How Language Processing Constrains (Computational) Natural Language Processing: A Cognitive Perspective*

Prakash Mondal

Cognitive Science Lab, International Institute of Information Technology,
Room No. C-8, NBH, Hyderabad 500032, India
prakash.mondal@research.iiit.ac.in

Abstract. In this paper I aim at sketching out in bare outline a new model/framework of language processing with its implications for natural language processing. Research in theoretical linguistics, computational linguistics and mathematical linguistics has delineated the ways natural language can be represented and modeled. Studies in cognitive (neuro)science, on the other hand, have shown how language processing is implemented in human cognitive machinery. Here an attempt will be made to integrate all these findings into a mathematical model of language processing which will point to some potential constraints on what natural language processing can achieve or do and what it possibly cannot.

Keywords: Language processing, mathematical model, theoretical linguistics, computational linguistics, cognitive science.

1 Introduction

A lot of research in mainstream theoretical linguistics (Chomsky, 1995, 2000; Pollard and Sag, 1994; Bresnan, 1982, 2001; Jackendoff, 2002), mathematical linguistics (Manaster-Ramer, 1987; Kracht, 2003), computational linguistics/natural language processing (Berwick, 1985; Barton *et al.*, 1987; Altman, 1995; Hausser, 2001; Roark and Sporat, 2007) has shown how the representational, structural and logical properties of natural language can be modeled. On the other hand, an emerging body of research in cognitive neuroscience of language (Levelt, 1989; Grodzinsky *et al.*, 2000; Brown, 2001; Pulvermüller, 2002; Hagoort, 2003; Paradis, 2004; Poeppel, 2004) has revealed the ways those properties of language are represented in human brain showing their connection to cognitive mechanisms of language processing. But there seems to be a gulf between these approaches on one hand, and between the integrated results from these studies and natural language processing in artificial intelligence on the other. Apart from that, mathematical linguistics does not seem to have committed itself to considering and reflecting these issues bordering on wider cross-connections. Against this context, it becomes a pragmatic need to see how all these issues can be linked to each other. The present proposal is aimed at bridging the gulf as specified above, which will probably project far reaching implications for natural language processing in artificial intelligence in general. A general architecture of language embedded within a cognitive architecture will first be drawn up to ultimately show how it might constrain natural language processing. In the middle, a mathematical exploration of language and language processing as falling out of the general architecture of language will also be taken up in order to derive significant formal generalizations that the entire architecture offers.

The aim is to strike out a synergistic confluence between the (formal) linguistic notions and concepts, and the emerging general architecture of language which will be shown to be bound up with each other. In sum, it will be argued that the nature of rules, constraints, function, processing, language universals and competence as derived from the overall general architecture of language has deep implications for computational intelligence underlying natural language processing.

2 A General Architecture of Language

Here a general functional architecture of language will first be presented with an eye on its potential implications for the overall framework of the mathematical model to be explored here. A certain generalizations will be drawn up from this general architecture of language to be implemented in and intertwined with a superimposed cognitive architecture. Such implications will be extrapolated for the cognitive architecture within which the mathematical framework will be positioned. A range of studies in linguistic phenomena (Neelman and Ackema, 2004; Hengeveld *et al.*, 2004; Mascaró, 2007) are increasingly pointing toward more interconnected, fluid interactions among the components of language- syntax, semantics, lexicon, morphology and phonology. Now let’s have a look at the general architecture as given below

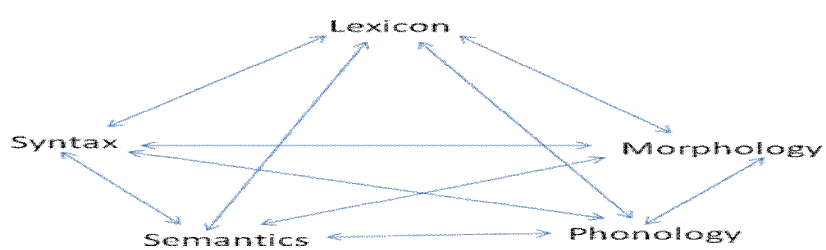

Fig 1: A multipartite unified functional architecture of language

This is a functional level architecture of language in mind/brain. The nodes are functional modules, not in Fodorian sense. They are functionally coherent units of representational resources realized in diffuse, often shared networks of brain regions (Gobet, 2005; Downing, 2005). Such representational resources can be characterized in terms of the activation patterns in neural regions which are involved in the components of language-namely, lexicon, syntax, semantics, phonology and morphology. And the bi-directional arrows regiment processing connections among them. Here we get in the architecture processing of representations, so processing is unified with representations. Processing interacts with and influences the representations in mind/brain in a mutual modulatory dynamics and processing applied to representations will constitute (linguistic) structures. In the proposed architecture above computation can run in any direction involving any number of language domains for the interpretation of any type of interaction effects- whether large or small- among the components of language. Different linguistic phenomena will involve different sets of interactions among the components. Here interface/interaction effects can be computed in parallel from and in any direction involving any number of domains or modules of language given that only some content of each as required by others is available to the others. These constraints will encode how each component of language will interact with the others, so for example, a specific constraint can determine how much information should be passed from syntax to phonology to which matters of recursion, argument structure, word order etc. may be irrelevant but syntactic constituency may well be relevant in specific situations. At best, such constraints may well be variable across different languages-some absolute and some relative. Before we proceed, let’s specify a number of concepts to be used for the formalizations in later sections below. A *linguistic phenomenon* is any phenomenon that involves a kind of interaction between any number of components of language for the observed regularities/irregularities found across languages. An *interaction* between any number of components of language involves a sort of interfacing between them such that such interfacing causes the components of language involved in the interaction to map required information onto each other, for example, the linguistic phenomenon of *split intransitivity*- a distinction between ergativity/unaccusativity- causes syntax, lexicon, morphology and semantics to work in tandem and have impacts on one another. *Computational relevance* requires the mapping of as much linguistic information at the interfaces between the components of language as is feasible and needed in a given linguistic phenomenon. *Adequacy of information* is the amount of linguistic information as demanded in the requirements of computational relevance, so for example, in the linguistic phenomenon of *middle constructions* information about word order/linearity, agreement from syntax are to be characterized as adequate information .

Stipulation For any χ in any L there must be an $I(\chi)$ such that there is a set Ω of L_c which are in parallel interaction with each other with $\tau(I(\chi))$ and $|L_c| = \Theta$ and $\Theta \geq 1$ and $\Omega \neq \emptyset$ iff $I(\chi)$ satisfies a set Φ of constraints that operate on $I(\chi)$.

What this means is simply that for any linguistic phenomenon χ in any language L, there must be some interaction involved in χ which is $I(\chi)$ determined by computational relevance and adequacy of information τ such that there is a set Ω of linguistic components L_c interacting with each other in χ and L_c can be of any number Θ but Θ must not be equal to zero, that is, must be equal to one or more than one and the set Ω of L_c must not be a null set with the set Φ of constraints that operate on the mutual interaction(s) of the linguistic domains.

3 Toward a Cognitive Architecture of Language Processing

It is necessary to fit the functional architecture of language into an architecture incorporating other cognitive domains so that how this architecture of language interacts with other cognitive domains can be linked to neural processing at the lower level. The extended picture in Figure 2 below is also symmetric with certain constraints on their mutual interaction such that not all information passes from one domain/system to another. Such constraints- computational, algorithmic and implementational in Marr's (1982) sense- are specified by our genome. They can each be specified in the manner below.

Computational Constraints:

Computational constraints will involve constraints that delimit the range of choices the modulatory interactive dynamics between the cognitive domains produces for adaptive problem solving. So for instance, such a constraint can determine how much memory (working) can be allocated for online construction of syntactic constituents or storage of them during comprehension.

Algorithmic Constraints:

Algorithmic constraints are those constraints that consist of restrictions on the algorithmic processes that inter-domain interactions can construct based on the available resources for specific tasks. It should also be noted that the algorithmic processes in each cognitive domain may be different from those in the other(s) (Roeper, 2007). Such constraints may, for example, affect how information can be transferred from visual perception to language so that language may reconstruct the visual perceptual space, so finer details of visual perception may not be available to language because of such algorithmic constraints.

Implementational Constraints:

Implemental constraints are constraints that fix up the layouts of neural connectivity among networks of brain regions such that cognitive functions are delimited by the manner of neural connectivity among neural networks. Constraints of this kind are mostly laid out genetically or epigenetically and modified to a large extent during developmental trajectory.

So these constraints fix up rather than determine how the cognitive domains or modules are going to interact with each other. Enumerating all of them will be a big problem, since we do not yet know how many constraints apply to the cognitive domains in question. Nor do we know in full details how they lay out the maps of interactional boundaries between the cognitive domains or modules, since a mapping from genes to behavior and cognition is still out of our reach (Karmiloff-Smith, 1992; Heschl, 2002).

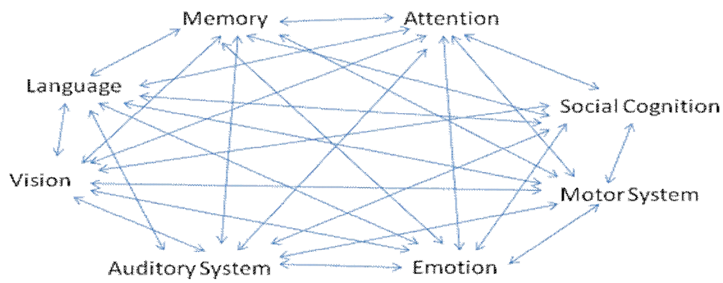

Fig. 2: Interconnection of language to other cognitive domains/systems.

Here all these domains have varying interactions (I(d)) with respect to each other; what this means is that all these cognitive domains have differential mutual interactions with respect to each other. And this can be represented through a distribution of joint probability in the following equation. This is in consonance with Mondal (2009). The interaction potential I(φ) can be characterized as

$$I(\varphi) = \sum_{i=1}^{N} P_i \int d_{c1, \dots} \, d_{cN} \, \delta(_{c1 \dots cN}).\ \Delta\psi \qquad (1)$$

Here *c1 … cN* are differential probability functions coming out of the interaction dynamics of language with other cognitive domains and N in *c1 … cN* must be a finite arbitrary number as required for the totality of cognitive domains; *P* is a probability function; $\Delta\psi$ is a temporal distribution. The cognitive domains are functionally coherent units of emergent self-organizing representational resources realized in diffuse, often shared and overlapping, networks of brain regions as mediated by the bidirectional interactions in Fig. 2.

This is also consistent with the dynamical view of cognition (Spivey, 2007). However, it may be objected that all the domains are connected to each other in a symmetric fashion, since such an architecture allows for transfer of information in any direction given that some constraints are maintained. But in biology such symmetric interconnections are being increasingly recognized (Gottlieb, 2002; Dunbar, 2007). Apart from that, research in developmental neuroscience and neurobiology of language (Neville and D. Bavelier, 1998; Stiles, 2000; Diamond *et al.*, 2004; Davidson *et al.*, 2006; Johnson, 2001; Johnson and Karmiloff-Smith, 2003; Lieberman, 2006) is also pointing toward a greater interconnectedness and neural overlap between the cognitive modules or domains. They all emerge in close synergy across the developmental scale. Research in the area of *Cognitive Linguistics* (Fauconnier, 1985, 1997, 2002; Lakoff, 1987; Langacker, 1987) has also increasingly emphasized that language, memory, vision, motor system and other cognitive domains are interlinked to each other in what we mean by language. These are not just disparate findings; they point toward something. Now the time has perhaps come to amass findings and results from fields as diverse as cognitive neuroscience, (developmental) genetics, evolutionary biology, developmental neuropsychology, psycholinguistics, neurolinguistics toward exploring language in its broadest possible spectrum so that the nature of language at all levels can be unlocked. The symmetric interconnected architecture of language is motivated on these grounds. Vision, memory, motor system, auditory system, emotion, social cognition are all linked to each other here in the architecture, as they all have something to do with language. It may not be the case that all will have equal influence and modulation on each other; rather they must have differential influence on each another. For example, memory might influence and modulate language more than vision. These factors will certainly be a part of the constraints on their mutual interaction. However, here the reasons for including emotion and social cognition are twofold. First, the role of emotion and social cognition in language has been relatively neglected in the literature. But recently the connection between language and emotion (Shanahan, 2007) has been raised. Shanahan (2007) argues that it is our emotions that projected the symbolic system that was later adapted to linguistic capacity. Given the possibility that emotion provided a boost to the growth of a symbolic capacity mediating the language-emotion interface, we find that language interacts with emotion at multiple levels with much stronger bond than has been perhaps presumed. Evidence of this comes from language acquisition research where it is becoming clearer that without emotional mechanisms language acquisition is almost impossible (Paradis, 2004). Emotional mechanisms as we understand them are also inseparable from social cognition. A range of findings (Johnson *et al*., 2005; Jackendoff, 2007; Kinzler, 2007; Olson, 2007) suggest that language, perception and social cognition go together in humans' cognitive development and help in cognitive tasks based on differential processing requirements. Such findings coupled with research in sociolinguistics (Hymes, 1972) can be integrated in support of the architecture above that posits that there is a symmetric interconnection between language and other cognitive domains including emotion and social cognition. Computationally speaking, social cognition will contain an ontology of knowledge about social categories, status, values, power, social setting etc.

All this indicates that there is much closer coupling between the cognitive domains than has been assumed. This is what makes the whole architecture susceptible to simplicity coupled with complexity as will be discussed in the fifth section where it will be linked to computational intelligence. This architecture possesses simplicity in the sense that there are fewer number of assumptions that are built into the architecture, it is complex in that the entire architecture is structurally composed of constituent subsystems each of which again having its own subsystems.

4 Toward a Mathematical Cartography of a Language Model

A range of mathematical generalizations by means of extrapolations from the assumptions made above can now be sketched out as following out from the overall picture of the architecture of language above. Such mathematical generalizations can be drawn up in such a way that such generalizations will have constraints on top from the superimposed cognitive architecture and at bottom from the general architecture of language embedded in that cognitive architecture. It is believed that such mathematical generalizations clothe the familiar notions about language in a newer way so that significant insights can be derived from an integration of mathematical modeling, a cognitive architecture and a generalized architecture of language. Hence, here a set of linguistic concepts in relation to neurocomputational processing of language will be mathematically formalized. Vectors will be used in the formalization as given below since they are perhaps better suited for the task of a mathematical model. Even if one may have certain reservations regarding their modeling capabilities in the framework being proposed here, parsimony dictates that the more compact and less elaborate characterization with much currency be used. The same principle will be regimented here. First, let us characterize the components of language L as a set of vectors representing syntactic, semantic, morphological and lexical features, processes and components.

$$L = \{\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n\} \quad (2)$$

So the same can be said about other cognitive systems, memory (M), attention (A), vision (V), audition (AD), emotion (E), motor system (MS), and social cognition (SC).

$$M= \{\beta_1, \beta_2, \beta_3, \ldots, \beta_n\} \tag{3}$$
$$A= \{\gamma_1, \gamma_2, \gamma_3, \ldots, \gamma_n\} \tag{4}$$
$$V= \{\omega_1, \omega_2, \omega_3, \ldots, \omega_n\} \tag{5}$$
$$AD= \{\xi_1, \xi_2, \xi_3, \ldots, \xi_n\} \tag{6}$$
$$E= \{\pi_1, \pi_2, \pi_3, \ldots, \pi_n\} \tag{7}$$
$$MS= \{\varphi_1, \varphi_2, \varphi_3, \ldots, \varphi_n\} \tag{8}$$
$$Sc= \{\mu_1, \mu_2, \mu_3, \ldots, \mu_n\} \tag{9}$$

Memory (M) will have components like working memory, long-term memory, semantic memory etc. in the vectors; attention (A) will have as components overt attention, covert attention etc.; vision (V) will have as components featural vision, spatial vision etc.; audition (AD) will have components like acoustic-phonetic audition, articulatory-auditory mapping etc.; emotion (E) will have as components affect, mood, conation of affect etc.; motor system (MS) will have schemas of selection, planning, execution, and social cognition (SC) will have ontologies of status, hierarchy, register, social setting etc. Undoubtedly, in many cases the vectors representing components, features, processes in each cognitive domain/system will be overlapping. However, here any vector (V_i) will have the following dynamics as specified in (Spivey, 2007).

$$\frac{\partial V_i}{\partial t} = f(V_i) \tag{10}$$

4.1 Mathematical Structure of Rules, Constraints, Function, Processing, Language Universals and Competence

The notions of rules, constraints, function, processing, language universals, and competence will be revisited here and illuminated in mathematical light. They form the core of language; so insights taken from the broader architecture of language as specified above can be applied to what constitutes the core of language. One may argue whether such new, if rudimentary, formalizations adequately model the notions as given below. However, one has to start de novo if one wants to explore familiar notions with a view to linking them to unfamiliar and unexplored paths of inquiry. At the end, such notions will be used to show what it is that we ultimately unveil through the interconnection of such notions as formalized below.

Rules: Rules $\boldsymbol{R}= (\mathbf{R_1},\ldots, \mathbf{R_n})$ in language will be formalized in the form below-

$$\boldsymbol{R} = \int \upsilon\, (P_i\,(\alpha_1.\ \alpha_2.\ \alpha_3, \ldots ,\alpha_n),\ P_j\,(\alpha_1.\ \alpha_2.\ \alpha_3, \ldots ,\alpha_n))\ \ f\,(P_i\,(\alpha_1.\ \alpha_2.\ \alpha_3, \ldots ,\alpha_n),\ t)\ d(P_i\,(\alpha_1.\ \alpha_2.\ \alpha_3, \ldots ,\alpha_n)) \tag{11}$$

Here υ denotes a probability weight recursively assigned to P_i and P_j which are differential probabilities having a complementary distribution with t denoting a particular point of time. In addition, υ also determines how strongly P_i and P_j may be correlated at time t. And f is a continuous probability function.

Constraints: Constraints (C_i) on rules and its processes abound in the literature (Kayne, 1994; Moro, 2000; Smolensky and Legendre, 2006). Such constraints operate on phonological, semantic and lexical-morphological rules and processes determining the possible mappings from input values of underlying linguistic forms to the potential values of surface linguistic forms or representations. So we get constraints like-

$$C_i = \cancel{R} \int \upsilon\, ((P_i\,(\alpha_1.\ \alpha_2.\ \alpha_3, \ldots, \alpha_m), P_j\,(\alpha_1.\ \alpha_2.\ \alpha_3, \ldots, \alpha_m))\ f\,(P_i\,(\alpha_1.\ \alpha_2.\ \alpha_3, \ldots, \alpha_m), t)\ d(P_i\,(\alpha_1.\ \alpha_2.\ \alpha_3, \ldots, \alpha_m)) \qquad (12)$$

Here $m < n$ from (10), m might be any arbitrary number between 2 and n-1 (2 being the lowest possible joint probability between the members of the set of vectors in (2)). And so m actually denotes differential interactive potentialities of vector components of language as they are deducted from rules.

Function: Functions have been defined and explained in functional theories in terms of communicative purpose of language use (Givón, 1995; Hengeveld and Mackenzie, 2006). Function is a statistical probability. So, for example, somebody utters 'a man' and stops, we cannot determine whether this NP is the subject or the object, this becomes the subject/object as the next word is processed; it becomes the subject in 'a man comes …', but becomes an object in 'a man I know…'. So in the same way, if we say 'Club..' and stop, the function predicate cannot be determined from this. As we go on and continue with 'Club them together…', it becomes the predicate; otherwise in 'Club is a social gathering…' it is the subject, not predicate. Let us formalize it in terms of conditional probability-

$$\frac{P(\alpha_i / \alpha_j)}{P\,(\alpha_j)} = P(\alpha_i \cap \alpha_j) \qquad (13)$$

α_j can be either syntactic category or phonological structure or even semantic structure in a vector. Taking all such potentialities, we get *functional potentiality* F_p. So it can be represented in the form of total probability as–

$$F_p = P\,(\alpha_1).\ P\,(F_p/\alpha_1) + P\,(\alpha_2).\ P\,(F_p/\alpha_2) + P\,(\alpha_3).\ P\,(F_p/\alpha_3) +, \ldots, + P\,(\alpha_n).\ P\,(F_p/\alpha_n). \qquad (14)$$

Processing: Now the aspect of cognitive processing of language as emerging out of the cognitive architecture of language in Figure 2. can have the following mathematical representation-

$$\frac{\partial\, {}^1C_d}{\partial t} \cdot \frac{\partial\, {}^2C_d}{\partial t} \cdot \frac{\partial\, {}^3C_d}{\partial t} \cdot, \ldots, \cdot \frac{\partial\, {}^iC_d}{\partial t} = f\,({}^1C_d.{}^2C_d.{}^3C_d., \ldots,.\ {}^iC_d) \qquad (15)$$

Here ${}^iC_d$ denotes a cognitive domain like language or vision etc. as described in Fig. 2. and $i \leq 8$, since we have eight cognitive domains interconnected to each other in Figure 2.

Language Universals: Language universals (U_L) as specified in Comrie (1981), Croft (1990) can also be mathematically expressed in the following fashion-

$$U_L = f\,(F_L.\ L_P, \Delta\lambda) \qquad (16)$$

F_L denotes language function, L_P language processing and $\Delta\lambda$ a vector field characterizing chance variation. This meshes fairly well with Kirby (1999).

Competence: At last Chomskyan I-language or competence (C_L) can be characterized as a state of processing in neurocognitive system at a point of time. It will be in the form of a triple of variable processing state ($\Delta\psi$), a point of time (t) and a probability density function (ρ).

$$C_L = (\Delta\psi, t, \rho) \qquad (17)$$

Where $\Delta\psi = f((\psi, t\text{-}1), (\psi, t\text{-}2), (\psi, t\text{-}3), \ldots, (\psi, t\text{-}m), (\psi, t+1), (\psi, t+2), (\psi, t+3), \ldots, (\psi, t + (n\text{-}1)), (\psi, t + n))$, when $m < n$.

(18)

What this suggests is that competence is not only dependent on previous processing states of the neurocognitive system, but also on those coming later to be expected. This may seem to be counterintuitive on the face of it; but a little more thinking will show that what we understand as competence is a bit elementary. Controversies abound in the literature. Competence is a function of what the cognitive agent did, does and can possibly do. Such characterization is quite consistent with the dynamical view of cognition as well (Spivey, 2007).

5 Emergent Computational Synergy, Natural Language Processing and Artificial Intelligence

Rules, constraints, function, processing, language universals, and competence as specified above form what may be called *emergent computational synergy* which characterizes self-organizing cognitive language processing, it is emergent in that all these components interact with each other in a complex self-organizing dynamics which can never be reduced to any of them; it is computational in that it can be characterized by Kolomogrov complexity and it is a synergy in that the emergent pattern grows out of dynamic, perhaps nonlinear, coupling among them. All these notions are inherently interwoven with one another. Constraints as we know them are defined and operate on rules; function and processing principles are also frozen and embedded in rules; function and processing are factors in the constitution of language universals and at last, competence is a function of rules, constraints, language universals and processing. Let's see why the notion of *emergent computational synergy* as emerging from the interwoven matrix of rules, constraints, function, processing, language universals, and competence is important here. They reveal some deeper patterns about language processing whether in humans or machines. Let's see how.

Tools like parser, language generator, language understanding system, morphological analyzer, tokenizer, stemmer etc. as are used in natural language processing tasks across a range of platforms are themselves products of this emergent computational synergy, as they are developed through human efforts. But they can never replicate or simulate this emergent computational synergy; since if hypothetically any of them can at all, the simulated/replicated synergy has to be able to produce another tool/ tools just like itself and that product has to produce yet another and so on ad infinitum- an infinite regress. This is logically impossible. So what our advanced natural language processing tools can achieve is only a fragment of what this emergent computational synergy is supposed to produce through individual or collective efforts. It is not just a constraint on natural language processing tools; rather, it is a constraint on machines in general which will be the products or gifts of artificial intelligence.

6 Conclusion

This is just a preliminary sketch of what natural language processing can offer for artificial intelligence given that our neurocognitive machinery imposes constraints on natural language processing tools which are just products of this machinery. They come naturally out of the mathematical framework sketched out here. What natural language processing tools can do is automatization of a few seemingly human tasks in barest depth, but perhaps nothing beyond that. The important thing is that such constraints have been couched in the preliminary mathematical framework described and delineated here. Further greater and richer insights into the fundamental nature of language processing can perhaps be extracted from the framework; this will be the goal of future research.

References

Altman, G. T. M. 1995. *Computational Models of Speech Processing*. Cambridge, Mass.: MIT Press.

Barton, E., R. Berwick and E. V. Ristad. 1987. *Computational Complexity and Natural*

Language. Cambridge, Mass.: MIT Press.
Berwick, R. 1985. *The Acquisition of Syntactic Knowledge.* Cambridge, Mass.: MIT Press.
Bresnan, J. 1982. *The Mental Representation of Grammatical Relations*. Cambridge, Mass.: MIT Press.
Bresnan, J. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.
Brown, C. M. 2001. *The Neurocognition of Language.* Oxford: Oxford University Press.
Chomsky N. 1995. *The Minimalist Program.* Cambridge, Mass.: MIT Press.
Chomsky, N. 2000. Minimalist Inquiries: The Framework. In R. Martin, D. Michaels, and J. Uriagereka, eds., *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik,* pp. 89- 155. Cambridge, Mass.: MIT Press.
Comrie, B. 1981. *Language Universals and Linguistic Typology*. London: Blackwell.
Croft, W. 1990. *Typology and Universals.* Cambridge: Cambridge University Press.
Davidson, M. C., D. Amso, L. C. Anderson, and A. Diamond. 2006. Development of Cognitive Control and Executive Functions from 4-13 Years: Evidence from Manipulations of Memory, Inhibition, and Task Switching. *Neuropsychologia*, vol. 44, pp. 2037 – 2078.
Diamond, A., L. Briand, J. Fossella, and L. Gehlbach. 2004. Genetic and Neurochemical Modulation of Prefrontal Cognitive Functions in Children. *American Journal of Psychiatry,* vol. 161, pp. 125-132.
Downing, P., A. W. Chan, M. Peelen, C. Dodds, and N. Kanwisher. 2005. Domain Specificity in Visual Cortex. *Cerebral Cortex,* vol. 15, pp. 12.
Dunbar, R. I. M. 2007. Brain and Cognition in Evolutionary Perspective. In S. Platek, J. P. Keenan and T. K. Shackelford, eds., *Evolutionary Cognitive Neuroscience*. Cambridge, Mass.: MIT Press.
Fauconnier, G. 1985. *Mental Spaces: Aspects of Meaning Construction in Natural Language.* Cambridge, Mass.: MIT Press.
Fauconnier, G. 1997. *Mappings in Thought and Language*. New York: Cambridge University Press.
Fauconnier, G. and M. Turner. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities.* New York: Basic Books.
Givón, T. 1995. *Functionalism and Grammar*. Amsterdam: John Benjamins.
Gobet, F. 2005. Modularity and Chunking. In W. Callebaut and D. Rasskin-Gutman. eds., *Modularity: Understanding the Development and Evolution of Natural Complex Systems.* Cambridge, Mass.: MIT Press.
Gottlieb, G. 2002. Form Gene to Organism: The Developing Individual as an Emergent, Interactional, Hierarchical System. In M. Johnson, Y. Munakata and R. O. Gilmore, eds., *Brain Development and Cognition*. London: Blackwell.
Grodzinsky, Y., L. Shapiro and D. Swinney, eds. 2000. *Language and the Brain: Representation and Processing.* San Diego: Academic Press.
Hagoort, P. 2003. How the Brain Solves the Binding Problem for Language: A Neurocomputational Model of Syntactic Processing. *Neuroimage*, vol. 20 (1), pp. S18-S19.
Hausser, R. R. 2001. *Foundations of Computational Linguistics.* Amsterdam: Springer, 2001.
Hengeveld, K. J., Rijkhoff and A. Siewierska. 2004. Parts of Speech Systems and Word Order. *Journal of Linguistics*, vol. 40(3): pp. 527-570.
Hengeveld, K. and J. L. Mackenzie. 2006. Functional Discourse Grammar. In K. Brown, ed., *Encyclopedia of Language and Linguistics. Vol IV.* Oxford: Elsevier.
Heschl, A. 2002. *The Intelligent Genome*. Amsterdam: Springer.
Hymes, D. 1972. On Communicative Competence. In J. B. Pride, and J. Holmes, eds., *Sociolinguistics*. Harmondsworth, Middlesex: Penguin.
Jackendoff, R. 2002. *Foundations of Language*. Oxford: Oxford University Press.
Jackendoff, R. 2007. *Language, Consciousness, Culture.* Cambridge, Mass.: MIT Press.
Johnson, M. H., R. Griffin, G. Csibra, H. Halit, T. Farroni, M. De Haan, L. A. Tucker, S. Baron-Cohen and J. Richards. 2005. The Emergence of the Social Brain Network:

Evidence from Typical and Atypical Brain Development. *Development and Psychopathology*, vol. **17**, pp. 599–619.
Johnson, M. H. 2001. Functional Brain Development in Humans. *Nature Reviews Neuroscience*, vol. 2, pp. 475-483.
Johnson, M. H. and A. Karmiloff-Smith. 2003. Neuroscience Perspectives on Infant Development. In G. Brenmer, and A. Slater, eds., *Theories of Infant Development*. Oxford: Blackwell.
Karmiloff-Smith, A. 1992. *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, Mass.: MIT Press.
Kayne, R. 1994. *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT Press.
Kinzler, K. D., E. Dupoux, and E. Spelke. 2007. The Native Language of Social Cognition. *Proceedings of the National Academy of Sciences*, vol. 104(30), pp. 12577-12580.
Kirby, S. 1999. *Function, Selection and Innateness*. Oxford: Oxford University Press.
Kracht, M. 2003. *The Mathematics of Language*. Berlin: Mouton de Gruyter.
Lakoff, G. 1987. *Women, Fire, and Dangerous Things: What Categories Tell us about the Mind*. Chicago: University of Chicago Press.
Langacker, R. 1987. *Foundations of Cognitive Grammar*. *2 vols*. Stanford, Calif.: Stanford University Press.
Levelt, W. J. M. 1989. *Speaking: From Intention to Articulation*. Cambridge, Mass.: MIT Press.
Lieberman, P. 2006. *Toward an Evolutionary Biology of Language*. Cambridge: Mass.: Harvard University Press.
Manaster-Ramer, A. 1987. *Mathematics of Language*. Amsterdam: John Benjamins.
Marr, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman.
Mascaró, J. 2007. External Allomorphy and Lexical Representation. *Linguistic Inquiry*, vol. 38, pp. 715-735.
Mondal, P. 2009. How Limited is the Limit? To appear in *Proceedings of International Conference on Recent Advances in Natural Language Processing*.
Moro, A. 2000. *Dynamic Antisymmetry*. Cambridge, Mass.: MIT Press.
Neeleman, A. and P. Ackema. 2004. *Beyond Morphology; Interface Conditions on Word Formation*. Oxford: Oxford University Press.
Neville, H. J. and D. Bavelier. 1998. Neural Organization and Plasticity of Language. *Current Opinion in Neurobiology*, vol. 8: pp. 254-258.
Olson, I. R., A. Plotzker, and Y. Ezzyat. 2007. The Enigmatic Temporal Pole: A Review of Findings on Social and Emotional Processing. *Brain*, vol. 130, pp. 1718 -1731.
Paradis, M. 2004. *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins.
Poeppel, D. 2004. Towards a New Functional Anatomy of Language. *Cognition*, vol. 92, pp. 1–12.
Pollard, C. and I. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: Chicago University Press.
Pulvermüller, F. 2002. *The Neuroscience of Language*. Cambridge: Cambridge University Press.
Roark, B. and R. Sproat. 2007. *Computational Approaches to Syntax and Morphology*. Oxford: Oxford University Press.
Roeper, T. 2007. *The Prism of Grammar*. Cambridge, Mass.: MIT Press.
Shanahan, D. 2007. *Language, Feeling and the Brain: The Evocative Vector*. New Brunswick: Transaction.
Smolensky, P. and G. Legendre. 2006. *The Harmonic Mind. Vol 1*. Cambridge, Mass.: MIT Press.
Spivey, M. 2007. *The Continuity of Mind*. New York: Oxford University Press.
Stiles, J. 2000. Neural Plasticity and Cognitive Development. *Developmental Neuropsychology*, vol. 18 (2), pp. 237-272.