Table of Contents

Foreword, by William S-Y. Wang

Introduction

Part 1 — Language Emergence

1. **Speech and language — a human trait defined by molecular genetics**
   — King Chow
2. **Conceptual complexity and the brain: understanding language origins**
   — P. Thomas Schoenemann
3. **Polygenesis of linguistic strategies: a scenario for the emergence of languages**
   — Christophe Coupé & Jean-Marie Hombert
4. **The emergence of grammar from perspective**
   — Brian MacWhinney

Part 2 — Language Acquisition

5. **Multiple-cue integration in language acquisition: a connectionist model of speech segmentation and rule-like behavior**
   — Morten Christiansen, Christopher M. Conway & Suzanne Curtin
6. **Unsupervised lexical learning as inductive inference via compression**
   — Chunyu Kit
7. **The origin of linguistic irregularity**
   — Charles Yang

Part 3 — Language Change

8. **The language organism: the Leiden theory of language evolution**
   — George van Driem
9. **Taxonomy, typology, and historical linguistics**
   — Merritt Ruhlen
10. **Modelling language evolution**
    — Felipe Cucker, Steve Smale & Ding-Xuan Zhou

Part 4 — Language & Complexity

11. Aspects of complexity and human language
— Murray Gell-Mann

12. Language acquisition as a complex adaptive system
— John H. Holland

13. Typology and complexity
— Randy LaPolla

14. How many meanings does a word have? Meaning estimation in Chinese and English
— Charles Lin & Kathleen Ahrens

15. Creoles and complexity
— Bernard Comrie

City University of Hong Kong Press

City University of Hong Kong
Tat Chee Avenue, Kowloon, Hong Kong
Phone: (852) 3442 6800
Fax: (852) 2779 3071
Email: uppkwong@cityu.edu.hk

Title of volume: ***Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics***

Editors: **James W. Minett & William S-Y. Wang**

Contributors:
Kathleen Ahrens, *National Taiwan University*
King Chow, *Hong Kong University of Science and Technology*
Morten Christiansen, *Cornell University*
Bernard Comrie,
Max Plank Institute for Evolutionary Anthropology
Christopher M. Conway, *Cornell University*
Christophe Coupé, *Institut des Sciences de l'Homme*
Felipe Cucker, *City University of Hong Kong*
Suzanne Curtin, *University of Pittsburgh*
Murray Gell-Mann, *Santa Fe Institute*
John H. Holland, *University of Michigan*
Jean-Marie Hombert, *Institut des Sciences de l'Homme*
Chunyu Kit, *City University of Hong Kong*
Randy LaPolla, *City University of Hong Kong*
Charles Lin, *University of Arizona*
Brian MacWhinney, *Carnegie Mellon University*
James W. Minett, *City University of Hong Kong*
Merritt Ruhlen, *Stanford University*
Steve Smale, *Toyota Technological Institute at Chicago*
P. Thomas Schoenemann, *University of Pennsylvania*
George van Driem, *Leiden University*
William S-Y. Wang, *City University of Hong Kong*
Charles Yang, *Yale University*
Ding-Xuan Zhou, *City University of Hong Kong*

Summary

The human language faculty has evolved at multiple levels: from changes in the cognitive processes by which language is acquired in the individual, to language change by diffusion of acquired linguistic features across populations of individuals, to the emergence of linguistic features over phylogenetic time scales. Evolution of language at each of these levels interacts with that at each other level. Furthermore, the language faculty has developed as a product of the complex interactions among the genetic codes that determine our physical and cognitive capabilities, and the environments — both physical and cultural — within which we live and interact with each other. In order to better understand how human language has come to be the way it is, a holistic approach to studying language evolution is now seen as essential.

This volume comprises sixteen essays on language evolution, originally presented in three Workshops on Language Acquisition, Change and Emergence (ACE I–III) organized by the Language Engineering Laboratory at the City University of Hong Kong in 2001/2. These essays reflect the aims of the workshops: to explore language evolution from an inter-disciplinary perspective in order to illuminate the interactions among the different features and mechanisms of language evolution. The volume is intended for researchers in fields such as linguistics, cognitive science, anthropology, and computer science who are working toward an inter-disciplinary view of language evolution. While most of the authors will be familiar to an audience of linguists, we are particularly pleased to be able to include here several contributors who are famous from achievements in other areas: **Gell-Mann** (physics), **Holland** (computer science), and **Smale** (mathematics). Language attracts the attention of many different types of minds. The contents of the book are collected into four sections.

The first section looks at the emergence of language from the perspective of genetics, anthropology and cognitive science, the common thread being that of incremental emergence of language. Chow summarizes the research by geneticists on the so-called "language gene", *FOXP2*. Chow argues that the language faculty cannot have emerged as a result of this single mutation, sometime in the last 120,000 years, but rather came about through the complex interaction of multiple genetic mutations. Schoenemann also argues for incremental emergence, basing his discussion on the relationship between brain size and cognitive ability. He concludes that the increasing capabilities of hominids in cognizing their environment led to the emergence of increasingly complex language, i.e., grammar. MacWhinney goes further to argue that the emergence of grammar was driven by perspective taking, his view being that grammar arose to support switching between different perspectives. Coupé and Hombert close this section by modeling the monogenetic and polygenetic emergence of linguistic features at multiple sites. Their simulations reinforce the position that language emerged as a re-

sult of complex interactions among linguistic sub-systems that each emerged polygenetically and then diffused gradually across the population.

Part Two examines models for language acquisition. In each essay, learning is seen to be achieved by distinguishing and encoding patterns of regularity. Christiansen et al. suggest a simple statistical learning mechanism by which children can acquire language from the utterances directed to them by care-givers. In particular, they demonstrate, by modeling the acquisition of word segmentation, that linguistic features can be acquired by extracting and processing cues from multiple sources, even when those cues have low validity. Kit treats the acquisition process as being driven by the distinction of regularities from irregularities using a compression-based learning algorithm. (His approach draws upon concepts of complexity, discussed by Gell-Mann in Part Four). Like Christiansen et al., he shows that a general purpose — i.e., non-language specific — learning algorithm can perform word segmentation on a corpus of actual child-directed speech. Yang, however, prefers a rule-based explanation, grounded within the approach of Chomsky's Minimalist Program, proposing an explanation for the widespread existence of irregularity in the lexicon, focusing here on irregularity in the past tense of English. In his Rule Over Words Model, irregularities are encoded by over-ridding default rules by more specific exceptions (much as in Holland's *default hierarchy*, see Part Four).

The four essays in Part Three deal with language change. Van Driem highlights the Leiden school theory of language evolution in which language is considered to be a symbiotic organism. In his account, language evolution is driven by the self-replicating components of language: *memes*. For van Driem, a meme is not a unit of imitation (cf. Dawkins's version of the meme) but a unit of linguistic meaning, whether lexical or grammatical, and so is a function of the structure of the brain. His views coincide with those of Schoenemann: that syntax arose from meaning. Ruhlen discusses language change from the perspective of historical linguists, his aim being to clarify three distinct fields — taxonomy, typology and historical linguistics — which he believes have come to be confused by linguists during the last century. Minett then describes a quantitative method for detecting language contact based on the skewing observed among the cognate counts for triplets of languages. Cucker et al. model language evolution using a formal mathematical model. For a population of speakers sharing a common language, they identify constraints on the degrees of influence of agents on each other for which a population can continue to maintain a common language. They also show that the model can be applied to the modeling of both language emergence and acquisition.

The final part of the book is concerned with the complexity of language. Gell-Mann begins by summarizing several aspects of complexity. He discusses unidirectional changes in the complexity of various linguistic sub-systems, such as word order and phonology, which may allow us to infer phylogenetic relationships among language

families at great time depths. Holland then examines the mechanisms by which language might evolve from the perspective of complex adaptive systems. He presents a general rule-based learning framework, called a classifier system, to model the process of language emergence and acquisition. Agents are initialized with default rules that encode simple, innate behaviors. He shows that exceptions, which over-ride the default rules, give rise to a default hierarchy which generates more efficient behavior than either the default rule or the exceptions alone. Over time, increasingly complex and efficient language emerges. LaPolla then considers complexification in linguistic sub-systems, commenting that one must consider *complex for whom*. For example, increased simplicity for the speaker may cause increased complexity for the listener. Lin and Ahrens then describe three methods for defining word meaning in order to investigate the psychological basis of lexical ambiguity. They conduct experiments on both Chinese and English to show that their method is psychologically sound. Comrie closes the section by discussing the complexity of creoles, focusing on morphological complexity. He distinguishes three aspects of morphological complexity by which creoles tend to differ from languages as a whole: complexity of affixes, fusing of semantic oppositions into a single morpheme, and irregularity of forms. Although creoles do tend to be simpler than other languages, Comrie concludes that there is no general trend towards simplification in creoles.

Synopsis of Contributions

In the twelve month period from May 2001 to May 2002, three Workshops on Language Acquisition, Change and Emergence (ACE I–III) were organized by the Language Engineering Laboratory at the City University of Hong Kong. The meetings brought together linguists, anthropologists, psychologists and computer scientists, among others, to discuss the acquisition, change and emergence of human language from an inter-disciplinary perspective. This book comprises essays by sixteen of the ACE Workshop contributors. The contents are collected into four parts: Language Emergence, Language Acquisition, Language Change, and Language & Complexity.

Part 1 — Language Emergence:

1. Speech and language — a human trait defined by molecular genetics by King Chow

Chow reviews the implications for our understanding of the emergence of human language based on the discovery of the so-called "grammar gene", *FOXP2*, which was found to be linked to the developmental verbal dyspraxia observed among members of an extended family of speech and language impaired individuals (known as the KE family).

Chow writes that the mouse *FOXP2* protein differs from the human protein at only three amino acid positions, putting this gene among the 5% of the most-conserved proteins in the mammalian genomes. However, he notes that the *FOXP2* gene changed in humans after the hominid lineage split off from the chimpanzee line of descent. This gene has been the target of selection during recent human evolution and was probably fixed less than 120,000 years ago. This date fits well with the sudden appearance of novel behaviors, such as art, ornamentation and long-distance trade, in human ancestors 50,000 years ago.

On the question of whether the *FOXP2* gene is the sole "language gene", Chow argues that language development has likely been influenced by numerous genetic factors, *FOXP2* simply being the first such gene identified.

2. Conceptual complexity and the brain: Understanding language origins by P. Thomas Schoenemann

Schoenemann argues for a tight relationship between increasing brain size during hominid evolution and increasing conceptual complexity. He views the universal features of grammar as reflecting our internal conceptual universe, manifested culturally in a variety of ways that are consistent with our pre-linguistic cognitive abilities. For Schoenemann, this explains both why grammatical rules vary so much across languages, as well as the fact that the commonalities appear to be semantic in nature.

Schoenemann proposes that language evolved by the incremental modification and elaboration of pre-existing cognitive mechanisms, not by the emergence of new, language-specific cognitive modules. As evidence, he points to the lack of clinical syndromes that effect only language. Rather, Schoenemann sees grammar and syntax as behavioral adaptations that take advantage of different pre-existing cognitive abilities to accomplish the task of representing higher-order conceptual complexity.

Schoenemann discusses the relationship between brain size and function, having previously modified a computer simulation introduced by Batali involving populations of interacting artificial neural nets. In his experiments, he found that populations of networks with larger numbers of hidden units evolve more complicated languages faster and with less error than populations with smaller numbers of hidden units.

Schoenemann assumes that language has molded itself to the hominid mind as much as the hominid mind has molded itself to allow increasingly sophisticated language. He concludes that the most parsimonious model is that increasing conceptual/semantic complexity drove the acquisition of syntax and grammar. The evidence of the relationship between brain size and conceptual complexity suggests, at a minimum, that fundamental changes in human cognition critical to language evolution had begun prior to ~2 million years ago.

3. The emergence of grammar from perspective by Brian MacWhinney

MacWhinney argues for an extension of current situational modeling theory to deal with perspective. He reviews a broad range of the literature on perspective taking, in terms of both cognitive processes and grammar, to argue that perspective taking is at the core of language structure and higher-level cognition, rather than adopting the typical

view that perspective is a pragmatic filter that operates only after hard linguistic constraints have been fulfilled.

MacWhinney begins by arguing that perspective taking operates online using images created in five systems: direct experience, space/time deixis, plans, social roles, and belief, which are bound together by language. Grammar arose to support switching between different perspectives. For example, whereas perspective shifting on the level of direct experience depends on imagery grounded directly on body maps, perspective shifting on the level of deixis depends on the projection of the body image across egocentric, allocentric, and geocentric frames.

MacWhinney believes that language emerged as a result of a series of gradual evolutionary adaptations that support perspective taking. The process of perspective shifting relies on at least four major neuronal systems that involve large areas of the cortex. Together, these systems allow us to store and produce images of previous direct experiences, spatial positions, plans, and social roles. Perspective allows us to thread together information from these three semi-modular sources into a coherent integrated cognitive view.

4. Polygenesis of linguistic strategies: a scenario for the emergence of languages by Christophe Coupé & Jean-Marie Hombert

Coupé and Hombert discuss arguments for and against the polygenesis of language and linguistic diversity. Rather than considering whether the human language faculty as a whole emerged at a single site or at multiple sites, they consider the emergence of multiple "linguistic strategies". They follow Freedman and Wang's approach of using probabilistic arguments to investigate the likelihood of polygenetic emergence under various conditions, modeling a wide range of plausible demographic states of pre-historic hominid populations. However, they go beyond Freedman and Wang's work by using simulation techniques to model contact among independent groups of hominids, each acquiring an arbitrary linguistic strategy with equal probability per unit of time.

In particular, they consider the density of the groups, the rate at which each group traverses a designated hunting zone, the geometric features of each group's movements, and the threshold distance (relative to the area of the hunting zone) between groups such that they come into contact. Significantly, they observe a phase transition in the time taken for a strategy to diffuse across a population by contact as a function of the

threshold distance. They also observe a phase transition for the evolution time as a function of population density. These two results suggest that the emergence of linguistic strategies might have been quite abrupt as hominid populations became progressively denser and/or increased the size of their hunting zones.

Although they conclude, much like Freedman and Wang, that polygenesis of arbitrary linguistic strategies is generally more probable than monogenesis, this probability is mitigated by the possibility of diffusion of a monogenetic innovation across an entire population.

Part 2 — Language Acquisition

5. Multiple-cue integration in language acquisition: a connectionist model of speech segmentation and rule-like behavior
by Morten Christiansen, Christopher M. Conway & Suzanne Curtin

Christiansen et al. examine a system for extracting and storing various statistical properties of language, based on a corpus, focusing here on probabilistic cues in caregiver speech to children. They explore the idea that combinations of cues provide evidence about aspects of linguistic structure that are not available from any single source. Even cues of low validity can be combined efficiently to illuminate linguistic structures that would otherwise remain opaque.

They suggest that such mechanisms are used by children to acquire language. However, before an infant can even start to learn how to comprehend a spoken utterance, the speech signal must first be segmented into words. Discovering word boundaries is a nontrivial problem as there are no acoustic correlates in fluent speech to the white spaces that separate words in written text. There are however a number of sub-lexical cues which can potentially be integrated in order to discover word boundaries. Christiansen et al. present simulation results that demonstrate how multiple-cue integration in a connectionist network, such as a simple recurrent network, can provide a solid basis for solving the speech segmentation problem.

The segmentation problem provides an appropriate domain for assessing this multiple-cue integration approach. Christiansen et al. also discuss how the process of integrating multiple cues may facilitate learning in general, reviewing evidence for the existence of probabilistic cues for the learning of word meaning, grammatical class and syntactic structure.

6. Unsupervised lexical learning as inductive inference via compression by Kit Chunyu

Kit models the unsupervised acquisition of word forms with no *a priori* knowledge using a "learning-via-compression" approach. Learning is formulated as a process of inferring regularities, in the form of string patterns, from a given set of data. The detected regularities allow the data to be compressed — the greater the compression, the better the model is in the sense that it captures more regularities in the data and thus reaches closer to the true machinery that has generated the data. Kit assumes that the learner has a very simple learning strategy: the learner simply attempts to derive a least-cost representation for the input data. In this sense, unsupervised learning is a process of inductive inference to derive that set of regularities from the data that can compress the data the most.

Kit develops algorithms to perform the optimal compression, invoking a *least-effort principle* that measures cost in terms of the number of bits required to represent the compressed data. His aims are: (1) "to test the hypothesis that there is a mechanism underlying language acquisition that seeks for the least-effort representation for the input data"; (2) "[to examine] how much a computer can learn from natural language data given that it has only a minimum innate capacity"; and (3) to "shed light on the mechanism of human language acquisition".

He focuses on a word segmentation task using input data derived from the Bernstein corpus of child-directed speech from the CHILDES collection. This corpus was chosen to ensure that the system is tested with language data that children actually receive in normal language-learning environments. Experiments on this corpus of child-directed speech show that its performance compares favorably with state-of-the-art models of unsupervised lexical learning.

7. The origin of linguistic irregularity by Charles Yang

Yang suggests that lexical irregularity results from the mechanisms of how words are learned, focusing on verbs with irregular past tenses; for example *hold-held*, which a child may say as *hold-holded*. He argues against Pinker's dual-route Words and Rule (WR) model, in which irregular verbs are memorized as associated stem-past pairs and regular verbs are computed by the rule 'add –d', in favor of a Rule over Words (RW) model, in which there is a default rule that can in principle apply to all words. In his approach, irregularity is achieved by applying the most specific rule when multiple

competing rules are present, a principle he refers to as the Elsewhere Condition. He presents developmental evidence, comprising how phonological rules are learned, how those rules are used in word learning, and the algorithmic processes underlying these two components, which Yang believes might have been present before the emergence of language since it is not unique to language. He also extends his model of word learning to model sound change over time, showing how irregularity in words is (almost) an inevitable outcome of how words are learned.

In particular, Yang points out that if Pinker's WR model were correct, irregular verbs heard more frequently than others would be remembered and used correctly more often — he illustrates by example that this is not the case. The RW model does much better, operating on verbs in classes, e.g. *throw-threw*, *know-knew*. The point is that although specific verbs belonging to a particular class might be rare, instances of verbs belonging to that class might be sufficiently frequent for that class of irregular verb to be learned — Yang calls this the Free-rider Effect.

Part 3 — Language Change

8. The language organism: the Leiden theory of language evolution by George van Driem

Van Driem presents a summary of the Leiden theory of language evolution which holds language to be a symbiotic organism. Fundamental to the Leiden school view of language is the meme. Dawkins first defined the *meme* as a unit of cultural transmission or imitation, with no reference to linguistic significance. Van Driem and his colleagues, however, refer to such a unit of imitation as a *mime*, reserving the term *meme* for meanings in the linguistic sense. Thus the meanings of grammatical categories, which are language specific, are considered as grammatical memes; the meanings of words, morphemes and idiomatic expressions are lexical memes.

Van Driem stresses that meanings are not subject to Aristotelian logic; rather meanings follow the mathematics of non-constructible sets and are independent of the law of excluded middle. Although the theory of fuzzy logic does away with the law of excluded middle, the Leiden school does not hold meanings to be fuzzy: "rather, meanings correspond to sets which are indeterminate in that there is no *a priori* way of saying whether a particular referent can or cannot be identified as a member of a set." He goes on to briefly summarize the Leiden view of the emergence of syntax from meaning by splitting of holistic utterances, and the phylogenetic evolution of humans from gracile australopithecines.

9. Taxonomy, typology, and historical linguistics by Merritt Ruhlen

Ruhlen aims to clarify the different goals and methodologies of three related, but distinct, fields: taxonomy, typology, and historical linguistics, which he believes have come to be confused by linguists in the 20th century. Ruhlen also argues that recent misunderstanding of the comparative method has led to a stagnation in the discovery of new genetic relationships among languages.

He begins by distinguishing taxonomy from historical linguistics, specifically reconstruction. Ruhlen defines linguistic taxonomy as the identification of the hierarchical structure of languages and their families, which should precede historical linguistics. Reconstruction, however, is the task of inferring the proto-language of a family that has already been identified by taxonomy.

Ruhlen equates taxonomy with classification, mass comparison and multilateral comparison. While he admits that the existence of language families such as Indo-European, Algonquian and Austronesian can be verified by the rigorous method of establishing regular sound correspondences among sibling languages, he maintains that these families were first recognized by linguists who observed the grammatical and lexical morphemes that characterize each, essentially by mass comparison as advocated by Greenberg and himself.

Ruhlen then turns to the issue of using word lists to identify genetic relationship. He discusses in particular the pattern *tVnV*, having meanings such as 'child', 'son' or 'daughter', which he finds to be widespread among the Amerind language of the Americas. Accepting that systematic sound correspondences for these putative cognates might be hard to come by, he maintains that the most parsimonious explanation for the pattern is genetic affiliation.

Ruhlen distinguishes typological classification and genetic classification: the former, he comments, is "based on historically-independent structural traits", while the latter is based on "historically-related genetic traits". He also discusses in some detail the controversy regarding the use of 2nd and 3rd person pronouns alone to detect genetic relationship, particularly the Eurasian M/T pattern and the Amerind N/M pattern.

10. Modelling language evolution
by Felipe Cucker, Steve Smale and Ding-Xuan Zhou

Cucker, Smale and Zhou develop a formal mathematical model for the evolution of language among a population of agents. The idiolect of each agent is treated as a matrix whose elements indicate the probability of association between a particular meaning and a particular signal. They further define a *communication matrix* which models the influence of each agent on the acquisition of language by other agents.

Their main result is to prove that an essentially common language will emerge within a finite number of iterations with non-zero probability given that agents update their language each iteration based on a sufficient number of meaning-signal pairs and that the communication matrix exhibits a certain property, known as *weak irreducibility*. They also prove that once a common language has emerged, a population will continue to maintain a common language. Thus although language change might temporarily introduce a degree of inhomogeneity among the idiolects of a population of agents, a globally (nearly) homogeneous language will eventually re-emerge.

Part 4 — Language & Complexity

11. Aspects of complexity and human language
by Murray Gell-Mann

Gell-Mann briefly describes the fundamental concepts of complexity, both in theory and as it pertains to language. He begins by discussing both *computational complexity*, which describes the minimum time a universal computer would take to solve problems of various sizes, and the more intuitive *effective complexity*, which can be thought of as a concise description of the regularities of the entity in question. In so doing, it is essential to distinguish the regular from the random, which both depend very much on the context.

Gell-Mann then sketches one particular approach to evaluating effective complexity by calculating the *algorithmic information content* of the regularities of an entity. He also discusses the concepts of *potential complexity*, which highlights that even a simple system (such as the game Go) can give rise to significant complexity due to the great number of possible states of the system, and *the arrow of time*, by which is meant the tendency of complex systems to evolve in a certain direction or manner.

Gell-Mann considers whether there are *arrows of time* (unidirectional processes) in human language from which the features of ancestral language can be deduced. The hopes of *long-rangers* — those who attempt to identify genetic relationships among language families at time depths beyond, say, 7,000 years — is that such features can be recovered. After discussing the proposed language families and super-families of the world, Gell-Mann discusses the possibility that word order is just such a feature, believing the word order of early human language to be the most prevalent word order of extant language, SOV.

Gell-Mann then considers the relative complexity and simplicity of extant languages in terms of various typological features. For example, he questions whether there might be a general tendency towards reduced phonological complexity. He poses similar questions with regard to delaryngealization and the loss of clicks, for example. On the other hand, he notes, processes such as palatalization, give rise to increased complexity. With a great number of unidirectional changes already known, for example the sound change /p/ to /f/ to /h/ to nothing, Gell-Mann suggests that there may be many such unidirectional processes that are less obvious. Finding such arrows of time and determining the relative rates of simplification and complexification may tell us a great deal more about the ancestral state of human language and how it came to be the way it is.

12. Language acquisition as a complex adaptive system by John H. Holland

Holland describes an agent-based model for language acquisition to investigate whether grammars can be acquired by agents having general cognitive mechanisms, but no language-specific learning strategies. The model that Holland adopts, known as a *classifier system*, is a rule-based, message passing system that he himself developed for investigating complex adaptive systems in general.

After giving a conceptual description of complex adaptive systems in general, Holland describes a classifier system used to represent the language users in some community. Agents move about a shared two-dimensional space in which are distributed the resources that enable the agents to survive The agents interact with the environment and each other by transmitting, receiving and acting upon messages. The survival of an agent is determined by its ability to acquire resources through appropriate action sequences. Agents are initialized with very general, default rules that provide a general utility, but which often lead to inappropriate action. When a default rule has a tendency not to fulfill designated needs and goals, an exception to it may occasionally be triggered. The resulting default hierarchy of rules has the effect that the two rules together perform more efficiently than either rule alone — Yang (in this volume) makes use of

much the same concept in his discussion of irregular verbs. In this way, increasingly complex and efficient language can emerge.

Holland's work raises some important questions:

1. What is the role of feedback and reinforcement in enabling a learner to develop the computational algorithms needed for robust acquisition of language?
2. What is the machinery for robust language acquisition in situations of reduced input and socialization?
3. What are the mechanisms used by parents to increase children's ability to communicate?
4. What are the factors in the evolution of language that increase the survivability of groups sharing a common language?

13. Typology and complexity
by Randy LaPolla

LaPolla considers the issue of complexity in language from several perspectives. Language as a complex system differs considerably from natural complex systems, such as the weather, and from intentionally man-made complex systems, such as tools and engines. Rather, language is an unintentionally created, man-made complex system — it developed to enhance communication effectiveness but without central planning.

Complexification of one system may lead, through extension, to simplification of another system. For example, speakers of the Qiang language in Northern Sichuan Province, China, have conventionalized a set of orientation marking prefixes related to their geophysical environment, e.g. 'up (the mountain)', 'down-river'. They have extended these metaphorically to mark perfectives and imperatives. The added complexity in the system for orientation marking has brought about increased simplicity for the marking of perfectives and imperatives because no separate set of markers need be developed.

One important question is to consider "complex for whom?" Chinese writing may be read from left to right, from right to left, even from left to right **and** from right to left, as well as from top to bottom in combination with the previous directions — in other words in just about any direction. Although this causes no added complexity for the writer, the reader's task is made more complicated because there is no standard direction of reading. On the other hand, with a standardized word order, the job of the

writer is more complex, but the task of the reader is simplified by the constraint to the inferential process.

LaPolla has the same view as Gell-Mann (this volume), that a language need not develop towards increased complexity. For example, the highly complex system of declension of nouns and adjectives in Old English has not been preserved in modern English. He also argues for the complexity of a language being considered in terms of its sub-systems, not the language as a whole. The complexity of a language sub-system is also tightly linked to the complexity of cognitive categories. For example, an English speaker will tend to perceive the initial voiceless unaspirated initial [p] of 'Beijing' as a voiced initial [b] since voiced stop initials and voiceless unaspirated initials are grouped together perceptually in English.

While some linguists argue that languages differ in terms of what you *can* say, LaPolla prefers the position that languages differ in terms of what you *must* say. For example, English requires explicit mention of the subject of a sentence due to grammaticalization of a set of obligatory constraints on referent identification that have come to be associated with 'subject'. Chinese, however, has not conventionalized these same constraints on referent identification, so the identification of the referent is not obligatory. Such conventions force particular interpretations of sentences, constraining what a language has to say.

14. How many meanings does a word have? Meaning estimation in Chinese and English by Charles Lin & Kathleen Ahrens

Lin and Ahrens investigate the psychological basis of lexical ambiguity. They describe three methods for defining word meaning: meanings listed in dictionaries, meanings provided by human subjects, and meanings analyzed by a linguistic theory.

Although dictionary meanings are favored by researchers because they are standardized, comprehensive, and easy to obtain, different researchers consult different dictionaries, which inevitably have distinct editing styles and meaning presentations. Furthermore, dictionaries are designed for language users' reference and so include archaic meanings and lack novel meanings that are emerging. An alternative approach, following Millis and Button, is to use semantic intuition, making use of the accessible polysemy — the number of different meanings that subjects are able to think of for a word — of many people. The third approach, adopted by Lin and Ahrens, is to identify and distinguish

senses of meanings — which involve different conceptual domains — from facets of meanings — which belong to the same semantic domain.

Conducting experiments for both Chinese and English, they find significant correlation between the numbers of meanings found by each method, although the actual numbers calculated for each method differ. They conclude that this method for distinguishing meaning senses from facets is psychologically sound.

15. Creoles and complexity
by Bernard Comrie

Comrie reviews the relationships between creoles and complexity to examine whether creoles are less complex than languages on average. He begins by discussing the difficulty that linguists have encountered in defining exactly what is a creole. Like pidgins, creoles arise as a result of contact between speakers having no common knowledge of each others' languages — the contact is insufficient for them to acquire a full command of each other's languages. However, creoles are distinguished from pidgins, which are the first language of no one. For creoles that have formed due to contact between speakers of a European language and one or more non-European language, the European language typically provides the majority of the lexicon, while the grammar is typically very different from the lexifier language. The longer the period of contact between the creole and the lexifier language, the greater the number of complexities of the lexifier language that tend to be found in the creole.

Rather than attempt to consider the complexity of creoles relative to other languages as a whole, Comrie focuses on the complexity of the morphology, obviating the need to weigh up the relative complexities of different linguistic sub-systems. Although most creoles have very low levels of morphological complexity, it is not always the case that they lack bound morphology completely: derivational morphology is widespread in some creoles, whereas some non-creole languages, such as Vietnamese, have very limited inflectional morphology.

Comrie distinguishes three types of morphological complexity. First, a language might be morphologically complex in that it permits the accumulation of a large number of affixes on a single root, for example Turkish. Although there is evidence that such complexity poses no problem for first language learners, borrowing of this feature does not occur. Second, a language may be complex in that it fuses a number of semantic oppositions together into a single morpheme, as observed in Italian, where no separate suffices to adjectives can be identified that encode number and gender. Such fusional morphology is rare in creole languages. Third, morphological complexity may be due

to the irregularity of some morphological forms. For example, German specifies several mechanisms for formation of the plural. Creole languages typically lack such complexities in their inflectional morphology. Nevertheless, creoles must be sufficiently complex to carry the full range of functions that are required of human language. Comrie concludes that there is no general trend towards simplification in creoles.