Constructions in action:
Learning motion event constructions for natural language understanding

Nancy Chang

University of California, Berkeley and International Computer Science Institute
1947 Center Street, Suite 600, Berkeley, CA 94704
nchang@cs.berkeley.edu

The problem

The task of building agents that communicate using natural language has proven to be one of the most enduring challenges in AI. People clearly exploit rich sources of knowledge during communication: knowledge about the world, the current situational and discourse context, and the goals and intentions of their interlocutors all interact with more specifically grammatical knowledge about the sounds and patterns of a particular language. In light of such complications, it is perhaps not surprising that natural language processing applications typically neglect semantics in favor of more tractable syntactic analyses, increasingly based on statistical properties of very large corpora. Where meaning plays any significant role, it often relies on hand-crafted domain- and task-specific representations — representations that are difficult and time-consuming to extend beyond their original domains. As a result, semantically informed, general-purpose approaches to language processing have lagged conspicuously in the wake of advances in speech recognition and parsing technology.

The proposed thesis research addresses three of the underlying sources of this imbalance: (1) the need for a general formulation of the language understanding process that broadly incorporates semantics and context; (2) the need for richer representations of linguistic constructions to support such a framework; and (3) the need for a principled approach to the automatic acquisition of such constructions from examples. More generally, it explores and makes precise the idea that linguistic knowledge at all levels can be characterized as mappings between form and meaning. This view serves as the basis for a computational model of the acquisition and understanding of simple phrasal and clausal constructions, with particular focus on constructions used in the expression of motion events. These semantically rich constructions drive a system in which active representations are executed (or simulated) in the process of interpreting an utterance and drawing appropriate inferences. The principles underlying this work are designed to be cross-linguistically sound, to be verified by conducting experiments in languages including English, Spanish and Mandarin Chinese, and broadly applicable to a wide variety of natural language tasks.

Simulatation in language understanding

One goal of this thesis is to describe a general framework for language understanding that integrates various forms of knowledge — including linguistic knowledge, world knowledge, situational or contextual knowledge and embodied knowledge — while still making motivated distinctions among them. The approach taken here is based on the assumption that many utterances can be understood through the simulation of the appropriate actions or events in some model of the current world state. Moreover, the same underlying representations used for executing an action are assumed to be useful in processing and understanding language about that action.

In this framework, language serves as the bridge between surface cues (such as a sound or text sequence) and the much more complex conceptual mechanisms needed to carry out such a simulation. The actual simulations require dynamic representations, which in our framework are based on Petri nets (Murata 1989) and motivated in part by motor and perceptual systems (Narayanan 1997). We assume, however, that linguistic analysis need not be concerned with the details of simulation, which can be as fine-grained as the active representations allow. Rather, linguistic knowledge at all levels helps to determine the parameters of the simulation, in the form of a simulation specification. This specification compactly provides all the information needed to simulate the minimal scenario described by a sentence, such as what kind of simulation will be run and which other concepts interact in it.

For our purposes, linguistic constructions of all kinds, whether corresponding to lexical, phrasal or clausal units, play the crucial role of specifying precisely the relation between surface cues and simulation parameters. This view of linguistic knowledge elaborates but is consistent with work on Construction Grammar (Goldberg 1995) that has demonstrated that linguistic analyses must incorporate diverse types of information that have traditionally been segregated. A major component of the thesis is thus a formal representation for constructions that can be used to analyze input utterances in the framework just described.

Constructions and their representations

The idea that language involves associations between form (the acoustic signal, or textual input) and meaning (some relatively richer set of features) is neither new nor surprising. Indeed, such an assumption is regularly made without controversy in models of the acquisition and representation of individual words, including those for objects, spatial relations, and actions. Thus far,
however, work on larger phrasal and clausal structures has been oriented toward symbolic syntactic patterns (e.g., context-free grammars), giving semantics either no role at all or only an indirect and strictly compositional one based on the lexical semantics of particular verbs. In such systems, the meaning of these larger structures is typically assumed to be predictable from the meaning of its constituents.

The current work adopts Construction Grammar’s less strictly compositional framework, in which larger phrasal and clausal constructions can also be described, like words, as mappings between form and meaning. The main qualitative difference is that the features being associated may involve more complex structural descriptions that describe relationships between constituents. Overall, however, this view allows a uniform treatment of all kinds of linguistic knowledge, ranging from intonational contours, morphological inflection and lexical items to clausal patterns, idioms and even discourse-level structures.

Parameters of form consist primarily of phonological (or orthographic) cues, including features of intonation, words and their inflections, and word order (and punctuation, in the case of text input). Parameters of meaning tend to encompass a much larger set of possibilities, including event structure, sensorimotor control, force dynamics, attentional state, perspective-switching ability, and communicative and social goals. Not surprisingly, the form parameters have received far more scrutiny in the literature than the meaning parameters, which are complex and require more sophisticated representations than those typically employed in computational linguistics.

Evidence from cognitive linguistic analyses suggests, however, that many basic clausal constructions are concerned with only a restricted subset of these features, and that the grammatical function they serve is to relate other, possibly more complex, concepts inconsistent ways. This is precisely the division captured by the simulation-based framework for language understanding described above: the simulation specification provides a clean and narrow interface between the particular meaning parameters affected by each grammatical construction and the more elaborate representations needed to model actions in context.

**Learning constructions**

Although the framework for language analysis using semantically oriented constructions described so far may be useful from a descriptive perspective, the scale and complexity of the constructions needed to handle even a fraction of the idiosyncrasies of any natural language remains daunting. An integral part of the proposed research is thus to develop automatic methods for acquiring constructions from examples, where each example is an utterance paired with a semantic representation in the form of a simulation specification.

The model described here assumes some initial set of constructions corresponding to already learned individual words or unanalyzed phonological units. The learning problem is to hypothesize, based on the current inventory of constructions and new input, the best new set of constructions to fit the data and generalize to future input. We extend previous work using Bayesian model merging as the basis for generalization (Stolcke 1994; Bailey 1997) to include composition, subset selection, and extension as possible operations on the set of existing constructions. Since these operations are not sufficient for expressing constructions that include binding relations between entities, we propose additional operations that allow a constrained form of unification in which relations detectable from the form representation can be associated with relations over corresponding entities in the meaning representation.

Besides addressing an instance of the general problem of learning relations and correlations over previously learned associations, this approach to grammar learning is also intended as a cognitive model that can make tractable the modeling of empirical data, such as that in the CHILDES corpus of parent-child interactions (MacWhinney 1991). The use of relatively rich structures for meaning places most of the generalization power in the semantic domain, creating a tradeoff between the (mostly semantic) drive to generalize and the map-enforced need to adhere to specific form-based correlates. The resulting learning behavior accords well with developmental evidence that children acquire constructions on a verb-specific basis before generalizing to more abstract constructions ( Tomasello 1992), demonstrating how a richer approach to meaning representation can facilitate the acquisition and use of natural language for both human and artificial agents.

**References**


