# STRUCTURED CONNECTIONIST MODELS OF LANGUAGE, COGNITION AND ACTION

NANCY CHANG, JEROME FELDMAN AND SRINI NARAYANAN
*International Computer Science Institute, 1947 Center Street, Suite 600*
*Berkeley, CA 94704, USA*
*{nchang,jfeldman,snarayan}@icsi.berkeley.edu*

The Neural Theory of Language project aims to build structured connectionist models of language and cognition consistent with constraints from all domains and at all levels. These constraints include recent experimental evidence that details of neural computation and brain architecture play a crucial role in language processing. We focus in this paper on the computational level and explore the role of embodied representations and simulative inference in language understanding.

## 1. Introduction

The original promise of neural computation derived from its potential to serve as a mechanism for formulating and testing theories that link brain and behavior. Although subsequent work has addressed a variety of specific technical problems, it has devoted relatively less attention toward building general models of intelligent behavior, especially in the domain of language. In this paper we summarize some progress made on this front by the Neural Theory of Language (NTL) project[a], an interdisciplinary group at ICSI and UC Berkeley. We focus on our efforts to build models of language understanding consistent with biological, psychological, linguistic and computational constraints.

The intellectual base for the NTL enterprise is a synthesis of findings from cognitively motivated approaches to linguistics (Lakoff 1987, Lakoff & Johnson 1980) and structured connectionist modeling (Feldman & Ballard 1982), linked by a three-part *Embodiment Hypothesis*:

1. Many concepts are directly embodied in motor, perceptual and other neural structures.
2. All other concepts derive their inferential structure via mappings to these embodied structures.
3. Structured connectionist models provide a suitable computational formalism for such neurally grounded representations and mappings.

---

[a] Previously the $L_0$ group; see also Feldman *et al.* (1996) (a summary of early work), Feldman & Narayanan (2003), and http://www.icsi.berkeley.edu/NTL.

These ideas have been extended with a *Simulation Hypothesis*—that language understanding exploits many of the same structures used for action, perception, imagination, memory and other neurally grounded processes, and that language provides parameters for simulations using such embodied structures. We explore these hypotheses using methods and convergent constraints from several related disciplines, including direct imaging and behavioral experiments, theoretical developments in linguistics and computation, and system implementations that illustrate important behaviors and may have practical applications as well.

To our knowledge, structured connectionist models (SCMs) comprise the best formalism for capturing the computational relationships between neural activity and complex language and thought. In some cases, the connection between language and neural computation has been direct: Regier's (1996) model of the acquisition of spatial relations terms captured cognitive linguistic phenomena directly with SCMs incorporating aspects of the human visual system. In other cases we have exploited an intermediate computational modeling level as a bridge between descriptions of behavior (cognitive and linguistic phenomena) and the SCM level, which can in turn be mapped to detailed neural architectures. This approach to describing language using neural computation is analogous to other layered abstractions in science, such as using the intermediate level(s) of chemistry to link biology and physics.

As exemplified by the models discussed below, the computational level of description makes use of various formalisms (e.g., feature structures, unification, probabilistic belief networks, Petri nets). But crucially, these formalisms have principled realizations using SCMs. Bailey's (1997) model of the acquisition of hand action verbs, for example, used feature structures encoding parameters for *executing schemas* (or *x-schemas*, described below) capturing many features of motor control; both feature structures and x-schemas also have implementations using SCMs (Bailey 1997, Shastri *et al.* 1998). The intermediate computational level thus allows a convenient level of abstraction while grounding linguistic and cognitive theories in more detailed and biologically inspired models.

This paper focuses on recent efforts to explore the Simulation Hypothesis at the computational level. The basic model is shown in Figure 1. Crucial to our model is the notion of the *construction* as the basic unit of linguistic representation, adopted from cognitive and constructional approaches to grammar (Goldberg 1995, Fillmore & Kay 1999). Constructions are mappings between schematic representations of *form* (phonological schemas) and *meaning* (conceptual schemas), as depicted in the figure. For our purposes, constructions are further designed to supply parameters for simulations.

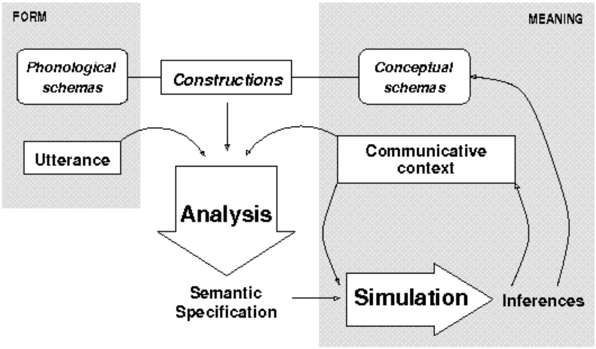

Figure 1. Simulation-based language understanding, consisting of an *analysis* process that draws on *constructions* (pairings of phonological schemas in the form domain with conceptual schemas in the meaning domain) to interpret an utterance in a communicative context. The result of analysis is a *semantic specification* that provides parameters for an active simulation.

Language understanding in this model proceeds in two phases. The *analysis* process determines which constructions are instantiated by a given utterance, drawing on linguistic knowledge (the set of known constructions and their associated forms and meanings), conceptual knowledge (entity and event types), and the current communicative context. The result of analysis is a *semantic specification* that provides the parameters for a *simulation* process that activates embodied conceptual structures represented as x-schemas, resulting in inferences that are then used to update the current context.

The paper is structured as follows. We begin with some linguistic and biological motivation for simulation (Section 1). We then elaborate on each component above: the construction-based grammatical formalism (Section 2); the analysis process linking surface forms with simulations (Section 3); the simulation process and our mechanism for simulation-based inference (Section 4), first proposed as part of a model of metaphorical inference in news stories (Narayanan 1999). Together these models show how embodied conceptual and linguistic structures can be integrated within a simulation-based framework to provide a common representational toolkit for language, cognition and action.

## 2. Evidence for simulation in language understanding

The idea that language draws on embodied structures is not new. Cognitive linguists (Lakoff 1987, Talmy 2000) have noted that patterns of sensorimotor experience, or *image schemas,* play a central role in semantic representation crosslinguistically. (For example, the prepositions *to* and *into* both involve motion of a trajector relative to a landmark, but *into* also involves containment.) This observation applies not only to literal language but also to abstract and metaphorical language (Lakoff & Johnson 1980). Recent neurobiological and behavioral findings also support the notion that perceptual and motor systems are activated during language understanding. (See also Bergen *et al.* 2004.)

Neurobiological evidence centers on experiments showing that areas of motor and pre-motor cortex associated with specific body parts are activated in response to motor language referring to those body parts. Other studies (Pulvermüller *et al.* 2001, Hauk *et al.* 2004) found that verbs associated with different effectors (e.g., mouth/*chew*, leg/*kick*, hand/*grab*) are processed at different rates and in different regions of motor cortex (i.e., areas responsible for the appropriate mouth/leg/hand motion display more activation). Tettamanti *et al.* (in press) have also shown that passive listening to sentences describing mouth/leg/hand motions activates corresponding parts of pre-motor cortex.

Several psycholinguistic experiments offer behavioral evidence for the automatic and unconscious use of perceptual and motor systems during language use. Richardson *et al.* (2003) found that sentences with visual semantic components can result in selective interference with visual processing. For example, subjects processing sentences encoding upwards motion (e.g., *The ant climbed*) take longer to perform a visual categorization task in the upper part of their visual field. Another experiment showed that subjects performing a physical action in response to a sentence take longer to perform the action if it is incompatible with the motor actions described in the sentence (Glenberg & Kashak 2002). These experiments, along with the imaging results above, provide suggestive evidence for an integrated, multimodal action representation that serves as a common substrate for action, perception and language.

## 3. Embodied Construction Grammar

The analysis process described in Section 1 relies on Embodied Construction Grammar (ECG) (Bergen & Chang, in press; Chang *et al.* 2002), a computationally precise formalism for representing constructions. As in other construction-based grammars (Kay & Fillmore 1999; Goldberg 1995; Croft 2001), constructions express generalizations linking the domains of form and meaning. Constructions vary in size (from morphemes and lexical items to larger phrasal and clausal units) and specificity (from frozen and partially frozen idioms to to more abstract grammatical constructions); and they encompass information that crosscuts traditional levels of linguistic analysis (e.g., phonological, morphological, syntactic, semantic and pragmatic).

ECG is designed to serve as an interface between language and simulation. It includes representations of meaning in terms of embodied structures (called *embodied schemas*) that specify parameters for simulation, and construction representations linking to embodied schemas. Some constructions directly specify which perceptual and motor schemas to deploy (e.g., *to* and *into* from Section 2, or action words like *walk* or *run*), while others specify how to combine subsidiary parameterized representations. The schema and construction formalisms include mechanisms for expressing type constraints, *identification* (or unification) constraints, self-reference, constituency and dependency relations. Computationally, both constructions and schemas are implemented using typed feature structures with unification constraints, organized in a typed inheritance hierarchy. Figure 2 shows some simple example ECG constructions, including the lexical HARRY and RAN constructions and a more complex DIRECTED-MOTION construction. We highlight some of their key properties; see Bergen & Chang (in press) for a more detailed description.

```
construction HARRY
   subcase of REF-EXPR
   form : "Harry"
   meaning : Harry
```

```
construction RAN
   subcase of MOTION-VERB
   form : "ran"
   meaning : Run-Action
      evokes Predication as p
      p.setting.time  past
```

```
construction DIRECTED-MOTION
   constituents
      x : REF-EXPR
      y : MOTION-VERB
      z : DIR-EXPR
   form
      x_f before y_f
      y_f before z_f
   meaning
      evokes Predication as p
      evokes Dir-Motion as d
      p  y_m.p
      p.schema  d
      d.mover  x_m
      d.means  y_m
      d.direction  z_m
```

**Figure 2.** ECG example constructions:.lexical constructions linking a specific form with embodied meanings and constraints, and a clausal DIRECTED-MOTION construction associating word order relations with a directed motion scene.

Each construction has **form** and **meaning** components (or *poles)*, and potentially a set of constructional **constituents**. It may be specified as a **subcase of** another construction (thus inheriting constraints). The HARRY construction is a subcase of the REF-EXPR (referring expression) construction, a general construction similar to the traditional NP. HARRY links its form (simplified as the orthographic form "Harry") to a referent whose type is constrained (using a colon) to instantiate the category Harry (of people known by the name “Harry”). Similarly, RAN is defined as a subcase of MOTION-VERB that links its form to the Run-Action schema, corresponding to an x-schema capturing perceptual and motor knowledge about running. It is this association that grounds the RAN construction in terms of its underlying motor-perceptual representations. RAN also constrains its associated evoked predication to take place in the past. The flexible **evokes as** relation allows schemas to be activated as part of the meaning pole without requiring inheritance or constituency.

The DIRECTED-MOTION construction corresponds to expressions describing the motion of some entity in some direction. The construction has internal **constituents** of types REF-EXPR, MOTION-VERB and DIR-EXPR (for direction specifiers, not defined here). The construction pairs word order constraints over its constituents’ form poles with *identification* constraints (using a double-headed arrow) over its constituents’ meaning poles, using subscripted f or m on the relevant constituent names to access their respective form and meaning poles. The meaning pole constraints specify the role fillers of the evoked Dir-Motion (directed motion) schema and link these to its associated predication.

These examples represent a particularly simple subset of English, but the formalism has also been applied to more complex crosslinguistic phenomena. Overall, the formalism provides means of linking linguistic structures with embodied simulations, through the processes to which we now turn.

## 4. Constructional analysis using embodied constructions

Constructional analysis is the process of determining which constructions are instantiated by a particular utterance in a situational context. It is thus analogous to parsing in traditional systems, but also incorporates meaning throughout, and its output is not merely a set of structures (analogous to a parse tree) but also a *semantic specification* (or *semspec*) that indicates which embodied schemas are evoked by the constructions and how they are related. This semspec serves as input to the simulation process to be described in Section 5.

We briefly illustrate the constructional analysis process for the sentence *Harry ran home* (Figure 3). The individual input words trigger lexical constructions HARRY, RAN and HOME, shown in the center column of the figure as linking their forms (left) with their accompanying semantics (right). (HOME is used here as a special path specifier that behaves like a directional particle (e.g. *in* or *out*).) These in turn trigger a search for a larger construction that can account for these items appearing in the specified order; the clausal DIRECTED-MOTION construction (from Figure 2) successfully matches both form and meaning constraints of the input and effects additional bindings among its constituents. The resulting semspec (schemas and bindings on the right side of the figure) indicates which perceptual and motor structures should be activated and how they are related, thus supplying the parameterization needed for the mental simulation of the described scene, i.e., a directed motion event in which the mover (Harry) moves by means of running in the direction of Home.

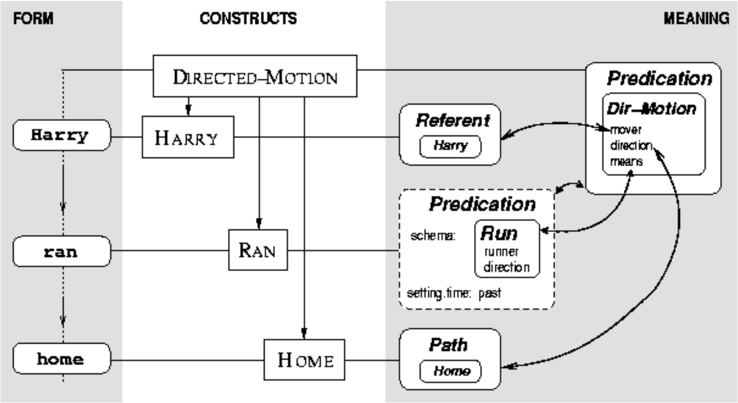

Figure 3: A simplified analysis of *Harry ran home*. The central boxes depict the constructs involved, each linking form (left) and meaning (right). The DIRECTED-MOTION construction asserts an ordering on its constituents' forms and bindings (curved arrows) on their meanings.

A construction analyzer has been implemented and tested on a corpus of English child-directed utterances involving simple motion events (Bryant 2003), including intransitive motion, directed motion, caused motion and dative constructions (Goldberg 1995). The analyzer extends partial parsing and unification-based chart parsing methods to check semantic constraints. It can thus find partial interpretations for sentences not covered by its grammar, and include only semantically coherent analyses in the chart. The best-matching set of constructions is chosen based on coverage of the utterance and semantic fit, favoring analyses with semspecs that are more complete.

## 5. Simulation-based inference

The semspec produced by constructional analysis specifies parameters for simulations. The model assumes that embodied representations of events and actions can be activated with respect to structures not linked directly to the body in its current physical context. Instead of being carried out directly, these actions can trigger simulations of what would happen in the imagined situation. This ability to simulate or imagine situations is a core component of human intelligence and is central to our model of language.

The computational formalism used in simulation is the *executing schema*, or *x-schema* (Narayanan 1999), a graph-based, token-passing formalism based on stochastic Petri nets that has an SCM interpretation (Shastri *et al*. 1998). X-schemas capture hierarchical structure, sequential flow, concurrency, resource consumption and other aspects of motor control and event structure. X-schemas can be *parameterized* (by the semspec) to execute with variable values and can represent both planned actions and events in the physical world. X-schemas are connected via input/output links; during execution they may make control transitions that modify state, leading to dynamic *activation*, *inhibition* and *interruption* of other x-schemas. X-schema inference consists of changes to the current system state, which is captured in the current graph marking and thus inherently distributed over the entire network of x-schemas and its parameters. A simulation of the semspec in Figure 3 would activate a RUN x-schema whose execution would result in the consumption of tokens representing the runner’s energy as well as a change in the runner’s location.

Narayanan (1999) describes how the simulation-based model can be applied to metaphorical language as well, which is understood as mapping to underlying embodied meaning, as hypothesized by Lakoff & Johnson (1980). In the computational implementation, the *source* domain of embodied spatial motion is encoded using the *x-schema* formalism described above. This basic model is extended so that simulation-based inferences in the source domain (e.g., resulting from executing a FALL x-schema) are projected via metaphorical mappings (e.g., FALLING IS FAILURE) to license inferences in more abstract *target* domains (e.g., international economics); the sentence “France fell into a recession” may thus be understood to involve an economic failure, where the state of recession is understood as a hole that contains France. The target domain is represented using a temporal belief network, with probabilistic belief update yielding additional inferences; this computational mechanism also plausibly models neural inference by spreading activation.

The system has been tested on narratives from the domain of international economics, using a source domain model of about 100 linked x-schemas in the domains of health and spatial motion and about 50 metaphor maps, all developed using a database of 30 discourse fragments from newspaper stories. The system made a surprising variety of subtle and informative inferences related to abstract plans and actions, involving *goals* (accomplishment, modification, subsumption, concordance, thwarting), *resources* (consumption, production, depletion, level), *aspect* (temporal structure of events), *frame-based* inferences, *perspectival* inferences, and inferences about *communicative intent*. These demonstrate how embodied representations facilitate inference in both concrete physical and more abstract domains.

## 6. Conclusions

Our explorations of embodied models of language and cognition demonstrate how a layered methodology has allowed us to reap the insights of linguistic theory while imposing the formal rigor of computational modeling. While we have focused on the computational requirements for realizing a simulation-based model of language understanding, the computational mechanisms employed also have plausible connectionist interpretations, as required by the NTL research paradigm. Additional support for the model's cognitive plausibility comes from two related projects: (1) a model of how children learn their earliest constructions, based on the ECG formalism and analysis process used here (Chang 2005); and (2) a Bayesian model of human online sentence processing assuming construction-based representations (Narayanan & Jurafsky 1998). Overall, we hope that both the methodological approach we have taken and the specific realizations of simulation-based language understanding we have described will bring us closer to fulfilling the promise of neural computation for illuminating links between brain and behavior.

## Acknowledgments

We thank the Klaus Tschira Foundation for support of this work and are happy to acknowledge that it describes the collective effort of the NTL group. We also thank an anonymous reviewer for helpful suggestions.

## References

Bailey D. (1997). *When push comes to shove: A computational model of the role of motor control in the acquisition of action verbs.* PhD thesis, UC Berkeley.

Bergen B., Chang N., Narayan S. (2004). Simulated Action in an Embodied Construction Grammar. *Proc. 26th Cognitive Science Society Conference.* (pp. 108-113)

Bergen B., Chang N. (in press). Embodied Construction Grammar in Simulation-Based Language Understanding. In Östman J.-O. and Fried M. (Eds.), *Construction Grammar(s): Cognitive and Cross-language dimensions.* Amsterdam: John Benjamins.

Bryant J. (2003). *Constructional Analysis.* M.S. report, UC Berkeley.

Chang N. (2005). *Constructing Grammar: A Computational Model of the Emergence of Early Constructions.* Ph.D. thesis, UC Berkeley.

Chang N., Feldman J., Porzel R., Sanders K. (2002). Scaling Cognitive Linguistics: Formalisms for Language Understanding. *Proc. SCANALU.*

Croft W. (2001). *Radical Construction Grammar.* Oxford: Oxford Univ. Press.

Feldman J.A., Ballard D.H. (1982). Connectionist models and their properties. *Cognitive Science*, 6, 205-254

Feldman J., Lakoff G., Bailey D., Narayanan S, Regier T, Stolcke A. (1996). $L_0$—The First Five Years of an Automated Language Acquisition Project. *AI Review,* 10, 103-129

Feldman J., Narayanan S. (2003). Embodied meaning in a neural theory of language. *Brain and Language,* 89, 385-392

Glenberg A.M., Kaschak M.P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9, 558-565

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

Hauk O., Johnsrude I., Pulvermüller F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41(2), 301-7

Kay P., Fillmore C.J. (1999). Grammatical constructions and linguistic generalizations: The What's X doing Y? construction. *Language,* 75/1, 1-33.

Lakoff G. (1987). *Women, fire, and dangerous things.* Chicago: Univ. of Chicago Press.

Lakoff G., Johnson M. (1980). *Metaphors We Live By.* Chicago: Univ. of Chicago Press.

Narayanan S. (1999). Moving Right Along: A Computational Model of Metaphoric Reasoning about Events. *Proc. AAAI.* (pp. 121-128)

Narayanan S., Jurafsky D. (1998). Bayesian Models of Human Sentence Processing. *Proc. 20th Cognitive Science Society Conference.* (pp. 84-90)

Pulvermüller F., Haerle M., Hummel F. (2001). Walking or Talking?: Behavioral and Neurophysiological Correlates of Action Verb Processing. *Brain and Language* 78, 143–168

Regier T. (1996). *The Human Semantic Potential.* Chicago: Univ. of Chicago Press.

Richardson D.C., Spivey M.J., McRae K.,, Barsalou L.W. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, 27, 767-780

Shastri, L., Grannes D., Narayanan S., Feldman J. (1998). A connectionist encoding of parameterized schemas and reactive plans. In Kraetzschmar G.K. and Palm G. (Eds.), *Hybrid Information Processing in Adaptive Autonomous vehicles.* Berlin: Springer-Verlag.

Talmy, L. (2000). *Toward a Cognitive Semantics*. Cambridge, MA: MIT Press.

Tettamanti M., Buccino G., Saccuman M.C., Gallese V., Danna M., Perani D., Cappa S.F., Fazio F., Rizzolatti G. (in press). Sentences describing actions activate visuomotor execution and observation systems.