Integrating Language and Cognitive Architecture Panel Discussion

John T. Hale [Organizer]
Cornell University
jthale@cornell.edu

Overview

Language is a fundamental human capacity: Any account of what it means to be human must explain (1) why language is structured as it is and (2) what it means to understand it. And any account of integrated cognition must specify the mental mechanisms that support language processing. For this reason, language has been at the core of artificial intelligence research since the inception of the field.

Against this backdrop, it is somewhat surprising that few accounts of integrated cognition and cognitive architecture give a prominent role to language. Although there may be some computational properties unique to natural language or human cognition, language integrates with a more general system for learning, perception and action. The purpose of this panel is to propose and discuss a variety of ways that language might be included in approaches to integrated cognition. We will focus especially on questions concerning *mutual constraints* between language processing and integrated cognition. These include:

1. How might language be integrated with other mental capacities? We will argue for a *cognitive integration* across the linguistic, perceptual, and memorial capacities of the mind. We argue for an *embodied integration* across the signaling systems of the body. Finally, we argue for a *social integration* across the minds of communicating language users, including readers and authors. We consider the implications of this proposal for modular versus interactive conceptions of mind.
2. How might the "rules of language" be shaped by the architecture of cognition? We consider the possibilities that they represent *a solution to the problem of communication*. This solution might be *optimal* given the operating characteristics – the mental representations and processes – of the mind. We argue that the solution is *adaptive* to the statistical structure of and the resources provided by the environment, both social and physical.
3. How should we understand the classic distinction between "competence" and "performance" in linguistics and cognitive science (Chomsky, 1992), in the context of an integrated approach to cognition? Competence is knowledge of language (e.g., whether sentences are grammatical or not) whereas performance is how that knowledge is applied to process language (e.g., recover the syntactic structure of grammatical sentences). We argue that the "rules of language" are (in part) the products of the architecture of cognition, and conversely, that the architecture of cognition can be understood (in part) as a set of mechanism capable of implementing the computations *required* by language. This is seen, for example, in the fundamental connection between the structure of language and the properties of short-term memory and long-term memory.

This panel is structured as follows: First, the organizer will introduce the study of the mutual constraints between language processing and integrated cognition (5 minutes). Next, each speaker will develop candidate constraints across different levels of language, different "modules" of the mind, and different social contexts (10 minutes for each of 5 speakers). This will demonstrate the productivity of this approach. The speakers will also identify challenges – concrete problems to be solved, phenomena worth puzzling over, and so on. Finally, the panelists will discuss the implications of the approach for artificial intelligence and integrated cognition (30 minutes). During this portion, audience members will be invited to ask questions of the panelists, challenge their claims, offer solutions to the problems identified, and suggest phenomena that might yield to analysis.

Integrating the WHAT and HOW of Language using Cognitive Architecture

John T. Hale
Cornell University
jthale@cornell.edu

Human-level performance in language is dazzling; we are all experts when it comes to understanding sentences. But what does this expertise amount to, and how does it relate to the various big-pictures of cognition that we have? We can answer this question in a way that integrates across Linguistics and Psychology, by leveraging cognitive architecture.

What is Performance?

Since the 1950s and especially since 2001, psycholinguists has become fascinated with the idea that the amount of work a person does understanding a word in a sentence could be related, somehow, to the amount of information the person gains at the word (e.g., Hale, 2001; 2006; Frank, 2013). But why? Why should words that are more unexpected lead to processing difficulty?

When we start to think of the degree of expectation as a degree-of-familiarity, we realize that *surprisal* – the correlation between processing difficulty and log conditional probability – is exactly the kind of practice effect that Rosenbloom, Laird and Newell studied in the early 1980s (Rosenbloom & Newell, 1987). The explanation that those authors proposed was called *chunking*. Related ideas were pursued under names like Production Compilation and Explanation Based Learning. The basic idea is that special macro-operators summarize the effects of multiple ordinary operators. The interesting part is that these macro-operators do not take as long to execute as the ordinary operators that they summarize.

What is Competence?

Generative grammar presents many problems. However, in 1982 David Marr endorsed a relatively straightforward interpretation: generative grammars are specifications of what sentences can look like. No more, no less. If we take a generative grammar as a starting point, we can then proceed to define parsing mechanisms using well-understood techniques such as left-corner parsing. This yields a pushdown automaton whose transitions can be straightforwardly embedded in a rule-based cognitive architecture like Soar.

Integrating the Two

Examining natural language corpora from the perspective of parsing, we find candidate macro-operators. For instance, short function words like prepositions and determiners strongly "predict" prepositional phrases and noun phrases. Empty subject positions in embedded clauses are usually followed by "to" and a small set of infinitival verbs = {make, do, get, see, take, go, have, ...}. In the talk I will compare examples from French and English.

The big picture is that language structures indeed govern sentence perception. But the performance characteristics of this perception are just like the rest of cognition: highly sensitive to frequent stimuli, and more likely to build outward from simple to more complex. By building integrated cognitive models in a way that leverages general linguistics – rather than repeating linguists' work – we can move forward more quickly to test these exciting universalist claims about architecture.

Understanding How Cognitive Architecture Shapes Language Capacity and Use ("Competence and Performance")

Richard L. Lewis
University of Michigan
rickl@umich.edu

In this presentation we consider a computational approach to understanding how the architecture of the mind shapes language use, language processing, and language structure. The approach is based on a simple idea for developing computational theory in cognitive science: capacities for language, thought and action, and their expression as behavior, can be understood as solutions to optimal control and state estimation problems that are defined and constrained by information processing resources and bounds. The antecedents of the idea trace back to work on signal detection theory and bounded rationality in the 1950s.

We illustrate the approach as it is applied to three problems that are concerned with the integration of language, cognition, perception and action: deriving the control of visual attention (eye-movements) in reading; deriving strategies for parsing; and deriving properties of grammar itself. Although each piece of work is in very early stages, in each case it is possible to see how assumptions about both cognitive architecture and environmental task demands are a source of explanatory and predictive power:

1. The work on visual attention in reading treats eye-movements as an adaptive control process shaped by

both architecture and prevailing linguistic goals. The accompanying eye-tracking experiments have yielded the first evidence that low-level saccadic control in linguistic tasks is precisely adapted to the joint constraints imposed by quantitatively varying task demands and the dynamics of the oculomotor control system. The approach provides a way to connect reading research with research on active (task-driven) perception.

2. The work on sentence processing treats parsing as an adaptive control process shaped by both architecture and linguistic task goals. In this view there is no fixed set of sentence parsing strategies; rather they are adapted to prevailing task demands, the statistical structure of the linguistic environment, and the fixed structure of supporting cognitive capacities (such as a bounded and noisy short-term memory). The approach provides a way to integrate probabilistic and architectural approaches to language processing.
3. The work on deriving language structure suggests that grammars can be understood as boundedly optimal policies in multi-agent control problems defined by specific environments, agent cognitive architecture, and task-oriented (vs. communication oriented) reward. Computational experiments that vary aspects of this control problem reveal how abstract properties of grammar, such as the presence of compositionality and local ambiguity, might be adaptive responses to these constraints. The approach thus provides a complement to standard computational explorations of language emergence that focus on simulating mechanisms of language evolution. The approach also bears strong relations to recent approaches in Chomksyan generative linguistic theory, possibly providing a way to deeply integrate generative linguistic theory and computational architecture approaches to language.

Cognitive Architectures that Integrate Language, Perception and Memory

David Reitter
Pennsylvania State University
reitter@psu.edu

There is much lively interaction between cognitive science and linguistics, and historically, many ideas about computational models of the mind originated in linguistics. Since then, however, models describing cognitive processes in general and models describing linguistic phenomena (grammar, discourse) have been distinct. In recent years, integrated models of cognition and language processing begin to capture a range of fascinating phenomena.

The Interactive Alignment Model, for instance, posits that speakers that talk to one another adapt their linguistic choices (Pickering & Garrod, 2004). If Alice talks to Bob, she will be more likely to use his choice of words (say, “vehicle” instead of “car”) - and not just because they tend to talk about the same topic. Bob will do the same, for instance by adapting sentence constructions, and say “I got delayed by traffic” rather than “Traffic delayed my trip”. Cognitive models (Chang, Dell, & Bock 2006; Reitter, Keller, & Moore, 2011, Jaeger & Snider, 2013) explain such effects as a cascade of basic memory effects that are computationally and quantitatively compatible with experimentally established simple recurrent network architectures, or other learning, cue-based memory retrieval and memory decay effects. Language processing is no longer isolated in models of computational competence, but it connects to the performance models of cognitive architectures.

I propose a research program that goes beyond that: cognitive architectures that model integration at the social level and in the individual.

1. Inter-personal (as in social networks, or aligned dialogue): standard memory models explain evolution of task lexica as well as invented graphical languages in networked groups of people (Chater & Christiansen, 2010 for an overview; Reitter & Lebiere, 2011 for a model).
2. Integration within the cognitive system: language and sensory perception, or language and thought. Examples of empirical work abound. People look at objects while and just before they are processing associated concepts when reading sentences - the “visual world paradigm” (Altman & Kamide, 1999). People adapt their category representations of colors to the categories associated with different languages such as English and Indonesian, if processed in one brain hemisphere (Gilbert et al., 2006; Regier & Kay, 2009). Eye-tracking can be used to understand in which order people process sentence structure, and computational models (Pollatsek et al., 2006) predict eye-movements by integrating with lexical processing and parsing. Computational models of discourse could also fit within this framework (see Varma, this symposium).

My talk will focus on the question what cognitive architectures can and cannot bring to the table when it comes to integrating language with other functions of human cognition: memory, perception, and social interaction.

Integrating Language with Other Methods and Modules of Meaning

Chung-chieh Shan
Indiana University
ccshan@indiana.edu

People signal to each other in three distinct ways (Buchler, 1940; Clark, 2003).

- *Describing*, which is to use symbols that are associated with meanings by conventional rules. For example, I might speak the words “a loud train rolled down the tracks”.
- *Depicting*, which is to use icons that are associated with meanings by perceptual resemblance. For example, I might mime a loud train rolling down the tracks by running down a corridor while yelling.
- *Indicating*, which is to use indices that are associated with meanings by agents’ attention. For example, I might point my finger or flashlight at a loud train rolling down the tracks.

Although many communication researchers like to study description as if it were a system that stands alone, it is well documented that in practice people switch among and take advantage of all three systems at once (Clark & Gerrig, 1990). For example, in the same breath I might combine running down a corridor with speaking the words “a loud train (went like...)” or with pointing. So any complete account of language must tightly integrate linguistic meanings with cognitive, perceptual, and motor representations, including coordinating with other agents on shared tasks. Even at the computational level, this integration requires a unified cognitive architecture.

It is an open question how the faculties of perception and attention can help an intelligent system depict and indicate to form meaningful signals, especially in concert with the conventional language faculty that has been the primary concern of linguistics (including computational linguistics). The traditional approach of writing or learning a grammar or parser that generates or processes strings (that is, sequences of symbol tokens) (Charniak, 1993) promises no integration with depiction or indication. Even for semantic interpretation to take and change pragmatic context as part of its input (Kaplan, 1989) and output (as in so-called dynamic semantics) (Heim, 1983; Kamp, 1981) would only allow an entirely descriptive utterance to interact with syntactically segregated acts of depiction or indication. That would not explain how speakers regularly switch from description to depiction to indication and back to description within the space of a single utterance, often with cues in syntax. This fluid switching occurs even in the domain of formal linguistics, such as in quotation (Davidson, 1979).

I find it a natural and appealing proposal to integrate description, depiction, and indication by associating each of them with the same meanings. In such a setup, the compositional semantics of conventional language can produce phrasal and utterance meanings not only from the meanings of constituent words but also, in the same uniform way, from the meanings associated with concurrent depicting or indicating. For example, the meaning of a proper name might be equal to the meaning of an index, and the compositional semantics that maps the sentence “is Alice asleep?” to its meaning would associate the sentence with pointing “is she asleep?” to its meaning in the same uniform way.

Two major questions arise from this proposal. First, it is nice and uniform that all signals, whether conventional or not, are associated with the same meanings, but what are those meanings? Second, it still seems that language, perception, and attention are separate modules (for example, they can fail separately, and there are many ungrammatical sentences that no amount of wild gesturing will fix), so in what sense of “modules” is that true?

On the first question, I am sympathetic to semantic theories that associate signals with things in the external world, not just neuron activation patterns. One way to justify such sympathy is to demand that we explain how signals such as utterances refer to things and may be true or false.

On the second question, I advocate that a module is a part of a description of a system (Gallistel, 1980), not necessarily a physical part. For example, a program that is stored only in compiled form by the computer running it may nevertheless be better described by source code, so the program may have source modules that are hard to recover at run time or by fMRI. This functional notion of modularity is relevant for organisms, species, and scientists because they all need to adapt to changes in the environment without re-learning, re-evolving, or re-discovering each new system from scratch (Parnas, 1972).

Lambda calculi and type systems offer expressive ways to describe semantics and thus carve out its modules. In particular, lambda calculi can express modules that operate on other modules, and type systems can circumscribe information flow among modules whose operation is tightly intertwined. These advantages are well exemplified by the meaningful interaction among language, perception, and attention as separate functional modules.

Integrating Language, Memory, and the Environment: An Emergentist Proposal

Sashank Varma
University of Minnesota
sashank@umn.edu

In *Sciences of the Artificial*, Simon (1996) describes an ant walking across a beach. It is tempting, he observes, to attribute the complexity of its path to a complex inner system: a hierarchy of navigational goals, spatial representations of the terrain, sophisticated geometrical inferencing capabilities, and so on. However, another analysis is possible: Complex behavior can result from the interaction between a simple inner system and a complex environment. Or more succinctly:

simple inner system + complex environment = complex behavior

An analogous, *emergentist* analysis of discourse comprehension is possible. A reader consists of a small number of simple memories. These are independently-motivated components of the cognitive architecture and not, for example, new components specific to language. This inner system is embedded in a complex environment: a textual artifact designed by a social actor, its author. Simple memories and structured texts interact during comprehension, producing the complex representations and processes that characterize understanding. Or more succinctly:

simple memory systems + structured texts = discourse comprehension

Emergentism shares with classical artificial intelligence an emphasis on the complexity of comprehension. It differs in locating the source of this complexity in structured texts rather than structured mental representations and structure-sensitive mental processes (Fodor & Pylyshyn, 1988). This paper develops the emergentist proposal with respect to cognitive science studies of discourse comprehension and models of long-term memory (LTM).

The structured representation we consider is the *mental model* or *situation model* that people construct of the global understanding of a text (Johnson-Laird, 1983; van Dijk & Kintsch, 1983). The structure-sensitive process we consider is *anaphor resolution*. Anaphors are expressions that refer to other expressions, called antecedents. When anaphors are distal – separated from their antecedents by relatively many sentences – resolution requires retrieval from LTM.

The simple memory system we consider is episodic LTM, which we understand with reference to the *exemplar models* prevalent in the memory literature (Hintzman, 1986). Exemplar models propose that the contents of short-term memory (STM) are continuously encoded as traces in LTM. Traces contain both semantic and contextual information. The contents of STM can serve as a cue to LTM, with the most similar trace having the highest probability of retrieval.

The richly structured texts we consider are long, naturalistic narratives containing late-occurring anaphors to early-occurring antecedents. Cognitive science studies have manipulated the spatial, temporal, and semantic structure of the narratives and examined anaphor reading times for evidence that this structure is reflected in readers' mental models.

We describe how an exemplar model can encode the spatial, temporal, and semantic structure of mental model representations, and support the structure-sensitive process of anaphor resolution. We evaluate this model against experimental findings on human discourse comprehension. We sketch how the model can be extended to understand how humans *read to learn* – acquire new knowledge through discourse comprehension.

References

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247-264.

Buchler, J. (Ed.) (1940). *Philosophical writings of Peirce*. London: Routledge.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*, 234-272.

Charniak, E. (1993). *Statistical language learning*. Cambridge, MA: MIT Press.

Chater, N., & Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cognitive Science*, *34*, 1131-1157.

Chomsky, N. (1992). Explaining language use. *Philosophical Topics*, *20*, 205-231.

Clark, H. H. (2003). Pointing and placing. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 243-268). Hillsdale, NJ: Erlbaum.

Clark, . H., & Gerrig, R. J. (1990). Quotations as demonstrations. *Language*, *66*, 764-805.

Davidson, D. (1979). Quotation. *Theory and Decision*, *11*, 27-40.

Fodor, J. A., & Pylyshyn, Z. W. (1988) Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3-71.

Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science*, *5*, 1-20.

Gallistel, C. R. (1980). *The organization of action: A new synthesis*. Hillsdale, NJ: Erlbaum.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences USA*, *103*, 489-494.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, *30*, 609-642.

Heim, I. (1983). File change semantics and the familiarity theory of definiteness. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, use and interpretation of language* (pp. 164-189). Berlin: Walter de Gruyter.

Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, *93*, 411-428.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*, 57-83.

Johnson-Laird, P. N. (1983). *Mental models: Toward a cognitive science of language, inference and consciousness.* Cambridge, MA: Harvard University Press.

Kamp, H. (1981). A theory of truth and semantic representation. In J. A. G. Groenendijk, T. M. V. Janssen, & M. B. J. Stokhof (Eds.), *Formal methods in the study of language: Proceedings of the 3rd Amsterdam Colloquium* (pp. 277-322). Amsterdam: Mathematisch Centrum.

Kaplan, D. (1989). Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481-563). New York: Oxford University Press.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Parnas, D. L. (1972). On the criteria to be used in decomposing systems into modules. *Communications of the ACM*, *15*, 1053-1058.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, *52*, 1-56.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169-190.

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, *13*, 439-446.

Reitter, D., & Lebiere, C. (2011). How groups develop a specialized domain vocabulary: A cognitive multi-agent model. *Cognitive Systems Research*, *12*, 175-185.

Reitter, D., Keller, F., & Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive Science*, *35*, 587-637.

Rosenbloom, P. S. & Newell, A. (1987). Learning by chunking: A production-system model of practice. In D. Klahr, P. Langley & R. Neches (Eds.), *Production System Models of Learning and Development* (pp. 221-286). Cambridge, MA: MIT Press.

Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.). Cambridge, MA: MIT Press.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. NewYork: Academic Press.