Bridging language with the rest of cognition: computational, algorithmic and neurobiological issues and methods

Shimon Edelman
Department of Psychology
232 Uris Hall, Cornell University
Ithaca, NY 14853-7601, USA
http://kybele.psych.cornell.edu/∼edelman

March 7, 2004

Abstract

The computational program for theoretical neuroscience proposed by Marr and Poggio (Marr and Poggio, 1977) calls for a study of biological information processing on several distinct levels of abstraction. At each of these levels — computational (defining the problems and considering possible solutions), algorithmic (specifying the sequence of operations leading to a solution) and implementational — significant progress has been made in the understanding of cognition. In the past three decades, computational principles have been discovered that are common to a wide range of functions in perception (vision, hearing, olfaction) and action (motor control). More recently, these principles have been applied to the analysis of cognitive tasks that require dealing with structured information, such as visual scene understanding and analogical reasoning. Insofar as language relies on cognition-general principles and mechanisms, it should be possible to capitalize on the recent advances in the computational study of cognition by extending its methods to linguistics.

Much of the discussion surrounding the integration of linguistics with the other cognitive sciences has traditionally been focusing on arguments for (Linebarger, 1989) and against (Karmiloff-Smith, 1992; Bates, 1994) the modular (Fodor, 1983) status of language. Even if language is a module, however, it may still rely on the same computational principles (in the sense of Marr (Marr and Poggio, 1977)), and may be supported by the same mechanisms, as the other cognitive functions. To explore this possibility, we need to bring together ideas from several fields. The first of these is cognitive linguistics (Langacker, 1987) — a natural home for the integration project. This discipline consistently produces valuable insights into the psychology of language, yet is little concerned with algorithmic or implementational issues. The second is computational linguistics (Jurafsky and Martin, 2000) (including statistical natural language processing (Manning and Schütze, 1999)), a field that examines the mathematical nature of language-related tasks and generates important applications, yet pays little attention to behavioral or neurobiological issues. Lastly, there is the Marr-Poggio computational framework (Marr and Poggio, 1977), which is used across cognition and which spans all the relevant levels of analysis, but has not yet been extended to the study of language.

This chapter discusses some of the general computational principles that emerge as useful for understanding cognition, focusing on those that are likely to be especially relevant in dealing with structured knowledge. It then brings these principles to bear on a theory of language that is rooted both in cognitive and in computational linguistics, and that views language as an incrementally learnable system of redundant, distributed representations akin to those found by neurobiologists in olfaction, audition and vision.

1 Common principles of cognitive representation and processing

The view that cognition hinges on internal representation of knowledge (Chomsky, 1957; Miller, 1962) is widely accepted in linguistics, psychology and neuroscience. A representational state in a cognitive system is characterized by its covariation with certain aspects of the relevant state of affairs in the world, and, crucially, by having counterfactually supported effects on the rest of the system (in linguistics, this corresponds to the requirement of psychological reality (Fodor et al., 1974; Edelman and Christiansen, 2003)). The reality of a representation can be indicated by double dissociation (in patients (Damasio and Tranel, 1993) or normal subjects (Pulvermüller et al., 1996)) or by priming (behavioral (Tulving and Schacter, 1990; Ochsner et al., 1994) or neural (Wiggs and Martin, 1998)), and its causal effectiveness — by direct intervention, such as microampere-level current injection at the appropriate brain site which brings about the predicted perceptual/behavioral change (Salzman et al., 1990).

Theories that posit distributed, overlapping, graded representations (Pouget et al., 2000) have enjoyed considerable explanatory success in contexts as diverse as olfaction, vision, reasoning and memory. In particular, researchers are now able to identify a few classes of computational processes, operating over distributed representations, that are common to a wide range of cognitive tasks. Some examples of such general-purpose computational building blocks of biological information processing are: (1) density estimation and hypothesis weighting in the service of probabilistic inference (e.g., in conceptual learning (Tenenbaum, 1999) and in vision (Kersten and Schrater, 2000)); (2) function approximation in the service of learning from examples (e.g., in vision (Poggio and Edelman, 1990) and in motor control (Poggio, 1990)), and (3) dimensionality reduction in the service of feature detection and categorization (e.g., in language (Landauer and Dumais, 1997) and in vision (Intrator and Edelman, 1997)).

In some cases, these abstract principles have been mapped onto the function of the brain and its circuitry, resulting in explanatory models that span all three levels of Marr's program. For example, in olfaction the anatomy and the physiology of the pathway leading from the sensory epithelium to the glomeruli in the olfactory bulb (Lancet, 1991; Shepherd, 1992) can be seen as filtering data through a bank of radial basis functions (Poggio, 1990). This operation implements what is known to be a universal approximation algorithm (Hartman et al., 1990) that can be used in learning from examples (Poggio, 1990). The same algorithmic approach can support visual object recognition, as demonstrated by the Chorus of Prototypes model (Edelman, 1999), in which the stimulus is represented by its similarities to (processed) memory traces of past stimuli. Recent single-cell studies in the monkey found neurons that are broadly and redundantly tuned to particular object categories (Freedman et al., 2001) and that embody an ensemble representation of inter-object similarities that is veridical with respect to the distal stimuli (Op de Beeck et al., 2001). Both these findings had been predicted by the Chorus of Prototypes model (Edelman, 1998; Edelman, 1999).

2 Dealing with structure: a special challenge?

To be relevant to language, the computational principles behind these findings must be extended to situations that require highly structured representations. Recent work in various areas of cognition has been pursuing precisely such an extension. For example, in complex analogy tasks a similarity-based model performs very well when the distributed representations it uses are made to reflect the structure of the input (Plate, 1995; Eliasmith and Thagard, 2001). Likewise, in vision the Chorus of Fragments model (derived from the Chorus of Prototypes), which aims at dealing with structured objects and scenes (Edelman and Intrator, 2003), is based on the twin principles of distributed representation by similarities (mentioned above) and of the use of visual space to anchor the various shape fragments (cf. (Edelman, 2002)), introduced next.

2.1 The role of space in representing structure

The idea that space should serve as a natural scaffolding for supporting structured representations, whose roots go back to the ancient mnemonic Method of Loci ((Neisser, 1976), p.137), is stated forcefully in Wittgenstein's *Tractatus* ((Wittgenstein, 1961), proposition 3.1431):

> The essential nature of the propositional sign becomes very clean when we imagine it made up of spatial objects (such as tables, chairs, books) instead of written signs. The mutual spatial position of these things then expresses the sense of the proposition.

In vision, sorting shape cues by their location in the visual field goes a long way towards solving the binding problem in the representation of object and scene structure (Edelman, 1999; Clark, 2000; Edelman, 2002; Edelman and Intrator, 2003). In particular, various components of a scene or an object need not be bound to each other in any special manner, as long as each of them is bound to its proper location in the visual space, merely by virtue of its appearance there.

The spatial scaffolding approach to the representation of visual structure is consistent with the omnipresence in the monkey inferotemporal and prefrontal cortex of *what+where* neurons, which are both shape-tuned (signaling *what* is the stimulus), and location-selective (signaling *where* it appears) (Rao et al., 1997; Op de Beeck and Vogels, 2000). On a larger scale, the neural substrate of the perceptually defined external space may be the cortical surface itself, as indicated by the ubiquity of map-like representations (Gallistel, 1990) in vision (Ward et al., 2002), olfaction (Joerges et al., 1997), and audition (Shamma, 2001).

2.2 Spatial representations for language

In those areas of the human brain that support language, the counterpart to the visual *what+where* neurons may be *what+when* neurons, tuned to particular structures appearing in a particular sequence (as illustrated in Figure 2). The possible role of temporal response properties of neuron assemblies in implementing sequence-sensitive processing has been discussed by Pulvermüller (Pulvermüller, 2002); parallels between the brain representations of space and time in vision and in audition have been pointed out, among others, by Shamma (Shamma, 2001).

Indeed, in recent years space has been conjectured to play a central role both in linguistics and in cognition in general. Consider, for example, the notion of iconicity in syntax (Simone, 1995): "...not only motor but also cognitive operations such as language, which do not appear to have any intrinsic spatial organization, are maintained in registration with spatial systems, and [...] this attention-requiring linkage confers a processing advantage" (Coslett, 1999). The iconicity hypothesis is intimately connected to Construction Grammar: linguistic freezes or prefabs (Landsberg, 1995) that are spatially (or, equivalently, temporally; cf. Figure 2) iconic become constructions when parameterized (Erman and Warren, 2000). Indeed, the patterns and their associated equivalence classes in the ADIOS model (Solan et al., 2003b) are just such parameterized constructions. Psycholinguistic and neuropsychological evidence in support of linguistic iconicity has been recently reviewed in (Chatterjee, 2001).

Thus, examples of space-based representations, which abound both in vision and in language, show that the goal of structure-sensitive processing by a distributed architecture is less forbidding than commonly thought, and that it is already within reach of cognition-general principles and mechanisms. This should not be surprising: functional (computation-level) analogies between language and vision suggest that the need to deal with structure is not unique to the former (Minsky, 1985). For instance, the treatment of a sentence with an embedded relative clause may be compared to the processing of a scene with occlusion

(Figure 1, left; additional parallels are illustrated and discussed in Figure 1, right). A cognitive approach to language, which is based on these foundations, and which casts the relevant computational, algorithmic and implementational issues in cognition-general terms, is outlined in the next section.

3 Treatment of structure in computational cognitive linguistics

3.1 Computational approach: the Chorus of Phrases and Construction Grammar

When applied to language, the idea of a distributed representation of structure based on similarities to multiple structured exemplars (called the Chorus of Fragments in the setting of visual scene processing (Edelman and Intrator, 2003)) translates into a *Chorus of Phrases*: a redundant ensemble of potentially overlapping, mutually reinforcing phrase fragments that, as Langacker puts it, "motivate" the sentence they cover:

> "...rather than seeing a composite structure as an edifice constructed out of smaller components, we can treat it as a coherent structure in its own right: component structures are not the building blocks out of which it is assembled, but function instead to *motivate* various aspects of it." (Langacker, 1987), p.453; italics in the original.

A simplified illustration of the Chorus of Phrases (COPH) in action is shown in Figure 2, where a stimulus (which could be an entirely novel sentence) evokes a cloud of associations, pointing to snippets of previously encountered phrases, each of which approximately matches parts of the input, and which together cover all of it.

On the abstract, computational (Marr and Poggio, 1977)) level, the view of language as based on structural generalization, exemplified by the COPH approach, differs radically from that of generative theories such as the Minimalist Program (Lasnik, 2002), which attempt to describe language in terms of syntax projected by the lexicon. We may recall that the basic theoretical challenge at the computational level is to specify what is it that needs to be done in the given task — in the present case, in language comprehension and production. According to the COPH framework, comprehension involves constructing a distributed representation of the stimulus in terms of its structure-dependent similarities to multiple stored exemplars, which convey information both about form (the exemplars are, generally, patterns with slots; see Figure 2) and about meaning. Production consists of letting a set of exemplars chosen for their semantics interact and constrain each other until a fully specified linear sequence of terminals is ready for output.

This distributed approach, which does not distinguish between syntactic and semantic representations and processes, is broadly compatible with the tenets of the Cognitive school in linguistics (Langacker, 1987), and, more specifically, with Construction Grammar (Goldberg, 1998; Goldberg, 2003; Croft, 2001). COPH is, however, more than a mere metaphor for constructions, for several reasons. First, COPH is deeply rooted in computational principles (multidimensional similarity spaces, distributed representations) and neural mechanisms (receptive fields and maps) that proved instrumental in analyzing other aspects of cognition. Second, by steering the goals of syntactic (and semantic) analysis towards those of cognition in general, COPH brings to the fore a collection of mathematical tools hitherto not considered by linguists (see **Appendix**: Mathematical Tools). Third, an implemented model of language acquisition and processing situated within the COPH framework provides empirical support and constraints for the construction grammar theories, as described briefly below.

3.2 Algorithmic and implementational issues: the ADIOS model

The pattern acquisition algorithm behind this working model of language acquisition and processing (Solan et al., 2003b; Solan et al., 2004a) (ADIOS, or Automatic DIstillation Of Structure) learns, in an unsupervised fashion, a streamlined representation of linguistic structures from untagged, large-scale natural-language corpora. The algorithm represents sentences as paths on a graph whose vertices are words. Significant patterns are defined as sets of paths in which a common prefix and suffix form a context surrounding a slot where distributionally equivalent (Harris, 1954) elements may appear. Such patterns, determined by recursive context-sensitive statistical inference, form new vertices. Linguistic constructions are encoded by trees composed of significant patterns and their associated equivalence classes. An entire utterance is typically represented by several such constructions (a Chorus of Phrases; cf. Figure 2), which may be activated to different degrees, depending on their fit to the input. Previously unseen inputs are processed by pursuing structural and lexical similarities to familiar patterns.

The probabilistic principle that drives the context-sensitive, hierarchical pattern abstraction process in the ADIOS model is closely related both to the notion of "suspicious coincidences" long thought to be the key to unsupervised learning in neural systems (Barlow, 1959; Barlow, 1989) and to the Minimum Description Length (MDL) criterion for representational efficiency (Bienenstock et al., 1997). Intuitively, two elements — such as two members of a potential linguistic construction or two fragments of a visual object — belong together to the extent that the probability of their joint appearance is higher than the product of the probabilities of their individual appearances; coding such elements as one results in a more concise representation. It has been conjectured that these principles, which can support structured learning in vision (Barlow, 1990; Edelman et al., 2002a; Edelman et al., 2002b) and in language (Bienenstock et al., 1997; Clark, 2001; Solan et al., 2003b), may provide "common foundations for cortical computation" (Phillips and Singer, 1997).

The implemented ADIOS model has been subjected to extensive tests, some of which focused on the acquisition of artificial languages generated by context-free grammars (CFG), and others — on learning from real natural-language corpora (CHILDES (MacWhinney and Snow, 1985)). The CFG experiments involved two ADIOS instances: a teacher and a student. In each of the multiple runs, the teacher was pre-loaded with a ready-made context free grammar (using the straightforward translation of CFG rules into ADIOS patterns), then used to generate a series of training corpora with up to 6400 sentences, each with up to seven levels of recursion. After training in each run i, a student-generated test corpus $C^{(i)}_{learned}$ of size 10000 was used in conjunction with a test corpus $C^{(i)}_{target}$ of the same size produced by the teacher, to calculate precision and recall. This was done by running the teacher as a parser on $C^{(i)}_{learned}$ and the student – as a parser on $C^{(i)}_{target}$. The results — nearly 100% precision and about 95% recall – indicate a substantial capacity for unsupervised induction of context-free grammars even from very small corpora (Solan et al., 2004b). Promising performance has also been demonstrated for real-life language: in the CHILDES experiments, an instance of the model that had been trained on transcribed speech of parents directed to small children performed at the level of an 8-year old in the CASL tests of grammaticality judgment (Carrow-Woolfolk, 1999). The same model also attained a level of performance considered to be "intermediate" for 9th-grade students when subjected to a standard test of English as a Second Language (ESL) proficiency (Solan et al., 2003b).

3.3 Open questions

Some of the exciting open issues in cognition that the proposed framework places within reach of empirical research are outlined next.

Linking psycholinguistics to visual psychophysics. Recent psycholinguistic evidence indicates that listeners and readers routinely settle for "good enough" representations of the linguistic material they face, rather than seeking an exhaustive parse or even just a fully disambiguated semantic interpretation (Bever et al., 1998; Ferreira et al., 2002; Sanford and Sturt, 2002); cf. Figure 1, right. How do these findings relate to the cluster of phenomena in visual psychophysics known as "change blindness" (Simons and Levin, 1997), which indicate that subjects do not fully parse visual scenes either?

From the "big picture" to the neural mechanisms. Marr's framework calls for equal attention to the computation- and the neurobiology-level understanding of cognition. On an abstract, computational level, construction-based approaches — in particular the Chorus of Phrases — readily integrate themselves into the rest of cognition, offering along the way a useful insight into the relationship between the final representational product of language and that of vision. Goldberg's thesis ("constructions all the way down" (Goldberg, 2003)) can be taken to imply that the Chorus of Phrases evoked by an utterance or a text is just about all there is to its interpretation. There is a clear parallel between this stance and the conjecture that in vision the Chorus of Fragments is an adequate, and in fact the only reasonable, bottom line (Edelman, 2002). On the level of mechanism, however, the details have yet to be worked out. A crucial question is whether or not it is possible to avoid altogether the need to manipulate constructions dynamically, rather than through a pre-wired network. This is important because the act of binding a variable to a value (or inserting a constituent into a construction) dynamically is deeply problematic in the context of a neural implementation (Edelman and Intrator, 2003).

Between the mechanism and the virtual machine. The ability of humans to do algebra or to program computers attests to the existence of some mechanism in the brain that at least creates the semblance of dynamic binding. People, however, must be trained for years before they become good at this kind of symbol manipulation, and it would be prudent to make it a means of last resort in a theory of any cognitive phenomenon that is more mundane than programming in Lisp. The same consideration applies to a related issue, recursion: although it has been recently reaffirmed by some linguists as the epitome of human uniqueness (Hauser et al., 2002), humans are notoriously bad at deep recursion (Gibson and Thomas, 1999) (and shallow recursion, as well as the manipulation of complexity-controlled constructions, can be handled by finite means such as the ADIOS representation (Solan et al., 2003b)). These considerations suggest that dynamic binding and deep recursion may both be supported in the brain by a virtual machine (cf. (Dennett, 1991), p.209) that is difficult to build and expensive to maintain and operate, and that is at least once removed from the neural mechanisms that are so good at supporting everyday cognition.

4 Conclusions

Computational cognitive science holds that a comprehensive theory of any information processing task must lead to its understanding on several levels of abstraction (Marr and Poggio, 1977). Although distinct, these levels cannot be studied independently, lest theorizing loses touch with psychological and neurobiological reality, or, conversely, the neurobiology becomes too myopic (Edelman, 1999). Accordingly, the framework for computational cognitive linguistics outlined here is informed by the top-down computational and algorithmic principles of context-dependent probabilistic learning and is based on a bottom-up implementational scheme that is ubiquitous in the brain: computing with connections, which carry dynamically unfolding neural activation patterns and which support low-dimensional, distributed, redundant, graded representations. The vision of language it offers should boost cognitively oriented theories such as Construction Grammar (Croft, 2001; Goldberg, 2003) and help connect them to rich repositories of computational knowledge (from learning theory, probability and information, and natural language processing) and empirical data (from psychophysics and neuroscience) about the brain.

Acknowledgments

I thank Zach Solan, Eytan Ruppin, David Horn, Roni Katzir, Rick Dale and Bo Pedersen for stimulating discussions.

References

Barlow, H. B. (1959). Sensory mechanisms, the reduction of redundancy, and intelligence. In *The mechanisation of thought processes*, pages 535–539. H.M.S.O., London.

Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, 1:295–311.

Barlow, H. B. (1990). Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Research*, 30:1561–1571.

Bates, E. (1994). Modularity, domain specificity and the development of language. *Discussions in Neuroscience*, 10:136–149.

Bever, T. G., Sanz, M., and Townsend, D. J. (1998). The Emperor's psycholinguistics. *J. of Psycholinguistic Research*, 27:261–284.

Bienenstock, E., Geman, S., and Potter, D. (1997). Compositionality, MDL priors, and object recognition. In Mozer, M. C., Jordan, M. I., and Petsche, T., editors, *Neural Information Processing Systems*, volume 9. MIT Press.

Bod, R. (1998). *Beyond grammar: an experience-based theory of language*. CSLI Publications, Stanford, US.

Carrow-Woolfolk, E. (1999). *Comprehensive Assessment of Spoken Language (CASL)*. AGS Publishing, Circle Pines, MN.

Chatterjee, A. (2001). Language and space: some interactions. *Trends in Cognitive Sciences*, 5:55–61.

Chen, S. and Donoho, D. L. (1994). Basis pursuit. In *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 41–44, Pacific Grove, CA. IEEE Comput. Soc. Press.

Chipere, N. (1997). Individual differences in syntactic skill. *Working Papers in English and Applied Linguistics*, 4:1–32.

Chipere, N. (2001). Native speaker variations in syntactic competence: implications for first language teaching. *Language Awareness*, 10:107–124.

Chomsky, N. (1957). *Syntactic Structures*. Mouton & Co., the Hague.

Christiansen, M. H. and MacDonald, M. C. (2003). Processing of recursive sentence structure: Testing predictions from a connectionist model. in preparation.

Clark, A. (2000). *A theory of sentience*. Oxford University Press, Oxford.

Clark, A. (2001). *Unsupervised Language Acquisition: Theory and Practice*. PhD thesis, COGS, University of Sussex.

Coslett, H. B. (1999). Spatial influences on motor and language function. *Neuropsychologia*, 37:695–706.

Croft, W. (2001). *Radical Construction Grammar: syntactic theory in typological perspective*. Oxford University Press, Oxford.

Damasio, A. R. and Tranel, D. (1993). Nouns and verbs are retrieved with differently distributed neural systems. *Proceedings of the National Academy of Science*, 90:4957–4960.

Dennett, D. C. (1991). *Consciousness explained*. Little, Brown & Company, Boston, MA.

Edelman, S. (1998). Representation is representation of similarity. *Behavioral and Brain Sciences*, 21:449–498.

Edelman, S. (1999). *Representation and recognition in vision*. MIT Press, Cambridge, MA.

Edelman, S. (2002). Constraining the neural representation of the visual world. *Trends in Cognitive Sciences*, 6:125–131.

Edelman, S. and Christiansen, M. H. (2003). How seriously should we take Minimalist syntax? A comment on Lasnik. *Trends in Cognitive Science*, 7:60–61.

Edelman, S., Hiles, B. P., Yang, H., and Intrator, N. (2002a). Probabilistic principles in unsupervised learning of visual structure: human data and a model. In Dietterich, T. G., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems 14*, pages 19–26, Cambridge, MA. MIT Press.

Edelman, S. and Intrator, N. (2003). Towards structural systematicity in distributed, statically bound visual representations. *Cognitive Science*, 27:73–109.

Edelman, S., Intrator, N., and Jacobson, J. S. (2002b). Unsupervised learning of visual structure. In Bülthoff, H. H., Wallraven, C., Lee, S.-W., and Poggio, T., editors, *Proc. 2nd Intl. Workshop on Biologically Motivated Computer Vision*, volume 2525 of *Lecture Notes in Computer Science*, pages 629–643. Springer.

Eliasmith, C. and Thagard, P. (2001). Integrating structure and meaning: A distributed model of analogical mapping. *Cognitive Science*, 25:245–286.

Erman, B. and Warren, B. (2000). The idiom principle and the open-choice principle. *Text*, 20:29–62.

Ferreira, F., Bailey, K. G. D., and Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11:11–15.

Fodor, J. A. (1983). *The modularity of mind*. MIT Press, Cambridge, MA.

Fodor, J. A., Bever, T. G., and Garrett, M. F. (1974). *The psychology of language*. McGraw Hill, New York.

Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, 291:312–316.

Gallistel, C. R. (1990). *The organization of learning*. MIT Press, Cambridge, MA.

Gibson, E. and Pearlmutter, N. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences*, 2:262–268.

Gibson, E. and Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, 14:225–248.

Goldberg, A. E. (1998). Patterns of experience in patterns of language. In Tomasello, M., editor, *The new psychology of language*, pages 203–219. Erlbaum, Mahwah, NJ.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences*, 7:219–224.

Harris, Z. S. (1954). Distributional structure. *Word*, 10:140–162.

Hartman, E. J., Keeler, J. D., and Kowalski, J. M. (1990). Layered neural networks with Gaussian hidden units as universal approximations. *Neural Computation*, 2:210–215.

Hauser, M., Chomsky, N., and Fitch, T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298:1569–1579.

Iliopoulos, C. S. and Smyth, W. F. (1998). On-line algorithms for k-covering. In *Proceedings of the 9-th Australasian Workshop on Combinatorial Algorithms*, volume 6, pages 97–106.

Intrator, N. and Edelman, S. (1997). Learning low dimensional representations of visual objects with extensive use of prior knowledge. *Network*, 8:259–281.

Jiang, T. and Li, M. (1996). DNA sequencing and string learning. *Math. Syst. Theory*, 26:387–405.

Joerges, J., Küttner, A., Galizia, C. G., and Menzel, R. (1997). Representations of odors and odor mixtures visualized in the honeybee brain. *Nature*, 387:285–288.

Joshi, A. and Schabes, Y. (1997). Tree-Adjoining Grammars. In Rozenberg, G. and Salomaa, A., editors, *Handbook of Formal Languages*, volume 3, pages 69 – 124. Springer, Berlin.

Jurafsky, D. and Martin, J. H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, New York.

Karmiloff-Smith, A. (1992). *Beyond Modularity: a Developmental Perspective on Cognitive Science*. MIT Press, Cambridge, MA.

Kersten, D. and Schrater, P. R. (2000). Pattern Inference Theory: A probabilistic approach to vision. In Mausfeld, R. and Heyer, D., editors, *Perception and the Physical World*. John Wiley & Sons, Chichester.

Lancet, D. (1991). The strong scent of success. *Nature*, 351:275–276. News and Views.

Landauer, T. K. and Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104:211–240.

Landsberg, M. E., editor (1995). *Syntactic iconicity and linguistic freezes*. Mouton de Gruyter, Berlin.

Langacker, R. W. (1987). *Foundations of cognitive grammar*, volume I: theoretical prerequisites. Stanford University Press, Stanford, CA.

Lasnik, H. (2002). The Minimalist Program in syntax. *Trends in Cognitive Science*, 6:432–437.

Linebarger, M. C. (1989). Neuropsychological evidence for linguistic modularity. In Carlson, G. N. and Tanenhaus, M. K., editors, *Linguistic structure in language processing*, pages 197–238. Kluwer Academic, Norwell, MA.

Lodhi, H., Shawe-Taylor, J., Cristianini, N., and Watkins, C. J. (2001). Text classification using string kernels. In Leen, T. K., Dietterich, T. G., and Tresp, V., editors, *Advances in Neural Information Processing Systems 13*, pages 563–569. MIT Press.

Maass, W., Natschläger, T., and Markram, H. (2003). Computational models for generic cortical microcircuits. In Feng, J., editor, *Computational Neuroscience: A Comprehensive Approach*. CRC-Press, Boca Raton, FL. to appear.

MacWhinney, B. and Snow, C. (1985). The Child Language Exchange System. *Journal of Computational Lingustics*, 12:271–296.

Manning, C. D. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA.

Marr, D. and Poggio, T. (1977). From understanding computation to understanding neural circuitry. *Neurosciences Res. Prog. Bull.*, 15:470–488.

Miller, G. (1962). Some psychological studies of grammar. *American Psychologist*, 17:748–762.

Minsky, M. (1985). *The Society of Mind*. Simon and Schuster, New York.

Neisser, U. (1976). *Cognition and reality*. Freeman, San Francisco, CA.

Ochsner, K. N., Chiu, C.-Y. P., and Schacter, D. L. (1994). Varieties of priming. *Current Opinion in Neurobiology*, 4:189–194.

Op de Beeck, H. and Vogels, R. (2000). Spatial sensitivity of Macaque inferior temporal neurons. *J. Comparative Neurology*, 426:505–518.

Op de Beeck, H., Wagemans, J., and Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nature Neuroscience*, 4:1244–1252.

Phillips, W. A. and Singer, W. (1997). In search of common foundations for cortical computation. *Behavioral and Brain Sciences*, 20:657–722.

Plate, T. A. (1995). Holographic Reduced Representations. *IEEE Transactions on Neural Networks*, 6:623–641.

Poggio, T. (1990). A theory of how the brain might work. *Cold Spring Harbor Symposia on Quantitative Biology*, LV:899–910.

Poggio, T. and Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, 343:263–266.

Pouget, A., Zemel, R. S., and Dayan, P. (2000). Information processing with population codes. *Nature Review Neuroscience*, 1:125–132.

Pulvermüller, F. (2002). A brain perspective on language mechanisms: from discrete neuronal ensembles to serial order. *Progress in Neurobiology*, 67:85–111.

Pulvermüller, F., Preissl, H., Lutzenberger, W., and Birbaumer, N. (1996). Brain rhythms of language: nouns versus verbs. *Eur. J. Neurosci.*, 8:937–941.

Rao, S. C., Rainer, G., and Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science*, 276:821–824.

Salzman, C. D., Britten, K. H., and Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 346:174–177.

Sanford, A. J. and Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences*, 6:382–386.

Shamma, S. (2001). On the role of space and time in auditory processing. *Trends in Cognitive Sciences*, 5:340–348.

Shepherd, G. M. (1992). Modules for molecules. *Nature*, 358:457–458. News and Views.

Simone, R. (1995). Iconic aspects of syntax: a pragmatic approach. In Simone, R., editor, *Iconicity in language*, pages 153–169. John Benjamins.

Simons, D. J. and Levin, D. T. (1997). Change blindness. *Trends in Cognitive Science*, 1:261–267.

Skiena, S. and Sundaram, G. (1995). Reconstructing strings from substrings. *Journal of Computational Biology*, 2:333–353.

Solan, Z., Horn, D., Ruppin, E., and Edelman, S. (2004a). Unsupervised context sensitive language acquisition from a large corpus. In Saul, L., editor, *Advances in Neural Information Processing*, volume 16, Cambridge, MA. MIT Press.

Solan, Z., Ruppin, D. H. E., , and Edelman, S. (2004b). Bridging computational, formal and psycholinguistic approaches to language. submitted.

Solan, Z., Ruppin, E., Horn, D., and Edelman, S. (2003a). Automatic acquisition and efficient representation of syntactic structures. In Thrun, S., editor, *Advances in Neural Information Processing*, volume 15, Cambridge, MA. MIT Press.

Solan, Z., Ruppin, E., Horn, D., and Edelman, S. (2003b). Unsupervised efficient learning and representation of language structure. In Alterman, R. and Kirsh, D., editors, *Proc. 25th Conference of the Cognitive Science Society*, Hillsdale, NJ. Erlbaum.

Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In Solla, S. A., Leen, T. K., and Mller, K.-R., editors, *NIPS (Advances in Neural Information Processing Systems)*, volume 12, Cambridge, MA. MIT Press.

Tomasello, M. (1998). Introduction: a cognitive-functional perspective on language structure. In Tomasello, M., editor, *The new psychology of language*, pages vii–xxiii. Erlbaum, Mahwah, NJ.

Tulving, E. and Schacter, D. L. (1990). Priming and human memory systems. *Science*, 247:301–306.

Valiant, L. G. (1984). A theory of the learnable. *Communications of the ACM*, 27:1134–1142.

Ward, R., Danziger, S., Owen, V., and Rafal, R. (2002). Deficits in spatial coding and feature binding following damage to spatiotopic maps in the human pulvinar. *Nature Neuroscience*, 5:99–101.

Wiggs, C. L. and Martin, A. (1998). Properties and mechanisms of perceptual priming. *Curr. Opin. Neurobiol.*, 8:227–233.

Wittgenstein, L. (1961). *Tractatus Logico-philosophicus*. Routledge, London. trans. D. F. Pears and B. F. McGuinness.

Zadrozny, W. (1994). From compositional to systematic semantics. *Linguistics and philosophy*, 17:329–342.

Appendix: Mathematical Tools for Computational Cognitive Linguistics

From the standpoint of methodology, a computationally motivated approach to cognitive linguistics does not imply pitching "mathematics versus psychology" (an expression used by Tomasello as a section heading in his introduction to *The New Psychology of Language* (Tomasello, 1998), p.ix). Rather, the usual tools of mathematical linguistics (such as formal languages, λ-calculus and symbolic logic) should be traded in for new ones. Some of these, which proved well-suited for analyzing distributed representations in various areas of cognition, are listed below.

Syntax: from constituent trees to string cover. Computational learning theory offers various tools capable of dealing with distributed, potentially overcomplete (Chen and Donoho, 1994) representations of sequence data. One of these is string kernels, a representation that tallies the occurrences of specific symbols in specific locations, and supports reasoning about global properties of the sequence probed in this manner and, in particular, about features that can help classify it (Lodhi et al., 2001). Similar methods are increasingly in demand in computational biology, because of the sequential nature of the data in both domains, and, specifically, because of the close analogy between text analysis by the identification of multiple, overlapping local patterns on the one hand, and hybridization approaches to DNA sequencing on the other hand. Recent developments in this field include derivations of the algorithmic complexity of specifying a string by its substrings (Skiena and Sundaram, 1995; Jiang and Li, 1996; Iliopoulos and Smyth, 1998). This approach is distinct from (and more relevant for our present purposes than) treebank-based parsing (combining multiple local or partial parse trees (Joshi and Schabes, 1997; Bod, 1998)) in that the cover it seeks need be neither precise nor exclusive.

Semantics: from functions to constructions and relations. According to the Chorus of Phrases metaphor (Figure 2), the representation of a sentence by an ensemble of active units can be approximately described as a *relation* (namely, as the subset of units whose activity exceeds some threshold). As in Construction Grammar (Goldberg, 2003), this representation captures both semantic and syntactic information about the input. Interestingly, recent work in computational semantics addressing various problematic aspects of compositionality suggests that systematicity of meaning is better served by defining meaning as a relation over sentence parts (Zadrozny, 1994) (rather than as a function of the parts, as stipulated by the classical, Fregean approach).

Acquisition: from parameter setting to structure discovery. The ascendancy of the generative grammar and its accompanying innateness postulate over competing distributional and behaviorist ideas in the 1960s can be ascribed in a large part to the inadequacy of the contemporary statistical inference methods and the perceived inability of association-based learning to handle recursion. Statistics, however, need not be limited to counting word frequencies: in computer science, the integration of advanced statistical inference (including Bayesian methods, the Minimum Description Length principle and other related information-theoretic tools), progress in computational learning theory, efficient algorithms, and cheap hardware led to important conceptual progress, as well as to practical achievements (Manning and Schütze, 1999). Likewise, learning need not be limited to the establishment of pairwise associations: bounded-depth recursively structured patterns can be learned from examples, by efficient algorithms that rely on modern statistical inference (Solan et al., 2003b; Solan et al., 2004a) (see (Clark, 2001) for an overview of the recent progress in this field).

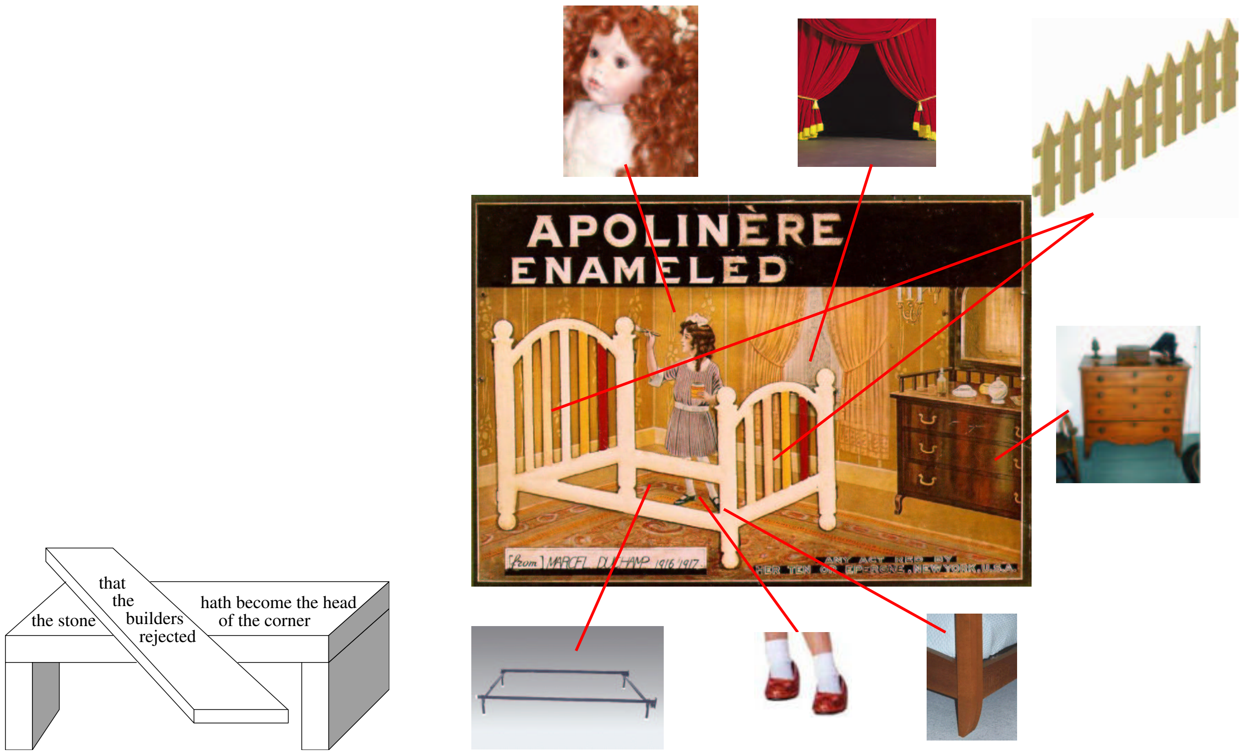

Figure 1: *Left:* there is a task-level analogy between interpreting a composite scene with occlusion and the processing of a sentence that contains an embedded relative clause (adapted from M. Minsky (Minsky, 1985), p.269). The existence of such analogies between vision and language on the abstract level of the computational tasks (Marr and Poggio, 1977) faced by the brain, along with the uniformity of the underlying low-level cortical mechanisms (Phillips and Singer, 1997), suggests that cross-cognition commonalities should be sought also on the algorithmic level. *Right:* the postcard shown here (*Apolinère Enameled* by M. Duchamp) can be used to make the same point about parallels between language and vision (cf. the occlusion of the girl's legs by the bed-frame), and more. For instance, Duchamp's painting could be represented (and understood) in terms of its local similarities to various familiar images (Chorus of Fragments (Edelman, 2002; Edelman and Intrator, 2003)), which need not match the scene perfectly (Edelman, 1999); likewise, an utterance could be represented (and understood) in terms of the cloud of constructions (Chorus of Phrases (Solan et al., 2003b)) it evokes, as illustrated in Figure 2. Furthermore, just as many viewers fail to notice that the bed in this scene would be useless (look closely at the frame), subjects exposed to ungrammatical sentences of moderate complexity may rate them as no less felicitous than similarly structured grammatical ones (e.g., the sentence "The apartment that the maid who the service had sent over was well decorated" tends to be rated as no worse (Gibson and Thomas, 1999) – and, in some settings, better (Christiansen and MacDonald, 2003) – than "The apartment that the maid who the service had sent over was cleaning every week was well decorated"; cf. (Gibson and Pearlmutter, 1998; Chipere, 1997; Chipere, 2001)).

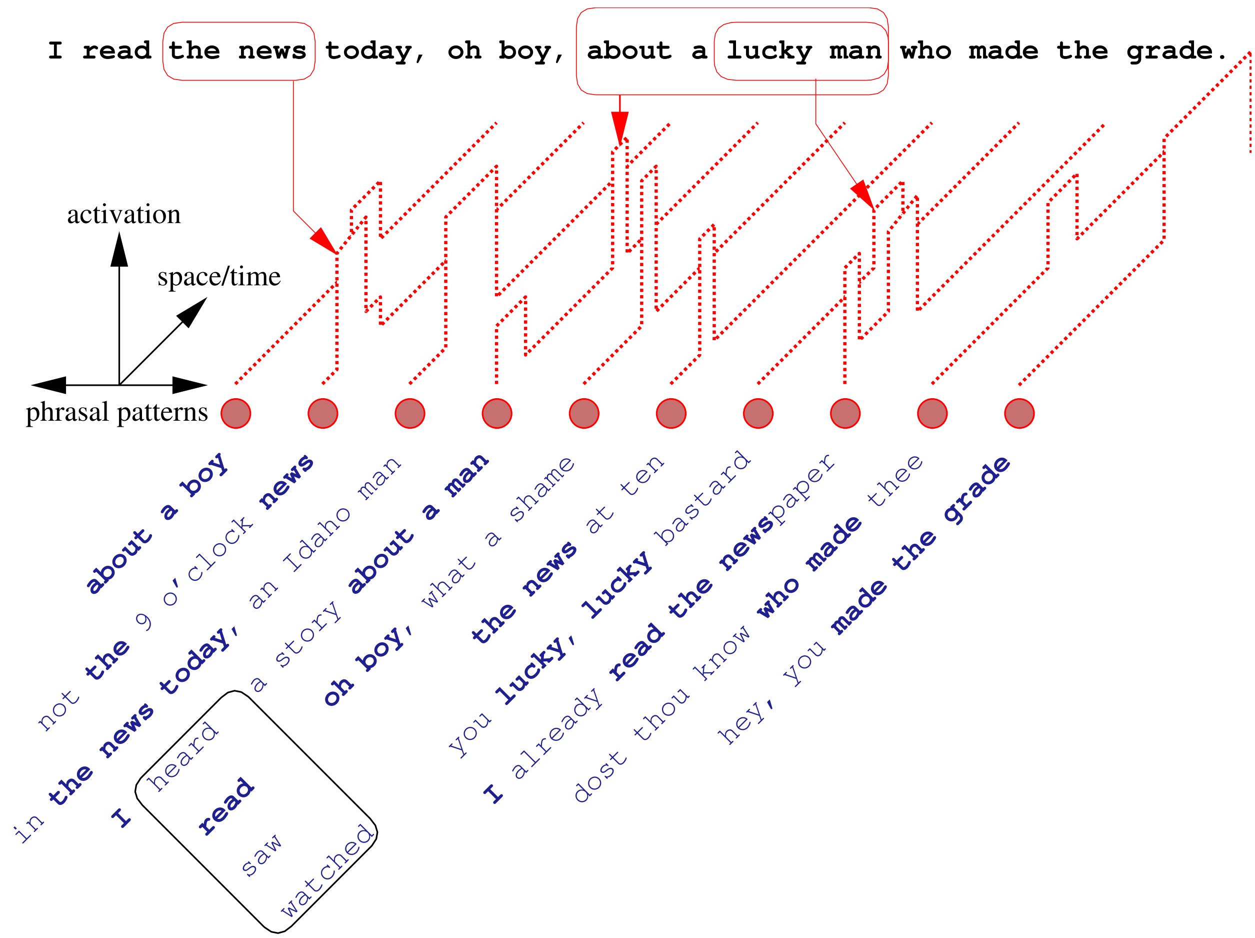

Figure 2: An illustration of the Chorus of Phrases in sentence processing (for actual examples produced by the ADIOS model, see (Solan et al., 2003a; Solan et al., 2003b)). An input sentence is shown along with a subset of phrases it evokes from memory, each of which matches some word, sequence of words, or, generally, a parameterized pattern (in cartouche: **I** heard, **read**, saw, watched a story **about a man**) in the input. The unfolding of each pattern's activation (which reflects its time-domain "receptive field") may be important (Pulvermüller, 2002), but even without it the ensemble of active patterns is a highly informative representation (just as its counterparts in vision are (Edelman, 1999; Edelman and Intrator, 2003)). The members of the ensemble disambiguate each other by supplying multiple interacting constraints on the interpretation. Consequently, it should be possible to process various queries about the input, both syntactic (voice, aspect, etc.) and semantic (thematic, connotational, conceptual). Moreover, it may be possible to use for that purpose generic cortical mechanisms (Phillips and Singer, 1997; Maass et al., 2003) that would map the distributed phrase activation patterns onto the corresponding required outputs, as in the scenario of function approximation found across cognition (Poggio, 1990; Intrator and Edelman, 1997). From the computational standpoint, it is interesting to observe that one can reconstruct the input sentence itself, should that be required for some reason, from a number of subsequence (phrase or pattern) queries that is on the order of $n \log \alpha + \alpha \log n$, where n is the length of the sentence and α is the size of the lexicon (Skiena and Sundaram, 1995). This computational complexity, which is quite benign in view of the α-fold parallelism inherent in a distributed lexicon, can be further reduced by allowing approximate (in Valiant's sense (Valiant, 1984)) matching (Jiang and Li, 1996).