example, the fact that the I construction's form pole comes *before* the THROW construction's form pole means that the meaning pole of I (i.e., the speaker in the situation) fills the thrower role in the Throw frame.

A more formal representation of the THROW-TRANSITIVE construction is given in Figure 2. For current purposes, it is sufficient to note that this representation captures the constituent constructions, as well as constraints on its formal, semantic and constructional elements. Each constituent has an alias used locally to refer to it, and subscripts f and m are used to denote the constituent's form and meaning poles, respectively. A designation constraint (in Langacker's (1987) sense) specifies a meaning type for the overall construction.

construction THROW-TRANSITIVE
  constituents:
    construct t1 **of meaning type** *Human*
    construct t2 **of type** THROW
    construct t3 **of meaning type** *Object*
  formal constraints:
    $t1_f$ *before* $t2_f$
    $t2_f$ *before* $t3_f$
  semantic constraints:
    $t2_m$.thrower $\longleftrightarrow$ $t1_m$
    $t2_m$.throwee $\longleftrightarrow$ $t3_m$
designates $t2_m$

Figure 2: Formal representation of the THROW-TRANSITIVE construction, with separate blocks listing constituent constructions, formal constraints (e.g., word order) and semantic constraints (role bindings).

Although this brief discussion necessarily fails to do justice to Construction Grammar and related work, we hope that it nevertheless conveys the essential representational demands on the structures to be learned.

[3]The definite determiner *the* explicitly depends on a representation of the situational and discourse context that supports reference resolution. For simplicity, we will ignore the internal structure of "the ball" and treat it as an unstructured unit.

[4]This example, like the rest of those in the paper, is based on utterances from the CHILDES corpus (MacWhinney, 1991) of child-language interaction.

Learning Constructions

We can now specify our construction learning task: Given an initial set of constructions C and a sequence of new training examples, find the best set of constructions C' to fit the seen data and generalize to new data. In accord with our discussion of conceptual prerequisites, a training example is taken to consist of an utterance paired with a representation of a situation, where the former is a sequence of familiar and novel forms, and the latter a set of frame-based conceptual entities and role bindings representing the corresponding scene.

Previous work on Bayesian model merging (Stolcke, 1994; Bailey et al., 1997) provides a suitable starting point. In that framework, training data is first incorporated, with each example stored as an independent model. Similar models are then merged (and thereby generalized); the resulting drop in likelihood is balanced against an increase in the prior. Merging continues until the posterior probability of the model given the data decreases. In the case of probabilistic grammars (Stolcke and Omohundro, 1994), structural priors favor grammars with shorter descriptions, and likelihood is based on the probability of generating the data using the grammar.

We apply a similar strategy to our current task by casting it as a search through the space of possible grammars (or sets of constructions), where the grammars are evaluated using Bayesian criteria. The operations on the set of constructions (merging and composition, described below as **reorganization** processes) extend previous operations to handle relational structures. Similarly, the evaluation criteria need not change significantly for the construction learning case: structural priors favor grammars with fewer, more general constructions that compactly encode seen data; this measure combats the inevitable corresponding drop in the likelihood of generating the seen data using the grammar. Again, the learning algorithm attempts to maximize the posterior probability of the set of constructions given the data.[5]

The main complication requiring a departure from previous work is the need to hypothesize structured maps between form and meaning like those described in the previous section. Essentially, incorporating new data involves both the **analysis** of an utterance according to known constructions and the **hypothesis** of a new construction to account for any new mappings present in the data. These processes, described below, are based on the assumption that the learner expects correlations between what is heard (the utterance) and what is perceived (the situation).[6] Some of these correlations have already been encoded and thus accounted for by previ-

[5]Model merging conducts a best-first search through the hypothesis space based on available merges. It is thus is not guaranteed to find the best model, which would require searching through an exponential number of possible grammars.

[6]The task as defined here casts the learner as primarily comprehending (and not producing) grammatical utterances. The current model does not address production-based means of hypothesizing and reinforcing constructions, which would be included in a more complete model.

ously learned constructions; the tendency to try to account for the remaining ones leads to the formation of new constructions. In other words, what is learned depends directly on what remains to be explained. The identification of the mappings between an utterance and a situation that are predicted by known constructions can be seen as a precursor to language comprehension, in which the same mappings actively evoke meanings not present in the situation. Both require the learner to have an analysis procedure that determines which constructions are potentially relevant, given the utterance, and, by checking their constraints in context, finds the best-fitting subset of those.

Once the predictable mappings have been explained away, the learner must have a procedure for determining which new mappings may best account for new data. The mappings we target here are, as described in the previous section, relational. It is important to note that a relational mapping must hold across arguments that are themselves *constructionally correlated*. That is, mappings between arguments must be in place before higher-order mappings can be acquired. Thus the primary candidates for relational mappings will be relations over elements whose form-meaning mapping has already been established. This requirement may also be viewed as narrowing the search space to those relations that are deemed *relevant* to the current situation, as indicated by their connection to already recognized forms and their mapped meanings.

Details of these procedures are best illustrated by example. Consider the utterance U_1 = "you throw a ball" spoken to a child throwing a ball. The situation S consists of entities S_e and relations S_r; the latter includes role bindings between pairs of entities, as well as attributes of individual entities. In this case, S_e includes the child, the thrown ball and the throwing action, as well as potentially many other entities, such as other objects in the immediate context or the parent making the statement: $S_e = \{$Self,Ball,Block,Throw,Mother,...$\}$. Relational bindings include those encoded by the Throw frame, as well as other properties and relations: $S_r = \{$Throw.thrower:Self, Throw.throwee:Ball, Ball.Color:Yellow, ...$\}$.

In the following sections we describe what the learner might do upon encountering this example, given an existing set of constructions C that has lexical entries for BALL,THROW,BLOCK,YOU,SHE, etc., as well as a two-word THROW-BALL construction associating the `before(throw,ball)` word-order constraint with the binding of Ball to the throwee role of the Throw frame.

Construction analysis and hypothesis

Given this information, the analysis algorithm in Figure 3 first extracts the set F_{known} = {`you`,`throw`,`ball`}, which serves to cue constructions whose form pole includes or may be instantiated by any of these units. In this case, C_{cued} = {YOU,THROW,BALL,THROW-BALL}. Next, the constraints specified by these constructions must be matched against the input utterance and situation. The form constraints for all the lexical constructions are trivially satisfied, and in this case each also happens to map to a meaning element present in S.[7] Checking the form and meaning constraints of the THROW-BALL construction is also trivial: all relations of interest are directly available in the input utterance and situation.[8]

Analyze utterance. Given utterance U in situation S and current constructions C, produce best-fitting analysis A:

1. Extract the set F_{known} of familiar form units from U, and use them to cue the set C_{cued} of constructions.
2. Find the best-fit analysis $A = < C_A, F_A, M_A >$, where C_A is the best-fitting subset of C_{cued} for utterance U in situation S, F_A is the set of form units and relations in U used in C_A, and M_A is the set of meaning elements and bindings in S accounted for by C_A.

   A has associated cost $Cost_A$ providing a quantitative measure of how well A accounts for U in S.
3. Reward constructions in C_A; penalize cued but unused constructions, i.e., those in $C_{cued} \setminus C_A$.

Figure 3: Construction analysis.

In the eventual best-fitting analysis A, the constructions used are C_A ={YOU,THROW,BALL,THROW-BALL}, which cover the forms and form relations in F_A = {`you,throw,ball,before(throw,ball)`} and map the meanings and meaning relations in M_A = {Self,Throw,Ball,Throw.throwee:Ball}. (Remaining unused in this analysis is the form `a`.)

We proceed with our example by applying the procedure shown in Figure 4 to hypothesize a new construction. All form relations and meaning bindings, respectively, that are *relevant* to the form and meaning entities involved in the analysis are extracted as, respectively, F_{rel} = {`before(you,throw)`, `before(throw,ball)`, `before(you,ball)`} and M_{rel} = {Throw.thrower:Self, Throw.throwee:Ball}; the *remainder* of these not used in the analysis are F_{rem} = {`before(you,throw)`, `before(you,ball)`} and M_{rem} = {Throw.thrower:Self}. The potential construction C_{pot} derived by replacing terms with constructional references is made up of form pole {before(YOU$_f$,THROW$_f$),before(YOU$_f$,BALL$_f$)} and meaning pole {THROW$_m$.thrower:YOU$_m$}. The final

[7] We assume the YOU construction is a context-dependent construction that in this situation maps to the child (Self).

[8] The analysis algorithm can be viewed as a version of parsing allowing both form and meaning constraints. More sophisticated techniques are needed for the many complications that arise in adult language – category constraints on roles may apply only weakly, or may be overridden by the use of metaphor or context. For the cases relevant here, however, we assume that constraints are simple and few enough that exhaustive search should suffice, so we omit the details about how cueing constructions, checking constraints and finding the best-fitting analysis proceed.

construction C_{U_1} is obtained by retaining only those relations in C_{pot} that hold over correlated arguments:

$(\{\texttt{before}(\mathrm{YOU}_f,\mathrm{THROW}_f)\}, \{\mathrm{THROW}_m.\text{thrower:}\mathrm{YOU}_m\})$

Hypothesize construction. Given analysis A of utterance U in situation S, hypothesize new construction C_U linking correlated but unused form and meaning relations:

1. Find the set F_{rel} of form relations in U that hold between the forms in the analysis F_A, and the set M_{rel} of meaning relations in S that hold between the mapped meaning elements in M_A.
2. Find the set $F_{rem} = F_{rel} \setminus F_A$ of relevant form relations that remain unused in A, and the set $M_{rem} = M_{rel} \setminus M_A$ of relevant meaning relations that remain unmapped in A. Create a potential construction C_{pot} = (F_{rem}, M_{rem}), replacing terms with references to constructions in C_A where possible.
3. Create a new construction C_U consisting of pairs of form-meaning relations from C_{pot} whose arguments are constructionally related.
4. Reanalyze utterance using $\mathcal{C} \cup \{C_U\}$, producing a new analysis A' with cost $Cost_{A'}$. Incorporate C_U into $\mathcal{C}$ if $Cost_A - Cost_{A'} \geq MinImprovement$; else put C_U in pool of potential constructions.
5. If U contains any unknown form units or relations, add (U,S) to the pool of unexplained data.

Figure 4: Construction hypothesis.

At this point, the utility of C_{U_1} can be evaluated by reanalyzing U_1 to ensure a minimum reduction of the analysis cost. As noted in Step 4 of Figure 4, a construction not meeting this criterion is held back from incorporation into $\mathcal{C}$. It is possible, however, that further examples will render it useful, so it is maintained as a candidate construction. Similarly, Step 5 is concerned with maintaining a pool of examples involving unexplained form elements, such as the unfamiliar article a in this example. Further examples involving similar units may together lead to the correct generalization, through the reorganization process to which we now turn.

Reorganizing constructions

The analysis-hypothesis process just described provides the basis for incorporating new examples into the set of constructions. A separate process that takes place in parallel is the data-driven, bottom-up reorganization of the set of constructions based on similarities among and co-occurrences of multiple constructions. Figure 5 gives a high-level description of this process; we refrain from delving into too much detail here, since these processes are closely related to those described for other generalization problems (Stolcke, 1994; Bailey et al., 1997).

Continuing our example, let us assume that the utterance U_2 = "she's throwing a frisbee" is later encountered in conjunction with an appropriate scene, with similar results: in this case, both the unfamiliar inflections and the article are ignored; the meanings are mapped; and constraints with appropriate correlations are found, resulting in the hypothesis of the construction C_{U_2}:

Reorganize constructions. Reorganize $\mathcal{C}$ to consolidate similar and co-occurring constructions:

1. Find potential construction pairs to consolidate.
   - **Merge** constructions involving correlated relational mappings over one or more pairs of similar constituents, basing similarity judgments and type generalizations on the conceptual ontology.
   - **Compose** frequently co-occurring constructions with compatible constraints.
2. Evaluate how possible merge/compose operations affect the posterior probability of $\mathcal{C}$ on seen data, performing operations on a greedy, best-first basis.

Figure 5: Construction reorganization.

$(\{\texttt{before}(\mathrm{SHE}_f,\mathrm{THROW}_f)\}, \{\mathrm{THROW}_m.\text{thrower:}\mathrm{SHE}_m\})$

C_{U_1} and C_{U_2} bear some obvious similarities: both constructions involve the same form relations and meaning bindings, which hold of the same constituent construction THROW. Moreover, the other constituent is filled in the two cases by SHE and YOU. As emphasized in our discussion of conceptual representations, a key requirement is that the meaning poles of these two constructions reflect their high degree of similarity.[9] The overall similarity between the two constructions can lead to a merge of the constructional constituents, resulting in the merged construction:

$$(\{\texttt{before}(\mathbf{h}_f,\mathrm{THROW}_f)\},\{\mathrm{THROW}_m.\text{thrower:}\mathbf{h}_m\})$$

where $\mathbf{h}$ is a variable over a construction constrained to have a *Human* meaning pole (where *Human* is a generalization over the two merged constituents). A similar process, given appropriate data, could produce the generalized mapping:

$$(\{\texttt{before}(\mathrm{THROW}_f,\mathbf{o}_f)\},\{\mathrm{THROW}_m.\text{throwee:}\mathbf{o}_m\})$$

where $\mathbf{o}$ is constrained to have an Object meaning pole.[10]

Besides merging based on similarity, constructions may also be composed based on co-occurrence. For example, the generalized *Human*-THROW and THROW-*Object* constructions just described are likely to occur in many analyses in which they share the THROW constituent. Since they have compatible constraints in both form and meaning (in the latter case even based on the same conceptual Throw frame), repeated co-occurrence eventually leads to the formation of a larger construction that includes all three constituents:

[9] The precise manner by which this is indicated is not at issue. For instance, a type hierarchy could measure the distance between the two concepts, while a feature-based representation might look for common featural descriptions.

[10] Although not further discussed here, examples with unexplained forms (such as the a in U_1 and U_2) may also undergo merging, leading to the emergence of common meanings.

$(\{\texttt{before}(\mathbf{h}_f,\text{THROW}_f),\texttt{before}(\text{THROW}_f,\mathbf{o}_f)\},$
$\{\text{THROW}_m.\text{thrower}:\mathbf{h}_m,\text{THROW}_m.\text{throwee}:\mathbf{o}_m\})$

Note that both generalization operations we describe are, like the hypothesis procedure, merely means of finding potential constructions, and are subject to the evaluation criteria mentioned earlier.

Discussion

We have described a model of the acquisition of grammatical constructions that attempts to capture insights from child language using the formal tools of machine learning. Methods previously applied to word learning are extended to handle grammatical constructions, which are claimed to require new representational and algorithmic machinery.

The model is compatible to the extent possible with evidence from child language acquisition. In particular, the tight integration proposed between comprehension and learning is consistent with usage-based theories of language acquisition: new constructions are hypothesized to capture form-meaning correlations not covered by known constructions, in a manner akin to some of Slobin's (1985) Operating Principles for mapping. The data-driven progression from lexically specific to more abstract constructions is also consistent with Tomasello's (1992) observation that the earliest verb-argument constructions are lexically specific and give way only later to more general argument structure constructions.

More broadly, since the algorithm produces constructions based on any utterance-situation pair and existing set of constructions represented as described above, it can apply equally well for more advanced stages of language development, when the learner has more sophisticated meaning representations and more complex constructions. The potential continuity between early language acquisition and lifelong constructional reorganization offers hope for the modeling of adaptive language understanding systems, human and otherwise.

Acknowledgments

We are grateful for the input and influence of Jerry Feldman and the NTL group at ICSI, as well as comments from the reviewers; opinions and errors remain ours alone. This work was supported in part by an IBM Research Fellowship granted to the first author, and a Praxis XXI Fellowship from the Portuguese "Fundação para a Ciência e a Tecnologia" to the second author.

References

Bailey, D. R., Feldman, J. A., Narayanan, S., and Lakoff, G. (1997). Modeling embodied lexical development. In *Proceedings of the 19th Cognitive Science Society Conference*.

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The Berkeley FrameNet P roject. In *Proceedings of COLING-ACL*, Montreal, Canada.

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.

Gasser, M. and Colunga, E. (2000). Babies, variables, and relational correlations. In *Proceedings of the Cognitive Science Society Conference*, volume 22, pages 182–187.

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

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. 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.

Regier, T. (1996). *The Human Semantic Potential*. MIT Press, Cambridge, MA.

Roy, D. and Pentland, A. (1998). Learning audio-visually grounded words from natural input. In *Proc. AAAI workshop, Grounding Word Meaning*.

Siskind, J. M. (2000). Visual event classification via force dynamics. In *Proc. AAAI-2000*, pages 149–155.

Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In Slobin, D. I., editor, *The Crosslinguistic Study of Language Acquisition*, volume 2, chapter 15. Erlbaum, NJ.

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

Stolcke, A. and Omohundro, S. (1994). Inducing probabilistic grammars by Bayesian model merging. In Carrasco, R. C. and Oncina, J., editors, *Grammatical Inference and Applications*, pages 106–118. Springer-Verlag.

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