TYPOLOGICAL STUDIES IN LANGUAGE 53

THE EVOLUTION OF LANGUAGE OUT OF PRE-LANGUAGE

Edited by

T. GIVÓN

BERTRAM F. MALLE

JOHN BENJAMINS PUBLISHING COMPANY

The Evolution of Language out of Pre-language

Typological Studies in Language (TSL)

A companion series to the journal *Studies in Language*

Volumes in this series will be functionally and typologically oriented, covering specific topics in language by collecting together data from a wide variety of languages and language typologies. The orientation of the volumes will be substantive rather than formal, with the aim of investigating universals of human language via as broadly defined a data base as possible, leaning toward cross-linguistic, diachronic, developmental and live-discourse data.

Volume 53

The Evolution of Language out of Pre-language
Edited by T. Givón and Bertram F. Malle

The Evolution of Language out of Pre-language

Edited by

T. Givón
Bertram F. Malle
University of Oregon

John Benjamins Publishing Company
Amsterdam / Philadelphia

Library of Congress Cataloging-in-Publication Data

The Evolution of Language out of pre-language / edited by T. Givón and Bertram F. Malle.
p. cm. (Typological Studies in Language, ISSN 0167–7373 ; v. 53)
Includes bibliographical references and indexes.
1. Language and languages--Origin. 2. Human evolution. 3. Language acquisition. I. Givón, T. II. Malle, Bertram F. III. Series.

P116 .E92 2002
401-dc21 2002027967
ISBN 90 272 2959 7 (Eur.) / 1 58811 237 3 (US) (Hb; alk. paper)
ISBN 90 272 2960 0 (Eur.) / 1 58811 238 1 (US) (Pb; alk. paper)

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Contents

Part 4. Language development

Introduction

T. Givón and Bertram Malle
University of Oregon

Language evolution has been an intellectual orphan in linguistics, often relegated to the soft, speculative periphery of the field. Some of the reasons for this trend involve deep-rooted epistemological prejudices that have licensed an artificial, rigid boundary between the biological and the cultural, the genetic and the learned, the modular and the distributive — with language taken to be a prime example of a cultural, learned, distributed capacity. Yet, however far language as a cultural phenomenon may have extended beyond strictly biological capacities, it remains a profoundly adaptive biological phenomenon. And as elsewhere in biology, a full understanding of such a phenomenon and, in particular, an explanation of some of its more puzzling features, is impossible outside an evolutionary framework.

There has also been a somewhat pragmatic reason for shunning the study of language evolution. Unlike the skeletal fossils of biological evolution, the fossils of language — those putative missing links between the pre-linguistic apes of ca. six million years ago and our own linguistic species — are not currently available and, most likely, never will be. Strictly speaking, we do not have extant samples of either primitive language or primitive brain tissue. And in the absence of such data, one is tempted to declare the entire enterprise hopeless.

Even though this conclusion is perhaps understandable, it is not necessarily defensible. What we currently have available are patterns of language ontogeny, second-language acquisition, and diachronic change — all analogical data that hint at whatever the intermediate fossils may have been. None of these patterns can be claimed to map directly onto language evolution, but they may still share enough fundamental properties with bona fide evolution to be valuable. Our predicament is not entirely unlike the one Aristotle found himself in ca. 300 BC. Out of the three data bases that prompted the theory of biological evolution — the diversity of extant life-forms, embryology, and the fossil record — the first two were already available to Aristotle. Had he been inclined to interpret them, they could have served him as early analogical guideposts for the putative evolutionary sequence, long before the fossil record emerged, since they both suggest the very same graduated progression from simple to complex organisms that became, eventually, the hallmark of Darwinian evolution.

What should ultimately defeat the resistance to studying language evolution is the logic of science itself. In any scientific discipline, an imperfect and tentative theory is always preferable to no theory at all, even if the imperfect theory is a daring extension of analogical data. To recoil from making use of analogical data is therefore to recoil from advancing promising theories, and that shows a lack of faith in scientific practice. If a new theory is advanced, the scientific community will properly scrutinize it, and the theory will be considered valuable only if it explains what would have otherwise remained inexplicable and makes predictions that would have otherwise been overlooked. Those predictions, of course, can themselves be put to the test, resulting in new data that will not only weigh the worth of the theory but also augment the evidence that any new theory must account for.

The contributors to this volume include linguists, psychologists, neuroscientists, primatologists, and anthropologists. Despite the diversity of their disciplinary backgrounds, they share the broad assumption that the human mind and brain as well as language and culture are the products of adaptive evolution. Accordingly, the rise of human language is not viewed as a serendipitous mutation that gave birth to an utterly unique language organ, but as a gradual, adaptively driven elaboration and extension of pre-existing mental capacities and brain structures. The evolution of human language thus follows the same route of gradual, adaptive extension (i.e., homoplasy) that is found elsewhere in biology.

Evolutionary thinking is undoubtedly more difficult to apply to mind, language, and culture than to anatomy, but this should not detract from the imperative for science to propose and examine hypotheses that elucidate connections between previously disparate facts. Within the diversity of contributions to this volume, the commitment to theory, scientific test, and critical discourse is therefore paramount.

The theme of Part 1 is language and the brain. T. Givón's chapter surveys neurological, cognitive, and linguistic evidence that suggests that many of the modules involved in language processing must have evolved out of the primate visual information-processing systems. Don Tucker's chapter deals with evolutionary connectivity and synchronic interaction between evolutionarily older midbrain systems and their more recent cortical collaborators, with particular reference to conceptual semantics. Charles Li's chapter discusses missing links in hominid evolution and brain anatomy and relates them to corresponding missing links in language evolution, particularly the transition from pre-linguistic to linguistic communication.

Part 2 deals with the broad relation between language and cognition. Joan Bybee tries to account for the hierarchic nature of language by relating it to sequential (temporal) structure. She argues that because of the cognitive processes of automaticity and chunking, linguistic elements that are frequently used together bind together into hierarchical constituents. Barbara Davis and Peter MacNeilage advance a hypothesis about the evolution of the syllable from constraints on jaw movement in mastication, citing supporting data from human ontogeny. A companion chapter by Peter MacNeilage and Barbara Davis explores the phonological, lexical and neurological ramifications of the same hypothesis. Marjorie Barker and T. Givón present an experimental study designed to test the hypothesis that the linguistic processing rates of words and clauses are an evolutionary product of the visual processing rates of objects and events, respectively. Gertraud Fenk-Oczlon and August Fenk then present another perspective on language processing rates, documenting and accounting for cross-language variability in word length and word processing rates.

Part 3 explores the interface between language and social cognition. Brian MacWhinney discusses the gradual emergence of language and focuses in particular on the emergence of perspective taking — the ability to represent the mental states of other people. This ability is part of a broader conceptual framework known as theory of mind. Bertram Malle lays out the possible evolutionary relations between theory of mind and language in light of their extant developmental relations and concludes that only a version of co-evolution fits the data at hand. Dare Baldwin focuses on a specific element of theory of mind, namely, young children's emerging grasp of intentions underlying human action. She explores possible relations between the development and evolution of intention understanding and the evolution of language.

Part 4 deals with various aspects of language development. Mike Tomasello outlines a "cut-and-paste" model of language learning, suggesting a possible analogy between language ontogeny and language evolution. Two related chapters, one by Susan Goldin-Meadow and the other by Jill Morford, document the spontaneous emergence of sign language in children, suggesting possible evolutionary analogies. As a counterpoint and final challenge, Dan Slobin argues against the validity of analogical data from child language for hypotheses about language evolution and, further, against the validity of analogical data from pidginization, diachronic change, or extant human brain morphology.

This volume sprang from an inter-disciplinary symposium on the evolution of language out of pre-language, held at the University of Oregon in the spring of 2001. The Institute of Cognitive and Decision Sciences at the University of Oregon and the Shaolin-West Foundation of Ignacio, Colorado, provided partial funding for this event. We are indebted to Vonda Evans for help in the local arrangement.

Part 1

Language and the brain

CHAPTER 1

The visual information-processing system as an evolutionary precursor of human language

T. Givón
University of Oregon

1. Introduction*

This paper surveys data from a wide range of disciplinary and methodological traditions, bringing them to bear on specific issues in language evolution. My central theme is fairly transparent — that a large chunk of the neurology that supports human language processing is an evolutionary outgrowth of the primate visual information-processing system. An ancillary theme, that the human peripheral lexical code must have arisen first as a visual-gestural code, has already been anticipated by others (Hewes 1973a, 1973b, 1975; Stokoe 1974; Armstrong *et al.* 1995; *inter alia*).

In the course of the discussion, a broader issue emerges, one that in my judgement will continue to haunt us for some time to come: Does the neural circuitry that supports language processing involve any language-specific mechanisms or modifications, or is it simply an assembly of pre-existing modules that have been recruited to perform a novel behavioral task? Theoretical tempers on this issue run high, and are inevitably clustered around the two extreme positions:

- *Extreme emergentist*: All language-processing modules continue to perform their older pre-linguistic tasks and reveal no special language-dedicated adaptations.
- *Extreme innatist*: All language-processing modules are either entirely novel, or at the very least have been heavily modified to perform their novel linguistic tasks.

However fond we may be of our theoretical predilections, they should not blind us to the possibility that the truth may lie somewhere in the middle. That is, that *some* modules may be still shared, while others are already dedicated (or re-dedicated). What is more, there may not be enough evidence on hand yet to resolve this issue conclusively, empirically, with the requisite analysis of the full pattern of associations and dissociations.[1] In principle at least, the pre-linguistic descent of some or even all neural sub-systems currently used in language processing remains compatible with either extreme position, though perhaps with a caveat:

- *Centrist caveat*: It is unlikely that pre-existing modules could perform their complex novel tasks without some novel coordinating-and-switching meta-circuitry.

Two distinct cycles of symbolization — or code-development — must have taken place in the evolution of human language. The first involved the rise of a well-coded *lexicon* and its coding instrument, phonology. The second involved the rise of *grammar* and its coding instrument, morpho-syntactic structure. In each cycle, an initial phase of a more natural (iconic, non-arbitrary) code must have been followed by a shift toward a more arbitrary, symbolic code. In each, both the early iconicity and the later shift toward symbolism had unimpeachable adaptive motivations. To understand the rise of both codes ('structures') as adaptive evolutionary processes rather than as serendipitous mutations, one must first describe human communication as a functional-adaptive instrument.[2]

2. A functional-adaptive overview of human communication

While human language undoubtedly performs affective, social and aesthetic functions, its two core functions are the representation and communication of experience ('knowledge', 'information'). The overall well-coded system[3] may be thus divided into two major components:

- The cognitive representation systems
  - Conceptual lexicon
  - Propositional information
  - Multi-propositional discourse
- The symbolic communicative codes
  - Peripheral sensory-motor codes
  - The grammatical code

2.1 The cognitive representation system

2.1.1 *The conceptual lexicon and words*

The human lexicon is a repository of relatively time-stable culturally-shared well-coded knowledge about our external-physical, social-cultural and internal-mental universe. By 'relatively time-stable' one means knowledge that is not in rapid flux, i.e. not unique episodic information. By 'culturally shared' one means that when launching into communication, speakers take it for granted that words have roughly the same meaning for all members of the same cultural/linguistic community.[4] By 'well-coded' one means that each chunk of lexically-stored knowledge is more-or-less uniquely — or at least strongly — associated with its own perceptual code-label.[5]

The conceptual lexicon is most likely organized as a *network* of nodes and connections. Within this network, nodes stand for individual concepts ('senses'), each with its own distinct code-label or ('word'). 'Concepts' are mental representations of *types* of conventionalized experience (rather than of tokens of individual experience). The conceptual lexicon is thus a repository of *generic* conventionalized information.

Lexical concepts may represent relatively time-stable entities (physical objects, landmarks, locations, flora, fauna, persons, cultural institutions or abstract entities), typically classified as nouns. They may represent actions, events, processes or relations, typically classified as verbs. They may also represent inherent qualities or temporary states, often classified as adjectives. Cognitive psychologists have long recognized the conceptual lexicon under the label *semantic memory* (Atkinson and Shiffrin 1968).

2.1.2 *Propositional information and clauses*

Concepts — thus words — are combined into propositions about the events, states, relations and qualities in which entities partake. Those propositions, coded grammatically as *clauses*, may again pertain to the external world, internal world, culturally-mediated world, or various combinations thereof. And cognitive psychologists have long recognized propositional information under the label of *episodic memory* (Atkinson and Shiffrin 1968; Squire 1987), the repository of uniquely-experienced *tokens* of states or events, or uniquely-encountered token entities.[6]

A somewhat reciprocal relation holds between the two longer-term memory systems:

- Developmentally, memory traces of unique but similar individual *tokens* of experience presumably give rise, over repeated exposure, to time-stable lexical concepts of *types* of experience.
- In processing unique *tokens* of experience, one recognizes them automatically as members of known lexical *types*.

2.1.3 *Multi-propositional discourse and clause-chains*

Individual propositions may be combined together into coherent discourse. Human discourse is predominantly multi-propositional, in that its coherence transcends the bounds of isolated propositions. The lowest multi-propositional unit of coherent discourse is coded grammatically as a *clause-chain.*

Many cognitive psychologists recognize this level as a distinct representational entity called 'text', albeit often conflated with propositional information under *episodic memory* (e.g. Kintsch 1974, 1982; Anderson *et al.* 1983; van Dijk and Kintsch 1983; Gernbsbacher 1990; Ericsson and Kintsch 1995; Trabasso *et al.* 1984; Mandler and Johnson 1977; Squire 1987; Squire and Zola-Morgan 1991; Mishkin and Petri 1984; *inter alia*)

To illustrate the combinatorial nature of conceptual meaning, propositional information and multi-propositional coherence, consider the simple-minded examples in (1), (2) and (3) below:

(1) **Concepts (words)**
  a. drive
  b. insane
  c. constant
  d. abuse
  e. maid
  f. kill
  g. butler
  h. knife
  i. hide
  j. fridge

(2) **Propositions (clauses)**
  a. The maid was driven insane.
  b. The butler constantly abused the maid.
  c. The maid killed the butler with a knife.
  d. The maid hid the knife in the fridge last night.

(3) **Multi-propositional discourse (clause-chain)**
  Having been driven insane
  by constant abuse,
  the maid killed the butler with the knife
  that she had hidden in the fridge the night before.

Taken by themselves, outside any propositional context, the words in (1a–j) can only have lexical–conceptual meaning. That is, one may only ask about them questions such as:

(4) a. What does *drive* mean?
b. Does *drive* mean the same as *abuse*?
c. If someone is a *maid*, can she also be a *butler*? Or a *woman*?
d. Is *kill* related in meaning to *murder*, and if so how?

Combined into clauses, as in (2a–d), the very same words now partake in the coding of propositional information. In addition to questions of meaning as in (4), the individual clauses in (2) may now prompt many questions of information, such as:

(5) a. Was the maid driven insane?
b. Who abused the maid?
c. Who killed the butler?
d. Who did the maid kill?
e. What did the maid kill the butler with?
f. Did the maid kill the butler?
g. Where did the maid hide the knife?
h. When did the maid hide the knife in the fridge?

Finally, the clause-chain in (3), in which the very same propositions of (2) are now combined into multi-propositional text, has discourse coherence. In addition to questions of meaning such as (4), and of information such as (5), one may now ask questions that pertain to that coherence, such as:

(6) a. Why did she kill him?
b. How come she had a knife?
c. Why had the maid hidden the knife in the fridge?
d. Could she perhaps have talked to him first before taking such a drastic step?
e. Was her action reasonable? Was it defensible in a court of law?

Questions (6) may appear deceptively like those in (5). However, each question in (5) can be answered on the basis of knowing a single proposition in (2). In contrast, none of the questions in (6) can be answered on the basis of such atomic propositional knowledge. Rather, the knowledge of several propositions in the connected discourse (3), or even of the entire coherent text, is required in order to answer such questions.

The partial dissociation between conceptual meaning and propositional information is easy to demonstrate by constructing grammatically well-formed sentences that make no sense; that is, sentences whose words are perfectly meaningful each taken by itself but still don't combine into a cogent proposition, as in Chomsky's ubiquitous example (7):

(7) Colorless green ideas sleep furiously

The meaning incongruities that make proposition (7) bizarre ('colorless green', 'green ideas', 'ideas sleep', 'sleep furiously') are all due to the considerable semantic specificity of individual words. The relation between lexical meaning and propositional information is thus one of inclusion relation, or a one-way conditional:

- One can understand the meaning of words independent of the proposition in which they are embedded; but one cannot understand a proposition without understanding the meaning of the words that make it up.

The partial dissociation between propositional information and discourse coherence is just as easy to demonstrate, by stringing together perfectly informative but incoherently-combined propositions. Thus, re-scrambling the coherent discourse in (3) yields the incoherent (8):

(8) a. The maid killed the butler with the knife
b. by constant abuse,
c. having been driven insane
d. that she had hidden in the fridge the night before.

One need not ascribe any propositional-semantic anomaly to the individual propositions in (8a–d) in order to explain the incoherence of the clause-chain (8). Its bizarreness is due to lack of cross-clausal coherence — at least given the grammatical form of the clauses, retained from the coherent (3).

The relation between propositional information and discourse coherence is thus also one of inclusion or a one-way conditional. That is:

- One can understand the meaning of clauses independent of the discourse they are embedded in; but one cannot understand the discourse without understanding the propositions that make it up.

2.2 The symbolic communicative codes

2.2.1 *Peripheral sensory-motor code*

As noted earlier, the sensory-motor coding system of human language serves primarily to map lexical-conceptual meaning onto phonological words. It involves both a perceptual *decoder* and an articulatory *encoder*, with possible variation in peripheral modality (auditory-oral, visual-gestural, tactile-motor, etc.).

2.2.2 *The grammatical code*

2.2.2.1 *Preamble* Grammar is, in all likelihood, the latest evolutionary addition to the arsenal of human communication (Givón 1979; Lieberman 1984; Bickerton

1980, 1990; Li, in this volume). While the evolutionary argument remains conjectural, it is supported by a coherent body of suggestive evidence. Ontogenetically, both hearing and signing children acquire the lexicon first, using pre-grammatical (pidgin) communication before acquiring grammar (Bloom 1973; Scollon 1976; Givón 1990c). Natural 2nd language acquisition follows the same course but often stops short of grammaticalization altogether (Bickerton 1981; Bickerton and Odo 1976; Selinker 1972; Schumann 1976, 1978, 1985; Andersen 1979; Givón 1990c).

In the natural communication of pre-human species, the existence of concepts of both entities (nouns) and events (verbs) must be taken for granted if one is to make sense of the behavior, communicative or otherwise. And such lexical concepts can occasionally become well-coded in natural pre-human communication (Cheney and Seyfarth 1990, Marler *et al.* 1991, *inter alia*).

Further, birds, dogs, horses, primates and other pre-human species can easily be taught auditory, visual or gestural lexical code-labels for nouns, verbs and adjectives (Premak 1971; Gardner and Gardner 1971; Fouts 1971; Terrace 1985; Savage-Rumbaugh *et al.* 1993; Savage-Rumbaugh and Lewin 1994; Savage-Rumbaugh, Shanker and Taylor 1998; Pepperberg 1991; Tomasello and Call 1997; *inter alia*).

The relative ease with which the teaching of a well-coded lexicon takes place in many pre-human species strongly suggests that the underlying neuro-cognitive structure — semantic memory — is already in place. In contrast, documenting the natural communicative use of grammar in pre-human species, or teaching it to them, is much harder to demonstrate (Premak 1971; Terrace 1985; Tomasello and Call 1997; Hauser 2000), despite some claims to the contrary (Greenfield and Savage-Rumbaugh 1991, Savage-Rumbaugh *et al.* 1993, Savage-Rumbaugh 1998).

2.2.2.2 *Grammar as a structural code* Grammar is a much more complex and abstract symbolic code than the sensory-motor codes of the lexicon. At its most concrete, the primary grammatical signal involves four major devices:

(9) **Coding devices of the primary grammatical signal**

- Morphology
- Intonation:
  - clause-level melodic contours
  - word-level stress or tone
- Rhythmics:
  - pace or length
  - pauses
- Sequential order of words or morphemes

Some coding devices (morphology, intonation/stress) are more concrete, involving the very same peripheral sensory-motor devices that code the conceptual lexicon. But even those concrete devices are integrated into a complex whole with the more abstract elements of the code (rhythmics, sequential order) that are probably second- or third-order inferences.

The use of the more concrete phonological code to signal grammatical morphology is a derivative of its primary use as the lexical code. This is so because grammatical morphemes develop historically out of lexical words (Givón 1971, 1979; Traugott and Heine eds 1991; Heine *et al.* 1991; Hopper and Traugott 1994; *inter alia*).[7]

From the primary signals in (9), more abstract elements of the grammatical code derived, perhaps as second-order inferences:

(10) **More abstract elements of the grammatical code:**
- Hierarchic constituency organization
- Category labels (noun, verb, adjective; noun phrase, verb phrase; subject, object)
- Scope and relevance relations (operator–operand, noun-–modifier, subject–predicate)
- Government and control relations (agreement, control of coreference, modality, finiteness)

The cognitive status of these formal entities is at the moment far from clear.

2.2.2.3 *The functional-adaptive correlates of grammar* The complexity — and thus abstractness — of grammar is probably due to its complex interaction with multiple language-processing modules. It interacts intimately with semantic memory (lexicon), propositional semantics, episodic memory (discourse pragmatics), working memory (surface form), Broca's area (automated rhythmic-hierarchic routines), and the attentional system (Givón 1991b). It is implicated massively in the many perspective-shifting and re-framing operations (MacWhinney 1999, 2002; Givón 1997). It is thus the quintessential *distributed*, interactive, multi-modular system, so much so that some take this as evidence for its neurological non-specificity (Bates and Goodman 1999).

Grammar codes simultaneously two cognitive representational domains — propositional semantics and discourse pragmatics. But the portion of grammar dedicated to coding propositional semantics is relatively small, confined largely to the marking of:

- semantic roles of the participant (agent, patient, recipient, location etc.)
- semantic transitivity (state, event, action).
- clause union and complementation

The bulk of the machinery of grammar is dedicated to coding discourse pragmatics, i.e. cross-clausal discourse coherence. The most common discourse-pragmatically oriented grammar-coded domains are listed in (11) below (Givón 1984b, 1990b, 2001).

(11) Major grammar-coded sub-domains

- grammatical roles (subject, direct object)
- definiteness and reference
- anaphora, pronouns and agreement
- tense, aspect, modality and negation
- pragmatic transitivity (voice)
- topicalization
- focus and contrast
- relativization (backgrounding)
- speech acts
- clause chaining and switch reference
- adverbial subordination (backgrounding)

2.2.2.4 *Pre-grammatical communication*

Humans can, under a variety of developmental, social or neurological conditions, communicate without grammar, using the well-coded lexicon with some rudimentary combinatorial rules. The difference between pre-grammatical and grammatical communication may be summed as follows:

(12) Pre-grammatical vs. grammatical discourse processing
(Givón 1979, 1989, 1993)

Properties	Grammatical mode	Pre-grammatical mode
Structural		
a. grammatical morphology	abundant	absent
b. syntactic constructions	complex/ embedded	simple/ conjoined
c. use of word order	grammatical (subj/obj)	pragmatic (topic/comment)
d. pauses	fluent	halting
Functional		
e. processing speed	fast	slow
f. Mental effort	lower	higher
g. Error rate	lower	higher
h. context dependence	lower	higher

Cognitive			
i.	processing mode	automated	attended
j.	acquisition	later	earlier
k.	evolution	later	earlier

The slow and analytic pre-grammatical communication is heavily vocabulary driven. This tallies with the fact that vocabulary is acquired before grammar, in both first and second language acquisition. Thus, pre-grammatical children, adult pidgin speakers and agrammatical aphasics comprehend and produce coherent connected discourse, albeit at slower speeds and high error rates than those characteristic of grammatical language. The identification of grammar with a more automated, speeded-up language processing system has been suggested (Givón 1979, 1989, 1991b; Blumstein and Milberg 1983; Lieberman 1984; Schnitzer 1989).

As an example of coherent but largely pre-grammatical child episodic text, consider the following two passages from a 29-month-old Spanish-speaking boy interacting with his father (Wise 1994; Givón 1990c). In the first passage, high-frequency verb inflections for first and second person already conform to the adult model:

(13) a. No sé.
NEG know-1SG
'I don't know.'
b. Ves?
see-2SG
'You see?'
c. Mira, así dos!
look-IMPER.2SG like.this two
'Look, like this, two!'
d. Sí, tuyo dos.
'Yes, yours two.'
'Yes, [there are] two [with] yours.'
e. Aquí hay campo.
here have/NEUT room
'Here there's room.'
f. Así, Papi!'
like.this, daddy
'Like this, Daddy!'

In the second passage, the same invariant high-frequency verb form (3rd person singular) is used regardless of number (14b) and person (14c), an oblique form of 'you' is substituted for the nominative, and the plural noun suffix is dispensed with (14b):

(14) a. Mi pantalón está . . .
my pant be-3sg
'My pants are . . .'
b. Aquí está dos carro aquí.
here be-3sg two car here
'Here there're two cars here.'
c. Aquí está tí, Papi.
here be-3sg you.obl daddy
'Here you are, Daddy.'
d. Aquí está Guiguis arriba!
here be-3sg G. up
'Here is Guiguis up there!'

As an example of coherent adult second-language pidgin, consider the following passage, spoken by a Hawaii Japanese-English pidgin speaker (Bickerton and Odo 1976; Bickerton and Givón 1976; Givón 1990c):

(15) . . . oh me? . . . oh me over there . . . nineteen-twenty over there say come . . . store me stop begin open . . . me sixty year . . . little more sixty year . . . now me ninety . . . nah ehm . . . little more . . . this man ninety-two . . . yeah, this month over . . . me Hawaii come-*desu* . . . nineteen seven come . . . me number first here . . . me-*wa* tell . . . you sabe gurumeru? . . . you no sabe gurumeru? . . . yeah this place come . . . this place been two-four-five year . . . stop, ey . . . then me go home . . . Japan . . . by-n-by . . . little boy . . . come . . . by-n-by he been come here . . . ey . . . by-n-by come . . . by-n-by me before Hui-Hui stop . . . Hui-Hui this . . . eh . . . he . . . this a . . . Manuel . . . you sabe-*ka* . . .

As an example of coherent narrative produced by an agrammatic aphasia patient, consider (Menn 1990: 165):

(16) . . . I had stroke . . . blood pressure . . . low pressure . . . period . . . Ah . . . pass out . . . Uh . . . Rosa and I, and . . . friends . . . of mine . . . uh . . . uh . . . shore . . . uh drink, talk, pass out . . .
. . . Hahnemann Hospital . . . uh, uh I . . . uh uh wife, Rosa . . . uh . . . take . . . uh . . . love . . . ladies . . . uh Ocean uh Hospital and transfer Hahnemann Hospital ambulance . . . uh . . . half'n hour . . . uh . . . uh it's . . . uh . . . motion, motion . . . uh . . . bad . . . patient . . . I uh . . .

> flat on the back . . . um . . . it's . . . uh . . . shaved, shaved . . . nurse, shaved me . . . uh . . . shaved me, nurse . . . [sigh] . . . wheel chair . . . uh . . . Hahnemann Hospital . . . a week, a week . . . uh . . . then uh . . . strength . . . uh . . . mood . . . uh . . . up . . . uh . . . legs and arms, left side uh . . . weak . . . and . . . Moss Hospital . . . two week . . . no, two months . . .

In the absence of morpho-syntax, the bulk of coded clues for establishing text coherence of pre-grammatical discourse are furnished by the lexicon. But a small component of *proto-grammar* is already evident in pre-grammatical communication (see below). Neither the lexical clues nor proto-grammar disappears in the grammaticalized fluent speech of native adults. Rather, vocabulary-cued processing remains a parallel processing channel alongside grammar (Kintsch 1992; Givón 1993). And the more iconic conventions of proto-grammar are integrated with more conventional rules of grammaticalized language Givón 1984a, 1990a).

2.2.2.5 *Pre-grammar as proto-grammar* Earlier discussions of pre-grammatical pidgin communication tended to suggest that it not only lacked grammatical morphology, but also displayed rather haphazard combinatorial behavior, thus lacking syntax (Bickerton 1975, 1977; Bickerton and Odo 1976). A closer examination of pidgin communication, however, reveals that it abides by several rather explicit regularities that may be called either pre-grammar or *proto-grammar*. The common denominator of those rules is that they are highly iconic and cognitively transparent, as compared with the more arbitrary symbolic conventions of full-fledged morpho-syntax.

The most common regularities of pre/proto-grammar are summarized in (17)–(20) below (Givón 1979, 1984a, 1989, 1990a; Haiman 1985a, 1985b):

(17) **Intonation rules**
  a. **Stress**: more information is stressed
  b. **Melody and relevance**: units of information that belong together conceptually are packed together under a unified melodic contour
  c. **Pause and rhythm**: the size of the temporal break between chunks of information corresponds to the size of the cognitive or thematic break between them

(18) **Spacing rules**
  a. **Proximity and relevance**: units of information that belong together conceptually are kept in close temporal proximity (see also (17b above)
  b. **Proximity and scope**: functional operators are kept close to the operands to which they are relevant

(19) **Sequence rules**
  a. **Order and importance**: the most important unit of information is placed first
  b. **Occurrence order and reported order**: The order of occurrence of events in reality should be mirrored in the linguistic account

(20) **Quantity rules**
  a. **Zero expression and predictability**: predictable — or already activated — information is left unexpressed
  b. **Zero expression and relevance**: unimportant or irrelevant information is left unexpressed

The most interesting observation about the rules of proto-grammar is that, without exception, they perform exactly the same functions in grammaticalized language as they do in pre-grammaticalized communication. Nothing has been lost, but rather some things have been added. In grammaticalized language, the rules of pre/proto-grammar may function by themselves. Most commonly, however, they are integrated together with the more arbitrary, symbolic elements of grammar — morphology and governed syntactic constructions.

3. The primate visual information-processing system

3.1 Directionality and hierarchy

In the primate cortex, visual information from the optic nerve and mid-brain reaches the primary visual cortex (area 17 in the occipital lobe). From there on, processing moves roughly forward toward more anterior cortical locations. But many feedback loops, backward pathways and sideways connections exist, so that the entire system — over three-score of centers at last count — is complex and interactive rather than linear (Maunsell and Van Essen 1983; Kaas 1989).

3.2 Iconic spatial ('retinotopic') maps

The 2D iconic spatial representation of the retina is preserved in all the visual processing centers in the cortex, in what is known as *retinotopic maps*. But it is also transformed during the movement toward more anterior locations. So that while the general spatial orientation is preserved, more anterior visual maps are more abstract, selective and context-dependent. They are, presumably, progressively more amenable to higher-level feedback from more anterior cognitive centers, as well as to cross-modal information (Kaas 1989).

3.3 The two streams of visual information processing

Somewhere still in the occipital lobe beyond area 19, visual information-processing splits into two main 'streams', the lower *ventral stream* leading toward the front of the temporal lobe via an inferior temporal route, and the upper *dorsal stream* leading toward the posterior parietal lobe (Ungerleider and Mishkin 1982; Ungerleider 1985; Kaas 1989). A schematic representation of the two pathways is given in Figure 1.

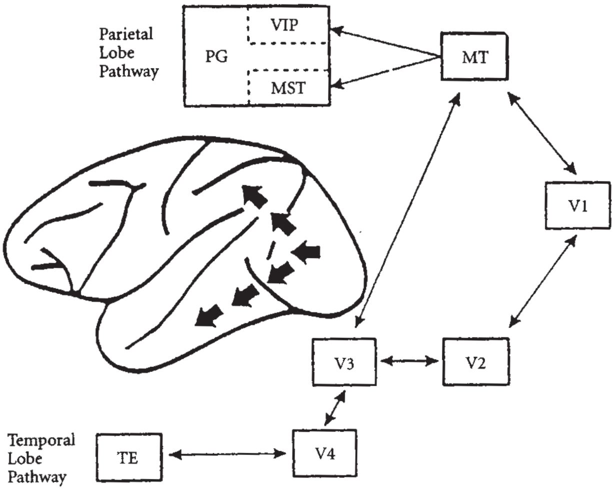


Figure 1. The two pathways of visual information in the primate brain (Kaas 1989). A simplified scheme of processing in visual cortext of macaque monkeys stressing two parallel pathways, with the ventral stream for object vision and the dorsal stream to PG (parietal area "G") for visual attention and spatial relations. Modified from Ungerleider (1985). See text and Ungerleider (1985) for details and abbreviation.

The ventral stream of visual information processing is the *object recognition* pathway, identifying visual objects as members of particular semantic *types*. The dorsal (upper) stream of visual information processing is responsible for episodic *tokens* of spatial relations, spatial motion, or visual objects.

3.4 The correspondence between visual, cognitive and linguistic information systems

The two streams of visual information processing correspond, rather transparently, to two core components of human cognitive and linguistic representation systems (sec. 2. above), of which they are but the visual precursors:

- object recognition = semantic memory = lexicon
- spatial relation/motion = episodic memory = propositions

This identification is surprisingly straightforward. First, recognizing a visual object as a member of some generic type according to its visual attributes is merely an early prototype of lexical-semantic identification.

Second, processing an episodic token of spatial relation between objects is but the early prototype of episodic-propositional information about *states*. And processing an episodic token of spatial motion of one object relative to another is but the early prototype of episodic-propositional information about *events*.

Other facts concerning the brain's physical and functional organization reinforce this identification. First, the dorsal visual stream feeds into the subcortical limbic region of the *hippocampus* (Tucker 1991, in this volume), site of episodic-declarative memory (Squire 1987; Squire and Zola-Morgan 1991). The ventral visual stream feeds into the limbic subcortical region of the *paleo-cortex* and *amygdala* (Tucker 1991). From the limbic region, both streams eventually project to front-cortical representation, one for long-term episodic memory (Squire 1987; Squire and Zola-Morgan 1991), the other to lexical-semantic representation (Tucker 1991, in this volume; Peterson *et al.* 1988, 1989, 1990; Snyder *et al.* 1995; Posner and Pavese 1997; Abdulaev and Posner 1997).

Already in pre-human primates, the ventral visual stream does not process only visual objects, but also visual events and actions — i.e. referents of *verbs*. Thus Perrett *et al.* (1989), in a study of single-cell activation in monkeys, identified single cells that respond to objects vs. those that respond to actions — both in the *superior temporal sulcus* of the ventral visual stream. The actions — e.g. moving an object by hand towards an open mouth — are concrete and spatial-visual, but they already involve more abstract computation of purpose and goal:

> . . . The coding of interrelationships that is inherent in goal-centered descriptions provides a framework through which the visual system can achieve a rich understanding of the world which embodies causation and intention . . . (Perrett *et al.* 1989: 110)

The transformation of the ventral object-recognition stream in the temporal lobe into a more generalized, conceptually richer — and eventually cross-modal — lexical representation pathway must have thus begun in pre-human primates.[8]

4. The neurology of human language

4.1 Preamble: Modularity vs. interaction

In this section I will sketch out what can be assumed prudently about the localization in the human brain's left-cortical hemisphere of automated modules involved in language processing. In saying that these modules are 'involved', I deliberately bypass the question of whether they are or aren't *dedicated* exclusively to language processing (Bates and Goodman 1999).

Since the use of the term 'module' is controversial and given to extreme reductive interpretations, I propose to follow the more neutral perspective adopted by Posner and Carr (1991), who suggest that a processing circuit is modular if:

- *Informational encapsulation*: It does not rely on help from other operation in its own input-output transformations.
- *Attentional self-sufficiency*: It has its own attentional resources.
- *Cognitive impenetrability*: It follows its own operational rules independently of higher-level strategic or volitional considerations.

Posner and Carr (1991) further note that the three conditions are logically (and often factually) independent of each other. So that rather than an all-or-none forced choice, a system's may be both 'modular' and 'interactive' — to a degree.

In adopting a more flexible approach to modularity, one has no choice but to recognize that — both functionally and neurologically — language processing in the left hemisphere is distributed over many distinct functional-anatomical modules. Such a modular organization is indeed the hallmark of the left hemisphere, which specializes in automated processing of both linguistic and non-linguistic information (Geschwind and Levitzky 1968; Geschwind 1970; Oujman 1991). Hierarchic organization of modules is also the hallmark of all automated processing systems. In this regard, the left hemisphere contrasts sharply with the less modular, less automated, less specialized right-hemisphere (Joanette *et al.* 1990; Olness 1991).

Do the language-related modules in the human brain constitute a global, dedicated 'language organ' in the Fodorian sense? A clean yes-or-no answer to this question may be unwise, for a number of reasons. First, the two older cognitive representation modules — semantic memory (lexicon) and episodic memory (propositions/discourse) — still serve their pre-linguistic functions and thus represent both linguistic and visual information.

Second, the two younger and more language-specific left-hemisphere coding modules — phonology and the grammar — also turn out to occupy brain locations that are directly adjacent to their immediate phylogenetic precursors.

The evolutionary functional extension of pre-existing 'modules' is thus a recurrent theme in the neurology of language.

Third, of the two coding modules, phonology is probably older and may have already become a language-dedicated module. But even that remains open, since it has been shown that even a relatively short life-time exposure to reading can give rise to an automated word recognition (Sierof and Posner 1988; Givón *et al.* 1990). What is more, the brain module involved in automated visual word-recognition is not innately dedicated to reading, but rather develops in the region of the left-occipital lobe that is dedicated to visual object recognition (Petersen *et al.* 1988, 1989, 1990; McCandliss *et al.* 1999).

4.2 The semantic modules

4.2.1 *Location*

Early studies of left-hemispheric lesions located all meaning in *Wernicke's area* of the left-temporal region (Geschwind 1970). But the linguistic performance data obtained from lesion studies leave open at least two interpretations as to what was disrupted in Wernicke's aphasics:

- *lexical-semantic* (word level) representation
- *propositional-combinatorial* (clause level) representation

Thus, Chomsky's celebrated example (7) (*Colorless green ideas sleep furiously*) is indeed a fair caricature of Wernicke's aphasic speech, but it leaves us no grounds for deciding whether the semantically aberrant 'colorless green', 'green ideas', 'ideas sleep' and 'sleep furiously' are due to disrupted lexical access to 'color', 'colorless', 'green', 'idea', 'sleep' and 'furiously', or to a disrupted combinatorial processor that cannot integrate lexical meanings into larger semantic projections.

More recently, PET (positron emission tomography) studies identified another meaning-related area at the *left-pre-frontal* cortical region, in addition to Wernicke's area (Petersen *et al.* 1988, 1989, 1990; Posner *et al.* 1988, 1989; Fiez and Petersen 1993; Raichle *et al.* 1993).

When ERP (evoked response potential) imaging was combined with PET studies, the time-course of activation of the two semantic regions was shown to be rather different: The pre-frontal area activates first, peaking at ca. *200ms* following stimulus presentation. The temporal (Wernicke's) region follows, peaking at ca. *800ms* (Snyder *et al.* 1995; Posner and Pavese 1997; Abdulaev and Posner 1997). What is more, activation of the pre-frontal activation was triggered by purely lexical-semantic tasks, while activation of the Wernicke region was triggered by more combinatorial tasks.

The combined results of the PET and ERP studies lead to a tentative identification of the pre-frontal site as that of lexical-semantic representation, and the

Wernicke medial-temporal site as that of propositional-semantic (combinatorial) representation (Posner and Pavese 1997; Abdulaev and Posner 1997).

The most interesting aspect of the time-course of ERP activation is that it converges rather closely with the time-course studies reported earlier in the cognitive literature for, respectively, lexical and propositional semantic activation (multiple-sense activation and clausal-context resolution, respectively; see Swinney 1979; Neely 1990; see discussion in Barker and Givón, in this volume).

4.2.2 *Cross-modal vs. bi-modal lexicon*

In the neuropsychological literature, two diametrically opposed schools of thought exist as to whether the human brain has two modality-specific lexical-semantic representation modules — one for vision, the other for language — or a single unified multi-modal module. Since the evolutionary hypothesis presented in this paper is compatible only with the second position, a brief summary of the two is in order.

- *Two modality-specific semantic lexicons*: this tradition, represented by two converging groups of lesion and PET studies, holds that the two modalities of information input — visual and verbal — have two distinct representations (Warrington 1975; Warrington and Shallice 1979, 1984; Warrington and McCarthy 1983; Shallice 1988a; Bub *et al.* 1988, Chertkow *et al.* 1990).
- *Single cross-modal lexicon*: this tradition, looking at the very same cortical lesion patients, argues for a unified cross-modal representation (Riddoch and Humphreys 1987; Humphreys and Riddoch 1988; Riddoch *et al.* 1988; Funnell 1989; Funnell and Alport 1989).

While I am not competent to pass on the fine technical details of the debate, the single lexicon position is much more compatible with the bulk of the cross-disciplinary evidence reviewed in this paper, as well as with the paper's central thesis.

The single cross-modal lexicon is also much more compatible with general trends in biological evolution (Mayr 1982; Futuyma 1986) as well as in neuro-evolution (Tucker 1991, in this volume). In both, the dominant mode of innovation is the extension of existing organs and modules to novel but closely-related functions (*homoplasy*), perhaps as the first step toward the *terminal addition* of newly re-configured modules. Much — though by no means all — of the connectivity of the cortex may be viewed in such evolutionary terms (Tucker 1991, in this volume).

4.3 Episodic memory

4.3.1 *Localization*

The brain localization of early episodic memory is relatively uncontroversial. Studies of both primate and human brain lesions implicate the *hippocampus* and *amygdala* under the left *medial temporal* lobe (Squire 1987; Squire and Zola-Morgan 1991; Mishkin 1978, 1982; Mishkin *et al.* 1982, 1984; Mishkin and Petri 1994). This system is responsible for declarative, propositional, non-procedural, non-semantic memory of token states/events or individuals. Episodes may be either single or long chains of states/events.

According to Squire and Zola-Morgan (1991), *early episodic memory* displays the following characteristics:

- It is extremely malleable and involves further processing and reprocessing and reorganization of stored information.
- It is a limited-capacity processor in which storage-space and processing activity compete.
- It is emptied periodically and thus remains available for the processing of incoming new information.
- It is thus a crucial intermediary between *working-memory* (Baddeley 1986; Shallice 1988; Ericsson and Kintsch 1995) and a more permanent *longer-term episodic-memory* module in a front-cortical region.
- Impairment in episodic recall due to hippocampal lesions is dissociated from both procedural and lexical knowledge.
- Input into the hippocampus-based system may come from multiple modalities (pictorial-visual, written-visual, gestural-visual, auditory-verbal, auditory non-verbal).

The anatomic location and connectivity of the sub-cortical hippocampus-based memory system hints at both its functional and evolutionary connectivity. To begin with, it is the destination of the dorsal visual-information stream, which is its evolutionary projection (Tucker 1991, in this volume). It also sits directly under the temporal lobe, and is connected outward toward the left-temporal surface by the amygdala and other left-cortical structures. This presumably allows connectivity to and input from the *auditory language module*, located near the *primary auditory cortex* in the left-temporal region.

4.3.2 *Modality of input*

That input into the hippocampus-based memory system is cross modal is uncontroversial. Access into this system is mediated via *working memory* ('short-term memory', 'current attention focus'). The latter involves, in addition to the conscious *executive attention*, several modality-specific sub-conscious channels (Baddeley 1986; Shallice 1988b; Gathercole and Baddeley 1993). The relationship between the modality-specific working memory modules, executive attention and episodic memory is represented schematically in Figure 2.

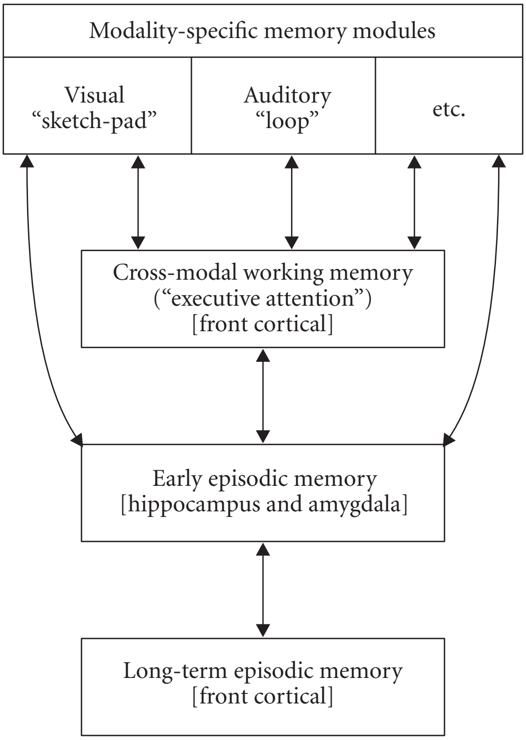


Figure 2. Relationship working memory (attention) and episodic memory (following Baddeley 1986; Shallice 1988b; Gathercole and Baddeley 1993; Squire and Zola-Morgan 1991)

4.4 The peripheral coders and de-coders

4.4.1 *Peripheral word-recognition modules*

The localization of an auditory (phonological) word-recognition module in the left-temporal lobe near Wernicke's area and the primary auditory cortex is well established (Geschwind 1970). This localization has been re-confirmed recently with added precision by two independent groups of PET-scan studies (Howard *et al.* 1992; Raichle *et al.* 1993). In addition, two locations of a visual written-word recognition module have been recently detected. First, a group of PET studies located that module in the *left-medial extra-striated* cortex, just anterior to the primary visual center in the occipital lobe (Petersen *et al.* 1988, 1989, 1990; Posner *et al.* 1988, 1989; Posner and Carr 1991; Carr and Posner 1992).

Raichle *et al.* (1993) comment on the apparent density and specificity of multiple language-related modules in the traditional Wernicke's area:

> . . .several areas in the left posterior temporal cortex are involved in the processing of words. A more dorsal area(s) (Petersen *et al.* 1988, 1989; Wise *et al.* 1991) appear to be concerned with the processing of auditory words, whereas a more ventral area observed in the present experiment appears to be concerned with the processing requirements of visually-presented words . . . (Raichle *et al.* 1993: 23)
>
> . . . The appearance of activity in the left posterior temporal cortex is consistent with its well-established role in language processes that began with Wernicke's seminal observations (Wernicke 1874). However, the current findings challenge the concept that a single area located in this region has, as its major function, the analysis of words regardless of modality of presentation. Rather, this region of the human brain almost certainly contains a more complex array of processing areas . . . (1993: 24)

The most interesting point for the purpose of this discussion is that both proposed visual word-recognition modules sit along the *ventral* object-recognition stream, one in a more posterior location, the other in a more anterior one:

> . . . From the present experiment there also appear a more ventral area which is active when subjects generate verbs from visually-presented nouns, but not during task involving the auditory presentation of words. The more ventral location of this latter area and its apparent involvement visually and not auditorily-presented words would suggests that it may be part of the postulated ventral visual processing stream in the middle and inferior temporal cortices (Ungerleider and Mishkin 1982) . . . (Raichle *et al.* 1993: 23)

4.4.2 *The peripheral articulatory modules*

The localization of the oral-articulatory motor control and arm-hand-finger control for writing and gestured sign is fairly well established, in various locations in the primary motor cortex, in the left-frontal lobe near Broca's area. The significance of this localization will be discussed further below.

4.5 Grammar and Broca's area

The coding-decoding module at the lower end of the left-frontal primary motor cortex — *Broca's area* — was earlier implicated with grammar. But this locus may well be just one out of many closely-interacting brain regions involved in grammar. The incredible complexity of the human grammatical code, its late development in both evolution and ontogeny, and most of all its multiple functional interactions with other modules — lexical semantics, propositional semantics, discourse pragmatics, attention and working memory, episodic memory, the right-brain perspective modulation — all suggest a *distributive* brain representation (Bates and Goodman 1999).

Agrammatic aphasia resulting from lesions in Broca's area has been described extensively (Zurif *et al.* 1972; Goodglass 1976; Bradley 1979; Kean 1985; Menn and Obler eds 1990; *inter alia*). In her review, Greenfield (1991) discusses the localization around or above Broca's area of a variety of automated complex skills. Their common denominator is roughly that the activities in question all involve habituated, automated, complex, sequential-hierarchic routines. Some of these are motor skills. Others are planned action routines, hierarchic-object combination routines, hierarchic visual-tracking skills, or hierarchic tool-using skills.

Greenfield's (1991) comparative survey of human, primates and children data is important for a number of reasons:

- The homologue of Broca's area that is responsible for automated complex routines, including coordinated manual-dextral routines and oral-facial skills, is already present in non-human primates.
- In human neonates, Broca's area matures around the age of two years, roughly when grammar kicks in. And children up to that age also lag in other complex hierarchic skills.
- In addition to the well-known strong neurological connection (via the *arcuate fasciculus*) to Wernicke's area, Broca's area also has strong connection to the pre-frontal lexical-semantic area.

The Broca-Wernicke connection is important in terms of the functional interaction between syntax and propositional semantics. The Broca-prefrontal connection is important in terms of syntax's functional interaction with verb-frame semantics ('argument structure'). But it is also important in terms of syntax's well-known dependency on grammatical morphology, which evolves diachronically from lexical words (Givón 1971, 1979; Heine *et al.* 1991; Traugott and Heine eds 1991; Hopper and Traugott 1994; Bybee *et al.* 1994) and may still be represented in the same brain location as the lexicon.

Finally, a series of studies by Rizzolatti and associates (Rizzolatti and Gentilucci 1988; Rizzolatti *et al.* 1996a, 1996b, 2000; Rizzolatti and Arbib 1998) show that Broca's area and related regions in both humans and non-human primates involve not only the motor neurons that control manual action routines, but also *mirror neurons* that are implicated in visual representation of observed manual action. Put another way, Broca's area is also involved in the visual representation of complex rhythmic-hierarchic activity. The evolutionary implications of these findings is immense, in that they provide for a representational precursor of syntax.

4.6 Interim summary

The distribution of the language-related cortical modules in the left hemisphere is given, somewhat schematically, in Figure 3.

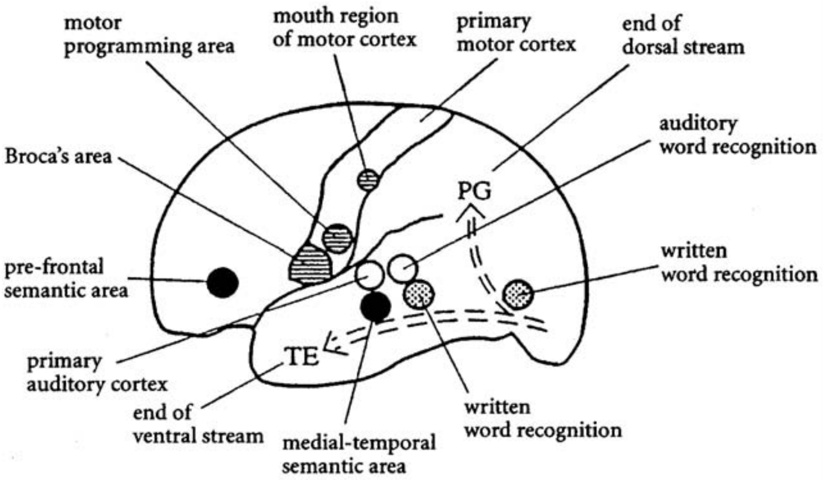


Figure 3. Localization of language-related cortical modules relative to their pre-linguistic precursors (following Posner *et al.* 1989; Car and Posner 1992; Howard *et al.* 1992; Raichle *et al.* 1993; Greenfield 1991; Kaas 1989)

What is most salient about this layout is that each language-specific module in the left-hemisphere is either a transparent direct adaptation of its respective pre-linguistic use, or is located in close proximity to its putative pre-linguistic sensory precursor. The vast majority of these modules have clear evolutionary connections to the visual information-processing system. In order:

- **The lexical-semantic module**
  The ventral visual object-recognition stream leads from area 17 through the pre-striate area and the medial-temporal lobe, projecting forward up to the inferior temporal lobe. Like the pre-frontal lexical-semantic area, it is an evolutionary projection of the limbic system (paleo-cortex and amygdala), which still supplies the connectivity between the temporal and pre-frontal regions.
- **Combinatorial-semantic module**
  Wernicke's area, now re-identified as a combinatorial-semantic module, sits directly astride the ventral object-recognition stream.
- **Episodic memory**
  The dorsal visual processing stream for spatial relations (states) and motion (events) projects back to the hippocampus. The medial-temporal cortical areas

above the hippocampus (amygdala) are strongly involved in its operations, and are adjacent to the primary auditory cortex and the language-specific auditory word-recognition module.

- **Visual word-recognition**
  Both the pre-striate (posterior) and medial-temporal modules of written word recognition sits directly along the ventral visual object-recognition stream.
- **Broca's area (grammar)**
  Broca's area is directly adjacent to its functional-evolutionary precursor that control automated, complex, hierarchic cognitive skills of varying degrees of abstraction, but are also involved in the more abstract, visual representation of such syntactic routines.

Only two modules, both involving the modality-specific peripheral codes, have no direct connection to the visual antecedents. Both, however, have some connectivity to adjacent pieces of the visual system:

- **Auditory word recognition (phonology)**
  The auditory word-recognition module in the medial-temporal lobe is adjacent to Wernicke's area and to its presumed pre-linguistic precursor, the primary auditory cortex. But it is also adjacent to the visual word-recognition module below it, astride the visual object-recognition (ventral) stream.
- The articulatory motor coders
  Both the oral-facial motor-skill module and the writing/ gesture hand-control module are located at the lower end of the primary motor cortex near Broca's area. Both display visual representation machinery — the mirror neurons.

5. Evolutionary hypothesis

Given the facts surveyed above, the conclusion that the neural circuits that support language processing in humans evolved out of their respective pre-linguistic precursors, primarily out of various components of the visual information-processing system, is inescapable. This hypothesis may be broken into the following parts:

- Cognitive representation modules: Both main modules of this system — semantic memory and episodic memory — arose pre-linguistically as visual information processing modules. Their cross-modal representational capacity is already attested in non-human primates, and has been simply expanded in humans.
- The peripheral lexical code: The early peripheral lexical code of human communication must have been visual-gestural, thus building upon two pre-existing capacities — the visual object-recognition stream, and the capacity for complex

manual routines and their visual representation. The shift to auditory-oral coding must have come sometime later.

- **The grammatical code**: The grammar code must have been the latest addition to the human communicative capacities, taking advantage of and integrating multiple pre-existing capacities:
  - a well-coded cross-modal lexicon;
  - a cross-modal episodic memory;
  - the combinatorial-semantics module (Wernicke's);
  - the complex manual dexterity module and its visual representation system (Broca's);
  - the modality-specific working-memory cum executive attention systems;
  - perspective-shifting right-cortical modules.

At least in principle, grammar could have arisen either before or after the shift from visual-gestural to auditory-oral lexical code.

6. Supporting arguments

6.1 The rise of a visual-gestural coding

Given that both rudimentary gestural and vocal communication is attested in non-human primates, and given the relative ease with which one could teach such primates either an auditory (Savage-Rumbaugh *et al.* 1993, Savage-Rumbaugh and Lewin 1994) or gestural (Fouts 1973) lexical code, it is reasonable to suppose that either modality was available for the evolution of the human peripheral lexical code. My conjecture that visual-lexical coding predominated during the early evolutionary stage of human communication is based on the following arguments.

6.1.1 *Neuro-cognitive pre-adaptation*

Visual information processing had already evolved as the main perceptual modality of vertebrates, mammals and primates. Both the perceptual and articulatory neurological mechanisms — the ventral object-recognition stream and the capacity for complex manual routines had been already in place. The vision-dominated cognitive representation system — the lexical and episodic visual-processing streams — were likewise in place. Virtually no change in the overall connectivity of the system would have been required.

6.1.2 *Ease of associative learning*

Gestural communication is, at least at its early onset, highly iconic, mapping onto concrete visual properties of its referents. The neuro-cognitive adjustment between recognizing a frequently-encountered visual object or action and recognizing their iconic gestural visual representations would have been relatively small.

Object recognition itself is already somewhat abstract and schematized, rapidly becoming even more so during ontogeny and learning. An iconic representation that preserves the most salient schematic features of objects or actions is thus highly transparent and requires minimal instruction.

6.1.3 *Concreteness of early human lexicon*

Given what is known of conceptual capacity and lexical learning of non-human primates (de Waal 1982; Savage-Rumbaugh and Lewin 1994; Tomasello and Call 1997), the early human lexicon must have been equally concrete, confining itself to, primarily, sensory-motor spatial-visual objects, states and actions. Early child vocabulary is equally concrete (Bloom 1973; Carter 1974; Lamendella 1976, 1977; Bates *et al.* 1979). Such concrete referents are thus maximally amenable to iconic, gestural coding. And even in the lexicon of extant human languages, the bulk of abstract vocabulary is derived by analogy and metaphoric extension from concrete core vocabulary (Lakoff and Johnson 1980).

6.1.4 *Natural sign*

The incredible ease with which the visual-gestural channel is recruited to take over the peripheral coding of human language in the case of hearing impairment, and the smooth integration of this peripheral component into the rest of the system, would be puzzling unless an old well-honed channel — the ventral visual object- recognition stream — had already been in place. There is no precedent for such a profound yet instant neurological adaptation, nor for the incredible facility with which it seems to occur (Golding-Meadows, in this volume; Morford, in this volume).

6.1.5 *Written word-recognition module*

As in the case of sign language, the ability of hearing humans to shift perceptual modality in the acquisition of literacy would have been unlikely unless it capitalized on the pre-existing viusual object recognition channel (ventral trend). The lifetime development of an automated visual word-recognition module along the ventral visual-object recognition stream (Petersen *et al.* 1988, 1989, 1990; Posner *et al.* 1988, 1989; Car and Posner 1992; McCandliss *et al.* 1999) attests to the enduring availability of this channel for recruitment to linguistic input.

6.2 The shift to auditory-oral coding

6.2.1 *Adaptive advantages*

While the visual-gestural channel may have had crucial advantages in pioneering the human lexical code, many of its initial advantages turn out to be eventual adaptive drawbacks. These drawbacks constitute supporting arguments for a subsequent shift to an auditory-oral lexical code.

6.2.1.1 *Ritualization and loss of iconicity* Frequent use of well-coded signals inevitably leads to automation, speed-up, signal reduction and ritualization (Haiman 1985a, 1985b, 1992, 1994; Givón 1989, 1991a; Bybee, in this volume). This is just as true of human sign languages (Klima and Bellugi 1979; Pettito 1991).

A similar process is also attested in the natural evolution of writing systems, whose time-course in all five known centers where literacy is known to have arisen independently — China, India, Mesopotamia, Egypt, Maya — followed the same gradual move from early iconicity to later abstraction, arbitrariness and symbolization.

Ritualization and symbolization of the communicative signal over time is also attested in primate dominance displays (Simonds 1974), chimpanzee recruitment signals (de Waal 1982), and equine dominance signals (Givón 1998).

The same broad trend from more iconic to more symbolic signals has also been noted in the evolution of honey-bee communication (von Frisch 1967; Gould and Towne 1987; Gould 1991), as well as in the development of human child pointing signals (Carter 1974; Bates *et al.* 1975, 1979; Bloom 1973; Givón 1998).

6.2.1.2 *The rise of abstract concepts* As the human lexicon expanded and acquired more abstract concepts, visual-gestural coding loses the advantage of its early iconicity, and in fact may become a drawback. A sign may suggest a concrete visual configuration, but its referent is already abstract.

6.2.1.3 *Freeing the hands for other activities* The shift from visual-gestural to auditory-oral communication frees the hands to perform other tasks, so that communication may now proceed simultaneously with manual activities, and can in fact support them.

6.2.1.4 *Transcending the immediate visual field* Auditory-oral communication may proceed in the dark, in thick bush, over physical barriers that prevent eye contact. The auditory signal is also much easier to decode from a distance than gestural signals, given that sign language typically involves facial expressions and hand-and-arm movements made primarily in front of the face and chest (Klima and Bellugi 1979).

6.2.1.5 *Parallel deployment of attention* Auditory-oral communication frees visual attention for strategic deployment elsewhere. When necessary, communication can now rely on auditory attention alone, while both parties attend visually to other tasks — emergencies, priorities, joint activities.

6.2.2 *Neurological plausibility of the modality shift*

6.2.2.1 *Adjacency* There is at least one well-known precedent for a visual-to auditory modality shift in the annals of neurology. While it involves the avian mid-brain and is thus not homologous to the primate cortex, it may nevertheless be cited by way of analogy.

In the mid-brain of the barn owl, four areas of spatial input analysis are directly adjacent to each other (Takahashi *et al.* 1984; Takahashi and Konishi 1986; Takahashi 1989):

- a non-iconic auditory input area (no spatial maps);
- an iconic auditory input area (spatial maps);
- a bi-modal (auditor-visual) iconic area (spatial maps);
- an iconic visual area (spatial maps);

A schematic representation of the structure of the optic tectum of the barn owl is given in Figure 4.

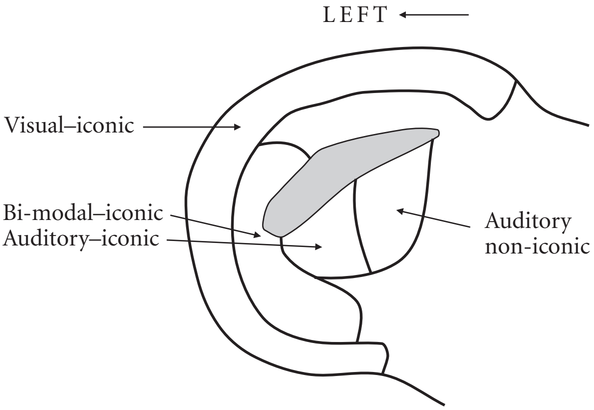


Figure 4. Approximate localization of visual, auditory auditory and cross-modal spatial representation areas in the optic tectum of the barn owl (T. Takahashi, i.p.c.)

As Knudsen (1985, 1989) and Knudsen and Knudsen (1990) have shown, the barn owl's auditory spatial-orientation capacity can only develop through early — pre-maturation — training by the visual system. Owl chicks raised without visual input during the first 90 days of their life are incapable of creating 2D iconic spatial maps from auditory input alone. In other words, it is the presence of visually-derived spatial maps in the visual layer of the optic tectum that trains the behavior of the adjacent auditory center — via an intervening bi-modal region that is adjacent to both.

By analogy with the modality-shift in the barn owl, the putative modality shift in the coding of the human lexicon may have been facilitated by the adjacency of the temporal-lobe auditory module to the ventral visual stream.

6.2.2.2 *Co-activation* Another necessary condition for retraining of one processing modality by another is temporal co-activation. In this connection, Neville (1995) notes that prior to neurological maturation, the human brain displays considerable cross-modal connectivity and co-activation of visual and auditory input. This connectivity is pared down markedly between birth and 3 years of age (Spelke 1987; Melzoff 1990; Rose 1990). It is at least plausible that this current human developmental trend recapitulates a phylogenetic one, so that co-activation of the auditory by the visual system may have facilitated the modality shift in human lexical coding.

6.2.2.3 *Sub-cortical connectivity* As Tucker (1991; in this volume) notes, the various perceptual modalities are virtual islands on the cortical surface. But they display much stronger cross-modal connectivity — thus mutual co-activation — in the sub-cortical limbic regions. Since both visual and auditory linguistic input is routed through, and presumably integrated in, the limbic system (hippocampus, paleocortex and amygdala), it is plausible that cross-modal connectivity and co-activation in the limbic system played an important role in facilitating the shift from visual to auditory lexical coding.

6.3 The evolution of grammar

There are many reasons why one should be inclined to suggest that grammar was the last major developmental step in the evolution of human language (Givón 1979; Bickerton 1981; Lieberman 1984). Many of the arguments are either conjectural, convergent or recapitulationist, so that they only make sense if taken *in toto*.

6.3.1 *Dependence on a precursor lexicon*

The evolution of grammar must have depended crucially on the prior availability of a well-coded lexicon, not only of concrete but also of some abstract concepts, and for the following reasons.

- *Communicative referents*: A propositional message containing grammar without a referential lexicon is contentless and thus adaptively nonsensical.

- *Abstractness*: Grammatical concepts are much more abstract and classificatory than the lexicon. Given the overwhelming concreteness of both primate and early childhood lexicons (Tomasello and Call 1997; Savage-Rumbaugh *et al.* 1993; Bloom 1973; Carter 1974; Bates *et al.* 1979), it is unlikely that the evolution of the grammatical code could have preceded the evolution of at least some rudimentary well-coded lexicon.
- *Grammaticalization I*: Grammatical morphology: Grammatical morphology arises, overwhelmingly, from the semantic bleaching of (often concrete) precursor lexical words (Givón 1979; Heine *et al.* 1991; Hopper and Traugott 1994; Bybee *et al.* 1994; Slobin 1997; Bates and Goodman 1999). In parallel, grammatical morphemes undergo phonological reduction and become clitics or affixes on lexical words. Both the functional and structural correlates of this process depend on the prior existence of a well-coded lexicon.
- *Grammaticalization II*: Syntactic constructions: Complex or embedded syntactic constructions arise diachronically via condensation ('syntacticization') of bi-clausal ('paratactic') precursors (Givón 1979, 1984b, 1990b, 1991c, 1995). But both primate communication and early childhood communication are overwhelmingly mono-propositional (Tomasello and Call 1997; Savage-Rumbaugh *et al.* 1993; Carter 1974; Bates *et al.* 1975, 1979).

6.3.2 *Declarative speech-acts*

Even when using a well-coded lexicon, early childhood communication and, more conspicuously, primate communication are predominated by manipulative speech-acts (Tomasello and Call 1997; Savage-Rumbaugh *et al.* 1993; Carter 1974; Bates *et al.* 1975, 1979). But the bulk of the grammatical machinery of human language is invested in the coding of declarative speech-acts (Givón 1984b, 1990b, 2001).

6.3.3 *Displaced reference*

Early childhood and primate communication is overwhelmingly about here-and-now, you-and-I, and this-and-that other referents accessible in the immediate speech situation. But the bulk of grammar is dedicated to communication about spatially and temporally displaced referents, states and events (Givón 1984b, 1990b, 2001).

6.3.4 *Multi-propositional discourse*

Early childhood and primate communication are overwhelmingly mono-propositional (Tomasello and Call 1997; Savage-Rumbaugh *et al.* 1983; Bloom 1973; Carter 1974; Scollon 1976; Bates *et al.* 1975, 1979). But the bulk of the formal machinery of grammar is invested in coding multi-propositional, cross-clausal coherence (Givón 1984b, 1990b, 2001).

6.3.5 *Perspective shifting*

Young children take their time learning to shift perspective to other times, other places and other persons. Their early communication, even with well-coded lexicon, is still peculiarly egocentric, in this way reminiscent of the communication of non-human primates (Tomasello and Call 1997; Savage-Rumbaugh *et al.* 1993; Baldwin, in this volume; Malle, in this volume). But the bulk of the formal machinery of grammar is invested in perspective-shifting operations (MacWhinney 1999, in this volume; Givón 1984b, 1990b, 1997).

6.3.6 *Complexity and distributiveness*

The lexicon, episodic memory and the peripheral sensory-motor codes of human language are relatively localized and modularized. In contrast, grammar is distributive, multi-modular and supremely interactive (Bates and Goodman 1999). In this, grammar bears the hallmark of a later, more complex, more global integrative faculty with a large number of evolutionary precursors.

6.3.7 *Concomitant cultural evolution*

Well-coded human language evolved in the cultural context of the foraging (hunting-and-gathering) *society of intimates*, which had been the exclusive cultural mode of social primates, hominids and humans until the advent of agriculture ca. 8,000 BC. The most salient characteristics of such a society are (Givón and Young 2002):

- *Restricted group size*: The typical group size of primate, hominid and human foraging societies remained stable at ca. 50–150 over the last several million years (Dunbar 1992).
- *Restricted gene pool and kinship-based cooperation*: Foraging societies exhibit high genetic uniformity and kin-based social structure and social cooperation.
- *Restricted cultural differentiation*: Foraging societies exhibit minimal occupational, technological and socio-economic differentiation, except along biologically-based parameters such as gender, age, descent-line and personality (dominance).
- *Non-hierarchic social organization*: Foraging societies are highly egalitarian and exhibit a fluid and contingent leadership structure, based on socially-recognized merit or charisma (de Waal 1982; Power 1991).
- *Restricted geographic range and communicative isolation*: Foraging societies occupy exclusive, relatively stable territories, and are largely in communicative isolation from — and hostile relations with — other social groups of the same species.
- *Informational stability and homogeneity*: In the society of intimates, generic-cultural (‘semantic’, ‘procedural’) knowledge is highly stable and uniformly

shared. Specific-personal ('episodic') information is rapidly disseminated and thus universally shared.

The change from the slow, error-prone, context-dependent pre-grammatical communication to the more streamlined, efficient and less context-dependent grammaticalized language took place in the context of the foraging society of intimates. The concomitants of this change — expanded universe of reference, declarative information, multi-propositional discourse, well-coded perspective-shifts — constituted the adaptive bonus of the evolution of grammar. This is the bonus that facilitated:

- more subtle and explicit cognitive representation;
- more efficient dissemination of information;
- more efficient social cooperation;
- more explicit social planning;
- more efficient social learning.

These concomitants of grammatical communication were in turn both cause and consequence of the gradual drift that eventually loosened up the closed and relatively rigid socio-cultural calculus of the society of intimates and propelled it, gradually but inexorably, toward the more open, socially complex, technologically sophisticated and informationally open-ended *society of strangers.*

Anatomically-modern humans date back to about 110,000 BC, and untill about 50,000 BC their presence outside Africa was confined to an early beach-head in the Middle East. About 40,000 BC, a series of explosive socio-cultural changes seem to have transformed that small population (Li, in this volume):

- rapid population explosion;
- far-flung geographic spread (Europe to Australia);
- explosive advance in stone-age technology;
- explosive rise in symbolic art.

It may be that the rise of more streamlined and precise grammar-coded communication was part and parcel of that explosion, indeed one of its major facilitators, as Li and Hombert (2000) suggest.

7. Recapitulation and analogy

7.1 Preamble

The suggestion that the grammatical code evolved later than the well-coded lexicon is part of the larger issue of the relevance of recapitulationist and analogical thinking to language evolution. Such thinking has been pursued by a number of authors as ancillary support for specific scenarios of language evolution (Lamendella 1976, 1977; Givón 1979, 1998; Bickerton 1981, 1990). All these authors argue that three linguistic developmental processes:

- child language development
- diachronic change
- pidginization-creolization

are relevant to language evolution.

In a spirited attack on recapitulationist and analogical arguments, Slobin (in this volume) rejects them all categorically, in part on purported factual grounds, but mostly on the principled ground that the neurology of current *Homo sapiens* — whether mature or neonate — cannot be fruitfully compared to the neurology of its hominid ancestors during the various stages of language evolution. In the following space I will argue that in spite of the *prima facie* logical validity of Slobin's main argument, an analogical, recapitulationist perspective on language evolution is both useful and legitimate, provided one is not dogmatic or reductionist about it, and provided it can be shown to converge with other types of evidence.

7.2 The philosophy of science argument

> ... At best, close attention to biology, development, and linguistic behavior can heighten the plausibility of those scenarios. But we can never have sufficient evidence to scientifically evaluate such narratives, however appealing and plausible some of them may be ... (Slobin, this volume: p. 387)

Science is not in the business of absolute certainty and incontrovertible proof, which are the province of deductive logic (Wittgenstein 1918; Popper 1934; Hanson 1958). In science, having an explicit hypothesis, however tentative and speculative, is vastly preferable to having no hypothesis at all. This is so because once an explicit hypothesis is on the table, one can derive its logical consequences and, with the hope that at least some of those are testable empirical predictions.

Testable empirical predictions are themselves neither conclusive nor absolute nor everlasting, nor are they always immediate. Some hypotheses wait for years before the facts that bear on their predictions are discovered. In other cases, some old facts that have been around forever turn out, on further inspection from a fresh perspective, to have a bearing on a hypothesis that has been going begging.

Scientific knowledge is by its very nature tentative, contingent and temporary. A hypothesis lasts as long as no better hypothesis — one that predicts more and enlarges the domain of explanation and understanding — has been advanced. Therefore, if by observing the three developmental processes and non-human communication one can get ideas, however suggestive and analogical, about the possible scenario of language evolution, this is still vastly preferred to refraining from making any hypotheses because 'we can never have sufficient evidence to scientifically evaluate such narratives'.

7.3 The not-the-same-neurology argument

Nobody denies that the neurology of current *Homo sapiens*, either neonate or adult, is different from the neurology of its hominid evolutionary antecedents. But this still leaves two questions wide open:

- How extensive are the differences?
- Are the differences relevant?

Since we know the precise answer to neither question, the argument loses a certain measure of its acuity.

As Bates and Goodman (1999) point out in their discussion of the emergence of grammar, and as I have suggested repeatedly above, the bulk of the neurological mechanisms relevant to human language are neither new nor language specific, but have been recruited into their novel linguistic use from multiple functionally-amenable pre-linguistic domains. Many of these still perform their older pre-linguistic functions. And in the relevant domains of language processing, these mechanisms may still retain much of their older properties.

7.4 The neural-mechanism argument

Given what was said directly above, it is entirely plausible that in spite of many differences, most neurological mechanisms that support language learning and change ('emergence') in humans are phylogenetically old and functionally generalized, as are the language-learning mechanisms proposed by Tomasello (in this volume).

The recruitment of Broca's area to partake in the distributed neurology that supports human syntax is a case in point, as is the syntactically complex structure of bird-songs, as are the dexterous manual routines of primates. Neither of the latter two map onto a communicative code, but the developmental neuro-cognitive mechanisms that support all three — chunking, hierarchization, categorization, automation — are phylogenetically ancient and domain-general (Hauser 2000).

Lastly, synchronic 'on-line' processing is central to the actual mechanism that drives both language ontogeny and diachronic change. This is so because both developmental processes are nothing but the cumulation of on-line processing behaviors. As Ernst Mayr has noted, the same is also true in the selection phase of biological evolution:

> . . . Many if not most acquisitions of new structures in the course of evolution can be ascribed to selection forces exerted by newly acquired *behaviors*. Behavior, thus, plays an important role as the *pacemaker* of evolution . . . (Mayer 1982: 612; italics added)

What is more, ontogeny — via neoteny — is likewise heavily implicated in the actual mechanism of phylogenetic change (Gould 1977).

7.5 Developmental trends

Rather than propose a full evolutionary scenario, recapitulated or otherwise, I will outline here the developmental trends that I think are also safe bets as evolutionary trends. In each case, a recapitulation of or analogy to other developmental processes is clearly manifest.

7.5.1 *Lexicon before grammar*

This developmental trend can be factored out into a number of sub-trends.

- lexicon before syntax
- lexicon before morphology
- pidgin before grammatical communication

In this connection, Bowerman (1973) observes:

> . . . early child speech is 'telegraphic' — that is consists of strings of content words like nouns, verbs, and adjectives, and lacks inflections, articles, conjunctions, copulas, prepositions and post-positions, and, in general, all functors or 'little words' with grammatical but not referential significance . . . (Bowerman 1973: 3–4; see also Slobin 1970)

This trend is also observed in natural adult second-language acquisition (Selinker 1972; Schumann 1976, 1978, 1985; Bickerton and Givón 1976; Bickerton and Odo 1976; Andersen 1979; Givón 1990c). The same trend was reported for 2nd language immersion programs (Cohen 1976; Campbell 1984; McLaughlin 1984).

The evolution of grammatical morphology from lexicon is the universal trend in grammaticalization (Heine *et al.* 1991; Traugott and Heine eds 1991; Hopper and Traugott 1994; Bybee *et al.* 1994).

7.5.2 *Iconic syntax before arbitrary syntax*

The more iconic regularities of pre-grammar (Section 2.2.2.5. above) are already present in second language Pidgin (Givón 1984a), and are also found in early child language (Keenan 1974, 1975; Bates and MacWhinney 1979). The more arbitrary, symbolic rules of grammar emerge later, but absorb and incorporate the conventions of pre-grammar (Haiman 1985a, 1985b, 1992, 1994; Givón 1989, 1991a; Bybee, in this volume).

In the development and evolution of animal communication, the move from natural-iconic to ritualized-symbolic signals is a major trend (sec. 2.2.2.4. above). In this context, the suggestion that the human peripheral sensory-articulatory lexical code evolved first as an iconic visual-gestural system and only later shifted to an auditory-oral code makes perfect sense.

7.5.3 *Simple clauses before complex clauses*

Simple clauses predominate in early child syntax (Bowerman 1973), where complexity kicks later and gradually (Limber 1973). Spontaneous primate communication is overwhelmingly made out of simple clauses (Savage-Rumbaugh *et al.* 1993; Tomasello and Call 1997). And 2nd language pidgin is conspicuously devoid of complex constructions (Selinker 1972; Bickerton and Odo 1976; Givón 1990c).

In addition, complex clauses most commonly arise diachronically from the conflation of two or more simple clauses; that is, syntaxis from parataxis (Givón 1979, 1984b, 1990b, 1991c, 1995; Traugott and Heine eds 1991, vol. II).

7.5.4 *Mono-propositional before multi-propositional discourse*

Early childhood discourse is predominated by short, mono-clausal turns (Bloom 1973; Scollon 1976). Only gradually do multi-propositional turns kick in (Ervin-Tripp 1970). Primate spontaneous communication is likewise overwhelmingly mono-propositional (Tomasello and Call 1997), as is animal communication in general.

7.5.5 *Manipulative before declarative*

Manipulative speech-acts predominate early childhood communication (Carter 1974; Bates *et al.* 1975), as it also does spontaneous primate communication (Savage-Rumbaugh *et al.* 1993; Tomasello and McCall 1997; de Waal 1982). The same is true, in the main, of all animal communication.

7.6 One-word before two-word before longer clauses

Early child syntax is predominated by one-word clauses (Bloom 1973; Scollon 1976), giving way gradually to two-word clauses (Bowerman 1973). Spontaneous primate communication is predominated by one-word clauses, with a smattering of two-word clauses (Savage-Rumbaugh *et al.* 1993; Tomasello and Call 1997).

8. Closure

As Slobin (in this volume) wisely cautions us, a certain element of speculation is inevitable in the discussion of language evolution. The fact that the bulk of the behavioral and neurological fossils are missing, most likely for good, cannot be glossed over. But such a situation has never stopped science from advancing explanatory hypotheses (Hanson 1958), and it is hard to see why it should stop us now. Like other biological phenomena, language cannot be fully understood without reference to its evolution, whether proven or hypothesized.

There is one conspicuous precedent in linguistics for an analytic gambit that is heavily dependent on hypothetical, analogical reasoning — the method of *internal reconstruction*. Much of what we have learned about grammaticalization in the last 30 years, let along much of the pioneering work of 19th Century linguistics, rests on our ability, or willingness, to interpret synchronic structural distortions as relic fossils of prior diachronic change (Givón 1979, 2000).

In the present context, I think the evidence strongly implicates various components of the primate visual information-processing system as evolutionary precursors of the human language-processing machinery. It would of course be absurd to suggest that my case has been proven beyond all doubt, given the analogical and circumstantial nature of the evidence. But the core hypothesis is explicit enough so that one can continue to derive specific predictions from it, and in some instances even test them experimentally (Barker and Givón, in this volume).

Does the neural machinery that supports human language processing involve any language-specific mechanisms, or is it but an assembly of pre-existing modules that have been recruited to perform novel behavioral tasks? Here I think the line of arguments advanced by Bates and Goodman (1999) can be followed — but only up to a point.

It seems at least plausible that the cognitive representation system that underlies human language has been recruited to language processing without major adjustments. On the other hand, both communicative codes — phonology, grammar — involve, at the very least, some highly specific re-organization of the pre-existing visual, auditory or articulatory modules. And if phonology is indeed phylogenetically older than grammar, it is more likely that this re-organization has been more profound.

Whether part of the re-organization has already taken place during evolution, and thus has enough neuro-genetic consequences to qualify the resulting structures for the lofty rubric ‘innate’, is to my mind an open question. Bates and Goodman (1999) reject this possibility with respect to grammar. Given the nature of grammar as an intricate, finely-coordinated, multi-modular network, I find this wholesale rejection premature. It may well be that the dedicated neurology of grammar is too subtle and distributive, involving temporal coordination and activation patterns. Whether such complex governing mechanisms can simply 'emerge' without at least some evolutionary reconfiguration and thus dedication of the attendant neurology remains an open question to me.

Notes

* I have benefitted immensely at various times from comments and suggestions from Elizabeth Bates, Derek Bickerton, Terry Deacon, Patricia Greenfield, Charles Li, Bertram Malle, Rich Marrocco, Helen Neville, John Orbell, Irene Pepperberg, Mike Posner, Sue Savage-Rumbaugh, Dan Slobin, Terry Takahashi, Nathan Tublitz, Don Tucker, Sherman Wilcox and Phil Young. Needless to say, they are absolved from any responsibility for the final product. Earlier precursors of this paper can be found in Givón (1995, 1998)

1. For an innatist position, even a non-extreme one, to be upheld, one must demonstrate that at least *some* language-processing modules are *never* activated during the performance of non-linguistic tasks. For an extreme emergentist position to be upheld, one must show that *every* language-processing module is still activated during the performance of at least *some* non-linguistic task.

2. In his *De Partibus Animalium*, Aristotle, the founder of empirical biology, observes that biological design ('structures') made no sense unless interpreted as purposive instruments ('functions'), an observation that remains foundational to modern biology:

> . . . if a piece of wood is to be split with an axe, the axe must of necessity be hard; and, if hard, it must of necessity be made of bronze or iron. Now exactly in the same way the body, which like the axe is an *instrument* — for both the body as a whole and its several parts individually have definite operations for which they are made; just in the same way, I say, the body if it is to do its *work* [= function], must of necessity be of such and such character . . ." (*De Partibus Animalium*, McKeon 1941: 650; italics and bracketed glosses added)

3. The use of well-coded ('linguistic') communication in either humans or non-humans does not preclude access to many less-well-coded information-gathering channels. Such parallel 'back-channels' extract relevant information from both the physical and social context, and are no doubt the evolutionary precursor of well-coded communication.

4. Both 'same meaning' and 'same community' are of course relative matters.

5. Given a certain level of homophony (lexical ambiguity), the uniqueness of code-meaning connections is not absolute, but rather a strong tendency. An ambiguous (homophonous) word automatically activates several senses ('multiple activation'), and an activated sense automatically activates conceptually-related senses ('semantic priming', 'spreading activation; Swinney 1979; Neeley 1990).

6. Some neuro-psychologists (Mishkin and Petri 1984; Mishkin *et al.* 1984; Squire 1987; Squire and Zola-Morgan 1991) re-cast the distinction between semantic and episodic memory *procedural* vs. *declarative*, respectively.

7. The lexical origin of grammatical morphology may not hold quite as absolutely in the case of gestural-visual communication (sign language), where at least some scholars report that most grammatical morphemes evolve directly from more iconic but non-lexical gestures (Pettito 1992). Dan Slobin (i.p.c.) suggests that this may not be universal (see also Hoiting and Slobin 2001).

8. Pepperberg's (1991) work on avian lexicalized communication suggests that a lexicon of nominal, adjectival and verbal concepts can be represented in the pre-mammalian mid-brain or limbic system. Given the column-like projection of mid-brain and limbic rings into corresponding higher cortical rings (Tucker 1991, in this volume), an evolutionary extension of both semantic and declarative-episodic memory from sub-cortical to cortical rings — both posterior and anterior — is plausible.

9. See also Goodenough (1990) for a discussion on the role of declarative propositions in language evolution.

References

Abdulaev, Y.G. and Posner, M.I. 1997. "Time-course of activating brain areas in generating verbal associations". *Psychological Science* 8.

Andersen, R. (1979 "Expanding Schumann's pidginization hypothesis". *Language Learning* 29.

Anderson, A.H., Garrod, S.C., and Sanford, A.J. 1983. "The accessibility of pronominal antecedents as a function of episodic shift in narrative". *Quarterly Journal of Experimental Psychology* 35A.

Armstrong, D.F., Stokoe, W.C., and Wilcox, S.E. 1995. *Gesture and the Nature of Language.* Cambridge: Cambridge University Press.

Atkinson, R.C. and Shiffrin, R.M. (1968) "Human memory: A proposed system and its control processes". in K.W. Spence and T. Spence (eds), *The Psychology of Learning and Motivation*, Volume II. New York: Academic Press.

Baddeley, A.D. 1986. *Working Memory.* New York: Oxford University Press.

Baldwin, D. This volume. "The rise of intentional understanding in human development: Analogies to the genesis of language".

Barker, M. and Givón, T. This volume. "On the pre-linguistic origins of language processing rates".

Bates, E., Camioni, L., and Volterra, V. 1975. "The acquisition of performatives prior to speech". *Merrill-Palmer Quarterly* 21.

Bates, E., Benigni, L., Bretherton, I., Camioni, L., and Volterra, V. 1979. *The Emergence of Symbols: Cognition and Communication in Infancy.* New York: Academic Press.

Bates, E. and Goodman, J.C. 1999. "On the emergence of grammar from lexicon". In *The Emergence of Language*, B. MacWhinney (ed). Mahwah, NJ: Lawrence Erlbaum Associates.

Bates, E. and MacWhinney, B. 1979. "A functionalist approach to the acquisition of grammar". In *Developmental Pragmatics*, E. Ochs and B. Schieffelin (eds). New York: Academic Press.

Bickerton, D. 1975. "Creolization, linguistic universals, natural semantax and the brain". Honolulu: University of Hawaii (ms).

Bickerton, D. 1977. "Pidginization and creolization: Language acquisition and language universals". In *Pidgin and Creole Linguistics*, A. Valdman (ed). Bloomington: Indiana University Press.

Bickerton, D. 1981. *Roots of Language.* Ann Arbor, Mich.: Karoma.

Bickerton, D. 1990. *Language and Species.* Chicago: University of Chicago Press.

Bickerton, D. and Givón, T. 1976. "Pidginization and syntactic change: From SOV and VSO to SVO". *CLS 12.* University of Chicago: Chicago Linguistics Society.

Bickerton, D. and Odo, C. 1976. *Change and Variation in Hawaiian English*, Volume I: *The Pidgin.* NSF Final Report (grant GS-39748), Honolulu: University of Hawaii.

Bloom, L. 1973. *One Word at a Time: The Use of Single Word Utterances Before Syntax.* The Hague: Mouton.

Blumstein, S.E. and Milberg W. 1983. "Automatic and controlled processing in speech/language deficits in aphasia". *Symposium on Automatic Speech.* Minneapolis: Academy of Aphasia.

Bowerman, M. 1973. *Early Syntactic Development.* Cambridge: Cambridge University Press.

Bradley, D.C. 1979. "Syntactic deficits in Broca's aphasia". In *Biological Studies of Mental Processes*, D. Caplan (ed). Cambridge, Mass.: MIT Press.

Bub, D., Black, S., Hampton E., and Kertesz A. 1988. "Semantic encoding of pictures and words: Some neuropsychological observations". *Cognitive Neuropsychology* 5.

Bybee, J. This volume. "Hierarchy vs. sequentiality: The making of linguistic structure".

Bybee, J., Pagliuca, W., and Perkins, R. 1994. *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World.* Chicago: University of Chicago Press.

Campbell, R. 1984. "The immersion approach to foreign language teaching". *Studies on Immersion Education: A Collection for US Educators.* Sacramento: California State Dept. of Education.

Carr, T.H. and Posner, M.I. 1992. "The impact of learning to read on the functional anatomy of language processing". *TR 92-1*, Institute of Cognitive and Decision Sciences, University of Oregon.

Carter, A. 1974. *Communication in the Sensory-Motor Period.* Ph.D. dissertation, University of California, Berkeley.

Cheney, D.L. and Seyfarth, R.M. 1990. *How Monkeys See the World.* Chicago: University of Chicago Press.

Chertkow, H., Bub, D., Evans, A., Meyer, E., and Marrett, S. 1990. "Processing words and pictures in the brain studied with positron emission tomography". *Canadian Journal of Neuroscience.*

Cohen, A.D. 1976. "The acquisition of Spanish grammar through immersion: Some findings after four years". *Canadian Modern Language Review* 23.

de Waal, F. 1982. *Chimpanzee Politics: Power and Sex among the Apes*, London: Unwin/Counterpoint.

Dunbar, R.I.M. 1992. "Co-evolution of neuro-cortex size, group size and language in humans". *Brain and Behavior Sciences.*

Ericsson, A. and Kintsch, W. 1995. "Long term working memory". *Psych. Review*, 102.

Ervin-Tripp, S. 1970. "Discourse agreement: How children answer questions". In *Cognition and the Development of Language*, J. Hays (ed). New York: Wiley & Son.

Fiez, J.A. and Petersen, S.E. 1993. "PET as part of an interdisciplinary approach to understanding processes involved in reading". *Psychological Science* 4.

Fouts, R.S. 1973. "Acquisition and testing of gestural signs in four young chimpanzees". *Science* 180.

Funnell, E. 1987. "Object concepts and object names: Some deductions from acquired disorders of word processing". In G.W. Humphreys and M.J. Riddoch (1987).

Funnell, E. and Alport, D.A. 1989. "Nonlinguistic cognition and word meanings". In *Language Perception and Production: Shared Mechanisms in Listening, Reading and Writing*, D.A. Alport, D.G. Mackay, W. Pririz, and E. Schearer (eds). London: Academic Press.

Futuyma, D.J. 1986. *Evolutionary Biology*. Sunderland, Mass.: Sinauer, 2nd edition.

Gardner, B.T. and R.A. Gardner. 1971. "Two-way communication with an infant chimpanzee". In *Behavior of Non-human Primates*, A. Schrier and F. Stollnitz (eds). New York: Academic Press.

Gathercole, S.E., and Baddeley, A.D. 1993. *Working Memory and Language.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Gernsbacher, M.A. 1990. *Language Comprehension as Structure Building.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Geschwind, N. 1970. "The organization of language and the brain". *Science* 170.

Geschwind, N. and Levitzky, W. 1968. "Human brain left-right asymmetries in the temporal speech region". *Science* 161.

Givón, T. 1971. "Historical syntax and synchronic morphology: An archaeologist's field trip". *CLS* 7. University of Chicago, Chicago Linguistics Society.

Givón, T. 1979. *On Understanding Grammar*. New York: Academic Press.

Givón, T. 1984a. "Universal of discourse structure and second language acquisition". In *Language Universals and Second Language Acquisition*, W. Rutherford (ed). Typological Studies in Language 5. Amsterdam: John Benjamins.

Givón, T. 1984b. *Syntax: A Functional-Typological Introduction*, Volume I. Amsterdam: John Benjamins.

Givón, T. 1989. *Mind, Code and Context: Essays in Pragmatics.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Givón, T. 1990b. *Syntax: A Functional-Typological Introduction*, Volume II. Amsterdam: John Benjamins.

Givón, T. 1990c. "Natural language learning and organized language teaching". In *Variability in Second Language Acquisition: Proceedings of the 10th Second Language Research Forum* (*SLRF*), H. Burmeister and P. Rounds (eds). Eugene: University of Oregon.

Givón, T. 1991a. "Isomorphism in the grammatical code: Cognitive and biological considerations". *Studies in Language* 15.

Givón, T. 1991b. “The grammar of referential coherence as mental processing instructions”. *Linguistics* 30.

Givón, T. 1991c. “The evolution of dependent clause syntax in Biblical Hebrew”. In E. Traugott and B. Heine (1991).

Givón, T. 1993. “Coherence in text, coherence in mind”. *Pragmatics and Cognition* 1.

Givón, T. 1995. *Functionalism and Grammar*. Amsterdam: John Benjamins.

Givón, T. 1997. “Toward a neuro-cognitive interpretation of ‘context’ ”. TR-97–08, *Institute of Cognitive and Decision Sciences*, University of Oregon.

Givón, T. 1998. “On the co-evolution of language, mind and brain”. *Evolution of Communication* 2.

Givón, T. 2000. “Internal reconstruction: As theory, as method”. In *Reconstructing Grammar: Comparative Linguistics and Grammaticalization*, S. Gildea (ed). Typological Studies in Language 41. Amsterdam: John Benjamins.

Givón, T. 2001. *Syntax: An introduction.* 2 vols. Amsterdam: John Benjamins.

Givón, T. and Young, P. 2002. “Cooperation and inter-personal manipulation in the society of intimates”. In *Causation and Inter-Personal Manipulation*, M. Shibatani (ed). Typological Studies in Language 48. Amsterdam: John Benjamins.

Givón, T., L. Yang and M.A. Gernsbacher. 1990 “The processing of second-language vocabulary: From attended to automated word-recognition”, *Proc. Second Language Research Forum* (SLRF), University of Oregon, Eugene.

Goldin-Meadows, S. This volume. “Getting a handle on language creation”.

Goodenough, W. 1990. “Evolution of human capacity for beliefs”. *American Anthropologist.*

Goodglass, H. 1976. “Agrammatism”. In *Studies in Neurolinguistics* 1, H. Whitaker and H.A. Whitaker (eds). New York: Academic Press.

Gould, J.L. 1991. “An evolutionary perspective on honey bee communication”. *Symposium on the Evolution of Communication: Cross-Species Comparison*, Institute of Cognitive and Decision Sciences, University of Oregon.

Gould, J.L. and Towne, W. 1987. “Evolution of the dance language”. *The American Naturalist* 130.

Gould, S.J. 1977. *Ontogeny and Phylogeny.* Cambridge, Mass.: Harvard University Press.

Greenfield, P.M. 1991. “Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior”. *Behavior and Brain Sciences* 14.

Greenfield, P.M. and Savage-Rumbaugh, S. 1991. “Imitation, grammatical development and the invention of proto-grammar by an ape”. In *Biobehavioral Foundations of Language Development*, N. Kresnegor, M. Studdert-Kennedy and R. Schiefelbuch (eds). Hillsdale, NJ: Lawrence Erlbaum Associates.

Haiman, J. 1985a. *Natural Syntax.* Cambridge: Cambridge University Press.

Haiman, J. (ed). 1985b. *Iconicity in Syntax.* Typological Studies in Language 6. Amsterdam: John Benjamins.

Haiman, J. 1992. “Syntactic change and iconicity”. To appear in *Encyclopedia of Language and Linguistics.*

Haiman, J. 1994. “Ritualization and the development of language”. In *Perspectives on Grammaticalization*, W. Pagliuca (ed). Amsterdam: John Benjamins.

Hanson, R.N. 1958. *Patterns of Discovery*. Cambridge: Cambridge University Press.

Hauser, M. 2000. *Wild Minds*. New York: Henry Holt.

Heine, B., Claudi, U., and Hünnemeyer, F. (eds). 1991. *Grammaticalization: A Conceptual Framework*. Chicago: University of Chicago Press.

Hewes, G.W. 1973a. "Primate communication and the gestural origin of language". *Current Anthropology*.

Hewes, G.W. 1973b. "An explicit formulation of the relationship between tool-using, tool-making and the emergence of language". *Visible Language* 7.

Hewes, G.W. 1974. "Language in early hominids". In R. Wescott (1974).

Hoiting, N. and Slobin, D. (2001) "Typological and modality constraints on borrowing: Examples from sign language of the Netherlands". In *Foreign Vocabulary in Sign Language: A Cross-Linguistic Investigation of Word Formation*, D. Brentari (ed). Mahwah, NJ: Lawrence Erlbaum Associates.

Hopper, P. and Traugott, E. 1994. *Grammaticalization*. Cambridge: Cambridge University Press.

Howard, D., Patterson, K., Wise, R., Brown, W.D., Friston, K., Weller, C., and Frakowiac, R. 1992. "The cortical localization of the lexicon: Positron emission tomography evidence". *Brain* 115.

Humphreys, G.W, and Riddoch, J.M. 1987. "Introduction: Cognitive psychology and visual object processing". In G.W. Humphreys and J.M. Riddoch (1987).

Humphreys, G.W. and Riddoch, J.M. (eds). 1987. *Visual Object Processing: A Cognitive Neuropsychological Approach*. London: Lawrence Erlbaum Associates.

Humphreys, G.W. and Riddoch, M.J. 1988. "On the case of multiple semantic systems: A reply to Shallice". *Cognitive Neuropsychology* 5.

Joanette, Y., Goulet, O., and Hannequin, D. 1990. *Right Hemisphere and Verbal Communication*. New York: Springer-Verlag.

Kaas, J.H. 1989. "Why does the brain have so many visual areas?" *Journal of Cognitive Neuroscience* 1.

Kean, M.-L. (ed). 1985. *Agrammatism*. New York: Academic Press.

Keenan, E.O. 1974. "Conversational competence in children". *Journal of Child Language* 1.

Keenan, E.O. 1975. "Making it last: The use of repetition in children's discourse". *BLS* 1. Berkeley: Berkeley Linguistics Society.

Kintsch, W. 1974. *The representation of Meaning in Memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kintsch, W. 1982. "Memory for text". In *Text Processing*, A. Flammer and W. Kintsch (eds). Amsterdam: North Holland.

Kintsch, W. 1992. "How readers construct situation models for stories: The role of syntactic cues and causal inference". In *Essays in Honor of William K. Estes*, A.F. Healy, S. Kosslyn, and R.M. Shiffrin (eds). Hillsdale, NJ: Lawrence Erlbaum Associates.

Klima, E. and Bellugi, U. 1979. *The Signs of Language*. Cambridge, Mass.: Harvard University Press.

Knudsen, E.I. 1985. "Experience alters the spatial tuning of auditory units in the optic tectum during a sensitive period in the barn owl". *Journal of Neuroscience* 1.

Knudsen, E. I. 1989. "Visual experience shapes auditory orienting behavior in developing owls". Paper read at the *Symposium on Orienting: Phylogenetic, Developmental and Cognitive Perspective*, University of Oregon, Eugene.

Knudsen, E. I. and Knudsen, P. F. 1990. "Sensitive and critical periods for visual calibration of sound localization by barn owls". *Journal of Neuroscience* 5.

Lakoff, G. and Johnson, M. 1980. *Metaphors We Live By*. Chicago University of Chicago Press.

Lamendella, J. 1976. "Relations between the ontogeny and phylogeny of language: A neo-recapitulationist view". In *The Origins and Evolution of Language and Speech*, S. R. Harnad, H. D. Stelkis and J. Lancaster (eds). New York: New York Academy of Science.

Lamendella, J. 1977. Neuro-Functional foundations of symbolic communication. Ms. San Jose State University.

Li, C. N. This volume. "Missing links, issues and hypotheses in the evolutionary origin of language".

Li, C. N. and Hombert, J.-M. 2000. "On the evolutionary origin of language". In *Mirror Neurons and the Evolution of Brain and Language*, M. Stamenov and V. Gallese (eds). Amsterdam: John Benjamins.

Lieberman, P. 1984. *The Biology and Evolution of Language*. Cambridge, Mass.: Harvard University Press.

Limber, J. 1973. The genesis of complex sentences". In *Cognitive Development and the Acquisition of Language*, T. Moore (ed). New York: Academic Press.

MacWhinney, B. 1999. "The emergence of language from embodiment". In *The Emergence of Language*, B. MacWhinney (ed). Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. This volume. "The gradual emergence of language".

Malle, B. F. This volume. "The relation between language and theory of mind in development and evolution".

Mandler, J. M., and Johnson, N. J. 1977. "Remembrance of things parsed: Story structure and recall". *Cognitive Psychology* 9.

Marler, P., Karakashian, S., and Gyger, M. 1991. "Do animals have the option of withholding signals when communication is inappropriate? The audience effect". In C. A. Ristau (1991).

Maunsell, J. H. R. and van Essen, D. C. 1983. "The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in macaque monkeys". *Journal of Neuroscience* 12.

Mayr, E. 1982. *The Growth of Biological Thought*. Cambridge, Mass.: Harvard University Press.

McCandliss, B., Posner, M. I. and Givón, T. 1999. "Brain plasticity in learning visual words". *Cognitive Psychology*.

McKeon, R. (ed). 1941. *The Basic Works of Aristotle*. New York: Random House [22nd edn].

Melzoff, A. N. 1990. "Toward a developmental cognitive science: The implication of cross-modal matching and imitation for the development of representation and memory in infancy". In *The Development and Neurological Bases of Higher Cognitive Functions*, S. Diamond (ed). New York: NY Academy of Sciences Press .

Menn, L. 1990. “Agrammatism in English: Two case studies”. In L. Menn and E. Obler (1990).

Menn, L. and Obler, E. (eds). 1990. *Agrammatic Aphasia* (3 vols). Amsterdam: John Benjamins.

Mishkin, M. 1978. “Memory in monkeys severely impaired by combined but not by separate removal of amygdala and hippocampus”. *Nature* 273.

Mishkin, M. 1982. “A memory system in the monkey”. *Philosophical Society of London* [*Biol.*], 298.

Mishkin, M., Malamut, B. and Bachevalier, J. 1984. “Memories andhabits: Two neural systems”. In *Neurobiology of Learning and Memory*, G. Lynch and J.L. McGaugh (eds). New York: Guilford Press.

Mishkin, M., Spiegler, B.J., Saunders, R.C., and Malamut, B.J. 1982. “An animal model of global amnesia”. In *Toward a Treatment of Alzheimer's Disease*, S. Corkin, K.L. Davis, H. Growdon, E.J. Usdin and R.J. Wurtman (eds). New York: Raven Press.

Mishkin, M. and Petri, H.L. 1984. “Memories and habits: Some implications for the analysis of learning and retention”. In *Neuropsychology of Memory*, N. Butters and L.R. Squire (eds). New York: Guilford Press.

Morford, J. This voilume. “Why does exposure to language matter?”.

Neeley, J.H. 1990. “Semantic priming effects in visual word recognition: A selective review of current findings and theories”. In *Basic Processes in Reading: Visual Word Recognition*, D. Besner and G. Humphreys (eds). Hillsdale, NJ: Lawrence Erlbaum Associates.

Neville, H.J. 1995. “Developmental specificity in Neurocognitive development in humans”. In *The Cognive Neurosciences*, M. Gazzaniga (ed). Cambridge, Mass.: MIT Press.

Olness, G.S. 1991. “Pragmatic-contextual difficulties as a correlate of deficits associated with right-hemisphere lesions”. Ms., University of Oregon, Eugene.

Oujman, G.A. 1991. “Cortical organization of language”. *Journal of Neuroscience* 11.

Pepperberg, I.M. 1991. “A communicative approach to animal cognition: A study of conceptual abilities of an African Grey Parrot”. In C.A. Ristau (1991).

Perrett, D.I., Harries, M.H., Bevan, R., Thomas, S., Benson, P.J., Mistlin, A.J., Chitty, A.J., Hietanen, J.K., and Ortega, J.E. 1989. “Framework of analysis for the neural representation of animate objects and actions”. *Journal of Experimental Biology* 146.

Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M.A., and Raichle, M.E. 1988. “Positron emission tomographic studies of the cortical anatomy of single word processing”. *Nature* 331.

Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., and Raichle, M.E. 1989. “Positron emission tomographic studies of the processing of single words”. *Journal of Cognitive Neuroscience* 1.

Petersen, S.E., Fox, P.T., Snyder, A.Z., and Raichle, M.E. 1990. “Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli”. *Science* 249.

Pettito, L.A. 1991. From gesture to symbol: The relation between form and meaning in the acquisition of personal pronouns. *Papers and Reports from Child Language Development* (ms).

Popper, K. 1934. *The Logic of Scientific Discovery*. New York: Harper and Row [1959].

Posner, M. I., and Carr, T. H. 1991. "Lexical access and the brain: Anatomical constraints on cognitive models of word recognition". *TR 91–5*, Institute of Cognitive and Decision Sciences, University of Oregon.

Posner, M. I. and Pavese, A. 1997. "Anatomy of word and sentence meanings". Paper presented at the *Colloquium on Neuroimaging of Human Brain Functions*. Irvine, Calif., May 1997.

Posner, M. I., Petersen, S. E., Fox, P. T., and Raichle, M. E. 1988. "Localization of cognitive operations in the human brain". *Science* 240.

Posner, M. I., Sandson, J., Dhawan, M., and Shulman, G. L. 1989. "Is word recognition automatic? A cognitive anatomical approach". *Journal of Cognitive Neuroscience* 1.

Power, M. 1991. *The Egalitarians: Human and Chimpanzee*. Cambridge: Cambridge University Press.

Premak, D. 1971. "Language in chimpanzee". *Science* 172.

Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, M. K., Pardo, J., Fox, P. T., and Petersen, S. E. 1993. "Practice-related changes in human brain functional anatomy during non-motor learning". *Cerebral Cortex*.

Riddoch, M. J. and Humphreys, G. W. (eds). 1987a. "Visual optic processing in a case of optic aphasia". *Cognitive Neuropsychology* 4.

Riddoch, M. J. and Humphreys, G. W. 1987b. "Picture naming". In G. W. Humphreys and J. M. Riddoch (1987a).

Riddoch, M. J., Humphreys, G. W., Coltheart, M., and Funnell, E. 1988. "Semantic systems of system? Neuropsychological evidence reexamined". *Cognitive Neuropsychology* 5.

Ristau, C. A. (ed). 1991. *Cognitive Ethology*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rizzolatti, G. and Arbib, M. A. 1998. "Language within our grasp". *Trends in Neuroscience* 21.

Rizzolatti, G., Fadiga, L., Fogassi, L., and Gallese, V. 1996a. "Premotor cortex and the recognition of motor actions". *Cognitive Brain Research* 3.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulescu, E., Perani, D., and Fazio, F. 1996b. "Localization of grasp representation in humans by PET: 1. Observation vs. execution". *Experimental Brain Research* 111.

Rizzolatti, G., Fogassi, L., and Gallese, V. 2000. "Cortical mechanisms subserving object grasping and action recognition: A new view on the cortical motor functions". In *The New Cognitive Neuroscience*, M. S. Gazzaniga (ed). 2nd edition. Cambridge, Mass.: MIT Press.

Rizzolatti, G. and Gentilucci, M. 1988. "Motor and visual-motor functions of the premotor cortex". In *Neurobiology of Neocortex*, P. Rakic an W. Singer (eds). Chichester: Wiley.

Rose, S. A. 1990. "Cross-human transfer in human infants". In *The Development and Neurological Bases of Higher Cognitive Functions*, S. Diamond (ed). New York: New York Academy of Sciences Press.

Savage-Rumbaugh, S. 1998. Language in nonhuman beings. Paper presented to the American Association for the Advancement of Science, Philadelphia, Febr. 1998.

Savage-Rumbaugh, S. and Lewin, R. 1994. *Kanzi: The Ape at the Brink of the Human Mind*. New York: Wiley and Sons.

Savage-Rumbaugh, S., Murphy, J., Sevcik, R.A., Brakke, K.E., Wiliams, S.L., and Rumbaugh, D.M. 1993. *Language Comprehension in Ape and Child*, *Monographs of the Society for Research in Child Development*, serial #233, vol. 58, nos 3–4.

Savage-Rumbaugh, S., Shanker, S.G., and Taylor, T.J. 1998. *Apes, Language and the Human Mind*. Oxford: Oxford University Press.

Scollon, R. 1976. *Conversations with a One-Year Old Child*. Honolulu: University of Hawaii Press.

Schnitzer, M. 1989. *The Pragmatic Basis of Aphasia*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schumann, J. 1976. "Second language acquisition: The pidginization hypothesis". *Language Learning* 26.

Schumann, J. 1978. *The Pidginization Process: A Model for Second Language Acquisition*. Rowley, Mass.: Newbury House.

Schumann, J. 1985. "Non-syntactic speech in Spanish–English basilang". In *Second Language Acquisition: A Cross-Linguistic Perspective*, R. Andersen (ed). Rowley, Mass.: Newbury House.

Selinker, L. 1972. "Interlanguage". *International Review of Applied Linguistics* 10.

Shallice, T. 1988a. "Specialization within the semantic system". *Cognitive Neuropsychology* 5.

Shallice, T. 1988b. *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.

Sierof, E. and Posner, M.I. 1988. "Cuing spatial attention during processing of words and letter strings in normals". *Cognitive Psychology* 5.

Simonds, P. 1974. *The Social Primates*. New York: Harper & Row.

Slobin, D. 1970. "Universals of grammatical development in children". In *Advances in Psycholinguistic Research*, W. Levelt and G.B. Flores d'Arcais (eds). Amsterdam: North Holland.

Slobin, D. 1997. "The origins of grammaticalizable notions: Beyond the individual mind". In *The Crosslinguistic Study of Language Acquisition*, D. Slobin (ed), Volume V. Mahwah, NJ: Lawrence Erlbaum Associates.

Slobin, D. This volume. "Language evolution, acquisition and diachrony: Probing the parallels".

Snyder, A.Z., Abdullaev, Y.G., Posner, M.I., and Raichle, M.E. 1995. "Scalp electrical potentials reflect regional cerebral blood flow responses during processing of written words". *Proceedings of the National Academy of Science, USA*, 92:1689–93.

Spelke, 1987. "The development of intermodal perception". In *Handbook of Human Perception: From Perception to Cognition*, P. Salapatek and L. Cohen (eds), Volume II. New York: Academic Press.

Squire, L.R. 1987. *Memory and Brain*. Oxford: Oxford University Press.

Squire, L.R. and Zola-Morgan, S. 1991. "The medial temporal lobe memory system". *Science* 253.

Stokoe, W.C. 1974. "Motor signs as the first form of language". In R. Wescott (1974).

Swinney, D.A. 1979. "Lexical access during sentence comprehension: (Re)consideration of context effects". *Journal of Verbal Learning and Verbal Behavior* 18.

Takahashi, T. 1989. Neuromechanisms of sound localization in the barn owl. Ms., *Cognitive Science Colloquium*, University of Oregon.

Takahashi, T. and Konishi, M. 1986. “Selectivity for interaural time differences in the owl’s midbrain”. *Journal of Neuroscience* 6.

Takahashi, T., Moiseff, A., and Konishi, M. 1984. “Time and intensity cues are processed independently in the auditory system of the owl”. *Journal of Neuroscience* 4.

Terrace, H.S. 1985. “In the beginning there was the ‘name’ ”. *American Psychologist* 40.

Tomasello, M. and Call, J. 1997. *Primate Cognition.* Oxford: Oxford University Press.

Tomasello, M. This volume. “How children produce utterances: The cut-and-paste model”.

Trabasso, T., Secco, T., and van den Broek, P. 1984. “Causal cohesion and story coherence”. In *Learning and Comprehension of Text*, H. Mandl, N.L. Stein and T. Trabasso (eds). Hillsdale, NJ: Lawrence Erlbaum Associates.

Traugott, E.C. and Heine, B. (eds). 1991. *Approaches to Grammaticalization* (2 vols), Amsterdam: John Benjamins.

Tucker, D.M. 1991. “Developing emotions and cortical networks”. In *Developmental Neuroscience*, *Minnesota Symposium on Child Psychology*, M. Gunnar and C. Nelson (eds), Volume 24. Hillsdale, NJ: Lawrence Erlbaum Associates.

Tucker, D.M. This volume. “Embodied meaning: An evolutionary-developmental analysis of adaptive semantics”.

Ungerleider, L.A. 1985. “The cortical pathways for object recognition and spatial perception”. In *Pattern Recognition Mechanisms*, *The Pontifical Academy of Sciences*, C. Chagas, R. Gattass and C. Gross (eds), 21–37.

Ungerleider, L.A. and Mishkin, M. 1982. “Two cortical visual systems”. In *Analysis of Visual Behavior*, D.G. Ingle, M.A. Goodale and R.J.Q. Mansfield (eds). Cambridge, Mass.: MIT Press.

von Frisch, K. 1967. *The Dance Language and Orientation of Bees.* Cambridge, Mass.: Harvard University Press.

Warrington, E.K. 1975. “The selective impairment of memory”. *Quarterly Journal of Psychology* 73.

Warrington, E.K. and McCarthy, R. 1983. “Category specific access disphasia”. *Brain* 166.

Warrington, E.K. and Shallice, T. 1979. “Semantic access dyslexia”. *Brain* 102.

Warrington, E.K. and Shallice, T. 1984. “Category specific semantic impairment”. *Brain* 107.

Wernicke, K. (1874) *Der aphasische Symptomencomplex*, Breslau. Trans. in *Boston Studies in Philosophy of Science* 4.

Wescott, R.W. (ed). 1974. *Language Origins.* Silver Springs, Md.: Linstok Press.

Wise, F. 1984. Word order and referential distance: A study of Spanish first-language acquisition. Ms., University of Oregon, Eugene.

Wise, R., Collet, F., Hadar, U., Friston, K., Hoffner, E., and Frackowiac, R. 1991. “Distribution of cortical neural networks involved in word comprehension and word retrieval”. *Brain* 114.

Wittgenstein, L. 1918. *Tractatus Logico Philosophicus.* Trans. by C.F. Pears and B.F. McGuinness. New York: The Humanities Press, 1968.

Zurif, E., Caramazza, A., and Myerson, R. 1972. “Grammatical judgements of agrammatic aphasics”. *Neuropsychologia* 10.

CHAPTER 2

Embodied meaning

An evolutionary–developmental analysis of adaptive semantics

Don M. Tucker
University of Oregon/Electrical Geodesics, Inc.

1. Overview

Language evolved to transcend evolution. With this structured communication, the complexity of behavior can now be achieved through cultural transmission, rather than through the slow selection for genetic mutations that bring change to instinctual behavior. Modern neuroscience research is providing new insights into the mechanisms of human neural plasticity that allow each child to self-organize, to respond to patterns of environmental information with adaptive changes in patterns of neural representation. The mammalian cortex, which has expanded considerably in the big apes and hominids, begins life with exuberant synaptogenesis, such that there is a massive connectivity across cortical regions. The specification of patterns in the cortex is then achieved through experience-dependent subtractive elimination (Innocenti 1983). Unused connections are lost. The differentiated brain then reflects the information content of adaptive exchange with the environment (Tucker 1992).

In humans, this sculpting of cortical anatomy by experience occurs not just for a few months or years as in most mammals, but over a decade or more. If there is a single biological fact that allowed language to evolve from prelinguistic cognition, it was probably not a unique gyrus of the brain, nor a particular segment of the vocal tract. Rather, it was the radical neoteny that allows culture rather than the genome to specify the connectional anatomy of the human brain. Neoteny, the retardation of ontogenesis to allow embryonic forms to extend into the juvenile period, is a common strategy in vertebrate evolution (Gould 1977). A simple mutation of a homeotic gene could shift the rate of maturation, allowing embryonic plasticity to continue to allow use-dependent specification of the morphogenetic process. If there was a pivotal development in prelinguistic hominid evolution, it may have been the co-evolution of neoteny with the parenting skills that allowed such immature juveniles to survive. The ones that survived maintained essentially embryonic mechanisms of neural plasticity over two decades of maturation. Given effective hominid family groups to allow this extended morphogenesis, given the inherent need for managing the behavior of incompetent juveniles, and given the emotional and attentional controls to allow the requisite communication of intentionality, language may have been the inevitable outcome.

Although this outcome has transcended the biological constraints of evolution in important ways, and probably ways we have yet to realize, it was nonetheless achieved within a framework of vertebrate neural evolution. In this framework, the neocortex is the last, specialized component added on to the the residuals of thousands of millennia of continuously adaptive, slowly evolving, nerve nets. Routine observations in clinical neurology show that the capacities of the human cortex depend on the successful operation of the entire evolved neural hierarchy, including memory control from limbic circuits, motor patterning from striatal-thalamic circuits, and arousal control from brainstem circuits. In this paper, I will outline the connectional architecture of the human brain that is emerging from modern anatomical studies, an architecture that is firmly rooted in the vertebrate plan. I will then propose that we can reason directly from the pattern of corticolimbic connectivity to interpret the structure of cognition. In this structure, both memory and motivation regulate the processes of corticolimbic consolidation to form the adaptive, embodied prelinguistic substrate that then allows the artifacts of language, words and their grammatical forms, to organize meaning.

1.1 Neural representation of language

Classical evidence from studies of aphasia suggested that language is represented primarily in the left hemisphere, with expressive capacities in the posterior inferior frontal lobe (Broca's area), and receptive capacities in the temporoparietal area (Wernicke's area) (Hecaen and Albert 1978; Lezak 1976). Both hemodynamic (PET; fMRI) and electrophysiological findings are largely consistent with this classical model, although right hemisphere contributions are often observed, such as in second languages or in sign language (Bavelier *et al.* 1998; Dehaene *et al.* 1997), or in interpreting emotional meaning of language (Borod 2000). The evolutionary origin of cerebral lateralization is an enduring problem that must hold answers for understanding the evolution of language (MacNeilage 1986).

Differentiation of expressive capacities to frontal, motor lobe, and receptive capacities to posterior, sensory, cortex has seemed unsurprising, given motor and perceptual requirements for language. However, this evidence does emphasize that the cognition of language is embodied (Johnson and Lakoff 1999). The representation of language is achieved not in some abstract cognitive space, but in the neural systems that evolved initial to support more elementary functions of perception and action (Givón 1998). This understanding of higher cognition as emergent from sensory and motor processes appears to be a consistent theme in recent PET and fMRI studies of cognition. Imagery, for example, is not a product of a disembodied mind. Sensory imagery is instantiated in visual and auditory networks, and motor imagery is instantiated in premotor and motor cortices (Kosslyn, DiGirolamo, Thompson, and Alpert 1998).

In addition to taking on the sensory and motor forms of speech, language must achieve meaning. Two components of meaning can be emphasized. The first is memory. Meaning must be abstracted from experience and represented in a way that can be contacted by words. A second component is emotional evaluation. Although the objectification of meaning may be a critical function of language, psychological studies have shown that, underlying the denotative, conventional level, verbal semantics rests on a foundation of affective evaluation (Osgood, Suci, and Tannenbaum 1957).

There is a common neural substrate for both memory and emotional evaluation: the limbic networks at the core of the cortex. Corticolimbic connections are essential for consolidating memory, and memory is in turn essential for language comprehension or expression. What seems to be overlooked in modern cognitive neuroscience research is that the control of memory is achieved by the same limbic circuits that are integral to emotional and motivational control. Since Broca first identified the "limbic lobe" forming the limbus or border of the medial core of the brain (Figure 1), it has been clear that the primitive limbic cortices are closely connected with subcortical structures. Since Papez's tracing of limbic circuits by introducing seizures (Papez 1937), it has been clear that these circuits are integral to hypothalamic regulation of emotional behavior. Modern neuropsychological researchers have rediscovered the limbic networks, but in relation to memory rather than emotion. It remains to be understood how cognitive representations are consolidated in memory in relation to motivational controls.

Memory control processes are essential to understanding both the evolution, and the current operation, of human language. When we look to the mechanisms of consolidation of memory in the cortex, we find that the limbic roots of these mechanisms evolved as visceromotor and viscerosensory control systems. Like other cognition, language is thus embodied within a context of adaptive, visceral, controls. To appreciate how language offers the capacity for objectivity, we might first appreciate its motivational basis.

The motivational controls on corticolimbic processes cannot be understood, in either an evolutionary or a developmental analysis, without a careful study of subcortical mechanisms. The human cortex represents an elaboration and extension of multiple levels of behavioral mechanisms that have appeared in vertebrate neural evolution. Considered in any detail, the prelinguistic roots of language run deep, to the fundamental self-regulatory strategies of vertebrate neural evolution that gave rise to the mammalian cortex.

Once the phylogenetic framework is appreciated, the classical biological analysis leads us to consider the ontogenetic mechanisms through which human language is self-organized within the extended plasticity of corticolimbic networks. The key issues from the phylogenetic analysis, the differential functions of the neocortical shell of the cortex and the limbic core, take on new forms in the ontogenetic analysis. Rather than the expected recapitulation of the phyletic order, with maturation of the limbic networks preceeding that of the neocortex, the ontogenesis of the human cortex inverts the developmental order, with early maturation of the neocortical shell of language (articulating the sensory and motor patterns of speech) and protracted neoteny and immaturity (and therefore plasticity) of the paralimbic core.

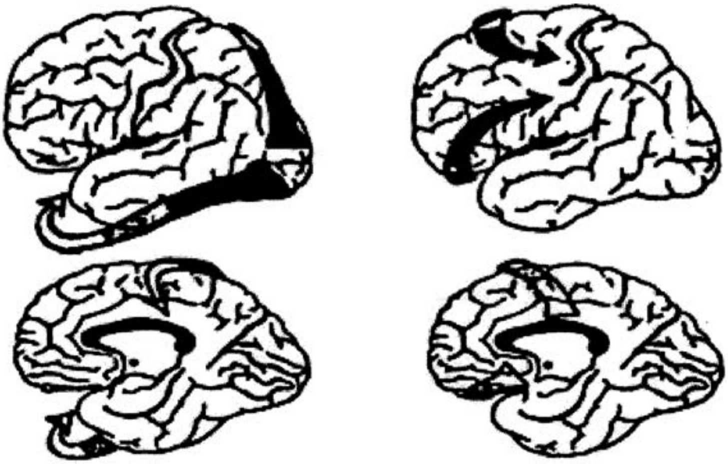


Figure 1. *Left*: Visual information appears to retrace the stages of phyletic differentiation in reverse, with the major visual projections from thalamic thalamic pathways targeting primary visual cortex. In the ventral trend, for example, V1 then projects to the adjacent visual association cortex, which projects to inferior temporal cortex, and on to temporal pole and the medial temporal lobe. *Right*: Motor control appears to recapitulate the phyletic stages directly. In the dorsal trend, for example, the initial motivational impetus and global organization of the action-in-context appears to begin in the anterior cingulate gyrus, which projects to the prefrontal cortex, to the premotor association area, and finally to primary motor cortex

2. Vertical integration in vertebrate evolution

Evolution of the brain has been a process of progressive reorganization through multiple extensions of primitive strategies of neural self-regulation. The early evolutionary principle was the “Law of Terminal Additions” in which more recent forms could be added only to the terminal stage of embryogeneis. Modern biologists have rejected this principle, because mutations can and do change the morphogenetic process at earlier stages (Gould 1977). However, this classical principle may continue to offer insight to an evolutionary-developmental analysis. Mutations of earlier stages of development are almost invariably fatal, so that the major variations for evolutionary selection are indeed the terminal forms.

This realization of the evolutionary history in the morphogenesis of each individual brain means that behavior, including language, must be achieved through vertical integration of all levels of the evolved neuraxis (Luu and Tucker 1998). The most complex functions of the frontal lobe (telencephalon), for example, require the recruitment of attentional regulatory influences from the thalamus, motivational influences from the hypothalamus (both in the diencephalon), and elementary arousal controls from brainstem reticular (mesencephalic) arousal systems.

2.1 Evolution of the telencephalon

Consistent with the Law of Terminal Additions, there is remarkable evolutionary conservativism of the lower brainstem, mesencephalic, and diencephalic structures in vertebrates, including humans (Sarnat and Netsky 1974). The most species variation in mammals is seen in the telencephalic structures, and even here, the species variations are subtle. In humans, the traditional theorizing on language has pointed to the differentiation and regional specialization of the massive cortex. A fundamental evolutionary question is then how the telencephalic structures, the striatum (or basal ganglia) and the limbic circuits (including amygdala and hippocampus) gave rise to the cortex.

Although most neuropsychological accounts of language emphasize the functions of the cortex, an important observation from clinical studies of aphasia is that the classical neocortical syndromes of expressive (Broca’s) and receptive (Wernicke’s) aphasia are limited to motor or sensory deficits when the lesions are limited to neocortex (Brown 1979). When semantic capacities are seriously impaired, the lesions typically encroach on limbic cortex (Brown 1988). It is therefore critical to understand not only the limbic base of the cortex, but the subcortical structures from which the limbic networks evolved.

If the evolutionary precursors of human language were the routinized and stereotyped communication of social animals, then these certainly pre-dated the evolution of the cortex. In avians, and probably in dinosaurs, neoteny set the stage for an extended plasticity of the developing brain. The requisite co-evolution was parenting behavior, coupled with the essential mechanisms of imprinting and attachment. These developments necessitated social communication. Thus, the most important preadaptations for human language — controlling neoteny, attachment, and communication — are almost certainly precortical. They can be seen to be well-developed in species with minimal cortical development, as well as in the mammals that went on to elaborate these themes into massive cortices and developmental products (cultures) that escaped the confines of the single evolutionary medium, the genome.

2.2 Reptilian limbic-striatal circuits give rise to mammalian cortex

We can approach the phylogenesis of the cortex with classical evolutionary-developmental reasoning. As he studied the neuroembryology of the early biologists' favorite neotenic (and thus generic root vertebrate) species, the salamander, Herrick (Herrick 1948) proposed that the functional circuits of the telencephalon could be seen by tracing their roots in the diencephalon. There are unique projections from the major divisions of the diencephalon (thalamus and hypothalamus) to cortex, and these provide an interesting perspective on the functional organization of the mammalian cortex (Tucker 2001). The thalamus appears specialized for the environment interface, with specific exteroceptive sensory nuclei (vision, audition, somatosensory) and specific motor nuclei. In contrast, the hypothalamus appears specialized for internal regulation, with representation both of interoceptive, viscerosensory processes and of visceromotor controls (Derryberry and Tucker 1990; Tucker, Derryberry, and Luu 2000).

In reptiles and avians, the specific sensorimotor circuits of the thalamus extend the representational function to a structure called the external striatum (Nauta and Karten 1970). In contrast, the hypothalamic controls appear to be elaborated preferentially within the limbic circuits. This general functional specialization within the telencephalon must have been critical to the key preadaptive communicative specialization of dinosaurs and avians, and it is retained in mammals. The striatal (basal ganglia) circuits form functional circuits with the thalamus to organize motor responses into routinized, hierarchic patterns (Alexander, Delong, and Strick 1986), achieved through the intrinsic plasticity of these circuits for modification through practice (Gabriel *et al.* 1996).

2.3 Telencephalic confluence at the neocortical shell

With the development of the mammalian cortex, the thalamic representational (somatosensory-somatomotor) circuits project to the neocortex, rather than external striatum, and the external striatum disappears (Nauta and Karten 1970). Where did this critical structure go? Without it, the basal ganglia of mammals are subordinated to the neocortex, and both sensory and motor projections from the specific thalamic nuclei are now taken up by the neocortex. Nauta and Karten (1970) speculated that the mammalian cortex absorbed the external striatum, and its embryological-evolutionary functional role. With this development, the specific somatosensory and somatomotor neocortices now become the external contact surface — the interface shell — for cortical evolution, communicating the brain to the environmental context.

2.4 Visceral base of the limbic core

As shown by Papez's studies (Papez 1937), the limbic circuits link the hippocampus and amygdala with hypothalamic circuits. Modern neuropsychological studies of amnesia have shown that the diencephalic extent of the limbic circuits is as important to memory consolidation as are the amygdala, hippocampus, and their adjacent (paralimbic) cortices (Zola-Morgan and Squire 1993). Modern anatomical tracer studies confirm that hypothalmic projections to cortex are confined to the limbic cortices at the core of the brain (Risold, Thompson, and Swanson 1997) Other crtical subcortical projections are to the arousal controls from brainstem (Mishkin 1990), and from the forebrain nucleus basalis (Mesulam, Mufson, Levey, and Wainer 1983). Through control over these arousal and regulatory systems, the limbic circuits appear to respond to events to provide an initial screening for novely and familiarity, thus coding the the current context as to whether it is appropriate for learning (Donchin and Coles 1988; Gabriel *et al.* 1996; Mesulam 1981). These limbic attentional controls thus appear to be closely linked to the hypothalamus's visceromotor control of the internal millieux.

2.5 Self-organization on the vertebrate plan

Why is this evolutionary structure of the brain important for understanding language? Because it is the plan within which human children must self-organize their intellectual capacity. Because of radical human neoteny, a long childhood extends the mammalian neuroembryonic process of massive synaptogenesis and then activity-dependent (i.e., experientially-sculpted) pruning (Tucker 1992). This plasticity may be fundamental to the human self-organization that creates the complex and flexible neural networks of language (Deacon 2000). To understand the essential self-regulatory mechanisms for activity-dependent pruning, we must look to the brain's systems for regulating arousal (Singer 1987). The mechanisms that regulate emotional arousal in the developing child are the same mechanisms that regulate the activity-dependent pruning of cortical networks in embryogenesis to achieve the coherent order of sensory and motor systems (Trevarthen 1984). Because of radical human neoteny, neural development may best be understood as self-organizing embryogenesis extended into the juvenile period.

In 1885 Pryer observed that the chick embryo shows spontaneous motor activity by the fifth day. Yet it will not show a response to sensation for many days after that (Hamburger 1977). The reason seems to be the vertebrate plan for acivity-dependent self-organizing neuroembryogenesis. In order for sensory systems to organize a coherent pattern of connectivity, they must be exercised through input. In the developing vertebrate embryo, the primary source of sensation comes from the spontaneous motor activity that provides the simultaneous input to the Hebbian synapses of sensory systems, thereby guiding the process of subtractive elimination and thus neuroanatomical specification (Hamburger, Balaban, Oppenheim, and Wenger 1965; Singer 1987; von de Malsburg and Singer 1988).

Activity-dependent neural plasticity is determined not only by coherent sensory input, but by the arousal systems that set the tone of neural activation (Singer 1993; Tucker 1992). Radical neoteny has produced human infants that are so immature as to be unable to self-regulate their neural arousal systems effectively. The arousal control processes therefore become entrained not only to brain growth and differentiation, but to the social communication through which parents help the infant stabilize immature and labile arousal controls (Trevarthen 1984). The internalized basis for self-control is therefore the child's experience as it is moderated by the assistance of the parental neural auxillaries. The communication is primarily nonverbal and emotional in the first year. This is the prelinguistic adaptive substrate of object relations upon which language is constructed in the second (Trevarthen 1984). The patterns of shared attention and social referencing formed by early emotional communication may form an essential basis for the child's learning of word meanings (Baldwin 1993).

As the child's brain responds to the communication process, the self-regulation of neural plasticity recruits adaptive mechanisms not just in the corticolimbic networks, but throughout the neuraxis (Tucker 1992). Language requires as much regulatory support from brainstem, diencephalic, and limbic circuits as it does from specialized neocortical modules (Brown 1979; Luu and Tucker 1998). To understand the vertical integration of language, across the evolved neural hierarchy, we can look to the neural mechanisms regulating primate vocalization (Luu and Tucker 1998).

3. Hierarchic anatomy of primate vocalization

Jurgens and Ploog have traced the control of vocalization in monkeys by combining lesion and electrical stimulation methods (Ploog 1981; Ploog 1992). The results illustrate the progressive elaboration of the vocalization process in each level of evolved circuitry, and, conversely, the dependence of higher systems on more elementary capacities of lower systems.

At the level of the lower brainstem (pons and medulla) are specific actions, such as control of respiration, that form the elements of vocalizations. At the midbrain level (periaqueductal gray and tegmentum) are representations of species-specific action patterns that organize the brainstem centers.

At the limbic level, Jurgens and Ploog observed that the limbic input allows vocalizations to be integrated with the animal's general emotional responses, rather than being stereotyped and isolated patterns. At the level of the cingulate gyrus (archicortex), the monkey is able to integrate the vocal pattern with general attention and goal-setting patterns, in what Jurgens and Ploog term "voluntary call initiation."

Finally, with the incorporation of neocortical (prefrontal) networks, the monkey appears able to carry out "voluntary call formation" in which vocalization is not limited to initiating the stereotyped vocalization patterns, but can be crafted into unique sequences from discrete vocal elements (Ploog 1992).

3.1 Volition and articulation

Consistent with the specialized neocortical networks for language articulation in humans shown by the aphasia evidence, Ploog (1992) proposes that human language evolved from increased specialization of the neocortical areas that allow monkeys the capacity for voluntary call formation, the articulation of specific vocal acts rather than holistic call patterns. In addition, however, Ploog points to the increased development of human limbic circuits, including projections through anterior thalamic nuclei to both dorsal (ACC, SMA) and ventral (orbital frontal lobe) limbic areas. These may be important to the complex motivational influences on human language. Although the Law of Terminal Additions emphasizes that viable mutations are most likely supported by the late stages of embryonic differentiation, i.e., the neocortical articulatory networks, human language is an integral component of a cognitive and behavioral system that must draw on effective operation of the entire neuraxis (Luu and Tucker 1998).

4. Corticolimbic architecture

Tracing the roots of higher cognition thus leads us to search out the primordial themes of regulation and representation that were reworked at each evolutionary level. In studying telencephalic circuitry in mammals, we can see that the functional divisions of the diencephalon seem to have given major direction to this circuitry, i.e., to the neostriatum and the limbic system. Similarly, the elementary diencephalic divisions, the interoceptive, visceral regulatory controls of the hypothalamus and the exteroceptive, somatic interface with the environment in the thalamus, seem to have continued to provide organizational influences during the continued evolution of the mammalian cortex. As we follow Herrick's (1948) advice and trace the diencephalic roots, we find them anchoring the internal and external poles of the architecture of the corticolimbic networks, the scaffolding upon which each memory, and thus each linguistic structure, must be constructed.

4.1 Cortical connectional anatomy

Modern studies of the cytoarchitectonics and the connectivity of the primate cortex have offered new insights into cortical evolution (Barbas and Pandya 1989; Pandya 1982; Pandya and Yeterian 1984). The cellular differentiation of the cortical layers is limited for the paralimbic cortices, with two or three layers and a poorly defined granular layer, consistent with the assumption that these are more primitive regions of the brain. In contrast, the greatest laminar differentiation, typically with six or seven distinct layers, is found for the primary sensory and motor cortices (Pandya 1982). Although it was assumed for many years that the primary sensory and motor areas were the first to evolve, and the more integrative "association areas" appeared in higher mammals, the view from the studies of cytoarchitectonics by Pandya and associates is that the paralimbic cortex was the first to evolve, followed by the heteromodal "association" cortex, followed by the unimodal association cortex, and finally by the primary sensory and motor areas (Mesulam 1985; Pandya and Yeterian 1985). Although the term "neocortex" was used for many years to distinguish the large expanses of laminar cortex in higher mammals from the primitive paralimbic cortices, it now appears that there were several major waves (or "growth rings") of cortical differentiation, with the first being the paralimbic cortices (archicortex and paleocortex), and the latest, and most laminar, neocortical regions being the primary sensory (visual, auditory, and somatosensory) and motor cortices.

The patterns of connectivity of these networks provide a complementary perspective to the levels of cytoarchitectonic differentiation. The following description, is taken from the studies of Pandya and associates (and is summarized in outline form in Tucker 1992). In each sensory modality, the general pattern is for most connections to be directed to the adjacent cytoarchitectonic region. Thus, in the ventral (object) visual pathway, the "forward" connections go from the primary visual area, to the visual association area, to the inferotemporal multimodal association cortex, to the temporal pole and paralimbic cortex (paleocortex). These network-to-network connections thus appear to reverse the phyletic order, given that the primary visual input from the thalamus targets the neocortical sensory area, and the processing (through memory consolidation) is directed toward the paralimbic cortex. Each area-to-area "forward" connection is reciprocated with "backprojections" which go from inferior cortical lamina of the relatively paralimbic network to the more superior cortical lamina of the relatively neocortical network.

The inter-regional connections provide another key piece of evidence to the global architecture. Networks of a given level of cytoarchitectonic differentiation tend to send projections to each other. The unimodal association areas, for example, send more projections to unimodal association areas of other modalities than to heteromodal association areas.

In addition, the general density of interregional connectivity decreases with neocortical differentiation. Paralimbic cortices are densely interconnected with each other, creating an integrated distributed network at the core of the cortex. The density of interconnectivity decreases with each cytoarchitectonic level, until the primary sensory and motor cortices are islands, isolated from the rest of the cortex except through connections to the immediately adjacent sensory or motor association areas.

If we interpret the connectivity directly, then the greatest integration of cortical function must occur in paralimbic regions, which are the most densely interconnected. It is of course relevant that these paralimbic regions interact closely with the hypothalamic interoceptive regulatory processes (Risold, Canteras, and Swanson 1994). In contrast, the primary sensory and motor neocortices are specialized modules, interfaced with the thalamic input and output channels. Although it may be tempting to focus on one network within this architecture, such as for lexical access, grammatical construction, or verbal comprehension, if it involves memory, then language in the human cortex must achieve some coherence across the multileveled corticolimbic architecture.

4.2 Dorsal and ventral roots of corticolimbic networks

The mammalian corticolimbic architecture appears to have evolved from two points of origin, the paleocortex at the ventral base of the brain and the archicortex at the dorsal surface of the brain (see figures in Mesulam 1985). The cytoarchitectonics vary fundamentally for these two cortical moieties. Whereas the paleocortical regions include a well-developed granular layer, the archicortical regions include a well-developed pyramidal cell layer. The functional differentiation of the paleocortical and archicortical networks can be understood at the limbic level from examining the subcortical connections, and at the neocortical level from examining the differing specializations for cognitive, sensory, and motor representations (Tucker *et al.* 2000).

From studying the connectivity of the limbic cortices not only with the hypothalamus but with autonomic centers of the brainstem, Neafsy and associates have suggested that there is a cortical representation of the visceral functions that parallels the cortical representation of motor activities and the sensory modalities. This representation includes cardiovascular, gastrointestinal, thermoregulatory, reproductive, adrenal and immune systems (Neafsey 1990; Neafsey, Terreberry, Hurley, Ruit, and Frysztak 1993). Importantly, there is a functional specialization of the archicortical and paleocortical regions for traffic with the internal milliuex. The archicortical networks, centered on the infralimbic layer of the cingulate gyrus on the medial wall of the hemisphere (anterior and subcallosal regions of the cingulate gyrus, CG), appear to control visceromotor processes. In contrast, the paleocortical networks, centered on the insular cortex between the temporal pole and orbital frontal cortex, appear to control viscerosensory processes (Neafsey 1990; Neafsey *et al.* 1993).

There is thus an input-output function of the cortical regulation of interoceptive processes that may be equally important to the organization of the cortex as the familiar input-output division of somatic networks (motor output in the anterior brain and sensory input in the posterior brain). The significance of the visceral regulatory functions is not widely appreciated, but the specialization of dorsal and ventral pathways has been recognized for both perceptual and motor functions, and to some extent the memory processes that integrate them.

4.3 Dorsal sensory, motor, and cognitive functions

In vision, the dorsal (archicortical) pathway (Figure 1, left; upper arrow) has been recognized to support spatial attention and memory (Haxby *et al.* 1991; Ungerleider and Mishkin 1982). Because this pathway connects the occipital and parietal neocortices to the cingulate and hippocampus, the studies of the dorsal pathway in monkeys and humans are consistent with the extensive evidence of hippocampal representation of spatial memory in rats (Nadel 1991). The anatomical evidence shows that the dorsal pathway extends the cingulate gyrus (archicortex), which evolved from hippocampal formation, to both posterior and anterior cortices.

Although there is no obvious connection between spatial cognition and the visceromotor function of the archicortex, the studies of contextual learning may provide a clue. Perhaps the dorsal pathway is representing not just the spatial context for memory, but context more generally. Gabriel and associates have observed that hippocampal input into posterior cingulate cortex appears to inhibit or gate the learning of new responses when the context is not appropriate (Gabriel, Sparenborg, and Stolar 1986).

The motor extension of archicortex is to anterior cingulate gyrus, medial frontal cortex, dorsolateral frontal cortex as well as supplementary motor area, and then primary motor cortex (Figure 1, right; upper arrow). Analysis of the unique contribution of this pathway in motor control has suggested that it contributes to a holistic, projectional influence on action, in which the action is launched in a vectoral direction, without discrete, ongoing feedback by sensory control (Goldberg 1985). This is an impulsive mode of control. Goldberg speculates that this dorsal motor pathway from anterior cingulate to supplementary motor area may be responsive to immediate affective outflow. This would be consistent with direction from the visceromotor functions of limbic archicortex. One interpretation may be that the projectional, impulsive control of action in the dorsal pathway is suited to the capacity of posterior archicortical networks to maintain an updated model of the context for action. The limbic contribution to action in the dorsal pathway may be projectional and impulsive because this is the operational mode of the visceromotor functions of the cingulate gyrus. An important question is whether this mode of control, seen by Goldberg in motor processes, may also be reflected in the dorsal limbic contribution to cognition.

4.4 Ventral sensory, motor, and cognitive functions

Similarly, both the sensory and motor operations of the ventral visual pathway (Figure 1, left; lower arrow) should provide clues to how these neocortical regions extend the regulatory controls of the ventral anchor of the limbic networks, the viscerosensory functions of the insula and paleocortex. In contrast to the contextual representation of the dorsal visual pathway, the ventral visual pathway appears to be specialized for object perception and memory (Haxby *et al.* 1991; Ungerleider and Mishkin 1982). How this cognitive specialization relates to viscerosensory controls is not immediately obvious. Perhaps there is a tighter constraint when viscerosensory signals direct memory consolidation than when visceromotor signals do.

In motor control, the translation of limbic regulation to neocortical representation seems easier to understand. The orbital and ventrolateral motor pathway from limbic to motor cortex (Figure 1, right; lower arrow) appears to guide the development of actions with a discrete, immediate guidance from ongoing sensory feedback (Goldberg 1985). This control may be well suited to direct regulation by viscerosensory signals. If so, then perhaps the tight motivational regulation of actions within the ventral trend (Tucker and Derryberry 1992) reflects the more general psychological operation of the viscerosensory functions that will be seen in more general patterns of self-control.

5. Constraints on the memory of language

It is thus possible to argue that a complete theoretical model of corticolimbic architecture must include not only the anterior and posterior divisions of the neocortex, defined in their input/output functions by their specific thalamic projections, but also by the dorsal and ventral divisions of the limbic cortex, defined in their input/output functions by their unique hypothalamic projections. The two major diencephalic strutures, thalamic and hypothalamic, seem to have served as the anchors for the evolution of the cortex (Herrick 1948). In the developing salamander, the thalamic projections are taken up by the lateral wall of the primtive hemisphere, and the hypothalamic projections are taken up by the medial wall. Given that the limbic cortex remains on the medial wall of the hemisphere in humans, and the neocortices are differentiated within the lateral convexities, Herrick's embryological-evolutionary observations on the salamander telencephalon remain quite relevant for the human brain.

5.1 Engines of neocortical evolution

Although the story of the evolution of the neocortex is controversial (Deacon 1990; Kaas 1989; Pandya and Yeterian 1984), and it is of course not easily resolved with observational evidence, we can draw on the studies of connectivity by Pandya and associates to apply a classical evolutionary-developmental analysis to guess at the process of evolutionary differentiation (Tucker 1992) . An important clue is the laminar projections from sensory neocortex to the adjacent, relatively more paralimbic, cortical network. This limbipetal pathway can be traced for the ventral visual pathway, for example, from the primary visual area,V1, to the parietal lobe and then cingulate gyrus in the dorsal pathway in Figure 1. In this direction, the projections are from layer III in the more neocortical network (e.g., V1) to layer IV of the more limbic network (e.g., VA) (Pandya and Yeterian 1985). Because the thalamic projections to sensory cortex terminate in layer IV, this organization appears as if the more neocortical network has interposed itself between the earlier, more limbic, network and the thalamic input (D.M. Tucker 1992). In this way, each new module of neocortex seems to have modeled the thalamic input in a more complex way before it was projected to layer IV of the existing sensory network.

If we consider the undifferentiated, primitive general cortex of reptiles, then neocortical differentiation appears to have occurred through four or five major waves of differentiation. In each wave, a network specialized for the sensory input of a given modality (e.g., vision), and took up the thalamic projections to the cortex. Of course, in mammals these projections were formerly the external striatal projections of reptiles, so it becomes necessary to theorize that the cortex is now encephalizing the striatal as well as thalamic circuits for sensory motor representations. After each wave, the center of that now specialized network appeared to re-differentiate further, at which point it took up the thalamic input, and the relayed this input to its predecessor, through projections from layer III to layer IV (the former recipient of thalamic input).

In this network hierarchy, neocortical differentiation is driven by progressive modeling and remodeling of the specific sensory and motor input and output representations of thalamic and striatal circuits, with sensory networks in the back of the brain and motor networks in the front. At the same time, all projections must not only reach the corebrain limbic networks, but they must be reciprocated by "back-projections" i.e., in the limbifugal direction, originating in the deeper layers (V and VI) of the more limbic network and terminating in the superficial layer (I) of the more neocortical network. In the ventral visual pathway, for example, this limbifgual progression is TP => IT => VA => V1. Memory cannot be instantiated unless bidirectional corticolimbic traffic links the evolved network hierarchy in a process of recursive, reentrant consolidation.

In this process, whereas we can look to the neocortical networks for the differentiation and articulation of the conceptual product, we must recognize that the limbic networks are the motivating engines. When a seizure begins anywhere in neocortex, it is the limbic cortices and circuits that become excited, and that lead to propagation of the seizure globally (Isaacson 1982; Papez 1937). When electrical stimulation is applied to the cortex of animals, there is a recruitment of discharge responses, such that smaller subsequent stimulations can lead to increasing discharge responses, and finally spontaneous seizures (Isaacson 1982). The recruitment, apparently mediated by long-term potentiation, tends to be focalized in limbic structures and adjacent paralimbic cortex (Racine and McIntyre 1986; Teyler 1986). These signs of electrophysiological excitement may be the best clues to the mechanisms of consolidation that synchronize the synaptic plasticity of the multiple adjacent maps in the corticolimbic hierarchy.

Given their connectivity with hypothalamic and brainstem regulatory systems, and given their excitability within the corticolimbic architecture, the limbic cortices must have provided integral formative constraints on the evolution of cortical differentiation, organizing each new wave of sensorimotor and cognitive representational capacity in relation to continuing, primordial adaptive needs.

5.2 Language bounded at the neocortical shell: sensorimotor articulation of the environmental interface

Although the evolution of the cortex is likely to remain a controversial topic, a theory of the evolution of language should be informed about the major dimensions of cortical connectivity, regardless of their evolutionary precedence. In addition to the essential representational capacities of the posterior neocortex implicated by Wernicke's aphasia, and those of the anterior neocortex implicated by in Broca's aphasia, it may be important to consider the adaptive controls on memory from the visceromotor and viscerosensory networks of the limbic system. With their hypothalamic connections, the limbic networks form the visceral core of the brain. With their specific thalamic connections, the neocortical networks form the interface shell with the external environment (Tucker 2001). At the surface shell, language is articulated within the speech apparatus that becomes capable of language-specific utterances. And it is perceived through specialized sensory modules that become capable of language-specific auditory discriminations. But anatomy teaches that these cognitive capacities are not isolated "modules" in some psycholinguistic space, but rather linked networks in a bounded corticolimbic hierarchy.

One bound of this hierarchy is formed by the differentiated sensory and motor cortices. Although it is often proposed that the capacity to articulate speech was important in human language evolution, it is less often emphasized that the perception of speech sounds was a causative factor (because perception skills are less obvious to a superficial analysis of language mechanisms than vocalization mechanics). Yet we can easily observe that an adult's difficulty in learning of a new language extend to perceptual as well as motor skills.

More generally, it is typically assumed that the sensory and motor mechanisms of speech, while important, are not immediately relevant to the mental capacities of language, which are of a different, psychological or linguistic form. Given the increasing evidence from cognitive neuroscience that mental operations are carried out within the sensory and motor cortices, and their allied association cortices, we have the opportunity to recognize that the organization of each linguistic construction is linked to its articulation in sensory and motor neocortices. We may come to recognize Hughlings Jackson's insight that the mind is "a great sensorimotor machine" (Jackson 1931).

Thus, Broca's area is motor association cortex of the ventrolateral, paleocortical corticolimbic pathway. The routinized access to the lexicon, and the automatized sequencing of the predictable forms of grammar (Givón 1989; Yang and Givón 1997) may take on their unique properties because of their place in the linked networks of the motor articulatory pathway. Indeed, as shown by the impairment of the articulatory process with disease of the basal ganglia, motor articulation is not just a cortical process, but may rely on the timing and sequencing capacities of subcortical circuits including the basal ganglia, brainstem, and cerebellum. The common evolutionary context of the mammalian brain may be important to explaining the flexible pattern generation apparatus that becomes so remarkable in human language. In reptiles, the striatum provided the ordered pattern generation that allowed more complex sensorimotor routines. In mammals, the evolution of the neocortex appears to have incorporated the external striatum (Deacon 1990; Nauta and Karten 1970) in what appears to be a radical encephalization of the striatal pattern generation capacity. With this encephalization, pattern generation was now extended by the extensive representation and memory capacities of the new cortical networks, allowing more abstract generation of the pattern systems of language.

If we consider the language centers of the anterior and posterior association cortex as connectionist networks, with functional capacities determined by their connections with other functional units, then, at the neocortical bound, articulation may be the essential function. Broca's area is just anterior to the motor areas for the face and mouth. It seems to have evolved its role in speech production, organizing the speaker's intended meaning into the acts of speech, because it is adjacent to and supportive of the speech articulatory networks. The process of differentiating the speech acts from the semantic intention is immediately precedent to the motor apparatus, and it takes its mechanistic form from the articulatory process itself. Language is embodied at the neocortical shell, as it is instantiated in sensorimotor form. Yet language articulation can be meaningful because it is also directed by networks with more global semantic capacities.

5.3 Language bounded at the limbic core: the visceral basis of meaning

The other bound of the corticolimbic hierarchy may be the visceral-limbic core networks. These exist at a level of "deep structure" in language that may be deeper than linguists typically consider, even for functional or pragmatic models. In a structuralist, e.g., Chompskian, analysis, the notion of deep structure refers to a generic frame of language that is more generalized than the "surface structure" of the specific utterance. Within the linked networks of the frontal lobe, this level might be more appropriately located to "middle structure" such as Broca's area. In a neural systems analysis, we can recognize that the deep structure must be in the paralimbic networks that provide semantics with an adaptive base, the constraints of motivational significance that ground concepts in personal meaning.

Psychological theorists have often proposed a level of cognition that is more holistic and primitive than the specified lexicon and grammar of language. Freud described primary-process cognition as fantasy, mental activity in which representational ability is immediately constrained by motivation (Freud 1953). Werner proposed that the cognition of the young child is syncretic, with feelings, perceptions, and actions represented in holistic, undifferentiated fashion in the "postural-affective matrix" (Werner 1957). From this primitive basis, linguistic representations are not entirely arbitrary, but retain their embeddedness in the holophrastic frame (Werner and Kaplan 1963).

If we interpret the anatomy directly, then a word achieves meaning through links across all levels of the corticolimbic hierarchy. A complex utterance takes form through the motor-linguistic sequencing apparatus which is continually constrained by the linked, nested semantic representations, from visceral to articulatory. Although we may not be able to apprehend the fact in experience, the neural architecture teaches that the meaning of language is multileveled, from the gut level that is inherently subjective to the surface articulation that is communicable within the articulatory conventions of the culture.

5.4 Conceptual arousal: The encephalization of the regulatory functions

As the corebrain limbic networks interface the massive cortices of the cerebral hemispheres with subcortical regulatory systems, they not only translate the demands of the cortical networks into commands for the hypothalamic and mesencephalic regulatory processes. They must also represent the arousal and regulatory functions to the cortex. In this process, both archicortical and paleocortical regions may have evolved to achieve an encephalization of regulatory control processes. The self-control mechanisms that were reflexive and automatic in simpler organisms may become represented in short-term memory with the extensive elaboration of the paralimbic as well as neocortical networks in recent hominid evolution.

One clue that human cognition has depended upon evolutionary changes in paralimbic networks is the elaboration of the spindle cell of the anterior cingulate gyrus (Nimchinsky *et al.* 1999). This is a type of neuron that is rare in the cingulate gyrus in monkeys, present in larger quantities in chimpanzees, present in more quantitiy and in clusters in Bonobos (pigmy chimpanzees), and present in dense clusters in humans. Thus, the 3-layered archicortex, long considered by neurologists to be a vestige of the ancestral brain, now appears to contain features that point to human uniqueness.

In recent times, the ubiquity of the cingulate gyrus in neurimaging results has brought it to center stage even before the discovery of its unique anatomy. Recognizing that the anterior cingulate becomes active in PET studies whenever cognitive tasks demand attentional control, Posner and Petersen proposed that the anterior cingulate gyrus forms the basis of the anterior attentional system (Posner and Petersen 1990). Observing anterior cingulate activity in fMRI studies when conflicting task demands must be resolved, Carter and associates emphasized the importance of this region of cortex to conflict monitoring (Carter *et al.* 1998). Documenting the linear increase in anterior cingulate activity in PET studies that demand increased effort, Paus and associates emphasized the importance of this region of limbic cortex in regulating arousal in response to cognitive demands (Paus 2000).

Through linking attentional control, and very likely memory consolidation, to arousal and bodily regulation, the cingulate gyrus may provide a bridge to bring the visceromotor function as a regulatory force into the cognitive domain. The importance of arousal regulation to higher cognitive processes has been apparent to theorists for many years (Luria and Homskaya 1970; Pribram and MacLean 1953; Tucker and Williamson 1984). Damasio (1999) has recently proposed that the neural mechanisms controlling bodily arousal are intimate components of conscious experience. Psychological theorists such as Thayer (Thayer 1970, 1989) have emphasized the importance of the experience of one's own arousal mechanisms. For example, for the "energy-vigor" dimension of arousal found in Thayer's factor-analytic studies, Thayer proposes that the experience of this form of arousal gives the person the feeling of being able to succeed with challenging tasks. In this way, the cingulate representation of the visceromotor function is not a reflex operating on homeostatic circuits, but a conceptual process, providing the networks of the cortex with information on the state of visceromotor operations. This information is immediately available to conscious experience, consistent with Damasio's (1999) emphasis on the importance of arousal control processes to the sense of self that is implicit within human awareness.

In most human experience, the subcortical arousal and visceral control functions are probably implicit, unconscious. But with the increasing evolution of the hominid cortex, the elaboration of necortical networks may have been accompanied by corresponding elaboration of the corebrain visceroregulatory networks. The increasing responsibility of limbic networks was not simply to translate sensorimotor traffic into demands for hypothalamic and reticular activating adjustments, but to bring visceroregulatory operations to bear on the process of consolidating and integrating the cognitive operations of the cortex itself. There was thus an increasing need for memory representation of the regulatory processes, such that an arousal change, or a visceral adjustment, has the quality not just of an immediate adaptive reflex, but of a concept. This form of bodily concept may be integral to each episode of consciousness, as in "the feeling of what happens" (Damasio 1999). The experience of energy-vigor may be an essential signal in personal decision-making, just as the experience of emotional depression may be a signal to decreasing coping efforts and recognizing the likelihood of failure in the current life context (Tucker 2001). The encephalization of the visceroregulatory functions may lead to a form of "adaptive semantics" in which meaning is evaluated, and experienced, through the implications that resonate at the gut level.

5.5 Visceromotor and viscerosensory contributions to self-control

The archicortical visceromotor and paleocortical viscerosensory networks may contribute to adaptive semantics in different ways. Keeping in mind the projectional, impulsive versus sensory feedback guidance of motor control formulated by Goldberg (1985) for the dorsal and ventral motor pathways, respectively, it may be possible to look to clinical observations for impairment of motivated cognition with lesions of the limbic base of these pathways.

With mediodorsal (anterior cingulate and dorsal pathway) lesions, patients often show the "pseudodepression" syndrome, in which there is loss of behavioral initiative even in the presence of intact motor, and at least superficially cognitive, function (Blumer and Benson 1975). These patients may appear to have language deficits, because they fail to speak. More severe mediodorsal frontal lesions lead to "akinetic mutism." When confronted, the patients are able to speak adequately, but left alone they remain mute. Interestingly, patients with lesions of the cingulate gyrus and associated meidal frontal cortex show a lack of concern about their problems, or even about pain they may eperience (Blumer and Benson 1975; Tow and Whitty 1953). Impressed by these observations, psychiatrists and neurosurgeons went on to lesion the cingulate gyrus in psychiatric and pain patients; the consistent effect was a lack of concern over previously distressing situations (Flor-Henry 1977; Laitinen 1972; Valenstein 1990).

Thus the visceromotor functions of the dorsal limbic networks may contribute to cognition and experience with the same feedforward, projectional, impulsive mode of control that they manifest in motor control. We can speculate that this is a mode of control that is also important to language, providing initiative, direction, and effort to communicate a motivational intention or significant experience.

With lesions of ventral limbic areas, including anterior temporal, insular, and particularly the orbital frontal cortices, patients are more likely to reflect a "disinhibition syndrome." Impulses are not appropriately constrained, and the result is manifested in the clinical interview by crude jokes and sexual advances (Blumer and Benson 1975; Tow and Whitty 1953). These patients appear to be impaired in the self-control of hedonic action in the same way that anxious and obsessive patients appear to be overly constricted (Tucker and Derryberry 1992). This unique component of self-control may reflect the adaptive function of the ventral pathway, in which the sensory feedback guidance of action (Goldberg 1985) has become elaborated from the viscerosensory function of the ventral limbic networks.

The lesions of the ventral pathway that produce personality deficits are typically more medial, inferior and caudal in the orbital frontal lobe (i.e., more paralimbic) than the lesions of Broca's area that cause expressive aphasia. Yet if there is an ordered hierarchy of function from limbic to neocortical networks, then it may be useful to consider the self-regulatory characteristics of the ventral limbic networks as they relate to the lexical and grammatical operations of the left frontotemporal cortices. The ventral cortex has evolved to organize perception into objects and features in the posterior brain, and to organize actions under sensory guidance in the anterior brain. Do these cognitive skills suggest memory consolidation that evolved from the viscerosensory function?

5.6 Core-shell dialectics

It may be possible to study language at its functional roots as well as its structural realization. Human speech may be formed by general adaptive semantics at the limbic core of the brain as it is then articulated within specific sensory and motor routines at the neocortical shell. This argument can be made on anatomical grounds, interpreting function through the density of connections. It can also be made on physiological grounds, interpreting the excitability of limbic-cortical connections as reflective of the process of memory consolidation. Putting the mechanism of consolidation together with the anatomy, we arrive at a functional interpretation such as that proposed by Pandya and associates (Pandya, Seltzer, and Barbas 1988; Pandya and Yeterian 1984). In the sensory pathways, cognition cannot be understood simply by one-directional processing, in which sensory data is processed through stages progressively from neocortex toward limbic cortex. Rather, consolidation must be balaned by traffic in the opposite direction, the so-called "back-projections" from limbic toward neocortical networks.

To the extent that we can interpret the limbifugal (from limbic) traffic as determined by memory, and the limbipetal (to limbic) traffic as determined by sensation, then the bidirectional nature of consolidation has been anticipated by at least one cognitive theorist (Shepard 1984). Shepard emphasized that perception must occur within the framework of memory, the framework of experience with the sensory information. Each act of perception is one of expectancy on the basis of both history and motivation. According to Shepard, perception could be described as hallucination constrained by the sensory data (Shepard 1984).

Evidence that we can interpret the relative contributions of limbic and neocortical networks in Shepard's terms has come from recent single unit studies in monkey temporal lobe (Naya, Yoshida, and Miyashita 2001). The monkey's task contrasted a perceptual signal with a signal to recall a perceptual event. Recordings were made simultaneously in paralimbic (perirhinal) cortex and neocortex (area TE of the lateral temporal lobe), and the timing of recorded activity was examined. When the perceptual signal was given, the neocortical area responded first, consistent with the direction of travel in the "forward" (limbipetal) direction. However, when the memory retrieval signal was given, the paralimbic area responded first, implying a "back-projection" or limbifugal progression. Interestingly, after the initial activity in the paralimbic network, the lateral temporal neocortex did respond, as if the perceptual network were gradually recruited in the process of reinstantiating the remembered event (Naya *et al.* 2001).

The integral role of adaptive control in paralimbic representations has been seen in studies of responses to food in orbital frontal cortex (Rolls, Critchley, Browning, Hernadi, and Lenard 1999). Neurons in sensory neocortex respond in an invariant fashion to sensory qualities such as smell, taste, or the feel of fat in the mouth (Rolls *et al.* 1999). In contrast, neurons in paralimbic areas of orbital cortex may respond to the sensory stimulus only when the animal is hungry. At the limbic base of the linked networks of the cortex, the perceptual representation appears to be adaptive, dynamic, and syncretic, a confluence of the exteroceptive data from the shell and the interoceptive representation of the encephalized corebrain regulatory function.

5.7 Objectivity and the emotional disorders

In order for the conceptual process to be coherent, the neocortical articulation of representational form may need to be congruent with the paralimbic conceptualization of the internal millieu and motivational context. However, it may also be that the specification of cognition in neocortical networks provides objective, structured control over the inherently labile and subjective limbic processes (Tucker 1981; Tucker and Liotti 1989). In PET studies of normal subjects reflecting on sad experiences, blood flow increases were observed in subgenual (paralimbic) anterior cingulate cortex, together with decreases in right dorsolateral neocortex (Mayberg *et al.* 1999). The opposite pattern was observed in depressed patients who had recovered from depression: blood flow increased in right dorsolateral areas and decreased in the ventral cingulate (Mayberg *et al.* 1999). Mayberg and associates proposed that the reciprocal relations between frontal neocortical and limbic cortices may be important to cognitive and affective imbalance in the emotional disorders.

Similar questions of corticolimbic balance have come up in other neuroimaging studies of self-control in psychopathology. When anxious subjects were asked to think of what made them anxious while in the cerebral blood flow scanner, flow increases were observed in the left inferior frontal area (Johanson, Risberg, Silfverskiold, and Smith 1986). These findings appear consistent with the tendency of anxious subjects to approach situations with the analytic, focused cognition of the left hemisphere (Tyler and Tucker 1982). Given the proximity of the flow increases observed by Johanson et al to Brocca's area, additional experiments were conducted to contrast the anxiety ruminations with non-affective verbal thought. Increased blood flow was still observed in the left orbital frontal region during the anxiety condition. Importantly, the anxiety-related flow increases were seen in the posterior, more paralimbic regions of the left orbital frontal lobe (Johanson, Smith, Risberg, Silfverskiold, and Tucker 1992). These studies provide an interesting perspective on the adjacent networks of the left inferior frontal lobe. Brocca's area is linked at the neocortical level to the motor control areas of the face and mouth, and at the limbic level to the affective, motivational functions of the posterior orbital cortex, which are in turn closely linked to the viscerosensory functions of the insular cortex and ventral limbic system (Tucker and Derryberry 1992).

As suggested by the importance of the right dorsolateral area in the findings of Mayberg *et al.* (1999), the frontal control of limbic responsivity may be important in the right hemisphere as well as the left. When Johansson and associates asked spider phobics to view videos of active spiders while in the blood flow scanner, the patients who reported being in control of their phobic response showed blood flow increases in the right dorsolateral frontal cortex (Johansson *et al.* 1998). Those patients who experienced panic showed decreases in the right dorsolateral area. Although the Xenon-inhalation measures of regional cerebral blood flow are superficial, and thus did not allow measures of limbic cortex activity in this research, the direct relation between frontal neocortical activity and the cognitive self-control of emotional response seems consistent with the interpretations of Mayberg et al (1999) that neocortical regions of the frontal lobe are able to provide inhibitory control over the affective responses of limbic regions. Although the asymmetries are clearly important (Liotti and Tucker 1994; Tucker 1981), a similar process of corticolimbic balance appears to occur for the nonverbal concepts in the right hemisphere and the verbal concepts of the left hemisphere.

On both sides of the brain, there may be an ongoing, microdevelopmental process which tends toward the differentiation and specification of discrete conceptual forms, such as occurs in the articulation of ideas in language. This process of articulation appears to emerge from a more primitive, holistic level of representation, in which motivational and emotional influences are fused with the semantic content. Judging from the evidence on emotional states and disorders, the allocation of cognition from one bounded constraint to the other is an important determinant of whether cognition is objective and rational or subjective and emotional. The dialectics of corticolimbic differentiation may be important not only to differing psychological states, but to enduring individual differences in cognitive and personality style (Shapiro 1965).

6. Retarded Embryogenesis

Classical evolutionary-developmental analysis looks to the ontogenetic recapitulation of the phyletic history to understand the relations among the evolved parts. Reasoning in this tradition, Piaget proposed in his *Genetic Epistemology* that the best way to understand the evolution of human knowledge may be to examine the emergence of intelligence in children. Although this approach continues to be informative, the notion of recapitulation cannot be applied in a simple way to explain mammalian cortical maturation. Primate and particularly human neoteny has involved early maturation of the recently evolved neocortices, while maintaining extended immaturity of the paralimbic regions.

6.1 Maturational gradients

In the human brain, the maturational process leads to myelination and functional fixity of the neocortical networks (primary sensory and motor cortices) early in development (Yakovlev 1962). This early maturation led early researchers in the 20th Century to believe that the sensory and motor cortices were evolutionarily primitive, wereas the "association areas" because they were late to mature, must have been more recently evolved. However, as described above, the study of both cytoarchitectonics and interregional connectivity suggests otherwise (Pandya 1987).

The early myelination of primary sensory and motor cortices is consistent with a loss of neural plasticity in these networks, compared with association and limbic cortices. Barbas (Barbas 1995), in fact, has observed a number of parallels between the plasticity of the infant brain and the characteristics of paralimbic cortex that are retained in the adult brain.

The effects of this maturational gradient may be seen behaviorally, in the difficulty of learning to speak a new language without an accent. Similarly, adults, but not young children, have difficulty discriminating the unfamiliar sounds of a new language. In contrast, the retention of juvenile plasticity in limbic cortices may be integral to the flexibility of adult human cognition, allowing adults to learn the meaning of the words in a new language, even though, to native speakers, they remain only marginally competent with the sensorimotor articulation of those words.

It is certainly possible that these effects are evolutionary accidents, the functional detritus of arbitrary selection. It is also possible, however, that the early myelination of the neocortical, articulatory shell may be essential to allow language to emerge. Fixation of the differentiated building blocks, the phonemes and graphemes, may be necessary to allow the complexity of the quasi-arbitrary representational forms to be assembled and reordered in ways that suit the organized patterns of meaning in linguistic intelligence.

Similarly, the remarkable feat of human language is not that we can make distinct speech sounds, nor even that we can order these in the motor sequence habits of grammar, but that we have anything interesting to say. The paralimbic retention of embryonic plasticity may be essential for the capacity for novel memory operations, for innovative conceptual constructions, which are continually evaluated and motivated by the prelinguistic mechanisms of adaptive semantics.

The enduring plasticity of the paralimbic core may thus be essential to flexible human cognition, but it is not without consequences. Furthermore, behavioral and experiential flexibility may become increasingly difficult for the older brain simply as a result of paralimbic plasticity (Tucker and Desmond 1998). Fully-interconnected networks, such as found at the limbic core, are subject to the stability-plasticity dilemma (Grossberg 1984), a limitation of distributed representation in which the accomodation of new input (plasticity) causes a disruption of prevous representations (McClelland, McNaughton, and O'Reilly 1995). When translated to the adaptive semantics forming the core of personality, the result of the stability-plasticity dilemma may be that any new experience becomes increasingly difficult to incorporate within the old self (Tucker and Desmond 1998). Although the corebrain networks retain their embryonic plasticity, they are also functionally active in representing a lifetime of deep structure. By early myelination, the neocortical networks may avoid the stability-plasticity dilemma in favor of stability and resistance to disruption by new information.

Through different mechanisms, then, cognitive representation gains a kind of developmental entropy at both the core and shell bounds. When constrained either to the visceral sensorimotor core, or to the somatic sensorimotor shell, cognition must remain concrete. Reframed in multiple levels through reentrant consolidation, however, cognition, and experience, may gain the capacity for abstraction. Perhaps it is in this stratification of conceptual form, linked across the nested corticolimbic networks, that we have the opportunity, however fleeting, for abstract knowledge.

6.2 Evolutionary strata and the structure of experience

In his evolutionary-developmental study of cognition, Werner proposed that the differentiation of specific cognitive form out of holistic and syncretic precursors is characteristic not just of the child's development, but the development, or microgenesis, of each idea (Werner 1957). As he considered the microgenetic process within the context of neuroanatomy, Brown recognized that the progression must be, at least in broad outline, phylogenetic. The development of an action, and a concept, must begin in the arousal and motivational centers in the upper brainstem, must recruit some adaptive resonance within diencephalic and limbic circuits, to be unfolded across paralimbic (e.g., cingulate) association (e.g., supplementary motor) and finally premotor cortices, before being instantiated in the motor cortex (Brown 1987).

As he considered the organization of motility in the vertebrate brain, Yakovlev emphasized that the internal core reflects the "sphere of visceration" patterned on respiration and homeostasis. At the next level of the exteriorization of movement are the axial movements of the body, reflecting the gross coordination of posture and the bodily basis for action. Next are the appendicular movements, of the limbs, coordinated in balancing gravity with the axial adjustments of posture. The final articulation from the visceral core is achieved by the oro-digital articulations of the distal extremities, in which language may be the final exteriorization (Yakovlev 1948).

We can extend this classical evolutionary-developmental reasoning to the modern evidence of neural plasticity within corticolimbic anatomy and physiology. If human cognition is indeed a kind of neuroembryogenesis extended unnaturally, then the evolutionary substrate of prelinguistic cognition may be of more than historical interest. It may be possible to recognize the recruitment of each level of the evolved neuraxis in the process of differentiating each linguistic construction.

References

Alexander, G.E., Delong, M.R., and Strick, P.L. 1986. "Parallel organization of functionally segregated circuits linking basal ganglia and cortex". *Annual Review of Neuroscience* 9: 357–81.

Baldwin, D.A. 1993. "Infants' ability to consult the speaker for clues to word reference". *Journal of Child Language* 20: 395–418.

Barbas, H. 1995. "Anatomic basis of cognitive-emotional interactions in the primate prefrontal cortex". *Neuroscience and Biobehavioral Reviews* 19: 499–510.

Barbas, H., and Pandya, D.N. 1989. "Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey". *Journal of Comparative Neurology* 286: 353–75.

Bavelier, D., Corina, D., Jezzard, P., Clark, V., Karni, A., Lalwani, A., Rauschecker, J.P., Braun, A., Turner, R., and Neville, H.J. 1998. "Hemispheric specialization for English and ASL: left invariance-right variability". *Neuroreport* 9: 1537–42.

Blumer, D., and Benson, D.F. 1975. "Personality changes with frontal and temporal lobe lesions". In *Psychiatric Aspects of Neurologic Disease*, D.F. Benson and D. Blumer (eds), 151–70. New York: Gruen and Stratton.

Borod, J. 2000. *The Neuropsychology of Emotion.* New York: Oxford University Press.

Brown, J.W. 1979. "Language representation in the brain". In *Neurobiology of Social Communication in Primates*, I.H.D. Steklis and M.J. Raleigh (eds), 133–95. New York: Academic Press.

Brown, J.W. 1987. "The microstructure of action". In *The Frontal Lobes Revisited*, E. Perceman (ed), 251–72. New York: The IRBN Press.

Brown, J.W. 1988. *The Life of the Mind: Selected Papers.* Hillsdale: Lawrence Erlbaum Associates.

Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., and Cohen, J.D. 1998. "Anterior cingulate cortex, error detection, and the online monitoring of performance". *Science* 280, 747–9.

Deacon, T.W. 1990. "Rethinking mammalian brain evolution". *American Zoologist* 30, 629–705.

Deacon, T.W. 2000. "Evolutionary perspectives on language and brain plasticity". *Journal of Communication Disorders* 33: 273–90.

Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D., van de Moortele, P.F., Lehericy, S., and Le Bihan, D. 1997. "Anatomical variability in the cortical representation of first and second language". *Neuroreport* 8:3809–15.

Derryberry, D., and Tucker, D.M. 1990. "The adaptive base of the neural hierarchy: elementary motivational controls on network function". *Nebraska Symposium on Motivation* 38:289–342.

Donchin, E., and Coles, M.G.H. 1988. "Is the P300 component a manifestation of context updating?" *Behavioral and Brain Sciences* 11:357–74.

Flor-Henry, P. 1977. "Progress and problems in psychosurgery". In *Current Psychiatric Therapies*, J.H. Masserman (ed), 283–98. New York: Grune and Stratton.

Freud, S. 1953. *The Interpretation of Dreams*. London: Hogarth Press. (First German edition 1900).

Gabriel, M., Kang, E., Poremba, A., Kubota, Y., Allen, M.T., Miller, D.P., and Steinmetz, J.E. 1996. "Neural substrates of discriminative avoidance learning and classical eyeblink conditioning in rabbits: a double dissociation". *Behavioral Brain Research* 82:23–30.

Gabriel, M., Sparenborg, S.P., and Stolar, N. 1986. "An executive function of the hippocampus: Pathway selection for thalamic neuronal significance code". In *The Hippocampus,* Volume 4, R.L. Isaacson and K.H. Pribram (eds), 1–39. New York: Plenum.

Givón, T. 1989. *Mind, Code and Context: Essays in Pragmatics*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Givón, T. 1998. "On the co-evolution of language, mind, and brain". Institute of Cognitive and Decision Sciences, TR-98–07.

Goldberg, G. 1985. "Supplementary motor area structure and function: Review and hypotheses". *The Behavioral and Brain Sciences* 8:567–615.

Gould, S.J. 1977. *Ontogeny and Phylogeny*. Cambridge, Mass.: Harvard University Press.

Grossberg, S. 1984. "Some psychophysiological and pharmacological correlates of a developmental, cognitive and motivational theory". In *Brain and Information: Event-related potentials. Annals of the New York Academy of Science* 425, R. Karrer and J. Cohen and P. Tueting (eds).

Hamburger, V. 1977. "The developmental history of the motor neuron". *Neuroscience Research Progress Bulletin* 15:1–37.

Hamburger, V., Balaban, M., Oppenheim, R., and Wenger, E. 1965. "Periodic motility of normal and spinal chick embryos between 8 and 17 days of incubation". *Journal of Experimental Zoology* 159:1–13.

Haxby, J.V., Grady, C.L., Horowitz, B., Ungerleider, L.G., Mishkin, M., Carson, R.E., Herscovitch, P., Schapiro, M.B., and Rapoport, S.I. 1991. "Dissociation of object and spatial visual processing pathways in human extra striate cortex". *Proceedings of the National Academy of Sciences*, USA, 88:1621–5.

Hecaen, H., and Albert, M.L. 1978. *Human Neuropsychology*. New York: Wiley.

Herrick, C.J. 1948. *The Brain of the Tiger Salamander*. Chicago: University of Chicago Press.

Innocenti, G.M. 1983. "Exuberant callosal projections between the developing hemispheres". In *Advances in Neurotraumatology*, R. Villani, I. Papo, M. Giovanelli, S.M. Gaini, and G. Tomei (eds), 5–10. Amsterdam: Excerpta Medica.

Isaacson, R.L. 1982. *The Limbic System.* 2nd edition. New York: Plenum.

Jackson, J.H. 1931. "The evolution and dissolution of the nervous system". *Selected writings of John Hughlings Jackson*, Volume II, 45–75. London: Hodder and Stoughton.

Johanson, A., Risberg, J., Silfverskiold, P., and Smith, G. 1986. Regional changes in cerebral blood flow during increased anxiety in patients with anxiety neurosis. In *The Roots of Perception*, U. Hentschel, G. Smith, and J.G. Draguns (eds). North-Holland: Elsevier Science Publishers.

Johanson, A., Smith, G., Risberg, J., Silfverskiold, P., and Tucker, D. 1992. "Left orbital frontal activation in pathological anxiety". *Anxiety, Stress, and Coping* 5, 313–28.

Johansson, A., Gustafson, L., Passant, U., Risberg, J., Smith, G., Warkentin, S., and Tucker, D. 1998. "Affective arousal and brain function in spider phobia". *Psychiatry Research: Neuroimaging* 84, 101–11.

Johnson, M., and Lakoff, G. 1999. *Philosphy in the Flesh: The Embodied Mind and its Challenge for Western Thought.* New York: Basic Books.

Kaas, J.H. 1989. "Why does the brain have so many visual areas?" *Journal of Cognitive Neuroscience* 1: 121–35.

Kosslyn, S.M., DiGirolamo, G.J., Thompson, W.L., and Alpert, N.M. 1998. "Mental rotation of objects versus hands: neural mechanisms revealed by positron emission tomography". *Psychophysiology* 35: 151–61.

Laitinen, L.V. 1972. "Stereotactic lesions in the knee of the corpus callosum in the treatment of emotional disorders". *Lancet* 1: 472–5.

Lezak, M.D. 1976. *Neuropsychological Assessment.* New York: Oxford University Press.

Liotti, M., and Tucker, D.M. 1994. "Emotion in asymmetric corticolimbic networks". In *Human Brain Laterality*, R.J. Davidson and K. Hugdahl (eds), 389–424. New York: Oxford.

Luria, A.R., and Homskaya, E.D. 1970. "Frontal lobe and the regulation of arousal processes". In *Attention: Contemporary theory and research*, D. Mostofsky (ed). New York: Appleton, Century, Croft.

Luu, P., and Tucker, D.M. 1998. "Vertical integration of neurolinguistic mechanisms". In *Handbook of Neurolinguistics*, B. Stemmer and H.A. Whitaker (eds), 55–203. San Diego, Calif.: Academic Press.

MacNeilage, P. 1986. On the evolution of handedness. Paper presented at the symposium *The Dual Brain: Unified functioning and specialization of the cerebral hemispheres.* Stockholm.

Martin, J.H. 1985. "Development as a guide to the regional anatomy of the brain". In *Principles of Neural Science*, E.R. Kandel and J.H. Schwartz (eds). New York: Elsevier.

Mayberg, H.S., Liotti, M., Brannan, S.K., McGinnis, S., Mahurin, R.K., Jerabek, P.A., Silva, J.A., Tekell, J.L., Martin, C.C., Lancaster, J.L., and Fox, P.T. 1999. "Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness". *American Journal of Psychiatry* 156: 675–82.

McClelland, J.L., McNaughton, B.L., and O'Reilly, R.C. 1995. "Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory". *Psychological Review* 102: 419–57.

Mesulam, M. 1981. “A cortical network for directed attention and unilateral neglect”. *Annals of Neurology* 10: 309–25.

Mesulam, M. (1985). “Patterns in behavioral neuroanatomy: Association areas, the limbic system, and hemispheric specialization”. In *Principles of behavioral neurology*, M. Mesulam (ed), 1–70. Philadelphia: F.A. Davis.

Mesulam, M., Mufson, E.J., Levey, A.I., and Wainer, B.H. 1983. “Cholinergic innervation of cortex by the basal forebrain: cytochemistry and cortical connections of the septal area, diagonal band nuclei, nucleus basalis (substantia innominata), and hypothalamus in the rhesus monkey”. *Journal of Comparative Neurology* 214, 170–97.

Mishkin, M., and Phillips, R.R. 1990. “A corticolimbic memory path revealed through its disconnection”. In *Brain Circuits and Functions of the Mind: Essays in Honor of Roger W. Sperry*, C. Trevarthen (ed), 196–210. Cambridge: Cambridge University Press.

Nadel, L. 1991. “The hippocampus and space revisited”. *Hippocampus* 1, 221–9.

Nauta, W.J.H., and Karten, H.J. 1970. “A general profile of the vertebrate brain, with sidelights on the ancestry of cerebral cortex”. In *The Neurosciences*, G.C. Quarton, T. Melnechuck and G. Adelman (eds), 7–26. New York: Rockefeller University Press.

Naya, Y., Yoshida, M., and Miyashita, Y. 2001. “Backward spreading of memory-retrieval signal in the primate temporal cortex”. *Science* 291, 661–4.

Neafsey, E.J. 1990. “Prefrontal cortical control of the autonomic nervous system: Anatomical and physiological observations”. In *The Prefrontal Cortex: Its Structure, Function and Pathology*, H.B.M. Uylings, C.G. van Eden, J.P.C. de Bruin, M.A. Corner, and M.G.P. Feenstra (eds), 147–66. New York: Elsevier.

Neafsey, E.J., Terreberry, R.R., Hurley, K.M., Ruit, K.G., and Frysztak, R.J. 1993. “Anterior cingulate cortex in rodents: Connections, visceral control functions, and implications for emotion”. In *Neurobiology of the Cingulate Cortex and Limbic Thalamus*, B.A. Vogt and M. Gabriel (eds), 206–23. Boston: Birkhauser.

Nimchinsky, E.A., Gilissen, E., Allman, J.M., Perl, D.P., Erwin, J.M., and Hof, P.R. 1999. “A neuronal morphologic type unique to humans and great apes”. *Proceedings of the National Academy of Sciencesy*, USA, 96: 5268–273.

Osgood, C.E., Suci, G.J., and Tannenbaum, P.H. 1957. *The Measurement of Meaning*. Chicago: University of Illinois Press.

Pandya, D.N., and Barnes, C.L. 1987. “Architecture and connections of the frontal lobe”. In *The Frontal Lobes Revisited*, E. Perecman (ed), 41–72. New York: IRBN.

Pandya, D.N., and Seltzer, B. 1982. “Association areas of the cerebral cortex”. *Trends in Neural Science* 5, 386–90.

Pandya, D.N., Seltzer, B., and Barbas, H. 1988. “Input-output organization of the primate cerebral cortex”. *Comparative Primate Biology* 4, 39–80.

Pandya, D.N., and Yeterian, E.H. 1984. “Proposed neural circuitry for spatial memory in the primate brain”. *Neuropsychologia* 22: 109–22.

Pandya, D.N., and Yeterian, E.H. 1985. “Architecture and connections of cortical association areas”. In *Cerebral Cortex*, Volume 4: *Association and Auditory Cortices*, A. Peters and E.G. Jones (eds), 3–61. New York: Plenum Press.

Papez, J.W. 1937. “A proposed mechanism of emotion”. *Archives of Neurology and Psychiatry* 38, 725–43.

Paus, T. 2000. “Functional anatomy of arousal and attention systems in the human brain”. *Progress in Brain Research* 126, 65–77.

Ploog, D. 1981. “Neurobiology of primate audio-vocal behavior”. *Brain Research Reviews* 3:35–61.

Ploog, D.W. (1992). “Neuroethological perspectives on the human brain: From the expression of emotions to intentional signing and speech”. In *So Human a Brain: Knowledge and Values in the Neurosciences*, A. Harrington (ed), 3–13. Boston: Birkhauser.

Posner, M.I., and Petersen, S.E. 1990. “The attention system of the human brain”. *Annual Review of Neuroscience* 13:25–42.

Pribram, K.H., and MacLean, P.D. 1953. “Neuronographic analysis of medial and basal cerebral cortex”. II. Monkey. *Journal of Neurophysiology* 16:324–40.

Racine, R.J., and McIntyre, D. 1986. “Mechanisms of kindling: A current view”. In *The limbic system*, B.K. Doane and K.E. Livingston (eds), 109–21. New York: Raven Press.

Risold, P.Y., Canteras, N.S., and Swanson, L.W. 1994. “Organization of projections from the anterior hypothalamic nucleus: a Phaseolus vulgaris-leucoagglutinin study in the rat”. *Journal of Comparative Neurology* 348:1–40.

Risold, P.Y., Thompson, R.H., and Swanson, L.W. 1997. “The structural organization of connections between hypothalamus and cerebral cortex”. *Brain Research Re*views 24:197–254.

Rolls, E.T., Critchley, H.D., Browning, A.S., Hernadi, I., and Lenard, L. 1999. “Responses to the sensory properties of fat of neurons in the primate orbitofrontal cortex”. *Journal of Neuroscience* 19:1532–40.

Sarnat, H.B., and Netsky, M.G. 1974. *Evolution of the Nervous System*. New York: Oxford University Press.

Shapiro, D. 1965. *Neurotic Styles*. New York: Basic Books.

Shepard, R.N. 1984. “Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming”. *Psychological Review* 91: 417–47.

Singer, W. 1987. “Activity-dependent self-organization of synaptic connections as a substrate of learning”. In *The Neural and Molecular Basis of Learning*, J.P. Changeux and M. Konishi (eds), 301–36. New York: Wiley.

Singer, W. 1993. “Synchronization of cortical activity and its putative role in information processing and learning”. *Annual Review of Physiology* 55:349–74.

Teyler, T.J. 1986. “Memory: Electrophysiological analogs”. In *Learning and Memory: A Biological View*, J.L. Martinez, Jr., and R.P. Kesner (eds), 237–65. Orlando: Academic Press.

Thayer, R.E. 1970. “Activation states as assessed by verbal report and four psychophysiological variables”. *Psychophysiology* 7:86–94.

Thayer, R.E. 1989. *The Biopsychology of Mood and Arousal*. New York: Oxford University Press.

Tow, P.M., and Whitty, C.W. 1953. “Personality changes after operations on the cingulate gyrus in man”. *Journal of Neurology, Neurosurgery: and Psychiatry* 16, 186–93.

Trevarthen, C. 1984. "Biodynamic structures, cognitive correlates of motive sets and the development of motives in infants". In *Cognition and motor processes*, W. Prinz and A. F. Sanders (eds). Berlin: Springer-Verlag.

Tucker, D. M. 1981. "Lateral brain function, emotion, and conceptualization". *Psychological Bulletin* 89: 19–46.

Tucker, D. M. 1992. "Developing emotions and cortical networks". In *Minnesota Symposium on Child Development: Developmental Neuroscience*, M. Gunnar and C. Nelson (eds), 75–127. News York: Oxford University Press.

Tucker, D. M. 2001. "Motivated Anatomy: A core-and-shell model of corticolimbic architecture". In *Handbook of Neuropsychology*, Volume V: *Emotional behavior and its disorders*, G. Gainotti (ed). Amsterdam: Elsevier. 2nd edition.

Tucker, D. M., and Derryberry, D. 1992. "Motivated attention: Anxiety and the frontal executive functions". *Neuropsychiatry, Neuropsychology, and Behavioral Neurology* 5: 233–52.

Tucker, D. M., Derryberry, D., and Luu, P. 2000. "Anatomy and physiology of human emotion: Vertical integration of brainstem, limbic, and cortical systems". In *Handbook of the Neuropsychology of Emotion*, J. Borod (ed). New York: Oxford University Press.

Tucker, D. M., and Desmond, R. 1998. "Aging and the plasticity of the self". In *Annual Review of Gerontology and Geriatrics, Vol 17: Focus on emotion and adult development. Annual review of gerontology and geriatrics*, K. W. Shaie and M. P. Laughton (eds), 266–81. New York: Springer.

Tucker, D. M., and Liotti, M. 1989. "Neuropsychological mechanisms of anxiety and depression". In *Handbook of Neuropsychology*, Volume III, F. Boller and J. Grafman (eds), 443–75: Elsevier Science.

Tucker, D. M., and Williamson, P. A. 1984. "Asymmetric neural control systems in human self-regulation". *Psychological Review* 91: 185–215.

Tyler, S. K., and Tucker, D. M. 1982. "Anxiety and perceptual structure: individual differences in neuropsychological function". *Journal of Abnormal Psychology* 91: 210–20.

Ungerleider, L. G., and Mishkin, M. 1982. "Two cortical visual systems". In *The analysis of visual behavior*, D. J. Ingle, R. J. W. Mansfield and M. A. Goodale (eds), 549–86. Cambridge, Mass.: MIT Press.

Valenstein, E. S. 1990. "The prefrontal area and psychosurgery". *Progress in Brain Research* 85: 539–54.

von de Malsburg, C., and Singer, W. 1988. "Principles of cortical network organization". In *Neurobiology of Neocortex*, P. Rakic and W. Singer (eds), 69–99. New York: Wiley.

Werner, H. 1957. *The Comparative Psychology of Mental Development.* New York: Harper.

Werner, H., and Kaplan, B. 1963. *Symbol Formation: An Organismic–Developmental Approach to Language and the Expression of Thought.* New York: John Wiley and Sons.

Yakovlev, P. 1962. "Morphological criteria of growth and maturation of the nervous system in man". *Research Publications of the Association for Research in Nervous and Mental Diseases* 39: 3–46.

Yakovlev, P. I. 1948. "Motility, behavior and the brain". *The Journal of Nervous and Mental Disease* 107: 313–35.

Yang, L., and Givón, T. 1997. "Benefits and drawbacks of controlled laboratory studies of second language acquisition: The Keki SL learning project". *Studies in Second Language Acquisition* 19: 173–93.

Zola-Morgan, S., and Squire, L. R. 1993. "Neuroanatomy of memory". *Annual Review of Neuroscience* 16: 547–63.

Chapter 3

Missing links, issues and hypotheses in the evolutionary origin of language*

Charles N. Li
University of California, Santa Barbara

1. Missing links in hominid fossils

In 1995, T. White and his colleagues published a correction of their 1994 paper announcing the discovery of a set of hominid fossils at Aramis, Ethiopia. With the addition of a nearly complete skeleton, they proclaimed a new hominid genus Ardipithecus ramidus (White et al 1995). Subsequently in an interview with the National Geographic, White stated that Ardipithecus ramidus constituted "the link that's no longer missing." Ardipithecus ramidus rightfully generated great excitement. Its age approximated 4.4 million years, making it the oldest hominid at the time. Its magnus foramen is located considerably more forward then that of any living great apes, suggesting that Ardipithecus ramidus might be bipedal. Its teeth have flattened crowns and non-slicing canines characteristic of hominids rather than apes. There is little doubt that Ardipithecus ramidus constitutes a landmark discovery of hominid fossils.

At about the same time when White et al published their findings on Ardipithecus ramidus, Meave Leakey and her colleagues (1995) unveiled a new species, Australopithecus anamensis from northern Kenya. At 4.2 to 3.9 million years of age, A. anamensis remains the oldest known species of Australopithecus. The structure of its tibia bone suggests bipedalism. But its dentition shows a U-shaped outline of tooth rows which is characteristic of Miocene as well as modern apes. Thus, A. anamensis displays an unusual combination of ape and hominid features.

Two years after the unveiling of A. anamensis, White (1997) contributed another major discovery, the Australopithecus ghari. This 2.5 million-year-old hominid from Ethiopia displays a mixture of the physical characteristics of both gracile and robust Australopithecus. The gracile and robust lines of Australopithecus are a long-standing and well-accepted evolutionary bifurcation of the Australopithecines with the understanding that the genus Homo evolved from the gracile branch. A. ghari threw a wrench in this conventional wisdom in paleoanthropology. Furthermore the fossil record from the site of A. ghari showed that this hominid used stone tools and carried the tools from a site miles away to the location where they were put to use. Prior to this discovery, it was believed that Australopithecines did not have stone tools, and the Homo genus to which modern human belongs had a monopoly on lithic industry. Thus, with the discovery of A. ghari, we could no longer use the manufacturing of stone tools as one of the defining characteristics of the genus Homo.

On the heels of White's discovery came Kenyanthropus platypos, a major find by Meave Leakey (1998) from Lake Turkana in the rift valley of Kenya. At 3.5 million years of age, Kenyanthropus is only slightly older than Lucy and the First Family, the world famous Australopithecus afarensis. But Kenyanthropus has small molars and a large flat face, placing its cranial appearance in much closer resemblance to anatomically modern human than Australopithecus afarensis which is marked by a small projecting face and relatively large molars. This is the main reason why Kenyanthropus was accorded the status of a new genus among the hominid taxon. It differs significantly from all of the extant hominid genera: Ardipithecus, Australopithecus, Paranthropus and Homo.

Upheaval in our knowledge of hominin taxonomy continued after the discovery of Kenyanthropus. In 2000, B. Senut and M. Pickford announced the unearthing of the Millennium Man, Orrorin tugenensis, from Kenya's Tugen Hills, with a set of bones including three femurs and a small human-like molar which is square and thickly-enameled, at approximately 6 million years of age. The age of O. tugenensis is simply astonishing. It approaches the estimated time of origin of hominids, when our ancestors evolved away from the common progenitors of modern chimpanzees and bonobos. Pickford and Senut (2001) claim that Orrorin tugenensis is, similar to the Australopithecenes, a bipedal hominid who is equally at home in trees. It has thickly enameled and square-shaped small molars that resemble the molars of modern humans rather than those of Australopithecenes. The molars of Ardipithecus ramidus, at 4.4. million years old, the second oldest hominid fossil, however, are thinly enameled. If Senut and Pickford's claim of O. tugenensis as a hominid is correct, it poses many challenges to our conceptualization of the hominid family.

This past summer, Y. Haile-Selassie (2001) published a description of a new set of fossil bones from Ethiopia dated between 5.8 to 5.2 million years ago. He claims that these bones come from a bipedal hominid, which he considers a subspecies of Ardipithecus ramidus. It is named Ardipithecus ramidus kaddaba. Both the postcranial and dental features of the fossils suggest that the hominid is phylogenetically close to the common ancestor of chimps and hominids. However, Ardipithecus shares with all other hominids the derived, elevated crown shoulders on the upper canine, and the cross section of the lower canine teeth is diamond-shaped like that of Australopithecus, not V-shaped like that of apes.

A point with significant implication is that both Orrorin tugenensis and Ardipithecus ramidus kaddaba appear to have lived in forest/woodland rather than wooded grassland. If the bipedal hominid status of these two species are confirmed, the cherished hypothesis that bipedal hominids emerged because of climatic change due to the formation of the East African rift valleys could be in serious jeopardy.

The flurry of new fossil discoveries during the past few years projects an image of a tangled web of hominid evolution. Instead of filling in the missing link from hominoids to Australopithecenes, Ardipithecus has become the first of a parade of fossil discoveries that suggest numerous missing links in hominid evolution. The old hypothesis of a straight-line phylogeny is no longer viable. The phylogenetic tree for hominids seems to have many more branches than we thought it had less than ten years ago at the time of the discovery of Ardipithecus. We are not sure how the known branches are connected or not connected to each other, and we don't know how many and what other branches might exist. However, thanks to the contributions of paleoanthropologists, we do have an overall picture of the evolution of hominids:

We know hominids emerged through the gradual evolutionary development of bipedalism. We also know that by the time Homo appeared a little more than two million years ago, hominids became obligatorily bipedal in the sense that their anatomical features such as the relative length of their arms vis-a-vis the length of their legs, the orientation of the shoulder sockets, the shape of their toes and fingers, no longer allowed them to climb trees easily.[1]

As a consequence of the inter-related evolutionary phenomena of increasing encephalization, the lengthening of maturation and consequently a significantly prolonged premature infancy, members of the Homo genus have no choice but to adapt to life on the open savanna grassland. Stanley (1996) eloquently designated the consequence of these inter-related evolutionary processes the "terrestrial imperative", meaning that members of the Homo genus are obligatorily terrestrial at all times. The bipedal hominids that preceded the Homo could escape predators by nesting in the trees in the night and climbing trees during the day. Homo's lost the tree-climbing option of their gracile Australopithecene forebears because of a chain of reasons:

1. In order to evolve a big brain and still allow the new born to pass through the birth canal, Homo's must have lengthened maturation so that most of the large brain can develop after birth.
2. Lengthened maturation implies prolonged pre-mature infancy, i.e. the new born will remain physically uncoordinated for a prolonged period of time after birth because the brain needs time to mature after birth.
3. Since the newborns are uncoordinated and helpless, they have to be held by their parents on the move.

Tree-climbing is, therefore, not a possible choice for Homo's on the move whether they are escaping predators or seeking refuge in the night. Notice that modern ape infant can hold on to its mother within one day after birth. Hence apes can climb up trees with their infants holding on to them. Human infants are not physically capable of holding on to their mothers until age two or three. The survival strategy of our Homo ancestors on the great plains of East Africa, where formidable predators and nimble preys had honed their respective skills over many million years before the emergence of Homo's, is to form larger and better organized social groups with improved means of communication. Thus, our Homo ancestors embarked on the pathway of evolving a more and more efficient and adaptive type of communicative behavior. In other words, the terrestrial imperative imposed upon hominid evolution the co-evolutionary expansion of cognitive capacity and the development of efficient and effective communicative behavior, which ultimately led to the emergence of language among anatomically modern humans.[2] The enhancement of hominid cognitive capacity and the evolution of their communicative behavior however, were not the only co-evolutionary development. It was accompanied by an array of concomitant anatomical changes summarized in Li and Hombert (2002).

As for the emergence of language, the fossil records yield little or no information. Fossil bones revealed three evolutionary developments related to the emergence of language: (i) the enlargement of the thoracic vertebral canal (MacLarnon and Hewitt 1999); (ii) the descent of the larynx (Arensburg 1989); (iii) increase in encephalization. Both the enlargement of the thoracic vertebral canal and the descent of the larynx are adaptations to enhance hominids' capability for vocalization. Increase in encephalization was in part made possible by the decrease of the gastrointestinal tract in hominid evolution. The decrease of the hominid gastrointestinal tract, however, is not directly linked to the evolution of hominid communicative behavior. Increase in encephalization during hominid evolution allows us to infer the enhancement of cognitive capacity. The details of the enhancement of cognitive capacity, however, is not known. Indentations on the endocranial casts may shed some light on the evolution of the meningeal venous system (Saban 1995) and the surface organization of the neocortex (Hollaway 1995). The evolutionary trend in the development of meningeal venous system and the neocortical organization in terms of surface gyri and sulci, as one would expect, progressed in the direction toward the complex meningeal venous system and neocortical parcellation of modern humans. Nevertheless, this evidence provides no clue to the evolution of hominid communicative behavior. It remains the case that the evolutionary origin of language can only be inferred by integrating all of the relevant information from paleoanthropology, neuroscience, animal communication, theories of evolution and linguistics.

2. The origin of language as a tool of communication

Scholars who recognize the merit of functional linguistics agree that language emerged evolutionarily first and foremost as a vehicle of human communication, *not* as an instrument of thought. The claim of language as an instrument of thought is an intuitive and a priori claim that defies scientific verification. Even from the perspective of intuition, it is difficult to imagine language as an instrument of thought before the dawn of civilization which is marked by such features as written language, urbanization, advanced technology, expanded population and complex social organization. As for pre-human hominids or early anatomically modern humans (AMH) one hundred thousand years ago, we don't know and probably will never know how they engaged in 'thinking'. But we do know that all hominids, like all animals, communicate. We can also be sure that at some point of the evolution of hominid communicative behavior after some radical innovations, language emerged. I will, therefore, assume that language emerged phylogenetically as a human vehicle of communication. Language may serve as an instrument of thought for modern humans. It is an interesting issue that seems more philosophical than empirical. The focus of this paper, however, revolves around the 'missing links' in the evolution of hominid communication from the perspective of neuroscience, animal communication and linguistics.

3. Pre-language vs. post-language communicative behavior of hominids

A critically important issue in the study of the origin of language is ascertaining the features and characteristics of hominid communicative behavior that would qualify it as language. From the perspective of generative linguistics, the definitive characteristics of human language is the recursive function in its grammar. Hence, from the perspective of generative linguistics, the critical point of transition from hominid communicative behavior to language would be the acquisition of the recursive function (Pinker 1994). Li and Hombert (2002) point out that the recursive function is a method of description. One may use it to describe certain aspect of human language as well as some animal communicative signals such as the songs of the humpback whale and the mockingbird. If hominid communicative behavior evolved into language gradually,[3] the point of demarcation should not be based on when the communicative behavior is susceptible to a description involving recursive functions. We must seek other means to establish the point in hominid evolution that separates pre-language communication from post-language communication.

Li and Hombert (2002) propose that an important criterion separating pre-language communication from post-language communication in hominid evolution is the difference between linguistic change and the evolutionary change of communication in animal kingdom. The pre-language communicative behavior of hominids, like animal communicative behavior, was subject to the constraints of Darwinian evolution. It involved natural selection and sometimes, genetic mutation. A change of the pre-language communicative behavior of hominids was adaptive in the sense that the change enhanced the hominids' life expectancy and reproductive success. Those hominids who made the change achieved a higher level of fitness than those hominids who failed to make the change. Hence, those hominids who made the change would have a competitive edge for both survival and reproduction over those who failed to make the change. Linguistic change, which began in the post-language communicative era of hominid evolution, however, is by and large tied to society and culture. It has nothing to do with life expectancy, reproductive success, genetic mutation or natural selection. Language changes constantly and rapidly. Our pronunciation, vocabulary, ways of speaking and even grammar change significantly within our life time.[4]

If the difference between linguistic change and the evolutionary change of communication in animal kingdom serves to demarcate pre-language communication from post-language communication in hominid evolution, we still need to know the nature of the communicative behavior of hominids at the point of that demarcation. I have designated this point of demarcation the **Crystallization of Language** from the perspective of evolutionary time. Let us explore this state of hominid communication.

Hominid communication, like higher primate communication, typically involves the simultaneous use of several sensory channels: visual, auditory and tactile. At the outset, we will leave behind the evolution of facial expressions, body postures and tactile communicative behavior among hominids. On the basis of comparative studies among mammals, we may conjecture that there is no radical change in the evolution of facial expressions, body postures and tactile communicative signals among hominids. Facial expressions, for example, are largely dependent on facial muscular and skeletal architecture. Most facial expressions among humans are controlled by emotional states and not subject to voluntary control (Ekman 1984). If we are interested in the origin of language, we need to understand the emergence of symbolic signals as Deacon (1997) points out. At the most elemental cognitive level, a symbolic signal must possess two properties:

1. It refers to a concrete object
2. The reference is context-independent

This is the beginning of 'meaning' and vocabulary. The emergence of the first symbolic signal referring to a concrete object also represents the crossing of the first 'missing link' in hominid evolution. Animal communicative signals are not symbolic according to our definition of elemental symbolic signals, and animal communicative signals do not have 'meaning'. What they have is 'function' such as threat, appeasement/submission, courtship/copulation, warning/alarm, recruitment, assembly, dispersion, identification, territoriality, feeding, etc. Functions must not be confused with 'meaning". Each of the following linguistic expressions has the function of 'threat':

(1) I'm going to bite your ears.
I'll kill you.
I'll knock you over.
I'm going to beat you up.

But they all have different meanings, and they have different meanings because they are linguistic expressions. In animal communicative behavior, threat signals are just threat signals, no more and no less. Different species have different threat signals. Some signals are graded according to the intensity of the signaler's emotional state, and some, discrete. Some are structurally more complex involving various components and several channels of communication, and some are simple. All members of the same species use the same threat signals, and all threat signals are emitted by signalers to threaten intended receivers. They have no meaning. We can make similar statements about any other functional category of animal communicative signals.

A particular set of animal communicative signals that have received a great deal of attention is the frequently cited warning signals of vervet monkeys. These warning calls are not symbolic signals because they do not possess property (2). Nevertheless, the warning calls of the vervet monkeys are halfway toward a symbolic signal, because they differentiate, for example, reptilian, avian, mammalian and other predators. The differentiation, however, holds only in the context of warning. It is not context-free.

Li and Hombert (2002) as well as Givón (this volume) have argued that symbolic communication begins with concrete nouns in hominid evolution. However, unlike Givón who claims that lexical concepts must be "taken for granted" but "are usually not well-coded in natural pre-human communication", Li and Hombert argue that "the creation of each new symbol (for a concrete object) represents a stroke of genius by a hominid, and the establishment of each newly created symbol in the repertoire of the communicative signals of the social group to which the creator belongs, requires social and cultural transmission." They go on to state that,

> Having a few communicative symbols for concrete objects, however, **is not tantamount** to being aware of the abstract principle of associating symbolic communicative behavior with concrete objects, even though the symbol itself is a token of this principle. In other words, there is a significant difference between

> using a communicative symbol for a concrete object and being aware of the principle underlying that act of creation. Thus, the appearance of communicative signals that signify concrete objects 1.5–2 million years ago did not imply the dawn of language. As we have stated earlier, the addition of each new communicative signal that symbolizes another concrete object is a significant step along the evolutionary pathway toward the emergence of language. (Li and Hombert 2002)

The appearance of a communicative signal that symbolizes an abstract entity such as an event or action represents a quantum leap toward the crystallization of language. It is the second most important milestone in the evolution of hominid communication toward language. Symbolizing an action or event presupposes an understanding of the relation between an actor and an action or an agent and an activity. Hence, the concatenation of an actor with an action, if and when a communicative situation calls for it,[5] should be naturally emergent. It does not constitute a quantum leap toward the crystallization of language. Its appearance represents an overt manifestation of an awareness of the relation between an agent and activity or an actor and an action.

The third most important landmark in the evolutionary process leading toward the crystallization of language was reached when the repertoire of symbolic signals, i.e. the size of lexicon, attained a **critical mass**. The magic number stands at approximately a few hundred items. This estimate is based on the following:

A. Language is a vehicle of communication.
B. The crystallization of language preceded the dawn of civilization and significant material culture by tens of thousands of years. A full-fledged and adequate tool of communication would not have required as large a lexicon at that time as it did later.
C. Pidginists (Hall 1953, Samarin 1971) have noted that a pidgin with a set of approximately 1,000 morphemes functions adequately as a communicative tool in a community.[6]

If the demarcation between pre-language communication and post-language communication in hominid evolution is the difference between linguistic change and the change of animal communication, then the crystallization of language must have occurred shortly after the size of lexicon reached a critical mass. A lexicon with a critical mass contains just about all of the bare essentials necessary for communication. Once words are sequenced to form larger linguistic units, grammar emerges naturally and rapidly within a few generations. The emergence of grammar in the first generation of speakers of a creole attests to the speed of the process. The speed of the emergence of the first grammar at the inception of language is astronomical in comparison to the speed of Darwinian evolution. It is a matter of a few generations after the lexicon has reached a critical mass. The first grammar must not be equated with the grammar of some contemporary languages. It is a grammar with a few word order principles and grammatical markers. It has few, if any, morpho-syntactic devices. In the study of the origin of language, the importance of morpho-syntax of a language tends to be overemphasized. It is true that many contemporary languages display highly complex morpho-syntactic structures. Perhaps presence of complex morpho-syntax in many modern languages has induced many scholars to believe that even at the origin of language, complex morpho-syntaical structures are a necessity. I do not belittle the importance of the study of morpho-syntax. What I wish to clarify are four points:

(i) The long process of the evolution of hominid communicative behavior leading toward the crystallization of language involves primarily the development and accumulation of symbolic communicative signals, i.e. the lexicon.
(ii) A large enough lexicon without a grammar will go a long way toward achieving communicative efficacy.
(iii) If creolization is a hint, a few generations probably constitute sufficient time for a grammar to emerge as speakers begin to sequence words together to form larger communicative units.
(iv) At its origin, language only appeared in what may be called "task-oriented" communication. Task-oriented speech is shorter and simpler than casual conversation. But it belongs to the genre of casual spoken language.[7] The grammar and structure of language used in casual conversation differs significantly from that of formal spoken language or written language (Chafe 1982). In languages with highly complex morpho-syntax, the complexity tends to diminish in casual conversation. Task-oriented speech tends to rely even less on grammar than casual conversation.

By (iii), I do not imply that most of the grammatical structures observed in contemporary languages emerged within a few generations. A first generation creole with a few simple word order conventions and some grammatical particles is a full-fledged language that will serve as an adequate vehicle of human communication, and in possession of the most important property as all languages do in comparison with animal communication: change rapidly due to social and cultural reasons, not natural selection. We tend to be impressed by the complexity of morphological phenomena in modern languages involving such phenomena as agreement, derivation, inflection, declension, mood, tense, aspect, number, gender, case, affixation, etc. If linguistic morphology of the types just cited were essential components of the grammars of **all** languages, one could conclude that it would require hundreds or thousands of years for a grammar to emerge after a large enough lexicon had come into existence. However, it is not true that all languages possess such morphological structures. Many languages in the world have little or no grammatical agreement, derivation, inflection, declension, tense, number, gender, case, and many languages with long histories of both written and spoken traditions rely primarily on some word order conventions and grammatical particles with little or no morphology as their grammar. The Chinese language family is but one of numerous examples.[8]

One of the issues that dominated twentieth-century linguistics is the recursive property of language structure exemplified by embedding and conjunction. In light of this issue, Frits Newmeyer raised the question: How did embedding emerge from task-oriented speech? The following example taken from my research on historical Chinese syntax illustrates a diachronic pathway for the emergence of an embedding structure from task-oriented speech.

In modern Chinese, there is a construction called "The Descriptive Clause" (Li and Thompson 1989). The following is an example,

(2) Wo you yi-ge meimei hen xihuan yinyue.
I have one-CLASSIFIER sister very like music
'I have a sister who happens to like music a lot.'

(1) contains an embedded subordinate clause: *hen xihuan yinyue* 'likes music a lot'. However, the embedded clause is not a relative clause. In Chinese, relative clause precedes the head noun and it is marked with the particle *de*, whereas the Descritpive Clause follows the head noun and it is unmarked. For example, (2) contains the same embedded clause as (1). But in (2) the embedded clause is a restrictive relative clause marked by the particle *de*:

(3) Wo you yi-ge *hen xihuan yinyue de* meimei.
I have one-CLASSIFIER very like music PARTICLE sister
'I have a sister who likes music a lot.'

In addition to differing from the relative clause in terms of its position vis-a-vis the head noun, the descriptive clause also differs from the relative clause semantically. The descriptive clause indicates an incidental feature of the referent signified by the head noun. Hence, the translation of (1) : 'a sister who **HAPPENS** to like music a lot'. The semantic function of the Chinese relative clause is the same as any restrictive relative clause in other languages. It establishes a sub-category of the entity signified by the head noun, and the subcategory is determined by the property stated in the relative clause.

The rise of the Descriptive Clause occurred in the 18th–19th century. It arose through the concatenation of two consecutive utterances in speech. The two consecutive utterances are represented by (3):

(4) Wo you yi-ge meimei. Hen xihuan yinyue.
I have one-CLASSIFIER sister very like music
'I have a sister. (She) likes music a lot.'

Absence of the subject (zero anaphora) in the second utterance of (2) referring to the same topic as the first utterance is standard practice in Chinese speech. The incidental meaning of the second utterance of (3) can be inferred. But the inference is not obligatory. Each of the utterances in (3) has a full-fledged intonation pattern of a declarative utterance. (1), on the other hand, has only the incidental reading, and it is uttered with one unbroken declarative intonation pattern. Both (1) and (3) are fully functional in modern Chinese. The mechanism for the rise of the Descriptive Clause involves changes in prosody, and these changes gave rise to a new embedding structure with new meaning. Structurally the prosodic changes entail a boundary change, i.e. an utterance boundary, which is typically signified by a relative long pause and the beginning of a new intonation unit, has been changed into a word boundary. The literature on diachronic morpho-syntax contains a wealth of examples of the emergence of new morpho-syntactic structures through prosodic and boundary changes which occur naturally and frequently in casual speech.

4. The emergence of symbolic signals and cognitive capacity

Symbolic signals have never emerged in the natural communicative behavior of animals throughout the entire history of evolution. Hominids of the genus Homo are the only ones that developed symbolic signals. Why? In a series of publication culminating in Dunbar (2001), Dunbar has argued that increased group size provided the immediate impetus for the evolution of a large brain and an effective and efficient communicative behavior among hominids. There is no communicative behavior that is more effective and efficient than one that is based on symbolic signals. Dunbar is correct in so far as singling out group size as the most important selectional force favoring the evolution of a large brain and symbolic communicative behavior. Early hominids, in order to survive on the savannas of east Africa where powerful predators and agile preys had been co-evolving for millions of years like an arm race, had to rely on better organized social groups than herds of ungulates. However, singling out the most important selectional force in the evolutionary development of the hominid brain and communication does not map out the process of evolutionary development of hominid communication that led to the emergence of language. In order to probe the intricate co-evolutionary development of the brain and hominid communication, we may pose the following questions: "What is the cognitive prerequisite for the emergence of language?" "Is there a type of cognitive capability that humans have and other primates don't have?"

Tomasello (2000) hypothesizes that the understanding of intentionality, causality and conspecific's mental states is unique to humans. To bolster his hypothesis, he cites that "in their natural habitats, non-human primates:

- do not point or gesture to outside objects for others;
- do not hold objects up to show them to others;
- do not try to bring others to locations so that they can observe things there;
- do not actively offer objects to other individuals by holding them out;
- do not intentionally teach things to others.

They do not do these things, in my view, because they do not understand that the conspecific has intentional and mental states that can be potentially affected." (Tomasello 2000: 170).

Byrne (2001), however, holds the opposite view from Tomasello. In Byrne's view, chimpanzee understands the mental states and the intention of another chimp. The experiments cited by him involve monkeys in one and chimps in another. Both experiments aim to find out if the animal's behavior suggests some understanding of the mental state of another. For monkeys, the mother gave an alarm call whether or not the infants could see the danger, according to an experiment conducted by Cheney and Seyfarth (1990). This experiment suggests that the mother monkeys have no concept of the mental state of their infants. For chimps, an individual gave intense alarm when it saw danger approaching behind the back of its friend, i.e. the friend couldn't see the approaching danger. When both could see the approaching danger, they did not utter alarm calls. This experiment, conducted by Boysen (1998), indicates that chimps distinguish a conspecific's mental state of ignorance versus its mental state of being aware of a danger.

Whether or not we could reconcile Tomasello's hypothesis with Byrne's claim rests with the following issue: Is the ability to understand a conspecific's mental state and intention a graded or discrete phenomenon? If it is graded, there will be a gradation of the ability. If it is discrete, then an animal either has it or doesn't have it. I lean toward the former option, i.e. an animal's ability to understand a conspecific's mental state and intention is graded. However, the obstacle in correlating the natural occurrence of symbolic communicative behavior with the ability to understand a conspecific's intentions and mental states is that in animal kingdom, the former is discrete, but the latter is graded. In other words, a species either has symbolic communicative behavior or doesn't. Humans have symbolic communicative behavior. The rest of the animal kingdom does not. The distinction between animals and humans in terms of having symbolic communicative signals seems discrete. Concerning the knowledge of a conspecific's intention and mental states, humans have it to the fullest extent, but chimps and possibly other great apes may have only partial understanding. Here the distinction between humans and great apes in this domain is not discrete but graded.

Let us illustrate with another example of the chimps' partial ability to understand a fellow chimp's intentions and mental states. Consider the five phenomena that according to Tomasello do not occur in non-human primates' natural behavior. They are directly correlated with issues in the questions posed earlier: What is the cognitive prerequisite for the emergence of language?, and, Is there a type of cognitive capability that humans have and animals don't have?

Humans regularly engage in all these five behaviors: pointing out an object to someone; holding up an object to demonstrate it to someone; facilitating another person's observation of something by bringing the person to the location; holding something out in order to offer it to someone; intentionally teaching someone. In fact, being able to engage in these behaviors is tantamount to having symbolic communicative behavior. Chimps, of course, do not have natural symbolic communicative behavior, yet they have been observed to engage in some rudimentary form of teaching. Boesch reports the following observation of the chimps of Taï National Park:

> On the 22nd February 1987, Salome was cracking a very hard nut species (Panda oleosa) with her son, Satre. He took 17 of the 18 nuts she opened. Then, taking her hammer, Satre tried to crack some by himself, with Salome still sitting in front of him. These hard nuts are tricky to open as they consist of three kernels independently embedded in a hard wooden shell, and the partly opened nut has to be positioned precisely each time to gain access to the different kernels without smashing them. After successfully opening a nut, Satre replaced it haphazardly on the anvil in order to try to gain access to the second kernel. But before he could strike it, Salome took the piece of nut in her hand, cleaned the anvil, and replaced the piece carefully in the correct position. Then, with Salome observing him, he successfully opened it and ate the second kernel. (Boiesch 1993: 176–77)

No matter how one defines "teaching", the observation reported by Boesch constitutes partial teaching. It suggests, at the least, Salome's partial understanding of Satre's intention and mental state.

In view of the fact that the other four behaviors cited by Tomasello have never been observed among non-human primates in the wild,[9] I believe that we can use a slightly modified version of Tomasello's hypothesis to answer our original questions: "What is the cognitive prerequisite for the emergence of language?" "Is there a type of cognitive capability that humans have and other primates don't have?"

The cognitive prerequisite for the emergence of language is the ability to understand fully intentionality, causality and the mental states of conspecifics. Humans are in full command of this ability. All other animals do not have it, with the exception of chimps and possibly other great apes that may possess a much diluted or weaker version of this ability.

The reason for this cognitive prerequisite of the emergence of language stems from the nature of symbolic signals which permeate language. Symbolic signals are the most effective instrument for one to influence and affect the mental states and intentions of another with respect to what the signals symbolize. For example, if one wishes to influence the intention and mental state of another person with respect to cats, the forthright approach is to communicate with the person about cats. In other words, one needs symbolic signals referring to cats and other concepts. If an animal does not fully understand that a conspecific has intention and mental states that can be affected and influenced, there will be no impetus for symbolic signals to evolve. After all, what is the point of having a communicative signal that symbolizes, for instance, a concrete object such as 'cat', if it cannot be used to affect the mental state of another conspecific?

While ethologists continue to probe the cognitive capacity of non-human primates,[10] an issue of great interest that calls for clarification concerns the extent to which chimps and other apes do understand their conspecific's mental states. In connection with this issue we may also pose the following question: To what extent is the understanding of the mental states of a con-specific a prerequisite of the evolutionary origin of symbolic signals, and to what extent did the emergence of symbolic signals co-evolve with the cognitive ability to understand the mental states of con-specifics?

Like the state of affairs in hominid fossils, this question remains a missing link in the evolution of hominid communicative behavior. The key, however, may rest in neuroscience. The following section attempts to relate the current advances in neuroscience to the origin of language.

5. Neuroscience and the origin of language

There are four pieces of information about the hominid brain that shed light on the origin of language. Some of them represent initial breakthroughs in important areas of investigation. Citing them at this juncture in relation to a study of the origin of language necessarily requires some degree of speculation and conjecture. I can only hope that some of my conjectures and speculations will be eventually verified as the "links" that are no longer missing.

5.1 Increase in encephalization and cognitive memory

Fossil records demonstrate that the allometric increase in encephalization plays a major role in the evolution within the genus Homo. The allometric enlargement of the Homo brain, in particular, the neocortex, is inextricably connected with the emergence of symbolic communicative behavior and therefore, language. A critically important aspect of this connection is the expansion of cognitive memory which directly correlates with the expansion of the neocortex. The need for a large cognitive memory in order to process language goes far beyond the fact that we need a vocabulary and a grammar. To be fluent in a language, one needs to know how to say things in that language, and how to say things, even the most common-place things, in one language usually differs from how to say the same things in another language. A person can master all of the grammatical principles of a language, command a large vocabulary in that language, but if that person has not learned the myriad ways of saying things in that language, s/he is likely to utter weird and unacceptable expressions. The myriad ways of saying things in a language are primarily cultural conventions. For example, an English speaker conveys his/her hunger by saying, "I'm hungry". A French speaker, however, conveys his/her pang of hunger by saying "J'ai faim" which literally means "I've hunger". There is nothing ungrammatical about the English expression "I've hunger". It is simply not an English cultural convention for expressing hunger!

Consider another example contrasting two unrelated languages: A Chinese, who was proficient in English grammar and vocabulary but unfamiliar with the English ways of saying things, once said "I am ascending the street". What he tried to say was, according to the English ways of saying things, "I am going out!" But the Chinese way of saying "I'm going out!" is precisely "I'm ascending the street" as shown in the following Chinese utterance in which each morpheme is glossed:

(5) Wo shang jie le.
I ascend street PARTICLE

An important and relevant point here is that the Chinese expression, *shang jie* 'ascend street', is not an idiom. It consists of the verb, *shang* meaning 'ascend' plus an object, *jie* meaning 'street'. The verb *shang* 'ascend' may take on different objects to have different meanings. For example:

(6) a. Shang shan
ascend mountain
'to climb mountain'
b. Shang chuang
ascend bed
'to get in bed'
c. shang chuan
ascend ship
'to board a ship'
d. Shang feiji
ascend airplane
'to board an airplane'
e. Shang ke
ascend lesson
'to attend a class or have a lesson'

These examples constitute just a few items from a large set of possible forms involving the verb, "*shang*" 'ascend'. They are not a set of idioms, and their different meanings can not be attributed to the polysemy of the verb *shang* 'ascend'. If one wishes to become a fluent speaker of Chinese, one must learn these Chinese ways of saying "to attend a class", "to board a ship/plane/train/bus", "to get in bed", etc. We won't be speaking Chinese if we translate the English words into Chinese and then string them together according to Chinese grammar. We would be speaking "English Chinese". Similarly one would be speaking "Chinese English", if one said "ascend the street" instead of "going out", or "ascend the plane" instead of "board the plane" or "ascend a class" instead of "attend a class". Pawley (1991) and Grace (1987) have written eloquently about this aspect of language. I will quote Pawley:

> A language can be viewed as being (among other things) a code for saying things. There are a number of conventions that constrain how things should be said in a language generally or in particular contexts. Here I will mention only the general maxim: *be idiomatic*. This means, roughly, that the speaker (author, translator, etc.) should express the idea in terms that native speakers are accustomed to. For example, if you ask me the time and my watch shows the little hand pointing just past the 5 and the big hand pointing to the 2, an idiomatic answer would be 'It's ten past five', or 'It's five ten'. A reply such as 'It's five o'clock and one sixth' or 'It's five-sixth of an hour to six' or 'It's six less than fifty' would not count as idiomatic. To break the idiomaticity convention is to speak unnaturally. (Pawley 1991:433)

The implication of this important characteristic of language is that linguistic behavior requires a prodigious cognitive memory. The neocortex of our brain must be able to store a vast amount of knowledge acquired through learning: the vocabulary, the grammar, and the myriad ways of saying things. Hence increase in encephalization is inextricably connected with the emergence of language in the sense that it is a prerequisite for the emergence of language in hominid evolution.

5.2 The prefrontal lobe and planning

Ever since Phineas Cage, the first documented person who survived an inadvertent frontal lobotomy due to an accident, neurologists have known that the prefrontal lobe plays a critical role in planning and decision-making. Planning presupposes freedom from the present in both time and space. One of the most conspicuous differences between animals and humans is that animals, not humans, are condemned to react to the present at all times. For example, if food appears in front of an animal and there is no other over-riding distraction, eating will follow immediately. Animal communication, as Hockett (1960) notes, with the exception of the waggle dance of honey bees, does not have the "displacement" feature. In other words, humans are unique in being able to communicate about things remote in time and place from the communicative interaction. In fact, the very structure of conversation presupposes planning. Since speech consists of linear sequences of words and expressions, i.e. serial behavior, a speaker must continuously plan ahead in order to decide what to say and how to organize what one wishes to say if s/he wishes to speak coherently. However, there is no neurological evidence suggesting that the prefrontal lobe plays a direct, frontline role in subverting language behavior. Furthermore, there is no record of Phineas Gage suffering from aphasia or speech deficiency after his injury. Hence, the connection between the evolutionary development of the prefrontal lobe and the evolutionary development of hominid communication leading to the crystallization of language is parallel, not causal. Given that language facilitates all our cognitive capabilities including planning and the executive function, the enhancement of hominid communication must have co-evolved with the expansion of the frontal lobe. We may view this co-evolution from the perspective of automobile development. If the engineers develop an engine and a suspension system for a car to cruise at high speed, they will also have to develop a steering mechanism that can handle a car cruising at high speed. The engine and suspension system is analogous to the neural circuitries designed to acquire language, and the steering mechansim is analogous to the prefrontal lobe. Both a highly developed prefrontal lobe and the neural circuitries for linguistic behavior are essential for humans to have our unique cognitive capabilities. It is interesting that the prefrontal lobe of the neocortex is probably the last phase of encephalization in hominid evolution. Our closest hominid cousin, the Neanderthal, remained slope-headed, suggesting an underdeveloped prefrontal lobe, even though their endocranial vault appeared larger than that of the average anatomically modern humans.

5.3 Mirror neurons and the capacity for learning

I have pointed out in Section 5.1 that acquiring a language requires a prodigious memory, implying that a child must learn a great deal of information before achieving fluency in his/her native language. Thus, becoming a fluent speaker in a language requires not only a prodigious memory but also an impressive ability to learn. The impressive human ability to learn is characterized not only by the complexity and difficulty of what can be learned but also by the speed of learning. Humans are lightening fast in learning simple tasks by imitation when compared with other primates.[11] For examples: it took eight years for the potato-washing and wheat-throwing behavior to spread among a small troop of macaques confined to Kojima island in Japan (Visalberghi and Fragaszy 1990). Boesch (1993) reports that no chimpanzee less than eight years of age in the Taï forest can master the nut-cracking behavior of adults. In contrast, a human infant is capable of imitating shortly after birth (Meltzoff and Moore 1977) and a child is usually fluent in his/her native language by 3 years of age.[12] The speed of learning with which humans are uniquely endowed most likely derive from all aspects of the evolutionary improvement of the brain ranging from its size, its physiology to its circuitry. One particularly aspect of the brain that plays an important role in speeding up learning could be the number of mirror neurons in the human brain. Mirror neurons are discovered in the pre-motor cortex, namely, area F5, by Gallese *et al.* (1996) and Rizzolatti *et al.* (1996). In the words of Rizzolatti *et al.* (2000):

> When a specific population of these neurons becomes active, an internal copy of a specific action is generated. This copy may be used for two purposes: (1) planning and executing goal-directed actions or (2) recognizing actions made by another individual. The action recognition is based on a match between an observed action and its internal motor copy. (p. 539)

Even though mirror neurons were discovered in F5, transcranial magnetic stimulation and positron emission tomography experiments suggest that a mirror system may also exist in the Broca area (Rizzolatti and Arbib 1998). If mirror neurons are not confined to F5, the quantity of mirror neurons will constitute a critical component for the speed of learning.

The discovery of mirror neurons is recent. To date, there is no cross-species comparative study on the relative quantity of mirror neurons. However, it seems reasonable to assume that with the increase in encephalization in hominid evolution, the quantity of mirror neurons have increased proportionally in hominid brain. As quantity increases to certain threshold, the total capacity for learning engendered by the mirror neurons may take a quantum leap. Of course, without experimental and comparative evidence, I am formulating a conjecture that the prodigious human ability to learn is partially induced by the presence of a large quantity of mirror neurons in our neocortex. There is no doubt that in order to acquire a language, a human baby must possess a uniquely human ability in terms of speed and complexity to learn from the speech of the adults constituting his/her human environment. Hence, a significant quantity of mirror neurons may serve as the third prerequisite for the origin of language in hominid evolution.

5.4 The migration of neurons and higher-level cognition

The fourth prerequisite for the emergence of language involves a newly discovered mechanism in the ontological development of the human brain. Letinic and Rakic (2001) report that unique to the development of the human brain, neuronal precursor cells from the ganglionic eminence of the telencephalon migrate to the thalamus in the diencephalon. This migration occurs during the eighteenth to the thirty-fourth week of gestation during which the specific thalamic region targeted for study in the experiment (the pulvinar nucleus) grows significantly. Yet during the same gestation period there is no cell proliferation within the diencephalon. Hence, the growth of the thalamus is primarily due to the migration of precursor cells from the telencephalon. Letinic and Rakic (2001) provides direct evidence demonstrating this migration.

The thalamic nuclei transmit all somatosensory (except olfactory) information and internally generated information such as emotions to the neocortex of the telencephalon. Bundles of axons project from the thalamus to the neocortex and bundle of axons project from the neocortex back to the thalamus. The pulvinar nucleus of the thalamus has several subdivisions connected with the neocortex as well as some subcortical areas. It plays a central role in attentional processing because of its heavy interconnectivity with the posterior parietal lobe and the ventral projection pathway of the visual cortex. Its neurons are visually responsive and one of its functions is to filter distracting information. In short, the migration of telencephalon cells to the thalamus probably enhances the interconnectivity between the thalamus and the neocortex. The interconnectivity, in turn, facilitates the activity of the neocortex responsible for higher level cognitive activity. The details of the relevance of this new neurological discovery remain to be elucidated. Since linguistic behavior is not only a high level cognitive activity but also involves the integration of numerous neurological functions and neuronal circuitries, its evolutionary origin may depend partially on this unique migration of neuronal precursors from the telencephalon to the thalamus. This hypothesis is supported in part by the fact that both the human thalamus and the neocortex are disproportionally larger than those in other primates. In other words, the expansion of neocortex probably co-evolved with the expansion of the thalamus. The newly discovered neuronal migration in human ontological development suggests this co-evolution.

It is well known that chimps and humans share 99% percent of their DNA. It is also well known that the significant phenotypic differences between chimps and humans trace their origin to the different regulatory genes and the cascades of interactions among regulatory genes, structural genes as well as between regulatory genes and structural genes. It is estimated that one third of all of the human genes play a role in the expression of the human brain. Thus, in spite of the 99% overlap between chimp and human genomes, the enormous difference between the human brain and the chimp brain is not surprising. However, the precise nature of this cerebral difference in relation to the fundamentally different communicative behavior of the two species remains largely a mystery. For example, what is the neurological basis for the ability to understand a conspecific's mental state and intention? What is the neurological basis for the emergence of symbolic communicative signals? Questions of this type bring to mind the missing links in human fossil records. In this regard, Letinic and Rakic's discovery is trail-blazing and enlightening. The same may be said of Gallese and Rizzolatti's discovery of mirror neurons.

6. Conclusion

I have discussed the various missing links relevant to a probe into the evolutionary origin of language. These missing links are found in hominid fossil records, in comparative study of animal and human cognition/communication, and finally in neuroscience. We have a reasonably clear picture depicting the evolutionary origin of language if one agrees that language began phylogenetically as a tool of communication. But we are far from clarity in our attempt to understand the cognitive and neurological prerequisites for the evolutionary emergence of language. Based on current discoveries in various disciplines, I have offered some hypotheses and speculations in the hope that they will stimulate further inquiries into the issues concerning the cognitive and neurological prerequisites for the evolutionary emergence of language.

Notes

* I am grateful to Wallace Chafe, Fritz Newmeyer, Paul Schachter and Sandra Thompson for their comments and suggestions. The mistakes and deficiencies are all mine.

1. I agree with Wood and Collard (1999) that Homo habilis should be classified in the genus Australopithecus rather than Homo. *Habilis* did have a larger brain than other Australopithecenes. But it retained most of the features of Australopithecenes. These features include a significant sexual dimorphism, large teeth, facileness at tree-climbing, and most important of all, an ape-like ontological developmental clock which implies that *habilis* was not bound by the "terrestrial imperative". Furthermore, Louis Leakey's original motivation for designating *habilis* a species of Homo is no longer valid, i.e. at the time of the discovery of *habilis*, it was the oldest hominid associated with stone tools. Australopithecus ghari has displaced *habilis* for that honor. The first species of Homo, therefore, should be *ergaster*.

2. Li and Hombert (2002), citing four pieces of evidence, hypothesize that the 'crystallization' of language occurred several tens of thousands of years after the emergence of anatomically modern humans. The most critical piece of evidence is the arrival of humans in Australia at approximately 60,000 years ago. It entailed the crossing of many miles of deep and fast-moving ocean water from Asian mainland to Australia, implying the existence of language as a means of communication. As Li (2002) notes, since the emergence of anatomically modern human preceded the crystallization by several tens of thousands of years, the assumption of the monogenesis of language may or may not hold. This is a topic that calls for further research.

3. A system of communicative behavior in animal kingdom never emerges suddenly, although anatomical and physiological features may have a sudden origin due to the mutation of regulatory genes such as the Homeobox genes (Schwartz 1999). A system of communicative behavior evolves through behavioral and anatomical changes, interactions among conspecifics, genetic mutations and natural selection according to geological time.

4. Certain animal communicative signals changes rapidly also. For example, the humpback whale song sung by males during mating season for courtship function changes almost daily. At the end of the mating season, the males stop singing until it returns to the breeding ground the following year to start another mating season. Then, it will pick up the exact song it sang at the end of the last season and the song changes from there on. But the rest of the communicative repertoire of the humpback whale does not change this way. It observes the constraints of Darwinian evolution and if it changes at all, it changes in evolutionary time. One may argue that while human language may change rapidly without any relevance to life expectancy and reproductive success, it constitutes only a sector of the entire human communicative repertoire. However, language is the main vehicle of human communicative behavior that serves all communicative functions and purposes even without the aid of visual and tactile communication. The song of the male humpback whale constitutes only one signal among the humpback whale's repertoire of communicative signals and serves primarily the communicative function of courtship. Another point worth noting is that the vast majority of animal communicative signals are not known to change rapidly like the humpback whale song and some bird songs. Most animal communicative signals change at an evolutionary pace.

5. In most spoken languages of the world, actors or agents are rarely explicitly expressed. The reason is that in most conversations, when a speaker wishes to describe a particular activity or action, the actor or agent has already been established in the working memory of both the speaker and the hearer as a known entity. There is no need in such a discourse context to mention it again, even in a pronominal form. English is rather exceptional in that the mention of an actor or agent either in nominal or pronominal form is required in most grammatical contexts. This unusual grammatical feature of English cannot be attributed to the fact that English verbs are no longer inflectional. Many languages of the world that are not inflectional are zero-anaphora languages, i.e. languages in which the default grammatical principle is not to mention an actor or agent which is already an established entity in the working memory of the conversationalists. Chinese, Tai, Hmong, Vietnamese are examples of zero-anaphora languages (Li 1997).

6. M. Swadesh (1971) considers 1,000 morphemes an adequate number for natural languages.

7. Written language is a very recent cultural invention. The oldest written records are no more than six or seven thousand years old. Present day spoken languages have many registers or genres. The formal register such as an academic lecture resembles the written from in structure and grammar. At the origin of language, language should be seen as a tool for task-oriented communication.

8. The Chinese language family contains hundreds of languages many of which are not mutually intelligible. For a comprehensive survey of Chinese languages, see Ramsey (1987).

9. Some species of bower birds, which build elaborate bowers as courtship signals, are known to hold colorful objects in their beak in order to show them to a potential mate. This breach of Tomasello's observation about the 5 behaviors that are unique to humans brings to mind the waggle dance of honey bees as the only exception to the observation that humans are unique in being able to communicate about things that are remote in time and space from where the communication occurs. These exceptions are not meaningful from a comparative perspective because birds and bees are too distant from humans in terms of brain structure and function. As we probe the origin of language, it may be more sensible to confine a comparative study of the communicative behavior and the cognitive capacity to primates.

10. I have refrained from bringing up the enormous effort and literature to explore the limit of the cognitive capacity of great apes by training them to manipulate symbolic communicative signals. As Li and Hombert (2002) points out, all higher animals have a **cognitive reserve**, implying that an animal's cognitive capacity is not fully manifested in its entire behavioral repertoire. While training great apes to manipulate symbolic signals has its own scientific merit, it does not shed light on the evolutionary development of the hominid brain and hominid communicative behavior. Humans can train various animals to perform astonishing feat of activities. The best illustration is Irene Pepperberg's success in training Alex, an African grey parrot, to manipulate symbolic signals (Pepperberg 1991).

11. Some scholars question non-human primates' ability to learn by imitation, e.g. Tomasello *et al.* (1993) and Galef (1998). Much of the controversy, however, depends on the definition of imitation.

12. While a 3-year old might be fluent, the learning of vocabulary, grammar and ways of saying things continues into the teens even in societies without a written tradition.

References

Arensburg, B. 1989. "A Middle Paleolithic hyoid bone". *Nature* 338:758–60.

Boesch, C. 1993. "Transmission of tool-use in wild chimpanzees." In *Tools, Language and Cognition in Human Evolution*, eds. K.R. Gibson and T. Ingold (eds). Cambridge: Cambridge University Press.

Boysen, S.T. 1998. Attribution processes in chimpanzees: Heresy, hearsay or heuristic? Paper presented in the 17th Congress of the International Primatological Society at the University of Antananariro, Madagascar, 10–14 August.

Bryne, R.W. 2001. "Social and technical forms of primate intelligence". In *Trees of Origin*, F. De Waal (ed), 145–72. Cambridge, Mass.: Harvard University Press.

Chafe, W.L. 1982. "Integration and involvement in speaking, writing and oral literature". In *Spoken and Written Language: Exploring Orality and Literacy*, D. Tannen (ed), 35–54. Norwood, NJ: Ablex Publishing Corporation.

Cheney, D.C. and Seyfarth, R.M. 1990. "Attending to behavior versus attending to knowledge: Examining monkeys' attribution to mental states". *Animal Behavior* 40:742–53.

Deacon, T.W. 1997. *The Symbolic Species: The Co-evolution of Language and the Brain.* New York: W.W. Norton.

Dunbar, R.I.M. 2001. "Brains on two legs: Group size and the evolution of intelligence". In *Trees of Origin*, F. de Waal (ed), 173–92. Cambridge, Mass.: Harvard University Press.

Ekman, P. 1984. “Expression and the nature of emotion”. In *Approaches to Emotion*, P. Ekman and K. Scherer (ed). 319–43. Hillsdale, NJ: Lawrence Erlbaum Associates.

Galef, B.G. Jr. 1988. “Recent progress in studies of imitation and social learning in animals”. In *Advances in Psychological Sciences*, Volume II: *Biological and cognitive aspects*, M. Sabourin, F. Craick, and M. Roberts (eds), 275–97. Hove: Psychology Press.

Gallese, V., L. Fadiga, L. Fogassi, and G. Rizzolatti. 1996. “Action recognition in the premotor cortex”. *Brain* 119: 593–609.

Givón, T. This volume. “The visual information-processing system as an evolutionary precursor of human language”.

Grace, G. 1987. *The Linguistic Construction of Reality*. London: Croom Helm.

Hail-Selassie, Y. 2001. “Late Miocene hominids from the Middle Awash, Ethiopia”. *Nature* 412: 178–80.

Hall, R.A. 1953. *Haitian Creole: Grammar, Texts, Vocabulary*. American Folklore Society Memoir 43.

Hockett, C.F. 1960. “The origin of speech”. *Scientific American* 203: 88–111.

Holloway, R.L. 1995. “Toward a synthetic theory of human brain evolution”. In *Origins of the Human Brain*, J-P. Changeux and J. Chavaillon (eds), 42–54. Oxford: Clarendon Press.

Leakey, M.G., C.S. Feibel, I. McDougall, and A.C. Walker. 1995. “New four-million-year-old species from Kanapoli and Alia Bay, Kenya”. *Nature* 376: 565–71.

Letinic, K. and P. Rakic. 2001. “Telencephalic origin of human thalamic GABAergic neurons”. *Nature Neuroscience* 4: 931–36.

Li, C.N. 1997. “On zero anaphora”. *In Essays on language function and language type*, J. Bybee, J. Haiman, and S. Thompson (eds), 275–300. Amsterdam: John Benjamins.

Li, C.N. 2002. “Some issues concerning the origin of language”. To appear in *Complex sentences in grammar and discourse*, J. Bybee and M. Noonan (eds). Amsterdam: John Benjamins.

Li, C.N. and J.-M. Hombert. 2002. “On the evolutionary origin of language”. To appear in *Mirror Neurons and the Evolution of Brain and Language*, M. Stamenov and V. Gallese (eds). Amsterdam: John Benjamins.

Li, C.N. and S.A. Thompson. 1989. *Mandarin Chinese: A functional reference grammar.* Berkeley: University of California Press.

MacLarnon, A. and G. Hewitt. 1999. “The evolution of human speech: The role of enhanced breathing control”. *American Journal of Physical Anthropology* 109: 341–63.

Meltzoff, A.N. and M.K. Moore. 1977. “Imitation of facial and manual gestures by human neonates”. *Science* 198: 75–8.

Pawley, A. 1991. “Saying things in Kalam: Reflections on language and translation”. In *Man and a Half: Essays in Pacific Anthropology and Ethnobiology in Honor of Ralph Bulmer*, A. Pawley (ed), 432–44.

Pepperberg, I.M. 1991. “A communicative approach to animal cognition: A study of conceptual abilities of an African grey parrot”. In *Cognitive Ethology*, C.A. Ristau (ed), 153–86. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pickford, M., and B. Senut. 2001. “Contexte geologique et faunigue des restes d’hominides du Miocene superieur de Lukeino, Kenya”. *Compte Rendus de l’Academie des Sciences — Serie IIa — Sciences de la Terre et des Planetes* 332: 145–52.

Pinker, S. 1994. *The Language Instinct.* New York: William Morrow.

Ramsey, S. R. 1987. *The Languages of China.* Princeton, NJ: Princeton University Press.

Rizzolatti, G. and M. A. Arbib. 1998. "Language within our grasp". *Trends in Neuroscience* 21: 188–94.

Rizzolatti, G., L. Fadiga, L. Fogassi, and V. Gallese. 2000. "Cortical mechanisms subserving object grasping and action recognition: A new view on the cortical motor functions". In *The New Cognitive Neurosciences*, M. S. Gazzaniga (ed), 539–52, 2nd edition. Cambridge, Mass.: MIT Press.

Rizzolatti, G., L. Fadiga, L. Fogassi, and V. Gallese. 1996. "Premotor cortex and the recognition of motor actions". *Cognitive Brain Research* 3: 131–41.

Saban, R. 1995. "Image of the human fossil brain: Endocranial casts and meningeal vessels in young and adult subjects". In *Origins of the Human Brain*, J.-P. Changeux and J. Chavaillon (eds), 11–38. Oxford: Clarendon Press.

Samarin, W. J. 1971. "Salient and substantive pidginization". In *Pidginization and Creolization of Language*, D. Hymes (ed), 117–40. Cambridge: Cambridge University Press.

Schwartz, J. H. 1999. *Sudden Origins: Fossils, Genes and the Emergence of Species.* New York: John Wiley and Sons.

Stanley, S. M. 1996. *Children of the Ice Age: How a Global Catastrophe Allowed Humans to Evolve.* New York: W. H. Freeman.

Swadesh, M. 1971. *The Origin and Diversification of Language.* Chicago: Aldine.

Tomasello, M. 2000. "Two hypotheses about primate cognition". In *The Evolution of Cognition*, C. Heyes and L. Huber (eds), 165–84. Cambridge, Mass.: MIT Press.

Tomasello, M., A. C. Kruger, and H. H. Ratner. 1993. "Cultural learning". *Behavioral Brain Science* 16: 495–552.

Visalberghi, E. and D. Fragaszy. 1990. "Do monkey ape?" In *"Language" and Intelligence in Monkeys and Apes: Comparative Development Perspectives*, S. T. Parker and K. R. Gibson (eds), 224–73. Cambridge: Cambridge University Press.

White, T. D., Suwa, G., and Asfaw., B. 1994. "Australopithecus ramidus, a new species of early hominid from Aramis, Ethiopia". *Nature* 371: 306–12.

White, T. D., Suwa, G., and Asfaw., B. 1995. "Correction". *Nature* 375: 88.

Part 2

Language and cognition

CHAPTER 4

Sequentiality as the basis of constituent structure

Joan Bybee
University of New Mexico

1. Sequentiality and constituent structure

Recent investigations of the relation between grammar and language use, especially in the literature on grammaticization (Bybee *et al.* 1994, Hopper and Traugott 1993, among many others), have made it clear that grammar is not a fixed logical structure, but rather is constantly changing and evolving. Bybee (1998b) and Haiman (1994) have argued that the basic mechanisms of the grammaticization process are neither domain-specific nor species-specific. Thus the hope has arisen that the highly abstract and heretofore mysterious properties of grammar might be explainable in more general terms. In my earlier work I have addressed the process by which grammatical morphemes develop, emphasizing the role that repetition plays in the process. The current paper addresses the mechanism behind the hierarchical arrangement of linguistic elements into constituents, once again emphasizing the role of language use and repetition.

The existence of constituent structure and the hierarchical organization resulting from it has always been taken by linguists as prime evidence that linguistic behavior does not consist merely of linear strings of elements. It is further believed that the hierarchical organization of sentences is one of the most basic aspects of language, indeed, a defining feature of human language.

Linguists rarely ask why natural language has constituent structure;[1] they merely assume that it does, just as they assume that all phonologies will be organized into segments and features. In the spirit of Lindblom *et al.* (1984), I submit that structure can be explained, that form is emergent from substance, and that the larger design of language can be interpreted as the indirect consequence of local behavior (Lindblom *et al.* 1984: 186). My goal in this paper is to present and defend the hypothesis that sequentiality is basic to language and constituent structure emerges from sequentiality because elements that are frequently used together bind together into constituents.

This point is likely to seem rather obvious to some readers, and indeed follows directly from comments in Givón (1998, in this volume). To my knowledge, however, it has not been directly argued for or demonstrated. In fact, some linguists seem to be convinced that constituent structure is basic to language, and therefore might find the hypothesis surprising. Thus, towards the ultimate goal of identifying the mechanisms that create grammar and relating them as much as possible to domain-general cognitive processes, I undertake this demonstration.

2. Definition of constituent

A popular introductory textbook for linguistics, *Language Files*, has a nice discussion of the notion of constituent. It first begins with a discussion of linear order in sentences, and then points out that in addition to linear order, there is hierarchical organization. The book says that the semantically coherent groupings of words in sentences are constituents, thereby identifying both semantics and linear order as important to constituent structure. It goes on to give three tests for identifying constituents:

1. Constituents can sensibly be used alone, as in the answers to questions. (*Who chewed up your shoe? My new puppy.*)
2. Constituents can be replaced by pro-forms. (E.g. *I saw <u>him</u> <u>do it</u>* where *him* replaces *my new puppy* and *do it* replaces *chew up my shoe.*)
3. Constituents can occur in various places in the sentence, e.g. an NP can be the subject of the verb or object of the verb or a preposition; an NP can be extraposed (e.g. *My new puppy, he chewed up my shoe.*)[2]

3. Explanations for constituent structure

Of course various explanations for why language is organized into constituents are possible. One could claim that specific types of phrases, such as NP, VP and PP are innate and only need the presence of minimal input to trigger their language-specific organization. Or one could propose, as does Langacker (1987:310), that 'hierarchy is fundamental to human cognition' making it a general cognitive attribute that can be applied to language. Certainly I agree with this view. However, we must also ask just what is organized hierarchically and how languages come to be organized in the particular way that they are.

Langacker's (1987, 1997) particular proposal is that constituency reflects semantic relations. Again, it is easy to agree with this position since many linguists have pointed out the iconic relation between the conventionalized structures of languages and their semantic relevance (e.g. Bybee 1985). However, as we will see below, constituency and meaning are not always in a strictly iconic relationship and the human mind can apparently find constituents in re-occurring meaningless syllables (Saffran *et al.* 1996). Thus we search for an additional factor to fully explain the existence of constituent structure.

My hypothesis is that semantics, and to some extent, pragmatics and our experience with the world, will determine what elements tend to occur together in sequences in an utterance, but repetition is the glue that binds constituents together. Thus I hypothesize that hierarchies of constituent structure are derivable from frequent sequential co-occurrence. In this view, the more often particular elements occur together, the tighter the constituent structure. Thus low-level constituents such as a determiner, *the*, and a noun, such as *puppy*, frequently co-occur, while higher-level constituents, such as an NP, *the puppy*, and verbs such as *ran, licked,* or *slept* occur together less often. Note that in this view constituent structure can be gradient and two constituents which seem to have the same structure may have different degrees of cohesion due to the differences in their co-occurrence patterns (Bybee and Scheibman 1999).

4. Knowledge of grammar is procedural knowledge

To understand the role that frequency or repetition plays in the creation of grammar it is important to recognize that language production is a neuromotor behavior based on procedural knowledge rather than propositional knowledge. Propositional knowledge is 'knowing that' or knowing facts such as 'Santa Fe is the capital of New Mexico'. Procedural knowledge is 'knowing how' and includes knowing how to tie shoelaces or how to drive a car. Propositional knowledge is conscious knowledge which is easy to report on. Procedural knowledge is usually below the level of conscious awareness and while subjects can carry out the procedures, it is much more difficult for them to report what the procedure is. This distinction has an interesting parallel in the difference between lexical and grammatical knowledge. While speakers are often able to report on the meanings of words or phrases, it is much more difficult for untrained speakers to explain the meanings of grammatical morphemes or grammatical constructions. Thus we might conclude that lexical items involve at least some propositional knowledge, while grammatical constructions are largely procedural.

This conclusion is further supported by the manner in which procedural knowledge develops, as outlined by Anderson (1993) and Boyland (1996). Two properties of the development of procedural knowledge are important for our understanding of the way grammar develops. First, frequently used actions become fluent more quickly; that is, repetition increases fluency. For the purposes of language, our common-sense notion of fluency can be applied here: in fluent speech, words are strung together without inappropriate pauses. We can also go beyond this sense of fluency and note that with high levels of repetition, articulatory gestures can overlap one another and individual gestures can be reduced both in duration and in displacement. Thus grammaticizing constructions that undergo extreme frequency increases also undergo extreme phonological fusion and reduction, as for example, when *going to* reduces to *gonna* (see Bybee, Perkins and Pagliuca 1994; Bybee 2001). The second property is in a sense the mechanism that makes the first one possible: 'recurring sequences of actions come to be represented at a higher level as single actions, thus increasing fluency' (Boyland 1996: 10). Thus repeated sequences become fluent because they become automated into a single chunk that can be accessed and executed as a unit.

5. Linguistic evidence for chunking

Several types of evidence for the chunking of high frequency sequences can be cited. First, one unit of a chunk primes or automates the other unit. In the plaza at Old Town in Albuquerque I watched a boy of about five years spot an antique cannon. He said 'hey, Dad, there's a cannonball. Can I climb on the cannonball?' The father responded, 'that's not a cannonball, it's a cannon.' The boy insisted, 'Dad, can I climb on the cannonball' and the exchange repeated itself. The boy had learned *cannon* only in the context of the compound *cannonball* and that was the only context in which he could access the word. Hearers have automated chunks as well, with analogous priming effects. In the US, upon hearing *supreme*, one can expect *court* as the next word; or upon hearing *sesame* one can expect *street*.

Second, inside frequently used chunks, internal structure tends to be lost. Thus *gonna* no longer consists of the three morphemes *go, ing* and *to*. Third, the morphemes or words inside a chunk become autonomous from other instances. For example, speakers probably do not associate *go* in *gonna* with the lexical movement verb anymore. Sosa and MacFarlane (to appear) show that subjects have difficulty identifying the word *of* when it occurs in frequent chunks such as *sort of* or *kind of*. And fourth, the components of a chunk become compressed and reduced phonologically, as illustrated by *gonna,* but also by other verbs that are frequently followed by *to*, such as *wanna, hafta, gotta,* and so on.

The hypothesis of this paper is: **Items that are used together fuse together.**

This could be called the Linear Fusion Hypothesis. The explanation for this phenomenon is two-fold: first (and in my mind, foremost) is the automation of production that is typical for procedures; second, the fact that items are predictable in context allows the speaker to unleash the reductive processes that increase fluency.

6. Further examples of the fusion of high frequency combinations

Boyland (1996) argues that the sequence *would have* fuses due to the high frequency with which the two units occur together. The misanalysis apparent in the non-standard spelling *would of* shows that the identity of the two units has been lost. Also, the context after a modal is the only place that *have* reduces down to [ə]. In the example of *be going to* the most fusion is among the parts that are invariant and therefore occur together more often, that is, the *going to* part. The *be*-verb part, which varies among *am, is, are, was* and *were,* reduces and fuses only in the high frequency first person singular: *I'm going to* becomes not just *I'm gonna* but also [aimə̃ɾə̃] in casual speech. Krug (1998) studies the rate of contraction of English auxiliaries with the subject and finds that the more frequent a combination, the more likely it is to show contraction. For instance, the most frequently occurring contraction is *I'm* and that combination is the most frequent in the corpus. The basis for the high frequency of *I'm* is the strong subjectivity of speech, which makes first person singular the most frequent pronoun and the fact that *am* cannot occur with any other subject (Scheibman 2002). Bybee and Scheibman (1999) study the reduction of *don't* in American English and find that it reduces most in the contexts in which it is used the most, i.e. after *I* and before certain verbs such as *know, think, mean, care, feel,* etc.

While it is not necessarily true in the cases just cited, it is usually the case that high frequency, grammaticizing items that are used together are in the same constituent in the traditional sense. For example, a determiner will tend to show reduction and fusion that depends upon the noun with which it is in construction. Thus the English article *the*, has two pronunciations, [ði] and [ðə] as does *an/a*. The second variant in each case developed as a reduction before the consonant in the noun or adjective of the same NP. The French masculine definite article *le* loses its vowel before a vowel-initial noun: *le* + *ami* > *l'ami* 'the friend', etc.

Postpositions fuse with the nouns they operate on to become case suffixes. Thus the Turkish word *değin* meaning 'as far as' has fused with the preceding noun to make a new case suffix, as in ***köydek*** 'as far as the village' (Kahr 1976). Auxiliaries that are part of tense and aspect constructions and that follow the verb fuse with the verb to become suffixes, as in the famous case in the Romance languages, where the infinitive + *habere* constructions yields the future tense: Old Spanish *cantar ha* becomes Modern Spanish *cantará* 'he will sing'. Examples such as these are abundant in the grammaticization literature.

7. Constituency with less frequent combinations

The examples just cited are the extreme cases, since they involve combinations of extremely high frequency. But more subtle effects can also be found in cases of co-occurrence that are less frequent, leading me to hypothesize that chunking and constituency relate directly to frequency of co-occurrence.

Gregory *et al.* (1999) examine three measures of shortening of content words in a large corpus of conversation (the Switchboard corpus, see below). They find that the tapping of a word-final /t/ or /d/ is highly affected by the probabilistic variable 'mutual information', which is a measure of the likelihood that two words will occur together. If the final /t/ or /d/ occurs before a word that is highly likely given the first word, the rate of tapping increases. For instance, tapping would be more common in the word pair *lot of* than in *out alone*. They write 'The effect of mutual information on tapping suggests that tapping is a process that may also preferentially apply internally to highly cohesive pairs' (Gregory *et al.* 1999:9). Word-final deletion of /t/ and /d/ and the duration of a word are also highly affected by mutual information as well as other measures of probability. Jurafsky *et al.* 2001 find similar effects of probability on function words.

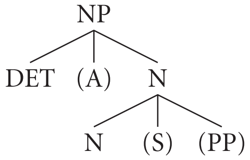
The usual interpretation of small phonological adjustments that occur between words is that they are due to on-line phonetic processes that apply as the words are strung together. However, if this is the case, it is difficult to explain why frequency of co-occurrence affects the probability that a phonological adjustment would take place. Part of the explanation is that the speaker, knowing that the next word has a high probability of occurring given the current word, allows reduction processes to occur. But for the speaker to know the probability of the occurrence of the next word, s/he must have a mental representation that includes knowledge about what words have occurred together in the past. Thus the occurrence of two words together strengthens the sequential link between them. It is this sequential link that is the local basis for the eventual emergence of constituent structure. That is, pairs of words with strong sequential links are regarded as occurring in the same constituent.

8. Sequentiality and constituency in the NP

To test the hypothesis that traditional notions of constituent structure correspond rather directly to frequency of co-occurrence, I elected to examine the English noun phrase in conversation. The noun phrase (especially one with a lexical noun) is a very good example of a constituent in English, as it has the same structure whether it is the subject or object of the verb, or object of the preposition. Further, it is sometimes used independently in conversation (Ono and Thompson 1994).[3]

The basic structure of an English NP was represented with the phrase structure tree in (1) in early models of generative grammar.

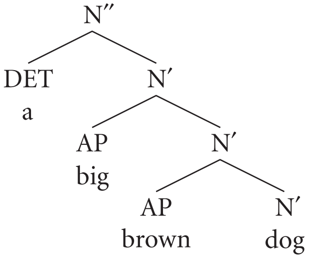
(1) NP structure:



The DET (determiner) and N are obligatorily present (although the DET can be a zero) and the other items are optional. The NP can contain an adjective which precedes the N, or a relative clause (S), or prepositional phrase (PP), which follows.

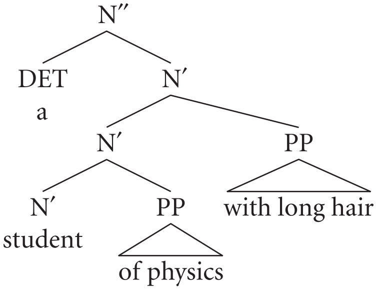
Another more recent representation of the NP in X-bar notation, provides an extra layer of constituency for each element added to the NP (see Radford 1988). Thus (2) shows an NP with two adjectives and (3) an NP with post-modifiers.

(2)



What predictions are made about levels of constituency by these phrase structure trees? In (1) the DET (A) N are all under the same node, NP, so they are will not be separated by any constituent boundaries except those that indicate that they are separate words. In (2) each adjective increases the distance between the determiner and the noun.

(3)



With post-modifiers in (1) or (3), there are two consequences for the constituent structure in the normal interpretation: first, the N will be separated from the DET (A) by an extra bracket, suggesting, perhaps erroneously, that an N with post-modifiers is not in the same constituent with the DET as the N is when it is final to the NP. Second, the post-modifier, being a constituent itself (as an S or PP) is separated from the N by several brackets, indicating a looser constituent structure.

While the second consequence seems fine, the first presents an odd prediction that a DET N combination has tighter constituent structure when no modifiers follow it than when they do. However, other mismatches between the bracketing generated by phrase structure rules and the surface constituency have been identified in the literature and a special component of rules, Readjustment Rules, were set up to fix these mismatches (Chomsky and Halle 1968). In addition, phonological phrasing, or the domain of phonological rules is sometimes at odds with the generated phrase structure and needs adjusting. Thus it is generally clear that the structures predicted by trees such as (1), (2) and (3) cannot be taken in a completely literal way. However, certain elements of the general structure presented in these trees are well-motivated, i.e. that the NP is a constituent, and that these elements, DET, pre-modifying adjectives and post-modifying PPs and Ss belong in the NP constituent. Thus I will compare this general structure to the quantitative facts of co-occurrence to determine whether or not generalizations about constituent structure can be said to be derived from multiple patterns of language use.

My hypothesis is that specific items that are used together frequently will form tighter bonds than items that occur together less often. The prediction is that items that occur within a traditional constituent are more likely to occur together in running discourse than are items that are in different constituents. One complication is the existence of optional items in a constituent. Strictly speaking, my hypothesis would predict that optional elements have weaker sequential links than obligatory elements, and thus looser constituency. While this prediction may be correct, I will not pursue it further here. Instead, I will focus on testing, in a general way, the predictions about co-occurrence made by the NP constituent. The question will simply be whether or not nouns occur more frequently with items inside or outside the NP.

To study the English NP, I selected eleven lexical nouns from the most frequently occurring nouns in the Switchboard corpus. This corpus consists of over 2.4 million words from recorded telephone conversations. The nouns used were: *husband, mother, computer, movie, school, car, house, money, idea, class* and *problem.*[4] These nouns were studied only in their singular forms. The goal of the study was to discover which items most frequently preceded and most frequently followed these nouns in the conversations and to compare the frequency of preceding vs. following items to see if the quantitative distributional facts correspond to traditional notions of constituency in the NP.

First consider where pauses occur in relation to these eleven nouns. Table 1 shows that the probability of a pause following the noun is much greater than that of a pause preceding the noun. In fact, more than one-third of the tokens of these nouns were followed by a pause, while fewer than one percent were preceded by a pause.[5] This suggests a much weaker constituency bond to the item following the noun, either because it is in another constituent or because it is a more loosely joined part of the NP. The reason for the low probability of a pause before the noun is presumably the obligatory presence of a determiner before the noun.

Table 1. Occurrence of pauses before and after eleven nouns from the Switchboard corpus (N=7870)

Before the noun	After the noun
74 (1%)	2696 (34%)

Table 2 shows the three most common linguistic items to precede and follow the noun. What is notable here is the greater predictability of the item preceding the noun compared to the one following it. In the Preceding column, *the* accounts for 17% of all tokens preceding the noun. Following the noun, *and* is the most common unit, but it only occurs after 7% of the tokens. Thus the co-occurrence patterns for X+N are stronger than for N+X, reflecting tighter constituency.[6]

Table 2. The three most frequent linguistic items preceding and following the eleven nouns

Preceding			Following		
the	1,348	17%	*and*	562	7%
my	958	12%	*that*	345	4%
a	767	10%	*is*	270	3%

In terms of understanding the constituent structure of the NP, it is instructive to view some of the most frequent items to precede and follow these eleven nouns. The items in Table 3 were selected by first counting the ten most frequent items to precede each noun, and then by selecting all the items that occurred in more than one list. Table 4 was constructed in the same way with attention to the items that followed the noun. Tables 3 and 4 are organized by grammatical function. The totals in the last column are followed by asterisks if they present apparent counter-examples to the hypothesis that elements following or preceding the noun are in the same constituent. These cases are discussed below.

Tables 3 and 4 confirm that the constituency of the NP as traditionally represented corresponds closely to the sequential co-occurrence patterns found in continuous speech. A large majority of the preceding and following items are part of the NP.

The position preceding the noun most frequently has a member of the class of determiners, including the high frequency articles, possessive pronouns (especially *my*), and quantifiers or demonstratives (a total of 47%). A few high frequency adjectives represent the adjective position in the NP. Finally, some items that are not in the NP made the list: the three prepositions *of*, *in* and *to* and the conjunction *and* with only six occurrences. A closer examination of the preposition + noun examples reveals that they occurred in particular phrases. *Of* occurred primarily with the mass noun *money* and was part of a quantifier phrase, e.g. *lot of* occurred 77 times, *amount(s) of* occurred 40 times, etc. Sinclair (1991:85) argues that in such phrases we do not have a head noun (*lot*) followed by a PP, but rather the second noun is the head of the phrase, and *lot of* or comparable phrases are modifiers to that noun. Under that analysis, which seems realistic for this data, *of* is not really a preposition and the phrase *lot of* is in the same constituent as the noun it modifies.

Table 3. Sets of items preceding the eleven nouns studied

Articles	*the*	134	3,703
	a/an	767	
Possessives	*my*	958	
	your	147	
	our	127	
	his	48	
	their	31	
	's	19	
Other determiners	*that*	137	
	any	53	
	no	76	
	one	11	
Prepositions	*of*	260	592*
	to	178	
	in	154	
Adjectives	*good*	152	248
	new	80	
	whole	16	
Conjunction	*and*	6	6

Table 4. Sets of items following the eleven nouns studied

Prepositions	*to*	120	445
	for	117	
	in	112	
	of	81	
	at	15	
Verbs	*is*	270	444*
	was	79	
	does	32	
	has	32	
	had	31	
Relative clause	*that*	345	409
	I	64	
N/NP Conjunction	*and*	342	342
S/VP-Conjunction	*and*	286	286*
Possessive	*'s*	145	145

The preposition *to* occurs with *school* in 170 out of 178 examples and with *class* in the other eight examples. *In* occurs with *school* in 139 out of 154 cases and with *class* and *computer* (e.g. *computer class)* in the other 15 cases. The usage suggests that *to*/*in school* and *in class* are locative adverbial rather than phrases involving full NPs. After all, it is quite idiosyncratic that these nouns lack a determiner in these expressions.

Even given that the prepositions *in* and *to* pose counter-evidence to my hypothesis, 93% of the items preceding the nouns investigated are in the same NP.

The position following the noun has no one category or item that is as strong as the articles or possessive pronouns in the preceding position. Prepositions, verbs and relative clauses follow the noun with approximately equal frequency. Prepositions are expected in this position and are considered part of the NP. Items that initiate relative clauses are also expected in this position, i.e. *that*. *I* turns out to be common at the beginning of a relative clause (as in *the movie I saw, the class I'm taking*). While the relative clause is in the NP, it also begins a new clause, so that it has a weaker constituent bond with the N than any items that precede the N. The rate of occurrence among these high frequency classes–relative clauses, prepositional phrases and high frequency verbs–is about the same, even though the verbs are not in the same constituent with the NP. This distribution suggests that the constituency bond between the noun and the modifying elements following the noun is weaker than with items preceding the noun. This relation is represented by additional constituency boundaries, as the relative clause would have a new S constituent and the PP itself is a constituent.

The fact that the frequency with which the N is followed by a verb, the first element in the other major constituent of the sentence, is comparable to that of the relative clause and PP, which are part of the NP, seems problematic for the hypothesis that frequency of co-occurrence corresponds to constituency. In particular, the high frequency of *is* would seem to be counter-evidence to this hypothesis. However, *is* and in some cases *has* and *had* are subject to contraction with the subject. Such cases of fusion where frequently co-occurring items are not in the same traditional constituents are the evidence I will use in sections 12 and 13 to argue that sequential co-occurrence is more basic than constituency.

None of the descriptions of the NP consulted before examining the conversational data prepares us for the high frequency with which the conjunction *and* follows a noun. In Table 4 I report a rough breakdown into those instances of *and* that conjoin another noun and those that introduce another clause or verb phrase. This breakdown was based on the item following *and*. Clause-conjunction was assumed if the following item was a conjunction, *then* or *so*, or an adverb. When a pronoun followed that was in the nominative case, the type of conjunction was assumed to be N-conjunction if the preceding noun was animate (*husband and I*) but clause-conjunction if the preceding noun was inanimate (*computer and we*). A following verb indicated verb or VP conjunction. All other cases could be reasonably assumed to be noun-conjunction. In a small number of cases these assumptions may have led to an erroneous assignment, but this breakdown allows us to get an idea of how many of these *and*'s introduce a major constituent break.

Overall the data suggest the predicted correspondence between sequential co-occurrence and traditional notions of constituency within the NP. They predict a stronger bond with the item preceding the noun than with the item following it, and most of the items found to precede or follow the noun are in the same traditional constituent. The primary counter-examples involve preceding prepositions, which seem to be restricted to certain high frequency phrases, and common verbs or auxiliaries following the noun.

In the next two sections, I discuss the cognitive mechanisms that underlie this general correspondence, and then in the following sections I proceed to the argument that the hierarchical structure of language is derivable from the more basic sequential nature of language.

9. Fragments and networks

Language learning in a natural setting involves the storage and processing of repeated fragments of speech. Such fragments might include parts of words,

words, or multiple words. Since human beings are sensitive to recurring sequences of stimuli and record them in memory, they learn to recognize repeated sequences of speech or speech-like stimuli (Saffran *et al.* 1996; Gomez and Gerken 1999). Similarly the automated neuromotor sequences that correspond to these perceptual units are recorded in memory and strengthened with use. Since the number of fragments of speech that must be stored in memory for a mature user of a language is extremely large, a tight organization of these fragments is necessary. I suggest for multi-word sequences, as I have for morphologically complex words, that whole complex units may be stored in memory, but they are typically associated with other units that are similar phonologically and semantically (Bybee 1985, 1995, 1998a). Figure 1 shows a simplified, schematic representation of a possible organization for some of the NPs we discussed above. The connecting lines between words indicate a relation of phonological and semantic identity.[7] Figure 2 shows a set of relations among NPs centering on the determiners rather than the N. The networks in Figures 1 and 2 interlock.

In this model representation is affected by language use. Each token of use of a word or sequence of words strengthens its representation and makes it more easily accessed. In addition, each instance of use further automates and increases the fluency of the sequence, leading to fusion of the units. Thus, as in morphology, high frequency combinations have a stronger representation. For instance, *my mother* (which occurred 182 times in the Switchboard corpus) has a stronger representation than *her house* (which occurred 22 times) or *this hospital* (which

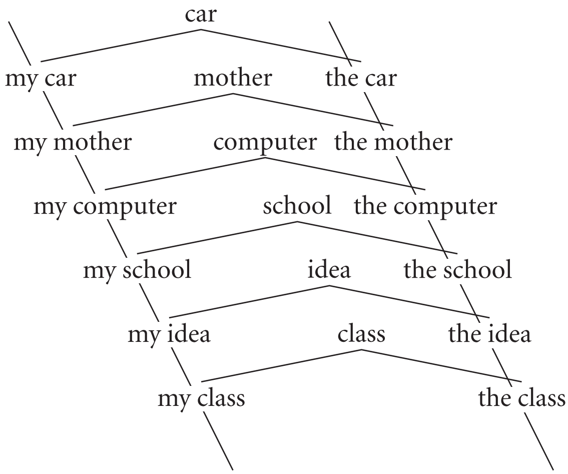


Figure 1. Possible organization of relations among NPs centering on frequently occurring nouns

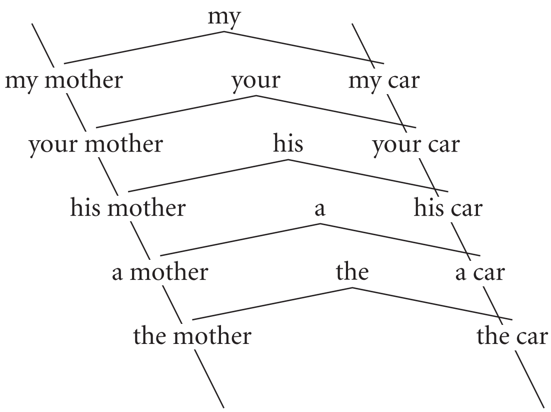


Figure 2. Possible organization of relations among NPs centering on dterminers

occurred only once). All of these NPs are interlocked in a network based on experience with language. From this network emerges the fact that any of these nouns, e.g., *mother, house* or *hospital* can occur with any of the set of determiners. Since particular combinations are also represented, we do not lose the information that some of these nouns occur more often with certain determiners (*my* with kinship terms, or *an* with *idea*) than others.

The hypothesis that very particular sequences of words from our experience are stored in lexical memory is supported by the findings of Erman and Warren 2000 that in the spoken and written texts they examined, 55% of the texts consisted of prefabricated units, that is, sequences of words that are conventionalized. Prefabricated sequences occur more frequently and have more idiomatically determined meaning than sequences that are put together afresh each time. In other words, prefabricated sequences have been experienced before. Pawley and Syder (1983) point out that native-like use of a language depends heavily upon knowing and using conventionalized word combinations.

10. Schemas emerge from the network

The organization apparent in Figures 1 and 2 is only possible if categories are established for strings or parts of strings. A relation between *my car* and *your car* is possible only if a category is established for *car*. Such a category is based on the similarity between tokens of use, which includes information not just about phonology and semantics, but also about contexts of use, both linguistic and non-linguistic. The linguistic context in which *car* is used will include information about what immediately precedes and follows it. Since, as we have seen, the preceding elements are more predictable, that is, the same ones occur more often, a category that includes the frequently occurring preceding element could be formed. The storage of multi-word strings and the categorization of their component elements leads to the formation of constructions. Constructions are schematic generalizations over sequences with shared parts. I agree with Langacker (1987) that schemas are formed at various levels of abstraction. For instance, for NP in English, we might find the following levels of abstraction:

1. Very specific: *my mother, my computer, the car, a problem, an idea*
2. Partially general: [*my* + NOUN], [POSS PRO + *mother*]
3. More general: [POSSESSIVE + NOUN]
4. Fully general: [DETERMINER + NOUN]

These levels of abstraction involve categorization at differing levels of generality. *Mother* is a collection of all the exemplars of this word encountered so far. POSSESSIVE PRONOUN is a level of categorization that ranges over specific words such as *my, your, our, his*, etc. NOUN obviously has an even broader range. Grammatical constructions arise, then, from the storage of frequently repeated sequences and the categorization of their parts at different levels of abstraction.

11. Sequentiality is more basic than hierarchy

In the previous sections I hope to have demonstrated that there is a correspondence between frequency of co-occurrence and traditionally-established notions of constituent structure. Now the question arises as to the directionality of that correspondence: do elements occur together frequently because they are in the same constituents or are constituents derived from clusters of items that occur together frequently? In this section I argue that sequentiality is basic and that constituents and hierarchies arise because frequently-used strings are chunked as single units.

First, observe that learning of sequences of behavior is part of implicit learning in other domains and that such sequences naturally fall into hierarchies with the more frequently-used sequences recombined into larger units to form higher-level constituents (Fentress 1983). Driving a car involves a number of automatic sequences such as shifting into reverse or drive, braking to slow down or stop, putting on a turn signal, turning right or left. On a frequently-driven path, such as from one's home to one's office, these chunks of behavior are sequenced in a particular way and constitute a larger constituent that itself can be automated, as evidenced by the mistake of following this frequent path even when the destination is elsewhere. Also, any of the subunits of the whole sequence, such as braking, accelerating, turning left or right, can be recombined infinitely so that one can drive anywhere one wants to go. For the experienced driver, such recombinations occur with great ease and fluency, just as the native speaker recombines the automated chunks of language to produce a sentence s/he has never produced before. Thus the creation of hierarchy out of automated sequences of behavior is a domain-general cognitive process.

It is by now well-established that Broca's area deals not just with language, but also with sequential neuromotor behavior (Greenfield 1991; Roland 1985). In addition, many researchers have stressed the importance of left-hemisphere dominance for both motor control and language (Armstrong, Stokoe and Wilcox 1995; Corballis 1989 and Kimura 1979, 1993). Thus the hierarchy in automated motor activities and the hierarchy in grammar could stem from the same neurological source. Moreover, recent research into the perceptual processing of predictable visual stimuli suggests that Wernicke's area processes predictable events in time and may not be exclusively associated with language (Bischoff-Grethe *et al.* 2000). Since predictability is the perceptual side of sequentiality, it may turn out that the mechanisms behind the ability to perceive linguistic sequences and perhaps group them into constituents may also be domain-general.

Humans from 12 months to adulthood can learn repeated sequences of meaningless syllables, as shown by Saffran *et al.* 1996; Gomez and Gerken 1999, 2000. Moreover, Gomez has recently shown that both babies and adults can learn sequences of two nonce words that are separated by a third 'word' chosen from a large class (Gomez 2001). Thus meaning is not necessarily involved in learning sequences, suggesting that the basis for constituent structure may be recurring sequences and not just semantics. (See also Santelmann and Jusczyk 1998.)

In addition, there is purely linguistic evidence for the dominance of sequentiality over hierarchy in cases where fusion between elements in different constituents occurs because the two elements occur frequently in sequence. Examples of such cases will be discussed in the next two sections (sections 12 and 13). Linguistic cases of the opposite type, that is, those that appear to show hierarchy dominating sequentiality, as when constituents are discontinuous are discussed in section 14, where it is argued that the linear constraints on discontinuous constituents, such as the Heavy NP Constraint and the Distance Principle, demonstrate that sequential organization underlies even discontinuous constituents.

12. Chunking in violation of 'constituent structure'

Linguistic units chunk together according to repeated sequences, sometimes in violation of usual notions of constituent structure (Bybee and Scheibman 1999). A very robust example is English auxiliary contraction, which occurs in *I'm, I've, I'd,* *I'll, he's, he'll, he'd,* etc. The auxiliary, which is the first element in the VP, contracts with the subject NP, which is in the other major constituent of the clause. Thus contraction occurs across the primary constituent boundary of the clause. The reason for this is that the pronoun + auxiliary or NP + auxiliary combinations are highly frequent. In a large corpus of British English, Krug (1998) finds that contraction occurs in the most frequently-occurring combinations. The most common contraction is between *I* and *am* and that is also the most frequent sequence.

But the auxiliary also occurs right before the verb, so why doesn't the auxiliary fuse with the following verb, where it belongs semantically and syntactically? In order to answer this question I examined the distribution of auxiliaries in spoken American English, using the Switchboard corpus again. In Table 5 we see the token count of the ten most frequent items to precede and follow the auxiliary *will,* in its full and reduced version, *'ll.*

Table 5. Ten most frequent items occurring before and after *will* and *'ll.* (Switchboard corpus)

Preceding		Following	
I	918	*be*	466
they	471	,	244
we	368	*have*	199
it	256	*get*	130
you	200	*go*	119
that	183	*do*	103
he	122	*probably*	90
she	53	*just*	81
,	47	*tell*	75
people	38	.	42

What we see is that, indeed, the most frequent items to precede *will* or *'ll* are pronouns and the most frequent items to follow are verbs. What governs the contraction is the asymmetry in the frequency of the preceding versus the following items. Note that the most frequent pronoun (*I*) preceding *will* is twice as frequent as the most frequent verb (*be*) following *will.* A similar pattern is found for all contracted auxiliaries in the Switchboard corpus. Not only are the fused items in different major constituents, but also they have no semantic relevance to one another. The fusion seems due entirely to frequency of co-occurrence.

Contraction is recorded in Switchboard for all the items listed here as preceding *will* except for *people.* The only other preceding items showing contraction are *there, this* and *who.* In other words, contraction only occurs between the most frequent combinations in the case of *will.* Apparently contraction can generalize from the most frequent items to a general class of items, as demonstrated by the contraction of *has* to *'s*, which appears in the data with a large class of full NPs.

Other details of the distribution of auxiliaries support the Linear Fusion Hypothesis. For all the modal auxiliaries except *can* and *can't*, the most frequent items to follow are *be, have* and the negative.

(i) When *have* is part of the Perfect construction, it contracts with the preceding modal, as in *could've, should've, would've, must've, might've.*
(ii) Of course the negative *not* also contracts in some cases, e.g, *couldn't, shouldn't, wouldn't.*
(iii) Interestingly, *be* also forms a unit with some of these modals, but of a different sort. The combinations *could be* and *maybe* become lexicalized units that serve as independent epistemics, while *would be* has become an adjective.

Thus we have ample evidence from the auxiliaries that high frequency sequences become chunked into units.

This paper is not the first to observe that the combination of subject pronouns or nouns plus the auxiliary in English behaves like a constituent. Halliday and Hasan (1976: 197) call this complex the Modal Element and distinguish it from the Propositional Element (roughly the remainder of the clause). Halliday (1985) regards this collocation of units to be a constituent in the structure imposed on the clause by the interpersonal metafunction where again the Modal Element (subject + finite element of verb group) is distinguished from the Propositional Element. This level of constituent analysis coincides, then, with the frequency of co-occurrence found in conversational data and evidence from contraction that we have just considered.

The distribution of pronouns and auxiliaries and verbs, along with the Linear Fusion Hypothesis, explain why English auxiliaries contract with a preceding pronoun (and in some cases, nouns) rather than becoming prefixes on the verb. If similar distributions occur in other languages it may explain why in languages where the auxiliary follows the verb (SOV languages) there is massive suffixation, while in languages where the auxiliary precedes the verb (VO languages) there is not a comparable trend toward prefixation (Bybee, Pagliuca and Perkins 1990).

13. Other cases of non-constituents fusing.

Another robust phenomenon demonstrable across a wide range of languages is the binding of a verb and preposition into a unit. Reh (1986) discusses this phenomenon in African languages as an explanation for why case affixes are almost always suffixes; that is, why prepositions do not tend to become case prefixes.

Reh (1986) points out that in several languages of the Southern Lwo family, a preposition following a verb becomes a suffix on the verb. This can be seen by examining the dative/benefactive preposition in Dhuluo, which is *nɪ* as shown in examples (4) and (5), where it is positioned between the verb and the NP that is its object.

Dhuluo (Southern Lwo; Western Nilotic)
(4) Otieno o-kelo nɪ Odhiambo kitabu.
O. PERF-bring DAT/BEN O. book
'Otieno has brought a book to Odhiambo.'

(5) Onyangɔ tiyo nɪ japuonj.
O. IMPF:work DAT/BEN teacher
'Onyango works for the teacher.'

In Lango, the cognate preposition and the pronoun that was its object have fused with the verb to form 'the benefactive stem.' Thus *-kèlò* 'bring' + *nɪ* > *-kèllì* 'bring for someone', as shown in (6).

(6) Lango
Ò-kèllì dàkô.
3SG-bring-BEN woman
'She brought it for the woman.'

In this case, then, the former preposition has fused with the preceding verb rather than with the following noun. I propose that the explanation for this is that that particular prepositions would tend to occur with certain verbs, such as verbs meaning 'bring' or 'give', while the noun that follows is much less predictable, presumably being drawn from the entire class of human nouns and perhaps some non-human ones as well.

Analogous situations, but usually without affixation, can be found in European languages. Second-language learners of Spanish and French must learn lists of verb + preposition combinations, as particular verbs select particular prepositions. For instance, Spanish *pensar en* 'to think about', *acabar de* 'to finish' and *comenzar a* 'to begin to'. Again, the verb + preposition sequence would be more frequent than any particular preposition + noun or preposition + infinitive sequence.

Another common fusion across constituent boundaries is the fusion of prepositions and determiners, European languages, e.g. Spanish and French. For instance Spanish: *a* 'to, at' and *el* 'the (MASC.SG) > *al*, *de* 'of, from' + *el* > *del*. In this case, as in the others, it is plausible to assume that the frequent co-occurrence of these grammatical items leads to their fusion. Note that there is no particular semantic relevance (in the sense of Bybee 1985) or semantic affinity between the meaning of a preposition and that of a determiner. This appears to be a case of pure sequentiality.

14. Discontinuous constituents and discontinuous dependencies

A major argument that the utterances of a language are not just linear strings of words is the fact that non-adjacent elements can be in the same construction or constituent. Thus in the English verb + particle combinations, the particle can be separated from its verb by a pronoun or short NP, as in *look the number up; look it up*. However, the separation of the constituents does not necessarily mean that their connection is not still linear or sequential. Other types of neuromotor behavior can be suspended at certain points and then resumed with the association still being sequential. Furthermore, the predictability of the second element from the first can still be maintained across intervening items, as when the phrase *look the number* leads to the expectation of hearing *up*. This is analogous to waiting for the other shoe to drop.

A purely linguistic argument for the importance of sequentiality even in these cases is the well-documented existence of constraints on the material intervening between the two parts of the constituent. The Heavy NP Constraint describes the fact that, for instance, an NP with a lot of modifiers does not do well between the verb and its particle. Chen (1986) shows that the separation of the verb and the particle is constrained by the number of syllables in the direct object. Separation of the verb and particle is practically non-existence in both spoken and written language for direct objects of more than five syllables. Thus (7) and (8) would be very rare or non-existent in discourse and thus have a very awkward feel:

(7) I need to look the number that I lost up.

(8) I need to look a word that I can't remember how to spell up.

Other kinds of discontinuous dependencies also rely on sequentiality. For example, in French certain main clause verbs can have Subjunctive verbs in their subordinate clauses. In Canadian French, the use of Subjunctive or Indicative is variable, and apparently is not meaningful. Poplack (1992) has studied the variables that affect the mood choice and finds that certain high frequency main clause verbs in combination with certain high frequency subordinate clause verbs are more likely to have the Subjunctive.

Examples (9) and (10) illustrate this variability. (*Faut* is a reduction of *il faut que*, which means 'it is necessary'.)

(9) Bien certain, faut qu'ils *aient* (*S*) une place eux-autres aussi pour vivre.
'Well, of course, they should have a place to live, too.'

(10) Faut j'*aille* (*S*) voir pour de l'ouvrage.
'I have to go look for a job.'

I would argue that such dependencies are basically sequential. In Poplack's analysis of Canadian French, it turns out that if a parenthetical expression intervenes between the main verb and the subordinate verb, the Subjunctive verb form is much less likely to appear. In other words, intervening material can disrupt the dependency, suggesting that the use of Subjunctive results from the selection of an automated chunk.

As mentioned above, infants are sensitive to sequential regularities in input even in the absence of semantic factors. Santelmann and Jusczyk (1998) found that 18-month-old infants can discriminate between grammatically correct English discontinuous constituents and ungrammatical ones. The construction in question was the Progressive, in which the finite form of the verb *to be* forms a construction with the suffix *-ing* with an open class of verbs intervening. Thus the infants demonstrated Head Turn Preference for the natural passages such as *Everyone is trying to be careful* vs. the unnatural one, *Everyone can trying to be careful.* When extra words were inserted between the discontinuous constituents, the infants no longer preferred the natural passages, suggesting that this is indeed a sequential effect that can be disrupted by intervening words.

Gomez (2001) also tested 18-month-olds using sequences of three nonce 'words'. The infants learned to discriminate sequences they had heard before even though other 'words' from a large set always came between the first and third word. Since the nonce words are meaningless, this experiment demonstrates that sequential dependencies can be detected and learned in the absence of meaning.

The sequentiality hypothesis makes general predictions about center embedding, i.e. that center embedding would be constrained by the length of the embedded unit. To my knowledge studies of the length of center embedded clauses in natural discourse have not been undertaken.

15. Conclusion

Linguists have been accustomed to viewing language as emanating from a mental structure that is autonomous from actual usage events. A more explanatory view is afforded by recent functionalist views of language as highly contextualized and embodied (Fox 2001). Most tokens of language use are routine events that respond to the environment — both social and physical, i.e., the people encountered and the utterances they produce. These responses are partially automatic, though they do involve an assessment of the environment and the choice of an appropriate response, as do other fine-tuned neuromotor behaviors. As with other neuromotor skills, language responds to practice. Perceptual skill also improves with repetition. Thus we have every reason to believe that repetition could be the main factor that builds up structure in language (Haiman 1994). Its importance in grammaticization has been well documented. Here I suggest that repetition of sequences of units is the main factor in the creation of linguistic patterns that have been identified as constituent structure.

My proposal is as follows. Constituents of the type proposed for generative grammar which are described by phrase structure trees do not exist. Instead, units of language (words or morphemes) are combined into chunks as a result of frequent repetition. Most of the time the units of these chunks bear a semantic and/or pragmatic relation to one another allowing them to fulfill the grammatical criteria for constituency: they can sensibly be used alone, as in the answers to questions; they can be replaced by a pro-form; and they can be used in various positions in a sentence (see the examples in section 2). In such cases, where frequency of co-occurrence corresponds to semantic relevance, we have traditional constituents. Indeed, the semantic coherence of such units may facilitate their establishment as chunks. However, other types of chunks also exist, as I have demonstrated in this paper, showing that frequency of co-occurrence is an independent factor. Thus pronoun + auxiliary, preposition + determiner, and verb + preposition sequences can form chunks but are difficult to describe in traditional frameworks since they do not meet the criterion of semantic relevance. For this reason, too, they do not fulfill the grammatical criteria of occurring alone or being replaceable by a pro-form. Thus constituency in this view is the convergence of two other factors and is itself not a basic structure. It is an emergent property of language.

A second point is that this emergent constituency differs from traditional constituency in that it can be gradient, since the factors determining it are themselves gradient. Gradience in constituency refers to the fact that different items of the same putative category might fuse less with one another. If frequency of co-occurrence is a major determinant of emergent constituency, then the two units in *my mother* are in a tighter constituent bond than the two units in *my appointment.* This difference has no overt consequences in English, but in some languages, it would be manifest as a difference between alienable and inalienable possession, where the latter always has a more fused expression (Nichols 1988). Other gradient differences in frequency of co-occurrence do have overt consequences, as seen in the fact that *I don't* is more fused than *they don't* and *hit'im* is more fused than *hit the ball.* These relations cannot be captured by using the same phrase structure tree for all instances of the same traditional constituent.

Considering now the evolution of language, the development of grammar may be easier to explain in terms of domain-general abilities than many researchers have supposed (e.g. Pinker and Bloom 1990). If constituent structure is epiphenomenal, then a theory of the evolution of language need not account for it directly, but rather by reference to the processes that underlie it and these appear to be domain-general. The abilities include (i) highly advanced motor skills, fine motor control and the associated neurological capabilities including the ability to compose, store and access relatively long sequences of behavior; (ii) the ability to combine concepts into communicatively coherent sequences which in turn is based on (iii) extensive categorization capacity that is applied to both phonological form and meaning; and (iv) the ability to store and categorize vast quantities of prefabricated sequences. These abilities interact in that one may facilitate the others. In particular, the semantic coherence of units in sequence may make it possible to compose longer sequences more fluently. And, as I emphasized throughout, the automation of lower-level sequences makes the composition of hierarchically complex sequences possible. Thus abilities that are neither domain-specific nor species-specific interact in current language processing to create the apparent structure that is grammar. As these abilities evolved from a more primitive to a more advanced stage, language might also have evolved from a set of relatively short utterances consisting of first one, then two units, to much longer utterances with apparent hierarchical structure via the concatenation of preformed chunks.

Notes

1. Langacker (1997), which is a discussion of the iconic relations between conceptual structure and syntactic structure, is an exception.

2. Interestingly, the best examples of syntactic constituents in English seem to be noun phrases. Noun phrases also have the best support as constituents in the discourse literature (Ono and Thompson 1994). Verbs and verb phrases and prepositional phrases present certain problems, as we will see below.

3. On the difficulties of identifying the VP as a constituent, see Givón (1995).

4. In selecting these nouns, there was a conscious bias towards count nouns. In addition, it was immediately obvious that kinship terms, such as *mother* and *husband* had the special property of occurring more with possessive pronouns, and thus only two of them were included. Bleached or grammmaticized nouns such as *thing, couple* and *stuff* were also passed over.

5. The Switchboard transcriptions distinguish several types of pauses, but I have added them all together for the purposes of Table 1.

6. One might expect a higher percentage for *the* as the item to precede a noun. Table 3 shows that there are quite a number of frequently-occurring determiners competing with *the.* In addition, some nouns frequently occur in compounds, so that the item preceding them is another noun. For instance, the item to most frequently precede *school* is *high.*

7. In the model developed in Bybee (1985), semantic and phonological connections can be made independently, but when they coincide a morphological relation is indicated.

References

Anderson, J.R. 1993. *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Armstrong, D., Stokoe W., and Wilcox, S. 1995. *Gesture and the Nature of Language*. Cambridge: Cambridge University Press.

Bischoff-Grethe A., Proper S.M., Mao H., Daniel K.A., and Berns G. 2000. "Conscious and unconscious processing of nonverbal predictability in Wernicke's area". *The Journal of Neuroscience* 20: 1975–81.

Boyland, J.T. 1996. *Morphosyntactic Change in Progress: A Psycholinguistic Approach*. Dissertation: University of California.

Bybee, J.L. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.

Bybee, J.L. 1995. "Regular morphology and the lexicon". *Language and Cognitive Processes* 10: 425–55.

Bybee, J.L. 1998a. "The emergent lexicon". *CLS 34: The Panels*, 421–35. University of Chicago: Chicago Linguistic Society.

Bybee, J.L. 1998b. "A functionalist approach to grammar and its evolution". *Evolution of Communication* 2: 249–78.

Bybee, J.L. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

Bybee, J.L., Pagliuca W., and Perkins, R. 1990. "On the asymmetries in the affixation of grammatical material". In *Studies in Diachronic Typology for Joseph H. Greenberg*, W. Croft, K. Denning, and S. Kemmer (eds), 1–42. Amsterdam: John Benjamins.

Bybee, J.L., Pagliuca W., and Perkins, R. 1994. *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Bybee, J.L. and Scheibman J. 1999. "The effect of usage on degree of constituency: The reduction of *don't* in American English". *Linguistics* 37: 575–96.

Chen, P. 1986. "Discourse and particle movement in English". *Studies in Language* 10 (1): 79–95.

Chomsky, N. and Halle M. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Cipollone, N., Keiser, S.H., and Vasishth S. (eds). 1998. *Language Files: Materials for an Introduction to Language and Linguistics*, 7th edition. Columbus: Ohio State University Press.

Corballis, M.C. 1989. "Laterality and human evolution". *Psychological Review* 96: 492–505.

Erman, B. and Warren B. 2000. "The Idiom Principle and the Open Choice Principle". *Text* 20: 29–62.

Fentress, J.C. 1983. "Hierarchical motor control". In *The Psychobiology of Language*, M. Studdert-Kennedy (ed), 40–61. Cambridge, Mass.: MIT Press.

Fox, B. 2001. "On the embodied nature of grammar: Embodied being-in-the-world". In *Complex Sentences in Grammar and Discourse: Essays in Honor of Sandra A. Thompson*, J.L. Bybee and M. Noonan (eds). Amsterdam: John Benjamins.

Givón, T. 1995. *Functionalism and Grammar*. Amsterdam: John Benjamins.

Givón, T. 1998. "Toward a neurology of language". *Behavioral and Brain Sciences* 21: 154–5.

Givón, T. This volume. "The visual information-processing system as an evolutionary precursor of human language".

Gomez, R. and Gerken, L. 1999. "Artificial grammar learning by 1-year olds leads to specific and abstract knowledge". *Cognition* 70: 109–35.

Gomez, R. and Gerken L. 2000. "Infant artificial language learning and language acquisition". *Trends in Cognitive Sciences* 4: 178–86.

Gomez, R. L. 2001. "Finding structure in language: Sometimes more variability is better." Presentation given at the annual meeting of the AAAS in the symposium: Tools Infants Might Use to Learn Language. San Francisco, February.

Greenfield, P. M. 1991. "Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior". *Behavioral and Brain Sciences* 14: 531–95.

Gregory, M., Raymond, W. D., Bell A., Fosler-Lussier E., and Jurafsky, D. 1999. "The effects of collocational strength and contextual predictability in lexical production". *CLS* 35.

Haiman, J. 1994. "Ritualization and the development of language". In *Perspectives on Grammaticalization*, W. Pagliuca (ed), 3–28. Amsterdam: John Benjamins.

Halliday, M. A.K. 1985. *An Introduction to Functional Grammar.* London: Edward Arnold.

Halliday, M. A.K. and Hasan, R. 1976. *Cohesion in English.* London: Longman.

Hopper, P. and Traugott, E. 1993. *Grammaticalization.* Cambridge: Cambridge University Press.

Jurafsky, D., Bell A., Gregory M. and Raymond, W. D. 2001. "Probabilistic relations between words: Evidence from reduction in lexical production". In *Frequency and the Emergence of Linguistic Structure*, J. L. Bybee and P. Hopper (eds), 229–54. Amsterdam: John Benjamins.

Kahr, J. C. 1976. "The renewal of case morphology: sources and constraints". *Working Papers on Language Universals* 20: 107–51. Stanford, Calif.: Stanford University Press.

Kimura, D. 1979. "Neuromotor mechanisms in the evolution of human communication". In *Neurobiology of Social Communication in Primates*, H. D. Steklis and M. J. Raleigh, (eds), 197–219. San Diego, Calif.: Academic Press.

Kimura, D. 1993. *Neuromotor Mechanisms in Human Communication.* Oxford: Oxford University Press.

Krug, M. 1998. "String frequency: A cognitive motivating factor in coalescence, language processing and linguistic change". *Journal of English Linguistics* 26: 286–320.

Langacker, R. 1987. *Foundations of Cognitive Grammar*, Volume I. *Theoretical Prerequisites.* Stanford: Stanford University Press.

Langacker, R. 1997. "Constituency, dependency and conceptual grouping". *Cognitive Linguistics* 88: 1–32.

Lindblom, B., MacNeilage P., and Studdert-Kennedy, M. 1984. "Self-organizing processes and the explanation of phonological universals". In *Explanations for Language Universals*, B. Butterworth, B. Comrie, and Ö. Dahl, (eds), 181–203. Berlin: Mouton de Gruyter.

Nichols, J. 1988. "On alienable and inalienable possession". In *In Honor of Mary Haas: From the Haas Festival Conference on Native American Linguistics*, W. Shipley (ed), 557–609. Berlin: Mouton de Gruyter.

Ono, Y. and Thompson, S. 1994. "Unattached NPs in English conversation". *BLS* 20.

Pawley, A., and Syder, F.H. 1983. "Two puzzles for linguistic theory: Nativelike selection and nativelike fluency". In *Language and Communication*, J.C. Richards and R.W. Schmidt (eds), 191–226. Longman: London.

Poplack, S. 1992. "The inherent variability of the French subjunctive". In *Theoretical Analyses in Romance Linguistics*, C. Laeufer and T.A. Morgan (eds), 235–63.

Pinker, S. and Bloom, P. 1990. "Natural language and natural selection". *Brain and Behavioral Sciences* 13: 707–84.

Radford, A. 1988. *Transformational Grammar: A First Course*. Cambridge: Cambridge University Press.

Reh, M. 1986. "Where have all the prefixes gone?" *Africanistische Arbeitspapiere* 5: 121–34.

Roland, P.E. 1985. "Cortical organization of voluntary behavior in man". *Human Neurobiology* 4: 155–67.

Saffran, J.R., Alsin, R.N., and Newport, E.L. 1996. "Statistical learning by 8-month-old infants". *Science* 274: 1926–1928.

Santelmann, L. and Juczyk, P.W. 1998. "Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space". *Cognition* 69: 105–34.

Scheibman, J. 2002. *Pont of View and Grammar: Structural patterns of subjectivity in American-English conversation*. Amsterdam: John Benjamins.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sosa, A.V. and MacFarlane, J. To appear. "Evidence for frequency-based constituents in the mental lexicon: collocations involving the word *of*". *Brain and Language*.

CHAPTER 5

The internal structure of the syllable

An ontogenetic perspective on origins

Barbara L. Davis and Peter F. MacNeilage
The University of Texas at Austin

0. Preamble

Neodarwinian evolution through natural selection can provide a basis for explanation of the nature of speech production in modern languages. Within this perspective, speech as a medium of linguistic transmission could have evolved from ancestral non-speech vocal capacities. In the absence of a fossil record, we have argued that the basic open-close mandibular "Frame" in speech had its origin in ingestive mandibular cyclicities (e.g. chewing, sucking, licking). From the perspective of the "frames then content" hypothesis (MacNeilage 1998; MacNeilage and Davis 1990), consonants and vowels, likely emerged from early rhythmic alternations based on elevation and depression movements of the mandible. The mandibular cycle is an oral movement pattern basic to all mammals. In addition, Broca's area in humans, long regarded as a major site for control of speech production, is the main cortical area for control of ingestive cyclicities in mammals as well (MacNeilage 1998). Further, the frame may have had an intermediate stage manifest in lipsmacks (Redican, l975), a visuo-facial communicative gesture common in other primates before it was used for speech.

The earliest protosyllabic vocalizations may have involved co-ordination of two output capacities present in mammals; laryngeal phonation, based on vocal fold oscillation, and articulatorily based syllable-like alternations between open and close mouth configurations. These characteristics are evident in earliest phases of canonical babbling (Oller 1980), in sequences that have syllable-like timing regularities based on jaw open and close cycles perceived as "speech-like" by adult listeners. In this paper, we will explore the hypothesis that syllables in modern languages have had an evolutionary origin that mirrors syllable-like sequences observed in early phases of speech ontogeny. If this claim is justified, then infants are recapitulating phylogeny at the first appearance of speech-like behavior in infancy when they co-ordinate mandibular oscillation with phonation at the onset of canonical babbling.

In contrast to this view of the syllable as based on production factors, a major conception of the syllable in linguistic theory is the "Sonority Hierarchy" (Blevins, l995). Sonority is considered to be a property of syllables whereby the syllable "nucleus" (the vowel) is maximally sonorous, and sonority decreases in "onsets" (intrasyllabic prevocalic consonants) and "codas" (intrasyllabic postvocalic consonants) in a linear relation with distance from the nucleus. Sonority is roughly synonymous with the phonetic concept of "loudness". However the sonority hierarchy, like distinctive features and markedness, is considered to be an *abstract* component of phonological competence (Blevins, l995). Vowels are louder than the consonants in babbling and the two forms alternate. The child phonologist Clartje Levelt's (1994) equivalent to our postulation of the frame in infants is to assert "universal sonority templates take care of the sequencing of segments in the word." (p. 84). However, sonority is a perceptually based abstract concept (Blevins, l995). As such, there is no obvious way that it could underlie the close-open alternation observable at the performance level of production.

1. Earliest phases of ontogeny

Outlining the facts of acquisition is an important first step to consideration of their importance in understanding the origins of syllable production patterns. Although consonants and vowels are probably not controlled as separate entities in babbling (7–12 months) and early speech (12–18 months), most basic facts about these stages have come from phonetic transcription studies using these categories. A number of findings suggest that these two early periods of ontogeny have a number of characteristics in common. (e. g. Oller, Wieman, Doyle and Ross 1975; Stoel-Gammon and Cooper, 1884; Vihman, Macken, Simmons and Miller 1985, Vihman, Ferguson and Elbert 1986). These studies have also shown strong similarities in sound preferences across different communities, suggesting a universal basis for babbling and earliest stages of speech production (e.g. Locke 1983; Roug, Landberg, and Lundburg 1989). Consonantal phones most frequently reported are stops [b], [d] and sometimes [g]; nasals [m], [n], glides [j], [w], and [h] (Davis and MacNeilage 1995; Locke 1983; Stoel-Gammon 1985; Labial (lip) consonants are more frequent in first words (Boysson-Bardies *et al.*, l992); Stoel-Gammon and Cooper 1984; Vihman *et al.* 1986). Dorsal (tongue body) consonants are relatively infrequent. Mid and low front vowels [ɛ], [e], [æ], and the central and low vowels [a], [ʌ], [ə], are most often reported in both transcription studies (Kent and Bauer 1985; Davis and MacNeilage 1990, 1995; Stoel-Gammon 1990) and acoustical studies (Bickley 1983; Buhr 1983; Kent and Murray 1982; Lieberman 1980). In English, this phonetic repertoire accounts for less than 1/2 of the approximately 40 odd phonemes in the language. Most frequent syllable types are CV, (i.e. "ba"), and CVCV (i.e. "baba") shapes (Kent and Bauer 1985; Stoel-Gammon 1985; Vihman 1992). In the time domain, the infant's development of both babbling and speech production patterns is characterized as operating within an overall rhythmic envelope consisting of close-open and movements with resultant alternations between non-resonant and resonant acoustic output (e.g. Bickley, Lindblom and Roug 1986; Koopmans, van Bienum and van de Stelt 1986).

2. "Frames then content" hypothesis

We (MacNeilage and Davis 1990a, 1990b, 1993) have proposed "Frames then Content" as a metaphor to describe spatio-temporal and biomechanical characteristics of babbling and early speech. Rhythmic mandibular oscillation (i.e. the "frame") accompanied by phonation is considered to be the main aspect of the infant movement system available for the initial approximation of the serial organization of adult speech at the onset of canonical babbling. Close and open phases of the cycle may often have no associated neuromuscular activity other than movement of the mandible. The unit at the onset of babbling, is considered to be the rhythmic mandibular cycle which yields simulations of "consonants" in the closure phase, "vowels" in the open phase and "syllables" as the consequence of these alternations between consonants and vowels. No sub-cycle independent movements of component articulators other than the jaw (i.e. tongue, lips, or soft palate) are typical when the infant produces early speech-like sequences.

With development, segmental "content" elements (consonants or vowels) are gradually differentiated from the "frame" to become separate entities as the infant acquires increasing control over the coordination of articulators in vocal sequences. As the tongue becomes relatively more independent of the jaw, the infant initiates place variegation (i.e. moving the tongue within consonant sequences) thus reducing syllable reduplication ("baba" becomes "badi" for "bottle"). Control over soft palate closure is evidenced by growth in alternation of nasals and orals rather than pervasive nasal or oral quality throughout sequences. Most importantly, action in the time domain is a basic tenet. In this view, the task facing the infant in ontogeny is to gain independent control of coordinated articulators in time, thus addressing for speech the general problem of serial order proposed originally by Lashley (1951): "How is any action organized serially in time?"

The initial stages in ontogeny during babbling and first words do not include differentiation of independent segmental content elements. In this paper, we will be focusing on this phase to explore the potential of our perspective for impacting understanding of phylogeny of speech production.

3. Intrasyllabic properties in ontogeny

A study of a single infant, 14 to 20 months of age (Davis and MacNeilage 1990) revealed particular co-occurrence patterns in contiguous consonants and vowels, based on mechanical characteristics of jaw movement and relative lack of lingual independence from the jaw. We termed this patterning "frame dominance" to refer to the over-riding influence of mandibular oscillations in the absence of other articulatory movement in explaining the patterns we observed. In this study, a large corpus of 13,500 vocalizations was analyzed, based on phonetic transcription of spontaneous vocalizations occurring in the infant's natural environment. For analysis of intrasyllabic patterns in her vocal output, consonants were grouped into labial, coronal and dorsal places of articulation; vowels into front, central and back categories. Chi square analysis was used to compare the observed frequencies with the expected frequencies of occurrence within each cell. Expected frequencies were derived from the overall frequency of that vowel in the corpus (i.e. if 40% of the vowels in the corpus were front vowels, then 40% of the vowels in each consonant environment were *expected* to be front vowels).

Two of the intrasyllabic patterns we observed were seen as a consequence of a non-resting tongue position within the close-open cycle: Coronal consonants such as [d] and [n] which involve tongue front closure tended to co-occur with front vowels such as [ɛ] and [ae] (e.g. "dad"). Dorsal consonants, such as [k] and [g] which involve a tongue body closure in the back of the mouth, and back vowels also tended to co-occur (e.g. "gogo"). We called these two patterns *Fronted* and *Backed* Frames respectively, and suggested that they might be achieved primarily by placing the tongue in the front or back of the mouth perhaps *before the utterance begins.* Consequently the tongue may play only a negligible active role in the generation of these patterns.

A third pattern found was a tendency for labial consonants ([b], [m]) to co-occur with central vowels ([a], and the two central vowels [ʌ], [ə] in English, e.g. "mama"). There is no reason to believe that the tongue makes an independent movement, in addition to depression of the mandible, in the production of a centralized vowel. Consequently, in infants at least, the central location of the tongue in these forms is presumably its resting position. In the absence of any mechanical reason for the tongue to be in the center of the mouth when the consonantal closure is at the lips, we called this pattern *Pure*

4. Frames

No lip rounding is predicted in the labial context for either the consonant or the vowel, as rounding would add an active lip movement to the mandibular cycle, suggesting that the lips are actively able to operate in the cycle along with the mandible. Munhall and Jones (1998) have shown in one infant that independent movement of the lips does not accompany babbling with labial consonants. Instead, contact between the upper and lower lips is simply achieved by elevation of the mandible.

The three predicted consonant-vowel (CV) co-occurrence preferences were found at significant levels, in contrast to other potential consonant-vowel co-occurrences. From a biomechanical perspective, these CV co-occurrences observed within syllables seem straightforward. The two lingual patterns (coronal-front and dorsal-back) may result from an extremely basic property of matter — the property of inertia. The tongue has a tendency to remain in a similar position in the mouth within a syllable rather than moving from one extreme position to another. In addition, like the tongue, the soft palate tends to not change position during early utterances making the entire babbling episode nasalized if the soft palate is at rest in its open position [mãmãmã] and non-nasal if the soft palate is elevated throughout the episode [bababa] (Matyear, MacNeilage and Davis, l998). Thus, the patterns observed for labial and coronal consonants tend to be similar whether the consonants involved are oral ([b], [d]) or nasal [m], [n]), except for some nasal resonance effects on vowels in nasal contexts. These forms in which the soft palate is at rest we have termed *Nasalized* Frames. As these various consonant and vowel co-occurrence patterns seemed to represent basic aspects of the vocal production system, and did not seem to be a result of perceptual experience (see Davis and MacNeilage 2000 for a discussion), we predicted that they would also be observable in babbling.

A subsequent study of one infant during canonical babbling 7–12 months (Davis and MacNeilage 1994) revealed the presence of the labial-central and coronal-front associations, although not dorsal-back associations (too few dorsals were produced to analyze). Davis and MacNeilage (1995) subsequently studied 6 normally developing infants during babbling. The three predicted intra-cyclic CV associations, involving stops, nasals, and glides, were confirmed at significant levels in all infants (18 total predictions for 6 infants). Only 9 out of 36 above chance associations were found in instances in which CV associations were not predicted. Gildersleeve-Neumann, Davis and MacNeilage (2000) also considered consonants with lower frequency of occurrence (fricatives, liquids, and affricates), which are often very late to appear in typical speech development, in four of the infants previously studied in babbling. With minor exceptions, CV co-occurrence patterns similar to the patterns for stops, nasals, and glides were found (i.e. coronal consonants with front vowels and labial consonants with central vowels).

Most recently, in a large scale study of 10 infants during the period of acquisition of single words), 4 of whom were studied during babbling (Davis, MacNeilage and Matyear, in press), we found the three predicted CV co-occurrence trends in 25 out of 28 instances, while non-predicted trends only occurred in 17 out of 58 cases (Chi square, $p < .01$). These results indicate that the intra-syllabic co-occurrence trends found in babbling do not appear to resolve during the single word lexical acquisition period in typically developing infants in an English environment.

Transcription studies. Most studies of this early period have used phonetic transcription methodologies employing IPA (International Phonetic Alphabet) symbols. These methodologies have the advantage of using the same types of descriptors for capturing vocal production patterns in earliest ontogeny for comparison with mature language characteristics. The drawback of transcription methods is difficulty with reliability of transcription in early phases of speech acquisition as well as the implication that use of consonant and vowel symbols implies independent *control* of segments employed to represent sequential patterns in infants. These studies (e.g. Boysson-Bardies 1993; Oller and Steffans 1993; Vihman 1992) have, for the most part, confirmed the existence of the three CV co-occurrence patterns in both babbling and early words. However, studies of both periods have, for the most part, been based on relatively small databases per infant or on diary studies in some instances (e.g. Ingram 1981). In addition, pooling of data across subjects masks variability, thus making it difficult to assess the generality of trends noted. One additional difficulty in comparing data across studies is lack of information regarding precise definition of consonant place and vowel height and front-back categories used. Consistent use of transcription-based categories would help to evaluate the frame/content hypothesis more consistently. Some of the discrepancies between results of various studies appear due to these methodological factors.

Small corpora have been typically analyzed A basic aspect of our method has been to investigate production patterns through analysis of large databases (more that 1,500 tokens) based on longitudinal data collection of spontaneous samples. Subsequent computer analysis for intrasyllabic trends is thus based on very large samples rather than on anecdotal reports that may not represent the overall pattern of output during the period of interest.

Vihman (1992) studied co-occurrence patterns in a group of 23 children in four different language communities in the period of the first 50 words. Labial central vowel and velar back vowel co-occurrence constraints were apparent, but the claim of a coronal front vowel association was not strongly confirmed. However, Vihman's data analysis included [ae] (as in "bat") as a central vowel, complicating comparison with our results on coronal-front vowel associations. [ae] is more frequently considered to be a front vowel and thus would indicate a coronal-front co-occurrence when it appeared. The source of negative association for coronals and front vowels was three English-speaking and two Swedish-speaking subjects. Vihman noted the strong role played by the individual child at this developmental stage of first word use, suggesting that lexical use might predominate over biological or physiological factors in contrast to the situation at the onset of canonical babbling.

Boysson-Bardies (1993) studied groups of five 10–12 month old infants from four different language communities (French, English, Swedish and Yoruba). She found predicted CV patterns to be influenced by characteristics of the target language (calculated “on the basis of words that 18 month old infants attempt in each language group”, p. 358). Labial-central vowel association in initial syllables was found for French, Swedish and Yoruba infants. American infants showed an association between labials and front vowels. A favored association between dentals and front vowels was found in English, Swedish, and French; between dentals and central vowels in Yoruba. Where infants differed from predicted patterns, Boysson-Bardies suggested influence of the ambient language as being characteristic of the resulting patterns. However, her definition of ambient language was based on targets attempted by the infants giving a potentially skewed estimate of overall phonological characteristics of the target language. A more appropriate estimate of target language properties could come from analyses of spoken discourse or of word patterns shown in dictionaries.

Oller and Steffans (1993) noted some association of consonants and vowels within syllables in a study of four children 10–12 months. Coronal consonants were more frequently associated with front vowels, dorsals with back vowels. Coronals showed the greatest association with high vowels, labials with low vowels. By 16–24 months, CV associations in the subjects had weakened, consistent with systematic development toward more independent segmental function, according to the authors.

Acoustic studies. Acoustic studies have also provided support for predominance of mandibular over lingual movement in early syllable-like vocal sequences. Bickley (1983) studied vowel development in 14 infants at 3-month intervals between 14 and 22 months. She found evidence for the development of vowel height before vowel front-back control and interpreted her results as a demonstration that infant control over degree of mouth opening (first formant, F1) precedes the ability to control anterior-posterior positioning of the tongue body (second formant, F2). Hodge (1989) studied 7;5–9;5 month old infants' spontaneous vocalizations, and imitated CV syllable productions in 3, 5, 9 year olds and in adult speakers. She proposed that patterns of displacement for first and second formants suggest younger children are accomplishing opening and closing gestures with greater relative contribution of mandibular than lingual movement, while older children and adults are using a greater relative contribution of tongue versus jaw movement. Specifically, she found less movement of F2 (tongue based) trajectories in younger age groups, suggesting a relatively smaller amount of change in place of major constriction in the vocal tract on the front-back dimension.

In a study of coarticulation in children 3 and 7 years old, Nittrouer, Studdert-Kennedy and McGowan (1989), found greater differences in the second formant within the fricative noise depending on the quality of the upcoming vowel in children's syllables than in adult syllables. In a subsequent study (Nittrouer 1999), spectral analysis showed that children 3, 5, and 7 years of age moved away from the consonant closure more slowly and initiated the vowel gesture sooner than did adults for stop-vowel sequences. Vocal tract close and open phases achieved adult-like patterns of movement by 3; tongue gestures were constrained by phonetic context until the age of 7. Nittrouer suggested that the second formant patterns might be envisioned as being flatter, broader, and as having more shared areas in children's than in adult syllable productions, reflecting less autonomy and more temporal coarticulation in the front-back dimension of tongue placement in children than in adults.

5. Evidence from infants with hearing impairment

Studies of hearing-impaired populations show that normal auditory perceptual input is an important causal factor in the production of frames. Infants with profound or even severe hearing loss sometimes do not babble (e.g. Oller and Eilers 1988; Oller, Eilers, Bull and Carney, l985). If babbling does occur, onset is often months later than normal. Output tends to be sparser, as well as abnormal in quality (Oller and Eilers, l988, Oller *et al.*, l985; Stoel-Gammon, l998; Stoel-Gammon and Otomo, l986). While output is like that of hearing infants, in having very few non-occlusive consonants, few dorsal consonants, and few high and back vowels, it is different in three specific ways. Hearing-impaired infants produce many more sounds with a nasal quality relative to hearing infants in whom oral quality is predominant (e.g. Lach, *et al.*, l970; Stoel-Gammon, l988). They produce a much higher frequency of consonants with a labial rather than coronal place of articulation, while hearing infants tend to do the reverse (e.g. Stoel-Gammon and Kehoe, l994; Yoshinago-Itano *et al.*, l992). They also produce mostly neutralized vowels with relatively few of the front vowels common in hearing infants (e.g. Apuzzo and Yoshinago-Itano 1998; Tobey, *et al.*, l994; Tye-Murray and Kirk, l993; Yoshinago-Itano *et al.* l992). Thus there is a marked skewing in the distribution of sounds produced by this population, which has been attributed to a different balance between visual and auditory information, heavily favoring the visual modality.

To explore our expectations of the role of CV co-occurrence constraints and consonant nasality effects on vowels, we studied one infant who received a multi-channel cochlear implant (MCI) at 25 months (McCaffrey, Davis, MacNeilage and von Hapsburg 2000). In our study, pre-implant data was derived from phonetic transcriptions of two 1-hour recording sessions 3 months and 1 month before implant, and from three 1-hour recording sessions 2, 7 and 9 months post-implant. Pre-implant the infant was typical of HI infants in producing relatively few syllabic utterances, mostly labial-nasal stop consonants and neutralized vowels in CV forms. She produced only 62 CV forms. There were too few coronals and front vowels and dorsals and back vowels to allow a comprehensive statistical analysis of CV co-occurrence patterns.

However, both labial-central and coronal-front CV co-occurrences were plentiful post implant. Combining 7 and 9-month post-implant data, the observed-to-expected ratio of labial-central CVs was 1.14 and the ratio for coronal-front CV's was 1.62. The coronal-front value was much higher than those observed in hearing infants or in languages. Both effects were significant in separate tests at 7 and 9 months.

There was marked improvement in the infant's output 9 months post implant. Nasal consonants reduced dramatically from almost 90% pre-implant (liquids and fricatives remained rare.) While 60% of vowels were mid-central pre-implant, post-implant figures were about 15%. Hearing infants do not produce a high proportion of nasal consonants at the onset of babbling. They have simulated an oral production mode for speech production, even though the oral production mode requires an active muscle contraction rather than the open velopharyngeal port typical for breathing. This hearing impaired infant only began to utilize this perceptual information regarding oral versus nasal mode of production after approximately 7 months of auditory exposure. In this regard, the elevated soft palate position has often been designated as allowing more contrast between sounds (e.g. Lieberman, l984) implicating an adaptive natural selection process favored for increase in message complexity. By the onset of babbling, hearing infants produce approximately 75–80% oral consonants (Davis and MacNeilage 1995; Locke 1983), simulating characteristic properties of adult languages (Lindblom and Maddieson 1994).

These results on a young CI infant confirm the robust nature of the labial-central and coronal-front CV co-occurrence patterns. When the infant had enough syllable type vocalizations post-implant for analysis, she produced the predicted intrasyllabic preferences we have identified based on characteristics of the production system found in typically developing infants.

6. Targets attempted by English-learning infants

We also investigated whether there were relationships between infant phonetic patterns and patterns in the early lexical targets they are attempting (Davis, *et al.* 1999). Targets attempted by eight English learning infants already studied in our lab who were in the single word period were analyzed. In addition, to gain a more

Table 1. Observed to expected ratios for the *CDI* and for lexical targets attempted by 8 infants

	Vowels	Consonants		
		Coronal	*Labial*	*Dorsal*
CDI	Front	**.97**	.33	.98
	Central	.99	**1.02**	1.02
	Back	1.09	.85	**1.01**
Infant Targets	Front	**1.09**	.97	.93
	Central	.96	**1.09**	.91
	Back	.85	1.09	**1.38**

general impression of the characteristics of targets attempted by English learning infants, lexical items on the *MacArthur Communicative Development Inventory* (*CDI*, Fenson *et al.* 1993), were also analyzed. The results (shown in Table 1) for the two infant databases show ratios above chance, except for the coronal-front (CF) association in the *CDI* database. However, although the trends are in the predicted direction for the eight of the nine potential intrasyllabic associations of interest, none achieved statistical significance using a chi square analysis. Non-predicted trends for intrasyllabic associations were found for labial-back (LB) vowels in the infant target data, coronal-back vowels (CB) and dorsal-central (DC) vowels in the *CDI* data. None of these were significant, although the ratios were above chance. No other counter trends were noted. These results indicate a general trend toward predicted intrasyllabic associations. However associations are not as strong in the targets being attempted as in these infant's actual productions during early words. This type of relationship between phonetic patterns and targets attempted indicates that the infants are not clearly *learning* from the environment, as the intrasyllabic associations are stronger in their productions than in the targets infants attempted or in the targets that are frequent in first words in English evidenced on the *CDI.*

7. Cross-language results

A second source of evidence of the generality of these intrasyllabic patterns can be found in studies considering these trends in infants in language environments other than English. Table 2 displays results of our survey of several available studies.

We re-analyzed the Japanese and Swedish data from the Stanford Child Phonology Project provided courtesy of M. Vihman. This data, first analyzed in Vihman (1992), is based on 4 data collection points during the earliest stages of

Table 2. Mean observer-to-expected ratios for the three predicted consonant-vowel co-occurrence patterns in infants in Swedish, Japanese, Ecuadorian-Quichua, Brazilian-Portuguese, Italian, Serbian, and Korean environments

	Swedish 5			Japanese 5			Quichua 7			Braz. Port. 2			Italian 1			Serbian 2			Korean 5		
	C	L	D	C	L	D	C	L	D	C	L	D	C	L	D	C	L	D	C	L	D
F	1.2			1.2			2.6			1.0			1.3			1.23			1.7		
C		1.2			1.2			1.6			1.2			1.4			1.3			1.3	
B			1.4			1.1			2.1			1.8			1.0			NA			2.0

the first word period. Native language transcribers in each language group originally transcribed the data. We entered the original transcriptions into our computer database for analysis of intrasyllabic trends. All predicted associations were present at significant levels for both languages. No consistent counter trends were present for any infant or for the group overall.

A study of 7 infants in an Ecuadorian-Quichua environment, (Gildersleeve-Neumann, unpublished dissertation, 2001) also showed the three predicted patterns at significant levels for all seven infants in both babbling and in first words. Transcription was completed by a transcriber with near-native fluency in the Quichua language. No non-predicted associations occurred at above chance level for any individual infant.

Two infants in a Brazilian-Portuguese environment were studied during the single word period (Teixeira and Davis, in press). Data were recorded and transcribed in a diary format by the infant's mother, a linguist and analyzed for intrasyllabic trends. All three predicted associations were found for both infants during the period in which they were producing single words. No consistent counter trends to those predicted were found for either infant.

One Italian infant, transcribed by an Italian speaker (Zmarich, and Lanni, l999), showed predicted associations in both babbling (10–12 months) and in first words (14–16 months) for labial-central vowel and coronal-front vowel associations. Intrasyllabic associations in first words were significant. During babbling they were above chance, but not significant. The dorsal-back vowel association was not noted in either babbling or first words in this infant, although trends for dorsal-back associations approached significance.

Bilingual twins were studied during the babbling period between 7 and 12 months of age (Zlatic *et al.* 1997). The infants were being reared in a Serbian-English environment with an English speaking father and a Serbian-English bilingual mother. The infant's mother, who is a linguist, transcribed canonical vocalizations. Significant CV associations for labial-central and coronal-front vowels were found for one of the two infants. Only the coronal-front effect was found for the second infant. Dorsals were not analyzed for either infant, as they occurred with very low frequency in both.

A study of canonical babbling vocalization in five infants in a Korean ambient language environment (Lee 2000) also showed the predicted effects at significant levels for all three types of associations. At an individual level, only one infant showed a lack of predicted association; the dorsal-back. Non-predicted effects were shown in 6 of 30 possible cells, no consistent patterns of non-predicted associations were observed.

Thus in 27 infants across 7 languages, CV associations were broadly confirmed with few exceptions and no consistent counter trends. Confirmation of the predicted associations in the majority of infants studied supports the predictions of mandibular "frame dominance" (MacNeilage and Davis 1990, 1993) as a general characterization of intra-cyclical properties of early vocalizations. Non-confirmations suggest the need for more attention to patterns produced by individual infants, including their tendencies to focus on particular lexical items during the first word period.

8. Languages

The possibility that patterns of babbling and early word acquisition are important to understanding of language origins is increased by the persistence of some babbling and early word preferences in modern languages. The CV syllable, the favored syllable type in babbling and early words is considered to be the only universal syllable type in languages (Bell and Hooper 1984). Consonants favored in babbling — simple stop consonants and nasals — are highly frequent languages as well (Maddieson 1984) and tend to dominate the repertoire of languages with small systems (< 15 phonemes) containing a few segments characterized as articulatorily "simple" (Lindblom and Maddieson 1988; Lindblom, Krull and Stark 1993).

For analysis of language patterns, dictionary databases were employed. In every instance, number of entries exceeded 1,000. Consonant vowel sequences in CV, CVC, and CVCV words were analyzed as well as initial consonant vowel consonant vowel strings that were a part of longer lexical items (CVCV . . .).

Our initial analyses (MacNeilage and Davis 1993) showed evidence of common tongue tongue-fronting and tongue-backing but not of the co-occurrence constraint involving labial consonants and central vowels in two sets of data containing five widely varying languages in each that were originally described by Janson (1968) and Maddieson and Precoda (1990). A further study including all English consonants (Davis, MacNeilage and Matyear 1999) based on analysis of 33,654 lexical types in the *Shorter Oxford English Dictionary* (SOED; Quinlan 1992) found the coronal-front and dorsal-back, but not the labial-central associations present at significant levels. MacNeilage *et al.* (2000) also found all three intrasyllabic properties to be largely present in an analysis of a group of 10 languages, suggesting that they are basic to speech in general, rather than simply being transient aspects of early speech development.

These results raise the possibility that the tongue-based constraints postulated for babbling and early speech may have an extremely fundamental status. The coronal-front and dorsal-back CV co-occurrence patterns involve the same mandibular oscillation pattern as the labial-central frames in infants in our view. A non-resting tongue position is characteristic in these two patterns. Consonants and adjacent vowels tend to share a similar tongue position in the front-back axis in coronal and dorsal CV forms, presumably reflecting an extremely basic aspect of movement systems in general — the presence of inertia in movement sequences. In principle the lack of lingual engagement leaves the vowel potentially freer to vary during the vocalic portion of the syllable to enhance diversification and some languages have taken advantage of this potential. Thus, the labial-central association is present in some languages, not others and tends to be somewhat weaker than the lingual associations.

While these inertial possibilities arise from a production perspective, another possibility has a perceptual rationale. Kawasaki (1982 — see also Ohala and Kawasaki l989) has found, in a count of 200 languages, that "there is a strong tendency in languages to *avoid* sequences of the sort (1) labial(ized) consonants + [w] and high back rounded vowels (and) (2) apical or palatal(ized) consonants + [j] and high front vowels." (Ohala and Kawasaki 1989: 121). Examples derived from their description include [bu], [wu], [di], and [ji]. Their explanation for these patterns is that "The essence of any communication channel is the production of modulations, i.e. differences in some carrier signal. This is true whether sign language, semaphore or speech is involved. The more difference created in the stimulus parameters in passing from one cipher in the code to the next, the better; the smaller the difference, the more likely it is that the sequence will not be detected (accurately or at all). Thus the problem of sequences such as [wu] and [ji] is that they create minimal modulations in amplitude, periodicity and spectrum. With [wi] and [ju], in contrast, at least the second and third formants show sufficient variation." (Ohala and Kawasaki 1989: 123).

In short, Ohala and Kawasaki are raising the possibility that in languages, at least for the coronal-front pairing, the need for phonotactic perceptual distinctiveness (PPD) overrides any possible inertial tendency for the tongue to remain high for the consonant and the following vowel in these cases. At first sight, the production-based inertial view and the perceptually based PPD view lead to opposite predictions in some regards-inertia is associated with a relative lack of change in formant frequencies, but PPD requires more than minimal change. However there is an alternative possibility. It is that the inertial view might be more appropriate to infants and the perceptually based view might be more appropriate to languages. That is, in the history of languages, the need for conveying large numbers of perceptual distinctions at high rates might lead to the demand that the modern mature producer of speech overcome some inertial forces in order to successfully transmit phonotactic distinctions.

In an attempt to explore the phlogenetic implications of these CV co-occurrences, we have analyzed a 27-word protoword corpus constructed by Bengston and Ruhlen (l994 (MacNeilage and Davis 2000a) to look for the presence of these patterns. The presence of these intrasyllabic regularities in first words, in languages, and in the protoword corpus, provides important evidence of the relevance of speech acquisition to speech evolution. In particular, our finding that labial consonants tend to co-occur with central vowels in babbling, first words, words of some modern languages and in the protoword corpus has profound implications for the nature of speech. We take it as evidence for our contention that the motor frame, produced by mandibular oscillation played a key role in the origin of speech. The existence of these preferred forms in modern languages and in the protoword corpus is further evidence for their basic nature. We have termed these intrasyllabic properties a "Motor Core" to describe basic characteristics of output patterning in infants and languages (MacNeilage, Davis, Kinney and Matyear 2000).

9. Summary

Speech ontogeny entails an initial frame stage involving basic capabilities available to the production system, followed by a frame/content stage, marked by increase in the sound inventory and in mastery of serial complexity (MacNeilage and Davis 1993). Our postulation of a parallel frame stage for the phylogenetic origins of speech is suggested by evidence that some of the main features of the intrasyllabic frame stage in infants are also present in syllables in modern languages and in a protoword corpus (MacNeilage *et al.* 2000, MacNeilage and Davis 2000). Characteristic intrasyllabic frame patterns found in infants are not clearly learned from the environment as they are typically stronger that the associations found in languages and they are present in some instances when they are not present in the ambient language (i.e. in the case of labial-central associations that are not consistently present in languages). In addition, they do not appear to be strongly present in the lexical targets attempted in early words based on analysis of actual targets attempted as well as analysis of the CDI, representing targets generally attempted by infants in this stage of lexical acquisition. These intrasyllabic patterns may a suggest a route to earliest syllable-like organization of serial output in early hominids; initiating vocal communication based on the jaw cycle and accompanying phonation co-opted from earlier functional routines. Maintenance of this characteristic of early ontogeny in mature language users indicates that it may be extremely basic to the operation of the movement control system and based on inertial constraints on independent movement of articulators within a syllabic sequence. In our view, evolution may likely have preserved functional aspects of the internal structure of the jaw cycle (i.e. intrasyllabic regularities) while elaborating over time with additional adaptations seen in increase of independent articulator movements within syllables in the service of increase in message complexity.

References

Apuzzo, M.-R.L. and Yoshinago-Itano, C. 1998. "The development of deaf and hard of hearing children identified early through the high risk registry". *American Annals of the Deaf* 143: 416–24.

Bell, A. and Hooper, J.B. 1978. *Syllables and Segments*. Amsterdam: North Holland.

Bengtson, J.D. and Ruhlen, M. 1994. "Global etymologies". In *On the origin of languages*, M. Ruhlen (ed). Stanford: Stanford University Press.

Bickley, C. 1983. "Acoustic evidence for phonological development of vowels in young infants". *Proceedings of the 10th International Congress of Phonetic Sciences*, Utrecht.

Bickley, C., Lindblom, B., and Roug, L. 1986. "Acoustic measures of rhythm in infant's babbling or 'all God's infants got rhythm'". *Proceedings of the 12th International Congress of the Phonetic Sciences*, Toronto.

Blevins, J. 1995. "The syllable in phonological theory". In *The Handbook of Phonological Theory*, J. Goldsmith (ed), 206–44. Oxford: Blackwell.

Boysson-Bardies, B., Vihman, M.M., Roug-Hellichius, L., Durand, C., Landberg, I., and Arao, F. 1992. "Material evidence of infant selection from the target language: A cross-linguistic study". In *Phonological Development: Models, Research, Implications*, C. Ferguson, L. Menn, and C. Stoel-Gammon (eds). Timonium, Md.: York Press.

Boysson-Bardies, B. 1993. "Ontogeny of language-specific syllable production". In *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, B. Boysson-Bardies, S. de Shonen, P. Jusczyk, P. MacNeilage, and J. Morton (eds). Dordrecht: Kluwer.

Boysson-Bardies B. and Vihman, M.M. 1991. "Adaptation to language: Evidence from babbling and first words in four languages". *Language* 67: 297–319.

Buhr, R.D. 1983. "The emergence of vowels in an infant". *Journal of Speech and Hearing Research* 23: 73–94.

Davis, B.L., MacNeilage P.F., and Matyear, C. In press. "Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production". *Phonetica*.

Davis, B.L. and MacNeilage, P.F. 2000. "An embodiment perspective on the acquisition of speech perception". Special issue of *Phonetica* 57: 229–41.

Davis, B.L., MacNeilage, P.F., Matyear, C. 1999. “Intrasyllabic patterns in babbling and early speech”. *Proceedings of the 14th International Congress of the Phonetic Sciences*, San Francisco.

Davis, B.L. and MacNeilage, P.F. 1995. “The articulatory basis of babbling”. *Journal of Speech and Hearing Research* 38: 1199–211.

Davis, B.L. and MacNeilage, P.F. 1994. “Organization of babbling: A case study”. *Language and Speech* 37: 341–55.

Davis, B.L. and MacNeilage, P.F. 1990. “The acquisition of vowels: a case study”. *Journal of Speech and Hearing Research* 33: 16–27.

Donald, M. 1994. *Origins of the Modern Mind*. Cambridge, Mass.: Harvard University Press.

Fenson, L., Dale, P., Reznick, D., Thal, D., Bates, E., Hartung, J., Pethick, S., and Reilley, J. 1992. *MacArthur Communicative Development Inventory*. San Diego, Calif.: Singular Press.

Gildersleeve-Neumann, C.E., Davis, B.L., and MacNeilage, P.F. 2000. “Contingencies governing production of for fricatives, affricates and liquids in babbling”. *Applied Psycholinguistics* 21: 341–63.

Gildersleeve-Neumann, C.E. 2000. *Production vs ambient language influences on speech development in Quichua*. Dissertation,The University of Texas.

Hodge, M.M. 1989. *A Comparison of Spectral Temporal Measures Across Speaker Age: Implications for an Acoustical Characterization of Speech Acquisition*. Ph.D. dissertation, University of Wisconsin-Madison.

Ingram, D. 1981. “The emerging phonological system of an Italian-English bilingual infant”. *Journal of Italian Linguistics* 6: 95–114.

Janson, T. 1986. “Cross-linguistic trends in the frequency of CV sequences”. *Phonological Yearbook* 3: 179–95.

Kawasaki, H. 1982. *Acoustic Basis for Universal Constraints on Sound Sequences*. Ph.D. dissertation, University of California.

Kent, R.D. and Bauer, H.R. 1985. “Vocalizations of one-year-olds”. *Journal of Infant Language* 12: 491- 526.

Koopmans-van Beinum, F. and Van der Steldt, J. 1986. “Early stages in the development of speech movements”. In *Precursors of Early Speech*, B. Lindblom and P. Zetterstrom (eds), 37–49. New York: Stockton Press.

Kent, R.D. and Murray, A.D. 1982. “Acoustic features of infant vocalic utterances at 3, 6, and 9 months”. *Journal of the Acoustical Society of America* 72: 353–65.

Lach, R. Ling, D., Ling, A., and Ship, H. 1970. “Early development in deaf infants”. *American Annals of the Deaf* 115: 522–6.

Lashley, K.L. 1951. “The problem of serial order in behavior”. In *Cerebral Mechanisms in Behavior: The Hixon Symposium*, L. Jeffress (ed). New York: Wiley.

Lee, S. 2000. Perceptual influences on speech production in Korean learning infant babbling. Ms., The University of Texas.

Levelt, C. 1994. *On the Acquisition of Place*. Ph.D. dissertation, State University of Leiden.

Lieberman, P. 1980. “On the development of vowel productions in young infants”. In *Infant Phonology,* Volume I: *Production*, G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (eds), 23–42. New York: Academic Press.

Lindblom, B., Krull, D., and Stark, J. 1993. "Phonetic systems and phonological development". In *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, B. Boysson-Bardies, S. de Shonen, P. Jusczyk, P. MacNeilage, and J. Morton (eds)., 399–409. Dordrecht: Kluwer.

Lindblom, B. 1992. "Phonological units as adaptive emergents of lexical development". In *Phonological Development*, C. Ferguson, L. Menn, and C. Stoel-Gammon (eds), 131–63. Parkton: York Press.

Lindblom, B. and Maddieson, I. 1988. "Phonetic universals in consonant systems". In *Language, Speech and Mind: Studies in Honor of Victoria A. Fromkin*, L. M. Hyman and C. N. Li (eds), 62–78. London: Routledge.

Locke, J. L. 1983. *Phonological Acquisition and Change*. New York: Academic Press.

Maddieson, I. 1984. *Patterns of Sounds*. Cambridge: Cambridge University Press.

MacCaffrey, H., Davis, B. L., MacNeilage, P. F., and Von Hopsburg, D. 2000. "Effects of multichannel cochlear implantation on the organization of early speech". *Volta Review* 5–29.

MacNeilage, P. F. 1998. "The frame/content theory of evolution of speech production". *Behavioral and Brain Sciences* 21: 499–548.

MacNeilage, P. F. and Davis, B. L. (2000a). "Origin of the internal structure of words". *Science* 288: 527–31.

MacNeilage, P. F. and Davis, B. L. 2000b. "Deriving speech from non-speech: A view from ontogeny". Special issue of *Phonetica* 57: 284–96.

MacNeilage, P. F., Davis, B. L., Kinney, A. and Matyear, C. M. 2000. "The motor core of speech: A comparison of serial organization patterns in infants and languages". *Child Development* 71: 153–63.

MacNeilage, P. F., Davis, B. L., Matyear, C. M., and Kinney, A. 2000. "Origin of speech output complexity in infants and in languages". *Psychological Science* 10: 459–60.

Maddieson, I. and Precoda, K. 1991. "Syllable structure and phonetic models". *UCLA Working Papers Phonetica* 78: 38–49 .

Matyear, C. MacNeilage, P. F., and Davis B. L. 1997. "Nasalization of vowels in nasal environments in babbling: Evidence for frame dominance". *Phonetica* 55: 1–17.

MacNeilage, P. F. and Davis, B. L. 1990. "Acquisition of speech production: Achievement of segmental independence". In *Speech Production and Speech Modeling*, W. I. Hardcastle and A. Marchal (eds), 55–68. Dordrecht: Kluwer.

MacNeilage, P. F. and Davis B. L. 1990. "Acquisition of speech production: Frames, then content". In *Attention and Performance*, Volume XIII: *Motor Representation and Control*, M. Jeannerod (ed), 452–68. Hillsdale, NJ: LEA.

Munhall, K. and Jones, J. A. 1998. "Articulatory evidence for syllable structure". *Behavioral and Brain Sciences* 21: 524–5.

Nittrouer, S., Studdert-Kennedy, M., and McGown, R. S. 1989. "The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel sequences spoken by children and adults". *Journal of Speech and Hearing Research* 32: 120–32.

Nittrouer, S. 1996. "The emergence of mature gestural patterns is not uniform: Evidence from an acoustical study". *Journal of Speech and Hearing Research*.

Ohala, J. J. and Kawasaki, H. 1989. "Prosodic phonology and phonetics". *Phonology* 6: 113–27.

Ohala, J.J. 1983. “The origins of sound patterns in vocal tract constraints”. In *The production of speech*, P. MacNeilage (ed), 189–216. New York: Springer.

Oller, D.K. 1980. “The emergence of the sounds of speech in infancy”. In *Child phonology*, Volume I: *Production*, G. Yeni-Komshian, J.F. Kavanagh, and C.A. Ferguson (eds). New York: Academic Press.

Oller, D.K., and Eilers, R.E. 1988. “The role of audition in infant babbling”. *Child Development* 59: 441–9.

Oller, D.K. and Steffans, M.L. 1993. “Syllables and segments in infant vocalizations and young child speech”. In *First and Second Language Phonology,* M. Yavas (ed). San Diego, Calif.: Singular Press.

Oller, D.K., Eilers, R.E., Bull, D.H., and Carney, A.E. 1985. “Prespeech vocalizations of a deaf infant: A comparison with normal metaphonological development”. *Journal of Speech and Hearing Research* 26: 47–63.

Oller, D.K., Wieman, L.A., Doyle, W.J., and Ross, C. 1976. *Infant Babbling and Speech. Journal of Child Language* 3: 3–11.

Ouinlan, P.T. 1992. *Oxford Psycholinguistic Database.* Oxford: Oxford University Press.

Redican, W.K. 1975. “Facial expressions in nonhuman primates”. In *Primate Behavior: Developments in Field and Laboratory Research*, L.A. Rosenblum (ed), New York: Academic Press.

Roug, L., Landburg, I., and Lundburg, L.J. 1989. “Phonetic development in early infancy: a study of four Swedish infants during the first eighteen months of life”. *Journal of Infant Language* 16: 19–40.

Stoel-Gammon, C. and Otomo, D. 1986. “Babbling development of hearing impaired and normally hearing subjects”. *Journal of Speech and Hearing Disorders* 51: 33–41.

Stoel-Gammon, C. 1988. “Prelinguistic vocalizations of hearing impaired and normal hearing subjects: A comparison of consonantal inventories”. *Journal of Speech and Hearing Disorders* 53: 302–15.

Stoel-Gammon, C. and Cooper, J.A. 1984. “Patterns of early lexical and phonological development”. *Journal of Child Language* 11: 247–71.

Stoel-Gammon, C, Kehoe, M.M. 1994. “Hearing impairment in infants and toddlers: Identification, vocal development and intervention”. In *Child Phonology: Characteristics, Assesment and Intervention with Special Populations*, J.E. Bernthal and N.W. Bankson (eds). New York: Thieme.

Stoel-Gammon, C. 1985. “Phonetic inventories, 15–24 months: a longitudinal study”. *Journal of Speech and Hearing Research* 28: 505–12.

Teixeira, E.R. and Davis, B.L. In press. “Early sound patterns in the speech of two Brazilian Portuguese speakers”.

Tobey, E.A., Geers, A. and Brenner, C. 1994. “Speech production results: Speech feature acquisition”. *The Volta Review* 95: 109–29.

Tye-Murray, N. and Kirk, K.I. 1993. “Vowel and dipthong production by young users of cochlear implants and the relationship between the phonetic level evaluation and spontaneous speech”. *Journal of Speech and Hearing Research* 36: 488–502.

Vihman, M.M., Macken, M.A., Miller, R., Simmons, H. and Miller, J. 1985. “From babbling to speech: A re-assessment of the continuity issue”. *Language* 61: 397–445.

Vihman, M. M., Ferguson, C. E., and Elbert, M. 1986. “Phonological development from babbling to speech: Common tendencies and individual differences”. *Applied Psycholinguistics* 7: 3–40.

Vihman, M. M. 1992. “Early syllables and the construction of phonology”. In *Phonological Development: Models, Research, Implications*, Ferguson, Menn, and Stoel-Gammon (eds), 393–422. Timonium: York Press.

Yoshinaga-Itano, C., Stredler-Brown, A., and Yancosek, E. 1992. “From phone to phoneme: What can we understand from babble”. *Volta Review* 94: 283–314.

Zmarich, C. and Lanni, R. 1999. “A phonetic and acoustic study of babbling in an Italian child”. *Proceedings of the 5th International Conference on Spoken Language Processing*, 2703–6. Sydney, Australia.

Zlatic, L. MacNeilage, P. F., Matyear, C. L., and Davis, B. L. 1997. “Babbling of twins in a bilingual environment”. *Applied Psycholinguistics* 18: 453–69.

Chapter 6

On the origins of intersyllabic complexity

Peter F. MacNeilage and Barbara L. Davis
University of Texas at Austin

0. Preamble

Languages strongly prefer to make the consonants and vowels in successive syllables different. But they may have started out with a simpler form. We believe that the first word patterns might have been similar to the preferred patterns of babbling and early infant words in which the *same* consonant and vowel occurs in successive syllables. In this paper we look at how infants eventually begin to make consonants and vowels in successive syllables different, and argue that hominids may have begun to do this in a similar way.

We begin with the claim in the paper immediately preceding this one, (Davis and MacNeilage, this volume) that the most basic component of speech, from its beginning, has been what we have called the "Frame"– the consonant-vowel (CV) sequence or syllable. We gave three reasons for this claim: 1. it is the primary syllabic form in the world's languages; 2. it is the primary syllabic form produced at the onset of babbling in infants; and 3. its claim to being a coherent package is enhanced by the finding that in both infants and language structures it tends to manifest three kinds of consonant-vowel co-occurrence tendencies, all of which result from the simple biomechanical property of lingual inertia: The tongue tends to stay in the central position that it adopts for rest during and following labial consonants ("pure frames"), and it tends to stay either in the front ("fronted frames") or the back of the mouth ("backed frames") for other CV sequences.

We believe that successive syllables tend to be the same in babbling and early speech because the neural control mechanism simply calls for more than one cycle of the same basic mouth close-open alternation, and it is achieved by an additional biphasic cycle of oscillation of the mandible. No changes in other articulators are necessary in order to produce this syllable "reduplication". We believe that, just as the basic frame types may have been used for the first monosyllabic words, reduplicated forms may have been used in the construction of the first multisyllabic words of language. How then do infants achieve the early stages of their transition from the reduplicative multisyllabic utterances that we have described here to nonreduplicative "variegated" multisyllabic sequences. After considering this we will then ask whether the infant's preferred solutions are echoed in the structure of multisyllabic sequences in languages, just as their preferred intrasyllabic solutions, the CV co-occurrence patterns, are echoed in syllables of languages.

1. Intersyllabic variation in infants

It has been widely believed for a long time that in the first part of the babbling stage, (7–10 months) babbling is primarily reduplicative, but that in a second stage (10–12 months) it becomes more variegated (e.g. Oller 1980). This has the common sense implication that infants were becoming more versatile in producing series of syllables as they proceeded through the babbling stage. However, we found, in looking at CVCV sequences in 6 infants in the first half of the babbling stage, that about 50% of sequences were reduplicative and 50% were variegated (Davis and MacNeilage 1995). Neither did variegation increase in the second half of the stage. Other studies (Mitchell and Kent 1991; Smith, Brown-Sweeney and Stoel-Gammon 1989) have also found that variegation does not increase across the babbling period.

Does the 50% of variegated sequences produced during the babbling period mean that infants are relatively versatile in inducing intersyllabic complexity at the onset of babbling, but that their versatility in this respect does not increase over the 5 month period. We think not. When we first found evidence that much of the variance in CV patterns in babbling may be produced by mandibular oscillation alone (Davis and MacNeilage 1990) we made the prediction that most of the variance in variegated sequences might also be attributable to the mandible, in the form of variation in the amplitude of a phase or two phases of mandibular oscillation across syllables. More specifically the prediction was that most intersyllabic variegation in babbling would be in what can be called the vertical dimension, rather than the horizontal dimension. In other words, most of the vocalic variation will be in tongue height rather than in the front-back axis of tongue positioning, and most of the consonantal variation will be in the aspect of manner of articulation that involves the amount of oral constriction, which is related to amount of mandible and tongue elevation, rather than in place of articulation, which involves the front-back dimension.

We have confirmed this prediction in utterances involving stop consonants, nasals and glides in both babbling (Davis and MacNeilage 1995) and first words (Davis *et al.*, in press). We have also found it to be true in babbling episodes that involve later developing consonants — fricatives, affricates and liquids (Gildersleeve-Neurman *et al.* 2000). We interpret these results to mean that most of the variance in early variegated sequences is being produced by variation in amplitude of phases of the mandibular cycle. The fact that there is no obvious tendency to repeat the same variegation pattern across successive utterances in babbling suggests that this variation might not be under the intentional control of the infant.

The question, then, is when and how does an infant begin to *systematically* increase intersyllabic complexity? A generation ago David Ingram (1974) drew attention to a favored pattern of infant consonant variegation which he called "Fronting" "where the sounds in the child's early words are ordered from those produced in the front of the mouth to those produced in the back." (p. 233). As an infant often has very few dorsals in the first word stage, the main manifestation of this phenomenon is a labial-vowel-coronal (LC) sequence as in [bado] for "bottle". The pattern also shows up in CVC words, which are, of course, monosyllabic. In order to discuss both disyllabic CVCV structures and CVC structures we will henceforth use the term *intercyclical* variegation rather than intersyllabic variegation to talk about infant performance.

In a review of 7 reports involving 5 language communities, we noted the LC tendency in 21 out of 22 instances (MacNeilage and Davis 2000b). There were even 2 reports of infants who attempted target words with the opposite (CL) sequence in this manner, as in "top" → "pot" (Jaeger 1997; Macken 1978). In our study of 10 infants in the first 50 word stage (12–18 months), we found that 9 infants showed the LC pattern and the other showed no trend (MacNeilage, *et al.* 1999). The mean ratio of LC sequences to CL sequences was 2.55.

Why does this particular sequence develop in first words? Before we attempt to answer this question it is necessary to step back and consider in more detail the relative role of the two main types of variables determining these early output patterns — basic production propensities and perceptually-based influences related to the ongoing process of simulation of speech patterns of the ambient language. Both of these types of influence are evident from the beginning of babbling. It has been found that profoundly hearing-impaired infants sometimes do not babble at all, (see McCaffrey *et al.* 1999 for a review) which suggests a crucial role of perceptual factors related to auditory input in inducing babbling to begin. Thus, even the earliest babbling seems to be to some degree a simulation of heard speech. However the simple biphasic alternation presumably reflects a basic production propensity rather than being copied directly because it is much the same in infants whether the adult language they are listening to has a simple or a complex syllable structure.

When hearing-impaired infants do babble, they also produce some biphasic cycles, but their overall output profiles are different from those of hearing infants, presumably because of their hearing deficit, which is a further indication of the importance of perceptual input in the formation of early output patterns (McCaffrey *et al.* 1999). With respect to the two main places of articulation of early consonants, hearing infants tend to produce more coronal than labial consonants in babbling. In contrast, hearing-impaired infants produce many more labials than coronals. It has been argued that the labial consonants are at least in part induced by the visual availability of labial movements in other language users (Osberger and McGarr 1982). But the finding that there are few coronal consonants suggests that under normal circumstances perceptual input regarding the auditory properties of coronals plays a role in inducing coronal production. It is possible that the typically greater frequencies of coronal than labial consonants in languages (Maddieson 1984) may play a role in the typical excess of coronals over labials in babbling. Perhaps the more often coronals occur in words, the more often an infant attempts to simulate them.

An additional perspective on this question is gained by considering an unusual class of infants — those who have undergone a tracheostomy at a stage of early speech development that spans the babbling and early word periods. (Tracheostomy involves insertion of a plastic tube into the trachea to permit breathing, which is otherwise endangered by some medical problem involving the laryngeal area). The result of this procedure is that the infant cannot activate the vocal tract in the usual way by vocal fold vibration, and therefore is not able to vocalize in anything like the normal manner. A surprising finding of three separate studies is that when the tracheotomy is removed, the first speech attempts overwhelmingly involve labial consonants rather than coronals (Bleile *et al.* 1993; Locke and Pearson 1990; Vaivre-Douret *et al.* 1995).

We have argued that it is easier for an infant to produce labials in the context in which they usually produce them, namely with central vowels, than to produce coronals in their usual manner, with front vowels. Only the mandible is required for the labial-central pairing in our view, while the coronal-front pairing requires an additional tongue fronting movement. Therefore we interpret the preference for labials over coronals in the initial attempts at vocalization of tracheostomized infants, in the presence of normal auditory input from other speakers, as further evidence for labials being easier to produce. Even though they are hearing more coronals than labials as infants without a tracheostomy do, they do not have the ability to simulate what they are hearing while the tracheostomy is in place. Apparently it is this opportunity for simulation, plus the presence of more coronals in the input that allows infants without tracheostomies to produce more coronals even though they are harder to make than labials.

There is one further piece of evidence that suggests that labials may be easier to produce than coronals. A number of studies have shown that the preference of coronals over labials in the babbling period tends strongly to reverse in the first word period. For example, this has been shown for groups of 5 English, French, Japanese and Swedish infants by Boysson-Bardies et al (1992). In our own work we found that while the mean ratio of labials to coronals in 6 babbling infants was

.46, the mean ratio in 10 infants producing first words is 1.76. (Davis *et al.*, in press). This is a dramatic reversal of preference. We interpret this trend toward more labials in first words as a regression to easier production forms when an infant begins the complex task of on-line interfacing of the hitherto autonomous output system with a new cognitive structure — the mental lexicon.

With the aid of these considerations, we wish to argue that the Labial-Coronal sequence effect in first words is a self-organizational phenomenon. We use the term self-organizational to mean that the LC pattern is the emergent result of the pattern of interaction of local effects of a number of variables without any contribution of a single superordinate master controller (See Clark 1997). This view can be contrasted with the view originally put forward by Ingram (1974) within the tradition of generative linguistics. He considered the phenomenon of fronting to be a result of a single linguistic cause termed "Markedness" according to which some phenomena are more basic to the language than others, and in this case the more basic the pattern, the more frequently it occurs. However, in markedness theory, which has been around for a number of decades "basic" has not yet been satisfactorily defined independent of frequency.

We believe that labials are easy than coronals, for reasons already given. Consequently we suggest that the LC form involves beginning with an easy movement pattern — the pure frame — and then increasing complexity by adding a second movement to the beginning of the second mandibular cycle — a tongue movement — thus producing a fronted frame. A factor that may make ease an important consideration in this context is the increase in functional load, mentioned earlier, that accrues when an infant no longer simply activates the output system, but now activates it as a result of an interface with a particular lexical concept, when a word is spoken. While any available output pattern can be activated in a babbling episode, in first word production, the particular speech pattern that needs to be activated depends on the concept that is being signalled. When the adult pattern for a particular word involves intercyclical variation, there is a pressure towards a simulation which also has intercyclical variation. Infants can often not respond effectively to this pressure, producing instead a reduplication of the first consonant (e.g. "babo" for "bottle"). And even when intercyclical variegation is produced, it is not always accurate, as evidenced by the two children who tended to produce LC sequences when the adult form of the word has a CL sequence. We believe that in these latter instances a simulation with an easy motor initiation is more available to the system than a more veridical one, whether or not the infant has a correct representation of the target word.

But why might it be advantageous to *start* with an easy action, rather than, say, end with one? The existence of functionally separable subsystems for initiation versus continuation of movements is sufficiently well known in motor system neurophysiology and clinical neurology for it to be included in a recent introductory textbook in cognitive neuroscience (Gazzaniga *et al.* 1999). Problems of movement initiation are well known to accompany various neurological syndromes, examples being an inability to initiate movement in many Parkinson's disease patients, and the inability to inhibit unwanted movements in Tourette's syndrome. The complexity of the voluntary movement initiation process in humans is suggested by the existence of the Bereitschaft potential, a frontal lobe negativity detectable as early as 1200 milliseconds before movement onset (Kornhuber 1987). We believe that it is an advantage to start with an easy action partly because the motor system needs to deal with a specific problem of initiating movement, a problem that may be proportional to the complexity of the initial movement/s. Initiating speech requires the coordinated action of three subsystems that otherwise have the capability of operating independently of one another — the respiratory, laryngeal and oral systems. This coordination should not be taken for granted. It presumably makes initiation more problematical than continuation for an infant, because much of a typical post-initial pattern in babbling and early speech can be produced by simply maintaining the actions that have been initiated — lung volume decrease to maintain subglottal pressure, vocal fold tension in the service of vibration, and frame production.

To review the argument, we believe that the tendency to favor LC sequences is a self organizational consequence of the following factors. Pure frames, with labial consonants and central vowels may be easier to produce than frames with coronal consonants. An addition to the functional load associated with vocalization in the babbling stage occurs at the first word stage when the infant has to interface a vocal episode with a specific lexical concept rather than simply producing an output episode. This produces a generalized bias towards increasing use of pure frames, with their labial-central pattern, in first words. The initiation of action is a separable functional role of motor systems, and the execution of this function may become more difficult, the more complex the output is. Consequently in instances in which infants are simulating an adult word with intercyclical variegation a tendency may arise to begin this simulation with the easier pure frame pattern and then add a tongue movement to the next frame.

2. Intercyclical variation in languages and a protolanguage corpus

The next question is whether the favoring of this LC pattern occurs only in infant speech, and could therefore be an indication of ontogenetic growing pains, irrelevant to phylogeny, or whether it is also present in languages, suggesting that it has a more fundamental status. Locke (1983) has presented some preliminary data suggesting its presence in English and French. We made dictionary counts involving words of 10 languages which began with CVC sequences, and in which the next sound, if there was one, was another vowel. The languages were English, Estonian, French, German, Hebrew, Japanese, New Zealand Maori, Quichua, Spanish and Swahili. We found that the LC preference was present in 9 languages, at significant levels in 8 (MacNeilage *et al.* 1999). The mean ratio of LC to CL sequences was 2.23, an extremely large ratio, though not as large as was found in infants.

The next question is whether this trend was present in the 27 word protoword corpus which Bengston and Ruhlen (1994) have suggested might have been words of a first language. There was a total of 8 instances of the LC pattern: "BUR", "MALIQ'A", "MANA", "MANO", "MENA", "PAL", "PAR", AND "PUTI". There was only 1 instance of the CL pattern: "TSUMA" (MacNeilage and Davis 2000).

These results suggest that the LC effect is at least as fundamental to language as the CV co-occurrence patterns are. We assume that 3 of the factors involved in self organizational contingencies governing the production of this pattern in early hominids were similar to what they are in infants, namely the greater ease of labials, the problem of movement initiation and the functional load of interfacing an output pattern with a lexical concept. However, one important difference between modern infants and the earliest speakers is that, regardless of how specifically they use it, infants do have an external model to simulate, while the earliest speakers did not . They initially had to invent words, not copy them. The presence of the LC pattern in languages and the protolanguage corpus suggests that it might have developed early in language history just as it develops early in infants. But it could not have originally developed by modelling. Nevertheless, being easier to produce than the reverse (CL) form, it may have simply occurred more often, therefore being more often made available to be linked with a concept to form an early word, assuming that there was pressure to produce new vocal forms that could serve for new words.

3. Intercyclical variation involving dorsal consonants

We left out dorsal consonants in our initial investigations of intercyclical complexity, because many infants during the first word stage either do not have any dorsals or have very few of them. However, together with labial and coronal, dorsal is one of the three most important places of articulation in languages. Stop consonants at each of these three places of articulation are almost universal in languages (Maddieson 1984) and we have found that, like labials and coronals, dorsal consonants are also subject to consonant-vowel co-occurrence constraints. Consequently, a consideration of the origin of intercyclical consonant organization should include dorsals.

Returning to Ingram's "Fronting" generalization, we note that it implies a relative frequency hierarchy of Labial > Coronal > Dorsal in initial position, and

Dorsal > Coronal > Labial in final position. There is some evidence that dorsals tend to favor final position in early words. For example, Stoel-Gammon and Cooper (1994) commented that one child's heavy use of final consonantal position might be related to his early and extensive use of dorsal stops. Vihman and Hochberg (1986) in a study of 8 children, also found evidence for a dorsal preference for final position. A number of studies have considered intercyclical consonant variegation since Ingram presented the fronting hypothesis. However, they have involved analysis of a small number of subjects, and overall sound inventories and overall syllabic production propensities have not been presented (e.g. Menn 1975; Cruttenden 1978; Levelt 1994; Stoel-Gammon and Stemberger 1994; Pater 1997). Here, we present an analysis of 2 large bodies of data on intercyclical consonant variegation in infants. They are the set of 4,324 consonant sequences produced in words from the age of 14 to 20 months by the infant studied by Davis and MacNeilage (1990) and a set of 1,570 sequences produced by the 10 infants studied by Davis et al (in press) during the 50 word stage. Table 1 shows the percentages of occurrence of the six possible variegated sequences, which are very similar in the two corpora.

Both of these patterns are confirmatory of the fronting principle, in that there is a relative frequency heirarchy of Labial > Coronal > Dorsal in initial position, and Dorsal > Coronal > Labial in final position. However, these data suggest a new interpretation of the fronting phenomenon beyond the suggestion that it is simply due to a linear front-to-back trend in place of articulation, a trend for which no satisfactory explanation has ever been presented.

To begin with, as an analysis of percentage of occurrence of each place in each word position (Table 2) shows, the trend does not appear to be linear. The most obvious pattern is that while labial consonants have a very strong trend towards initial position, the two lingual consonants, share an almost equally strong trend towards occurring in final position. We have argued that one aspect of the pattern shown here — the excess of LC over CL sequences — is due to an interaction between ease of production of labials, and the contingencies associated with lexical access and action initiation, which leads to labials being favored over coronals for

Table 1. Variation in place of articulation in consonant sequences in early words beginning with a CVC structure in one infant (Davis and MacNeilage), and a group of 10 infants (Davis et al). Percentages of occurrence of the 6 possible sequences. (L=labial, C=coronal, D=dorsal)

	LC	LD	CL	CD	DL	DC
Davis and MacNeilage 1990 (N=1)	53.0	19.4	18.0	4.3	3.0	2.2
Davis *et al.*, in press (N=10)	43.8	21.0	19.6	7.3	1.5	6.8

Table 2. Percentages of occurrence of the three major places of consonantal articulation sequences in the first and second positions of early words beginning with a CVC structure in one infant (D and M), and a group of 10 infants (Davis *et al.*).

D and M 1990 (N=1)				Davis *et al.*, in press (N=10)			
	First	Second			First	Second	
L	77.5	22.5	100	L	75.5	24.5	100
C	28.7	71.3	100	C	34.7	65.3	100
D	18.2	81.8	100	D	22.5	77.5	100

initiation. This data leads us to extend this hypothesis to the other major class of lingual consonants — dorsals. Like coronals, they require a tongue movement in addition to frame production, while labials are primarily produced by mandibular oscillation alone. Thus, the effect seems to be a labial-lingual effect rather than just a labial-coronal effect.

One reason to believe that the main effect is a labial-lingual effect is that virtually all of the differences between coronals and dorsals arise from pairings beginning with labials, rather than from pairings with each other. While there were more CD than DC sequences in the single infant, (61 versus 32) the number of CD and DC sequences in the 10 infant group was almost identical (30 versus 28). Furthermore, these lingual-lingual pairings constitute only an average of about 7% of the total number of variegated sequences in the 2 studies. In contrast, the labial-lingual pairings constituted an average of about 69% of the sequences. (Most of the remaining sequences were lingual-labial sequences, mostly involving coronals (19%), which violate the fronting principle.) Thus, if there is no fronting effect *between* the two lingual consonants, the fronting effect arises almost entirely from sequences initiated by a labial. Dorsals are simply relatively more subject to this trend than coronals. (While the ratios of labial-coronal to coronal-labial sequences were 2.9:1 for the single infant, and 2.6:1 for the 10 infant group, the corresponding ratios of labial-dorsal to dorsal-labial were 6.4:1 and 14.5:1 respectively.) It thus appears that what has been called the fronting effect in infants is due primarily to an ease-related labial initiation preference that dorsals are more subject to than coronals.

The next question is whether the patterns found here are also present in languages. We have already reported that there is a strong preference for LC sequences relative to CL sequences in languages. Does a similar pattern hold for labial-dorsal (LD) versus dorsal-labial (DL) sequences. It does not. Table 3 shows the average percentages of occurrence of the 6 forms in the 10 language sample.

LC sequences are not only about twice as frequent as CL sequences in our 10 language sample, but they are about twice as frequent as all 4 of the other possibil-

Table 3. Mean percentages of occurrence of the three major places of consonantal articulation sequences in the first and second positions of words of 10 languages

	LC	LD	CL	CD	DL	DC
10 Languages	27.0	15.1	13.4	13.1	15.2	16.2

ities as well (DL, LD, CD, and DC). Putting LC sequences aside, the mean frequencies of the other five sequences occupied an extremely narrow range — from 13.1 to 16.2%. LD and DL frequencies were almost identical. So we have the rather puzzling conclusion that while languages show the infant excess of LC over CL patterns they do not show the infant excess of LD over DL patterns. A topic for future research is the course by which infants move from a strong initial preference for LD over DL sequences to a lack of preference for one over the other. The course of development of the appropriate use of reduplicative pairings of consonants (LL, CC, DD) should provide an important context for further study of this question.

4. Evolving intercyclical variation: One general consideration

A proposition we have presented is that the evolution of modern languages involved major increases in intercylical variegation in order to develop the capability of transmitting a large message set. The situation in modern languages has been characterized in generative phonology by the Obligatory Contour Principle (e.g. Kenstowicz 1994) which bans, or at least severely frowns upon successive instances of the same segmental entity on the same tier (consonantal or vocalic — $\underline{C}V\underline{C}$ or $\underline{V}C\underline{V}$). In studying our set of 10 languages, we have obtained some interesting information regarding what this seems to have entailed (MacNeilage *et al.* 2000). From the point of view of our supposition that languages might have progressed from reduplication towards variegation, it might be expected that modern languages have reduced consonant harmony in the direction of chance levels under pressure to preserve perceptual distinctiveness in an increasingly large message set. However, a count of first and second stops and nasals in CVC, CVCV and CVCV words in our 10 languages showed repetition of consonantal place of articulation occurred, on the average, at only 2/3rds of chance levels, even when the stop nasal distinction is ignored. Overall values ranged from 44% of chance values in Hebrew (which is well known for its constraint against consonant repetition) to 89% in Swahili.

Why would what we believe was a basic hominid propensity to repeat consonants intercyclically, be reduced to below chance levels in modern languages? The result suggests that a problem might have arisen in modern high-speed speech that was not present when speech was produced at lower speeds using smaller inventories. The problem may lie in the confusing effect of frequent occurrence of the same sound in working memory, probably in both the stage of input analysis and that of output organization. A well known finding in working memory studies is the confusibility of simultaneously held items with similar pronunciations, e.g. the spoken letters "em" and "ef" or "bee" and "dee" (Conrad and Hull 1964). With respect to output, studies of speech errors show that they are potentiated by a "repeated phoneme effect" (MacKay 1987): the occurrence of two examples of the same sound in close proximity leads to an increase in serial ordering errors.

5. Beginnings and ends

A more complete treatment of the evolution of speech than one that focuses solely on its internal structure requires consideration of how utterances begin and end. It is well known that babbling utterances and early words tend to begin with a consonant and end with a vowel. Bell and Hooper (1978) have noted a similar trend for words in languages. They conclude that "sections" (which are roughly equivalent to words) "must begin with a consonant in about 20–40 percent of the world's languages (and) must end with a vowel in about 10–25 percent of the world's languages. No doubt it will eventually be shown, as we have argued for a number of other patterns shared by infants and languages, that infants are not copying these patterns from the ambient language. (This is already clear for the infant preference for final vowels in the English language environment, as English prefers to end words with consonants.) Instead they again are showing us patterns, presumably reflecting production constraints, that must have been present in the earliest language, though in the case of beginnings and ends we have no hypotheses as to why it is these particular patterns that are favored.

Even when non-preferred beginnings and endings (vocalic initiations and consonantal terminations) are considered, there may be important similarities between infant patterns and patterns of words in languages. We have found that English infants prefer central vowels when initiating utterances with vowels in babbling and first words, and a preliminary count of 5 languages which have 5-vowel systems show that this is true for these languages as well (MacNeilage and Davis 1999). We have also argued that the tendency of infants and languages to favor voiceless terminal consonants may be explicable in terms of a naturally occurring terminal energy decrease in the respiratory, and possibly the phonatory systems, that tends, statistically, to eliminate the conditions for voicing in infant productions (Redford *et al.* 1997). Although this terminal energy decrease does not occur at word boundaries in modern running speech, its effect, in the form of the tendency for final devoicing, may have gotten into language at the stage of single word utterances, and remained present in some modern languages.

6. Protolanguage

A few more comments on the structure of the protolanguage corpus of Bengston and Ruhlen might be of interest. There were some similarities between protoword structure and infant speech beyond the basic CV alternation pattern, the CV co-occurrence patterns and the LC sequence effect. There was also a large number of stop and nasal consonants, a paucity of other consonants, and a preference for the low vowel A. In contrast to these similarities there was a complete lack of inter-syllabic consonant reduplication and a large number of dorsal consonants, both of which differ from patterns of infants' first words.

The question of the evolution of individual languages is a highly controversial one. Our conclusion that the presence of the CV co-occurrence patterns and the LC pattern in the Ruhlen-Bengston corpus may be evidence for their presence in the earliest speech will no doubt be met with skepticism, as is true for any claims for deep reconstruction (meaning beyond 10,000 years). Our most basic contribution in this context is to argue for the probability that phonetic forces contributed to the design of the first words, and that some of these forces were sufficiently fundamental that their effects remain visible. An implication of this claim is that the assertion of the null hypothesis that languages completely cover up their historical tracks in as little as 3,000–4,000 years (e.g. Dixon 1997) might benefit from re-evaluation.

7. Conclusion

We have pointed out a number of similarities and differences between infant babbling and early speech, sound patterns of languages, and the sound patterns of one proposed protoword corpus. When all three sets of patterns were similar, we have argued that they were present in the earliest words, and we have attempted to provide performance-based explanations for some of them. A key supposition, basic to the discipline of phonetics, that is consistent with our claims, is that sound patterns of words are a result of two sometimes conflicting forces — production constraints, and the need for perceptual distinctiveness (e.g., Lindblom 1992). We believe the role of production constraints was dominant in the earliest speech, just as it is in the earliest speech of modern infants. However as languages evolved, the role of perceptual distinctiveness constraints became more important, forcing evolving speakers, and developing speakers out of their most comfortable production modes in the service of increases in the size of the communicable message set. With regard to the evolution of phonology, our conclusion is that phonology wasn't given in advance by a fortuitous mutation, providing a competence-related structure that just happened to instantly resolve the conflict between the productive and perceptual pressures. Instead, phonology evolved from the phonetic level of action by means of a self-organizing process of resolution of conflicts between production and perception constraints .

8. Summary

The thesis of this paper is that if one takes a broad enough approach to the purely functional aspects of serial organization of speech, it might be possible to understand the general outline of its phylogeny. Here we provide a summary of the main focus of the paper — the issue of word-internal organization. In our other paper we have argued that *intra*syllabic structure, in the form of consonant-vowel interrelationships shows a strong biomechanical effect of inertia, in infants and languages, and as inertia is such a basic biomechanical property, speech has probably been subject to this effect since it first began. The predominance of syllable *reduplication* in infants suggests that for infants this biomechanical inerital effect extends across syllables. But in contrast to infants, languages favor *inter*syllabic *variegation*. The question we raise here is how do infants progress from favoring reduplication to favoring variegation, and did earlier hominids follow what was presumably a similar path? The most well established infant step towards intersyllabic (for us, intercyclical) variegation in first words is to favor a labial-vowel-coronal (LC) sequence over its opposite. We consider this step in infants to be the self-organizational consequence of the confluence of four performance-related factors: 1. Greater ease of labial production; 2. The increasing functional load of interfacing the lexicon with the output system in the first word stage; 3. The basic neurological problem of movement initiation; 4. A generalized pressure towards simulation of the intercyclical diversity of adult word targets. As the LC preference is also present in our sample of 10 languages and in the protolanguage corpus we analyzed, we conclude that, as in *intra*syllabic organization, ontogeny may recapitulate phylogeny here, with the exception that in phylogeny, the pressure in factor 4 (above) was towards *creating* rather than *simulating* lexical diversity. Consideration of the third major class of place of articulation — the dorsal place — which is less frequently used by infants, suggests that the basic infant intercyclical pattern may be best regarded as labial-lingual rather than just labial-coronal. For infants, LD patterns were favored over DL patterns, just as LC patterns were favored over CL patterns. However, somewhat surprisingly, languages did not favor LD sequences over DL ones. In short, we have observed parallels between infants and languages in two sorts of patterns, suggesting both may have been always present in languages. The simplest (intrasyllabic) patterns seemed to governed by biomechanical contingencies so basic that they must have always been present in speech. The more complex (intersyllabic) patterns may be a result of a self-organizational interaction between biomechanical factors and other factors, most of which may be common to the process of generating words whether it be by infants or earlier hominids.

References

Bell, A. and Hooper, J. B. (eds). 1978. *Syllables and Segments.* Amsterdam: North Holland.

Bengtson, J. D. and Ruhlen, M. 1994. "Global etymologies". In *On the Origin of Languages*, M. Ruhlen. Stanford: Stanford University Press.

Bleile, K. M., Stark, R. E., and Silverman-McGowan, J. 1993. "Speech development in a child after decanulation: further evidence that babbling facilitates later speech development". *Clinical Linguistics and Phonetics* 7: 319–30.

Boysson-Bardies, B., Vihman, M. M., Roug-Hellichius, L., Durand, C., Landberg, I., and Arao, F. 1992. "Material evidence of infant selection from the target language: A cross-linguistic study". In *Phonological development: Models, research, implications*, C. Ferguson, L. Menn, and C. Stoel-Gammon (eds). Timonium, Md.: York Press.

Clark, A. 1997. *Being There: Putting Brain, Body, and World Together Again.* Cambridge: Cambridge University Press.

Cruttenden, A. 1978. "Assimilation in child language and elsewhere". *Journal of Child Language* 5: 373–8.

Conrad, R. and Hull, A. J. 1964. "Information, acoustic confusion and memory span". *British Journal of Psychology* 55: 429–32.

Davis, B. L. and MacNeilage, P. F. 1990. "Acquisition of correct vowel production: A quantitative case study". *Journal of Speech and Hearing Research* 33: 16–27.

Davis, B. L. and MacNeilage, P. F. 1995. "The articulatory basis of babbling". *Journal of Speech and Hearing Research* 38: 1199–1211.

Davis, B. L. and MacNeilage, P. F. This volume. "The internal structure of the syllable: An ontogenetic perspective on origins".

Davis, B. L., MacNeilage, P. F. and Matyear, C. L. In press. "Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches". *Phonetica.*

Dixon, R. M. W. 1997. *The Rise and Fall of Languages.* Cambridge: Cambridge University Press.

Gazzaniga, M., Ivry, R., and Mangun, G. 1999. *Cognitive Neuroscience: The Biology of the Mind.* New York: Norton.

Gildersleeve-Neumann, C. Davis, B.L., and MacNeilage, P.F. 2000. Contingencies governing the acquisitive of fricatives, affricates and liquids. *Applied Psycholinguistics* 21: 341–63.

Ingram, D. 1974. "Fronting in child phonology". *Journal of Child Language* 1: 233–41.

Jaeger, J. 1997. "How to say 'Grandma': the problem of developing phonological representations". *First Language* 17: 1–29.

Kenstowicz, M. 1994. *Phonology in Generative Grammar*. Oxford: Blackwell.

Kornhuber, H.H. 1987. In *Encyclopedia of Neuroscience*, G. Adelman (ed), 1302–3. Boston: Birkhauser.

Levelt, C. 1994. *On the Acquisition Of place*. Doctoral dissertation, Holland Institute of Generative Linguistics.

Lindblom, B. 1992. "Phonological units as adaptive emergents of lexical development". In *Phonological Development: Models, Research, Implications*, C. Ferguson, L. Menn and C. Stoel-Gammon (eds). Timonium, Md.: York Press.

Locke, J.L. 1983. *Phonological Acquisition and Change*. New York: Academic Press.

Locke, J.L. and Pearson, D. 1990. "Linguistic significance of babbling: evidence from a tracheostomized infant". *Journal of Child Language* 17: 1–16.

MacKay, D.G. 1987. *The Organization of Perception and Action: A Theory for Language and Other Cognitive Sciences*. New York: Springer Verlag.

Macken, M. 1978. "Permitted complexity in phonological development: One child's acquisition of Spanish consonants". *Lingua* 44: 219–53.

MacNeilage, P.F. and Davis, B.L. 1999. "Evolution of the form of spoken words". *Evolution of Communication* 3: 3–20.

MacNeilage, P.F. and Davis, B.L. 2000a. "Origin of the internal structure of word forms". *Science* 288: 527–31.

MacNeilage, P.F. and Davis, B.L. 2000b. "Evolution of speech: The relation between ontogeny and phylogeny". In *The Evolutionary Emergence of Language*, J.R. Hurford, C. Knight, and M.G. Studdert-Kennedy (eds), 146–60. Cambridge: Cambridge University Press.

MacNeilage, P.F., Davis, B.L., Kinney, A., and Matyear, C.L. 1999. "Origin of serial output complexity in speech". *Psychological Science* 10: 459–60.

MacNeilage, P.F., Davis, B.L., Kinney, A., and Matyear, C.L. 2000. "The motor core of speech: A comparison of serial organization patterns in infants and languages". *Child Development* 71: 153–63.

Maddieson, I. 1984. *Patterns of Sounds*. Cambridge: Cambridge University Press.

McCaffrey, H.L., Davis, B.L., MacNeilage, P.F., and von Hapsburg, D. 1999. "Effects of multichannel cochlear implant on the organization of early speech". *Volta Review* 101: 5–28.

Menn, L. 1975. "Counter-example to 'fronting' as a universal in child phonology". *Journal of Child Language* 2: 293–6.

Mitchell, P.R. and Kent, R.D. 1990. "Phonetic variation in multisyllabic babbling". *Journal of Child Language* 17: 247–65.

Oller, D.K. 1980. "The emergence of the sounds of speech in infancy". In *Child Phonology*, Volume I: *Production*, G. Yeni-Komshian, J.F. Kavanagh, and C.A. Ferguson (eds). New York: Academic Press.

Osberger, M.J. and McGarr, N.S. 1982. “Speech production characteristics of the hearing impaired”. In *Speech and Language*, Volume VIII: *Advances in research and practice*, N. Lass (ed). New York: Academic Press.

Pater, J. 1997. “Minimal violation and phonological development”. *Language Acquisition* 6: 201–53.

Redford, M.A., MacNeilage, P.F., and Davis, B.L. 1997. “Production constraints on utterance-final consonant characteristics in babbling”. *Phonetica* 54: 172–86.

Smith, B., Brown-Sweeney, S., and Stoel-Gammon, C. 1989. “A quantitative analysis of reduplicated and variegated babbling”. *First Language* 9: 175–90.

Stoel-Gammon, C., and Cooper. 1984. “Patterns of early lexical and phonological development”. *Journal of Child Language* 11: 247–71.

Stoel-Gammon, C. and Stemberger, J. 1994. “Consonant harmony and phonological underspecification in child speech”. In *First and Second Language Phonology*, M. Yavas (ed). San Diego, Calif.: Singular Press.

Vaivre-Douret, L., Le Normand, M.-T., and Wood, C. 1995. “Developmental outcome in a tracheostomized child: A case study”. *Early Child Development and Care* 112: 13–26.

Vihman, M.M. and Hochberg, J.G. 1986. “Velars and final consonants in early words”. In *The Fergusonian Impact*, Volume I: *From Phonology to Society*, J. Fishman *et al.* (eds). Berlin: Mouton de Gruyter.

CHAPTER 7

On the pre-linguistic origins of language processing rates

Marjorie Barker and T. Givón
University of Oregon

1. The temporal rate of natural language processing*

One of the most persistent facts about human language is the amazingly stable flow-rate of its two major information-processing units, words and clauses. Whether taken from natural oral communication, psycholinguistic experiments or measurements of neurological activation, the temporal flow of word-processing and clause-processing appears remarkably stable, averaging ca. 250 msecs per lexical word and 1.0 sec. per event clause.

1.1 Natural spoken language

As an illustration of the normal flow-rate of oral language, consider the transcript of a taped narrative describing a just-viewed 6-minute video clip. The language is Kalam (Papua-New Guinea Highlands), with SOV word-order, extensive verb-serialization, mostly bi-syllabic uninflected nouns, and mostly mono-syllabic verb stems with small yet complex portmanteaux suffixal inflections. All uninterrupted intonation units are counted clauses (Chafe 1994).

(1) **Rate of flow of Kalam oral narrative** (Givón 1991, Kalam subject #1)

	intonation units (**English glosses of lexical words only**)	seconds/ clause	words/ clause	average msecs per word
1.	man spade take	(1.0)	3	0.333
2.	there base put say	(0.9)	4	0.225
3.	there fall-hit	(0.9)	3	0.300
4.	again take	(0.4)	2	0.200
5.	take-put say-then	(0.8)	4	0.200
6.	axe piece carry-leave	(0.8)	4	0.200
7.	there go	(0.3)	2	0.150

	intonation units (**English glosses of lexical words only**)	seconds/ clause	words/ clause	average msecs per word
8.	wood	(0.3)	1	0.300
9.	cut-sever put	(0.5)	3	0.166
10.	*khamb-ak* (?)	(0.3)	1	0.300
11.	ah	(0.3)	1	0.300
12.	before cut-sever	(0.5)	3	0.166
13.	long-long cut-sever put	(0.9)	5	0.180
14.	later short-short cut-sever wait	(1.3)	6	0.216
15.	there.down take-put	(0.5)	3	0.166
16.	take-put there.and	(0.4)	3	0.133
17.	that wood-in wood-in wood-in there	(1.3)	8	0.162
18.	do.and woman-his come say	(1.0)	5	0.250
19.	wood cut-put or say	(0.9)	5	0.180
20.	yes say	(0.6)	3	0.300
21.	wood cut-put here carry-go say	(1.3)	7	0.185
22.	carry woman house go	(0.8)	4	0.200
23.	man-your	(0.5)	2	0.250
24.	return stay	(0.4)	2	0.200
25.	spade two carry	(0.6)	3	0.200
26.	go	(0.4)	1	0.400
27.	woman go house there go	(1.0)	5	0.200
28.	wood bunch take-throw	(0.8)	4	0.200
29.	ehm	(0.3)	1	0.300
30.	there take-come	(0.4)	3	0.133
31.	match hit light-burn	(1.4)	4	0.350
32.	wood-some take-light	(0.9)	4	0.225
33.	fire make	(0.4)	2	0.200
34.	again return say	(0.9)	3	0.300
35.	come be	(0.6)	2	0.300
36.	water pull take-go	(0.9)	4	0.225
37.	water before pull take-go put	(1.1)	6	0.183
38.	bucket pour take-go	(0.7)	4	0.175
39.	pot in put take-go there-over put heat	(1.9)	9	0.211
40.	house side here leave disappear	(0.8)	5	0.160
41.	chicken one pull do come-arrive	(0.9)	6	0.150
42.	take-go there	(0.5)	3	0.166
43.	knife small take	(0.6)	3	0.200
44.	throat cut	(0.5)	2	0.250
45.	put say	(0.6)	2	]0.300

46.	chicken there fall there fall	(0.9)	5	0.180
47.	afraid go	(0.4)	2	0.200
48.	there chase take-go there chase	(0.8)	6	0.133
49.	that leave	(0.5)	5	0.250
50.	return here come	(0.6)	3	0.200
51.	and go house-in there enter	(1.1)	6	0.183
52.	food bread put	(0.6)	3	0.200
53.	piece some tea-towel wrap-wrap do	(1.8)	7	0.257
54.	carry there come	(0.9)	3	0.300
55.	knife small take	(0.5)	3	0.166
56.	cut-sever be	(0.5)	3	0.166
57.	carry return	(0.6)	2	0.300
58.	ha!	(0.3)	1	0.300
59.	take-go take-go	(0.5)	4	0.125
60.	husband	(0.4)	1	0.400
61.	hoe take garden do	(0.9)	4	0.225
62.	go arrive	(0.4)	2	0.200
63.	go arrive be	(0.6)	3	0.200
64.	eh	(0.3)	1	0.300
65.	husband that give food take	(1.0)	5	0.200
66.	be say	(0.4)	2	0.200
67.	thing that this what say	(1.0)	5	0.200
68.	take-see	(0.2)	2	0.100
69.	thing thing bad bad some be	(1.0)	6	0.166
70.	man-your take sever-rise	(1.0)	5	0.200
71.	woman.his her hit put/I say	(0.8)	5	0.160
72.	how big take	(0.5)	3	0.166
73.	garden do thing big here carry round-round do	(1.7)	9	0.188
74.	tree base round-round go	(1.0)	5	0.200
75.	chase chase chase chase	(1.1)	4	0.275
76.	stay	(0.2)	1	0.200
77.	and	(0.3)	1	0.300
78.	place like.that nothing	(0.7)	3	0.233
79.	grass some rise	(1.2)	3	0.400
80.	wood-in wood-in there	(1.0)	5	0.250
81.	be	(0.3)	1	0.300
82.	then	(0.3)	1	0.300
83.	house there couple that river side there	(1.4)	7	0.200
84.	come pass	(0.3)	2	0.150
	Overall means:	**0.927**	**4.16**	**0.223**

The frequency distribution of clause-length in seconds and in number of words is given in (2) and (3) below, respectively.

(2) Frequency distribution of Kalam clause length in seconds: **seconds/clause**

	0.2	0.3–0.4	0.5–0.7	0.8–1.0	1.1–1.4	1.8–1.9	total
N	2	20	23	29	7	3	84
%	2.3	**23.8**	**27.4**	**34.6**	**8.3**	3.6	100.0

95% BET. 0.3 and 1.4 (mean = 0.927 secs)

(3) Frequency distribution of Kalam clause length in # of words.

words/clause	N	%
1	11	**13.1**
2	14	**16.5**
3	22	**26.1**
4	12	**14.3**
5	13	**15.8**
6	6	**7.1**
7	3	3.5
8	1	1.2
9	2	2.4
Total	84	100.0

93% bet. 1–6 words (mean: 4.16 words)

Since Kalam is a serial-verb language with an average of 2.0 verbs per clause and since serial verbs in Kalam clauses are typically co-lexicalized (Givón 1991), the corresponding ratio in the same narrative in English would have been ca. *3.0* words per clause.

1.2 Psycholinguistic studies of mental activation

As an example of the psycholinguistic experimental work on the rate of word and clause processing rates, consider the work of Neely (1977) or Swinney (1979), studying the automatic ('spreading') activation of multiple word-senses, as measured by semantic priming of subsequent target words (see also Neely 1990). In Swinney's (1979) study, two contrasting stimuli clauses such as (4a) and (4b) below were presented, each containing the same ambiguous *prime* word ('bank'), followed by either of two *target* words ('water', 'money'), one appropriate to the clausal context, the other inappropriate.

(4) Multiple activation of word-senses in differential contexts

prime words in context	target words approp.	inapprop.	unrel.
a. Joe cashed a check in the *bank*	money	water	pin
b. Joe walked along the river *bank*	water	money	pin

Beginning at ca. 100 msecs after the presentation of a clause containing an ambiguous prime-word, both target-words show a priming effect (in terms of shorter RT to a lexical decision task, as compared with a semantically unrelated target-word). In other words, both senses of the prime word were automatically activated regardless of their clausal context. The peak of this priming effect was shown at ca. 250 msecs following prime-word presentation. From that point onward, the priming effect of the ambiguous prime-word on the contextually inappropriate target-word decayed rapidly, so that at ca. 1.0 seconds after prime-word presentation, only the sense that was appropriate to the clausal context remained activated. The other sense, contextually inappropriate, was suppressed. Put another way, it took ca. 1.0 second for the clause to be processed, and then exert the context's inhibitory effect on the inappropriate sense of the prime-word.

1.3 Neuro-psychological studies of brain activation

The neurological data we would like to cite come from a group of studies by Snyder *et al.* (1995), Posner and Pavese (1997) and Abdulaev and Posner (1997) (see also Posner and Tudela 1997; Abdulaev and Posner 1998), reporting on the time-course of ERP activation of two meaning-related left-cortical brain areas:

- the *pre-frontal* semantic are (ventral-anterior to Broca's area); and
- the *medio-temporal-parietal* Wernicke's area.

In previous studies using PET imaging (cortical blood flow; Peterson *et al.* 1988 1989, 1990; Posner *et al.* 1988, 1989; Raichle *et al.* 1993; Snyder *et al.* 1995), both areas responded to lexical-semantic tasks. What the ERP studies found is that both the time-course of the activation and the semantic tasks that triggered it were rather different for these two cortical areas. A purely lexical-semantic task ('read word/noun aloud') activated only the pre-frontal area, with peak activation time at ca. 250 msecs. following stimulus presentation. A combinatorial-semantic task ('generate a use/verb associated with a stimulus/noun') activated *both* areas, with the pre-frontal area still showing its peak activation at ca. 250 msecs, and Wernicke's area at ca. 700–800 msecs following stimulus presentation.

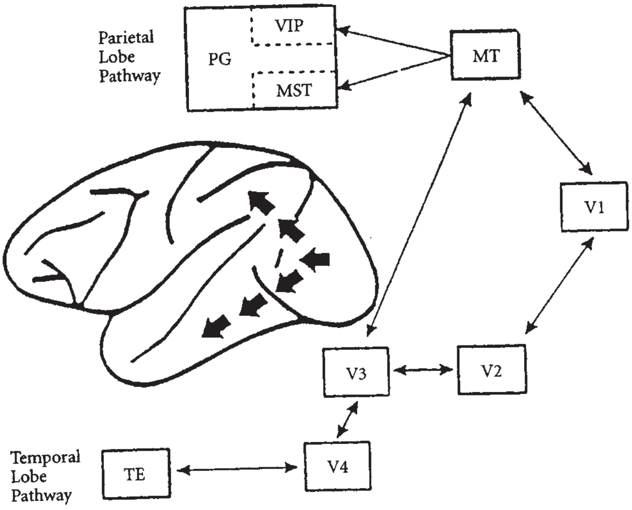
The results of these time-course ERP studies were interpreted, we believe correctly, to mean that the left pre-frontal area is involved the processing of *word-*

senses (lexical semantics), and Wernicke's area in the processing of *word-combinations* (propositional semantics). The one-way conditional association between the two tasks (combinatorial > lexical, but not vice versa) is fully consonant with this interpretation.

2. Language and visual information processing

In an earlier paper (Givón 1998) it was pointed out that several major components of the human language-processing apparatus must have arisen directly from specific precursors in the primate visual information processing system. To recapitulate briefly, in the primate brain the processing of visual information is split into two major parallel trends, the *dorsal* occipital-to-parietal trend, and the *ventral* occipital-to-temporal trend. The two trends split at a region anterior to *area 19* of the primary visual cortex (occipital lobe), though they remain in intimate cross-communication (Mishkin 1978, 1982; Mishkin *et al.* 1982, 1984; Mishkin and Petri 1994; Maunsell and van Essen 1983; Ungerleider and Mishkin 1982; Ungerleider 1985; Squire 1987; Squire and Zola-Morgan 1991; Kaas 1989).

(5) **The two trends of visual processing** (after Kaas 1989, modified from Ungerleider 1985)



In terms of processing function, the ventral trend is responsible for visual *object recognition*, categorizing objects *qua* types. That is, it is the analog of *lexical-semantic* representation of visual objects. The dorsal trend, in turn, is responsible for processing specific tokens of *visual events/states* as well as specific occurrence of individual objects *qua* tokens. That is, it is the analog of *episodic-declarative* representation of events/states.

In humans, both systems project backward (posteriorly) from distinct subcortical loci in the limbic system. The dorsal trend is projected from, and thus leads to, the *hippocampus*, site of both visual and verbal early episodic memory system. The ventral trend is projected from, and thus leads to, the *paleo-cortex* and *amygdala* (Tucker 1991).

What is more, from both subcortical areas to which the two trends lead, cortical frontal-lobe projections extend; respectively, to the longer-term anterior pre-frontal episodic storage locus (Squire and Zola-Morgan 1991), and to a pre-frontal of lexical-semantic locus (presumably the one identified by Petersen, Raichle and Posner and their associates).

In sum, both major cognitive representation systems that underlie human language — the lexical-semantic network (permanent semantic memory; Atkinson and Shiffrin 1968) and the propositional and multi-propositional representational (declarative-episodic memory; *ibid.*) are evolutionary extensions of the primate visual processing system.

In the earlier paper (Givón 1998), a number of reasons were given for conjecturing that the superficial coding of the human lexicon ('phonology') of human language began as a *visual-gestural* code; and that only later on, most likely gradually, it shifted toward the current *auditory-oral* code. This left grammar as the only major coding component of human language *not* to arise from precursors in the visual information-processing system. Rather, the earlier paper suggested, grammar was in large part an evolutionary extension of the control system of complex, rhythmic-hierarchic — syntactic — motor routines, located in Broca's area.

However, an important paper by Rizzolatti and Arbib (1998) suggests that even this component has firm visual-system precursors. In their review of activation patterns in Broca's area in humans and its F5 homolog in non-human primates, Rizzolatti and Arbib report that the primate F5/Broca's area has two types of cell. Those of the first type, *canonical F5 neurons*, are activated during the production of manual gestures. Those of the second type, *mirror neurons*, are activated during visual presentation of manual gestures by the experimenter, in the absence of any motor activity by the subject. The two systems can be co-activated, acting as a combined sensory-motor feedback and control system. But they can also be dissociated, with one being a system of *implementation*, the other of cognitive *representation*, of the same complex motor routine of manual action.

The evolutionary implications of this discovery are far reaching, and Rizzolatti and Arbib draw some of them explicitly. From our perspective here, their report adds one more piece of supporting evidence to the conjecture that the multimodule neurological apparatus that supports human language had its clear evolutionary precursors in various homologous modules in the primate visual information processing system.

3. Hypothesis

Given the wealth of evidence implicating various components of the primate visual information-processing system as precursors of various core components of human language, we were impelled to advance the hypothesis that the incredibly stable temporal rate of word and event-clause processing — ca. 250 msecs per word and ca. 1.0 sec. per clause — is itself a reflection of pre-linguistic constraints on visual information processing, of both visual objects and visual events/states.

To test this hypothesis, we designed an experiment that probes the temporal rate at which reliable cognitive representation of visual events and objects are formed. The independent variable we chose was the presentation speed of short event sequences, speeds ranging from 2 seconds to 35 msecs per visual state/event. As a dependent variable, we chose to measure the episodic memory traces of individual events/states, their correct sequence, and, implicitly, the participants ('objects').

The specific hypotheses we would like to offer at the start are:

> **Hypothesis I**: Episodic memory of individual events will decay rapidly at presentation rates below 1.0 second.
>
> **Hypothesis II**: Episodic memory of individual objects will decay rapidly at presentation speeds below 250 msecs.

4. Experiment I: event processing rates

4.1 Subjects

Fifty (50) subjects were recruited from the psychology department's undergraduates subject pool at the University of Oregon.

4.2 Stimuli and their presentation

Five separate cartoon-strip episodes were developed, each made out of 6 cartoons depicting highly concrete, simple and obvious visual state/events (Appendix 1), ordered in a coherent narrative sequence. Each episode had two central, active, animate participants plus 1–3 other objects or locations, none of them overlapping across stories. The cartoons were loaded on a computer hard-disk and presented in order at various speeds on the computer screen.

Each subject saw all five stories but each at a different presentation speed, to prevent referential contamination across stories. The presentation order of the 5 stories at various speeds was randomized (Appendix 2). For the first two speeds — 2.0, 1.0 — we obtained 6 subject-points per story. For the next three speeds — 0.500, 0.250, 0.125 — we obtained 11 subject-points per story. For the last two speeds — 0.064, 0.032 — we obtained 5 subject points per story.

The entire stimuli presentation sequence took, on the average, no more than 20 minutes. The presentation of each story was activated manually by the investigator. The schema of stimuli presentation to the 50 subjects, in terms of pairing of stories and speeds, is found in Appendix 2.

4.3 Instructions and post-viewing procedures

The subjects were instructed to attend to the computer screen. They were told that the experiment was about memory for visual stories, and were urged to attend to the presentations on the screen and memorize well what they saw (Appendix 3). At the end of the visual presentation of each story, the subjects were given a distractor task of about 3 minutes (counting backwards from 200), after which they were asked to tell all they could remember of the cartoon story they had just seen. Their verbal narratives were recorded and transcribed.

All subjects were given one practice trial of one story (see Appendix 1) at one speed 3.0 seconds per event frame. Their narration of the trial presentation was recorded but not analyzed.

4.4 Data coding and analysis

4.4.1 *Selection and identification of 'core' clauses*

In a previous experimental study of the verbal episodic memory of a 6.0-minute video presentation (Dickinson and Givón 1997), it was established that a group of verbal clauses in the narrative, termed *core* clauses, were recalled in an extremely stable fashion cross-subjects, and were very hard to dislodge from memory by various deceptive interventions by an interlocutor. These core clauses tended to depict concrete events that took place in sequential order in the stimulus video.

The core clauses contrasted with another group, termed *background* clauses, that were produced much less uniformly across subjects, and were much easier to dislodge from the recall narratives by various deceptive interventions. These background clauses tended to contain interpretations, perspective-laden observations, modal or evidential commentary, imputed mental states and motivation, or off-sequence static descriptions and elaborations.

In our coding procedures here, we took advantage of these earlier observations to select a group of core clauses for each episode, which then served as baseline for computing the effect of presentation speed on recall. We selected, for each story, the *five* (5) event clauses that appeared most frequently in the narratives collected at the two slowest presentation speeds (2.0 sec. and 1.0 sec.). The core clauses for all 5 stories are listed, in their *canonical* forms (see below), in Appendix 4.

4.4.2 *Computation of recalled events*

For each story, we first identified all verbal event/state clauses in the recorded narratives regardless of their grammatical form. Clauses with superficial grammatical variations were judged to be equivalent variants of the same canonical clause if they depicted substantially the same state/event type with the same participants. Thus for example, clauses (6b, c, d) below were all considered variants of the same 'canonical' existential-presentative clause (6a) that introduces the first participant:

(6) a. There was a penguin
    b. First I saw a penguin
    c. A little penguin was standing there
    d. It started with a penguin

Three types of subordinate clauses, relative clauses (7a) adverbial clauses (7b) and loose verb complements (7c) were counted as independent event clauses. In contrast, tightly bound verb complements with zero-marked equi-subjects, as in (7d), (7e), were not counted as independent clauses.

(7) a. I saw a penguin *standing next to a pond*
        (i) I saw a penguin
        (ii) The penguin was standing next to a pool
    b. *After he caught the fish*, he flipped it over his head
        (i) He caught the fish
        (ii) He flipped the fish over his head
    c. But the dog decides (that) *it wants to go for a walk*
        (i) the dog decides (something)
        (ii) the dog wanted to go for a walk
    d. It wanted *to go for a walk*
    e. He made the penguin *drop the fish*

Finally, reasonable variant interpretations of lexical items, be they the subject, object, verb or locative, were subsumed as variants of the same canonical clause. Thus compare:

(8) Object:

a. An elephant was chasing *a boy*
b. An elephant was chasing *a man*
c. The elephant was chasing *a little kid*

Subject/agent:

d. . . . and *a seal* then caught it
e. . . . and it was then caught by *an otter*
f. . . . and then *a sea lion* gets it

Verb:

g. The elephant *hits* the tree
h. The elephant *butts* the tree
i. The elephant *charged* the tree

Locative:

k. It was standing *next to a pond*
l. It was standing *near the water*
m. It was standing *at the beach*

For the purpose of computing the results of recalled events, we considered only core clauses, and expressed the results in terms of the average number and percent of core clauses recalled by subjects at each presentation speed.

4.4.3 *Computation of recalled participants*

Our experimental design for the recall of visual events has turned out to be rather unsuited for assessing the effect of presentation speed on the recall of event participants ('objects'). This is due to five factors that conspired to create potential interactions and/or confounds. First, the number of times a participant NP is referred to in each recalled narrative depends in part on how many core clauses — or cartoon frames — the participant appeared in.

Second, the number of times an event participant is recalled at each presentation speed is not wholly independent of the number of core-clauses that were recalled at that speed.

Third, the number of times an event participant is recalled also depends on whether it is an important active participant ('agent'), an important affected object ('patient'), or just an inanimate feature of the terrain ('locative').

Fourth, there appears to be a potential for a recency effect, so that at the highest presentation speed, the participant(s) appearing in the last-presented event frame are better recalled.

Fifth, the fact that participants appeared in many successive frames made it possible for subjects to recall them at least once even at the highest presentation speeds. This made the scoring of recalled participants by a simply binary method — "Was it or was it not mentioned *at all* as episode participant?" — rather impractical.

For these reasons, we chose to re-design the experiment to take account of all these difficulties, and will report results as *Experiment II*, below.

4.5 Results

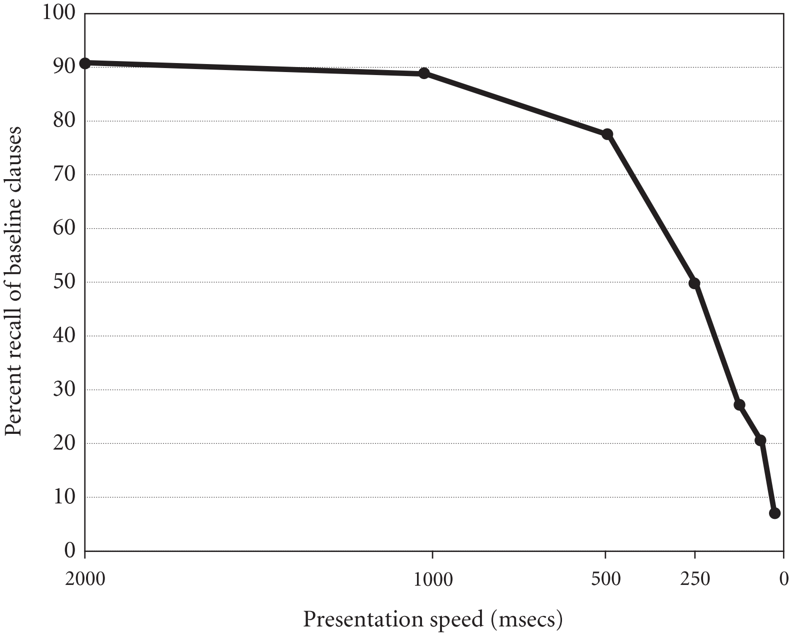
An example of the recorded narratives for one story (story #1) at all presentation speeds and one subject per speed is given in Appendix 5. The coding of recalled core clauses for all stories, speeds and subjects is given in Appendix 6. The number of clauses recalled per speed per subject for each story is given in Appendix 7. Those results, for the five stories at seven presentation speeds, were then collapsed over all subjects and expressed in terms of *average number* and *percent* of core-clauses recalled per subject, in (9) below.

(9) Average number and percent of core-clauses recalled per subject for all stories and all speeds

		Presentation speeds (in msecs)							
Story		2.0	1.0	0.500	0.250	0.125	0.065	0.032	Baseline
	N	4.2	4.2	2.8	2.0	1.5	1.8	0.6	5
1	%	80	80	56	40	30	36	12	100
	N	4.6	4.0	4.2	2.3	1.8	0.8	0.6	5
2	%	92	80	84	46	36	16	12	100
	N	5.0	4.3	4.7	3.4	1.3	0.8	0.2	5
3	%	100	86	94	68	26	16	4	100
	N	4.5	4.7	3.9	3.2	1.4	1.2	0.2	5
4	%	90	94	78	64	28	24	4	100
	N	4.6	4.8	3.3	1.7	0.8	0.2	0.0	5
5	%	92	96	66	34	16	4	0	100
Total	N	4.6	4.4	3.8	2.5	1.4	1.0	0.3	5
	%	92	88	76	50	28	20	6	100

The results of percent core-clauses recalled per subject, collapsed together for the five stories and plotted in a graph form against the various presentation speeds, are given in (10) below.

(10) Percent core-clauses recalled as function of presentation speed



4.6 Statistical analysis

A preliminary analysis was done to ensure that the subjects performed similarly across the five stories. A repeated-measures ANOVA with one between-subjects factor (story with five levels) showed no significant differences among the stories [F (4, 216) = 1.60, $p = 0.1759$] (Barker 2001).

For the main analysis, data from the two subject groups were analyzed separately. A repeated-measures ANOVA for the first subject group had one between-subject factor (presentation time) with five levels (2000, 1000, 500, 250, 125). The omnibus test was significant [F(4, 116) = 77.3; $p = <.0001$] (Barker 2001).

Polynomial contrasts were performed with spacing weighted according to presentation time at each level. The linear, quadratic and cubic trends are all significant. Linear [F(1, 29) = 33.89; $p = <.0001$]; quadratic: [F(1, 29) = 139.49; $p = <.0001$]; cubic: [F(1, 29) = 21.18; $p = <.0001$]. This suggests that at these five presentation-times, performance drops sharply after 1000 msecs along an S-shaped curve as presentation-time decreases (Barker 2001).

The data from the second group of subjects was submitted to the same repeated-measures ANOVA with one between-subjects factor (presentation time) with five levels (500, 250, 125, 064, 032). The omnibus test was significant [$F(1,24)=33.89$; $p=<.0001$] (Barker 2001).

Polynomial contrasts were performed with spacings weighted according to presentation-time at each level. The linear and quadratic trends were significant: Linear [$F(1,24)=147.28$; $p<.0001$]; quadratic: [$F(1,24)=13.20$; $p=.0013$]. This suggest that over this lower range of presentation-times, performance drops off in a downward-sloping parabolic curve as presentation-time decreases. The cubic trend was not significant (Barker 2001).

The over-all results suggest a major break at recall level occurring betwee 1000 msecs and 500 msecs presentation times, with performance dropping off exponentially below 1000 msecs, as predicted by Hypothesis I.

5. Experiment II: participants ('objects') procession rates

5.1 Subjects

Thirty five (35) subjects were recruited from the psychology department's undergraduate subject pool at the University of Oregon, as well as from one linguistics class.

5.2 Stimuli and their presentation

Fourteen (14) different cartoons were drawn (the first seven adapted from experiment I, above), each depicting highly concrete, simple and obvious visual events (Appendix 8), with no overlap in participants across the 14 cartoons. Each cartoon had two central, active, animate participants, and in some 1–2 other inanimate objects or locations. The cartoons were loaded on a computer hard-disk and presented on the computer screen at seven randomized orders, using the same seven speeds as in experiment I, above.

Each subject saw all 14 cartoons but each at a different presentation speed. The seven randomized presentation orders ('stimuli blocks') of the 14 cartoons are shown in Appendix 9. There were five subject-points per speed for each cartoon, thus 70 subject points per presentation-speed for the entire set of 14 cartoons.

The entire stimuli presentation sequence took, on the average, up to 30 minutes. The presentation of each cartoon was activated manually by the investigator. The schema of stimuli presentation to the 35 subjects, in terms of pairing of cartoons/speeds with subjects, is given in Appendix 10.

5.3 Instructions and post-viewing procedures

The subjects were instructed to attend to the computer screen. They were told that the experiment was about memory for visual events, and were urged to attend to the presentations on the screen and memorize well what they saw. At the end of the visual presentation of each cartoon, the subjects were given a distractor task of about 2 minutes (counting backwards by threes from anywhere above 200), after which they were asked to tell all they could remember of the cartoon they had just seen. Their verbal narratives were recorded and transcribed (see sample narratives and coding decisions is Appendix 12).

5.4 Data coding and analysis

5.4.1 *Selection of base-line participants for each event*

For each cartoon, we used the slowest presentation speed (2000 msecs) to identify the 2–4 participants that were recalled by at least 4 out of the five subjects that saw the cartoon at that speed. Those participants ('visual objects') were considered the *base-line* 100% recall level for that event, relative to which the rest of the results were then expressed as percentage.

5.4.2 *Lexical variation in identifying recalled participants*

Since the interpretation of drawn cartoons allows some variation in the use of plausible alternative lexical nouns, we allowed for each participant a range of reasonable lexical alternatives, all attested in the actual transcripts. Thus, for example, we counted the following as equally plausible alternatives, given the level of detail of the drawings and the cultural literacy of the subjects:

(11) a. bunny, rabbit, bunny-rabbit
b. ball, football
c. cat, mountain-lion, cougar, puma, jaguar, leopard, tiger, cheetah, lion,
d. boy, man, farmer, guy, Amish-guy
e. dragon-fly, fire-fly, moth, fly
f. snake, worm
g. crab, lobster
h. eagle, hawk, vulture, raptor
i. wagon, carriage, cart, buggy, coach, chariot
j. ocean, water, river, shore

For a number of cartoons, only the specific context of the depicted event determines the reasonableness of allowed variants. Thus, for example:

(12) a. dog, puppy, puppy-dog cat, kitten, kitty-cat
CONTEXT: girl putting small pet in box
b. bird-seed, bird-food, food, bread-crumbs, feed, popcorn, feeding
CONTEXT: girl throwing feed to birds
c. tree, branch, limb
CONTEXT: boy sitting on limb up a tree

The base-line participants all of 14 cartoons, including all accepted variants, are given in Appendix 11.

5.5 Results

The aggregated results for the 14 cartoons over all subjects are expressed in (13) below, as the *average number* and *percent* of base-line participants recalled at each presentation speed.

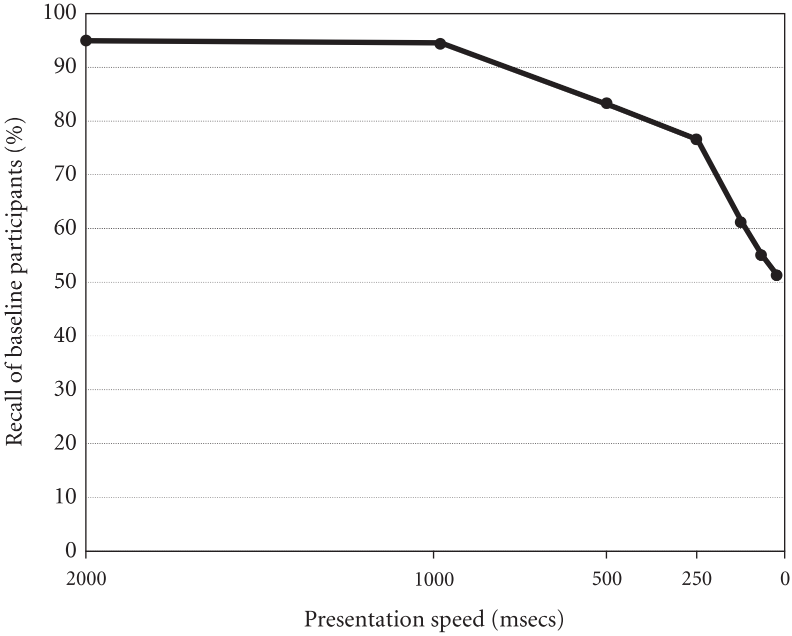
(13) Number and percent of base-line participants recalled at all speeds for all cartoons (collapsed for all subject)

		Presentation speed (msecs)							
Cartoon		2.0	1.0	0.500	0.250	0.125	0.065	0.032	Baseline
	N	15	14	14	10	8	4	5	15
1	%	100	93	93	66	53	26	33	100
	N	10	10	10	6	6	5	5	10
2	%	100	100	100	60	60	50	50	100
	N	10	9	7	8	9	7	5	10
3	%	100	90	70	80	90	70	50	100
	N	15	15	10	11	7	7	7	15
4	%	100	100	66	73	46	46	46	100
	N	17	17	15	18	14	16	14	20
5	%	85	85	75	90	70	80	70	100
	N	15	15	15	15	12	8	11	15
6	%	100	100	100	100	80	53	73	100
	N	15	15	15	14	11	14	15	15
7	%	100	100	100	93	73	93	100	100
	N	20	19	17	15	10	5	8	20
8	%	100	95	85	75	50	25	40	100

	N	10	9	6	8	6	0	0	10
9	%	100	90	60	80	60	0	0	100
	N	14	14	9	10	2	7	7	15
10	%	93	93	60	66	13	46	46	100
	N	15	15	12	10	11	9	8	15
11	%	100	100	80	66	73	60	53	100
	N	12	15	15	15	12	12	8	15
12	%	80	100	100	100	80	80	53	100
	N	7	9	7	4	3	4	3	10
13	%	70	90	70	40	30	40	30	100
	N	10	10	10	6	8	9	5	10
14	%	100	100	100	60	80	90	50	100
Total:	N	185	186	162	150	119	107	101	195
	%	95	95	83	77	61	55	52	100

The percent of base-line participants recalled at each presentation speed, taken from the totals in table (13), is plotted in a graph in (14).

(14) Percent core-clauses recalled as function of presentation speed



ANOVA tests were applied to the 70 trials at the 7 presentation times and showed the means to be significant (F=34.21; p=0.000). Fisher's Least Significant Difference post-hoc test was applied to the 7 means. Two steps along the presentation-speed curve — from 1.0 sec. to 5000 msecs and from 250 msecs to 125 msecs, prove significant. The latter conforms to our predictions. The former suggests a possible an interaction between participant ('object') recognition and event recognition. In hindsight, we find this a natural possibility, given that participants/objects are indispensable perceptual components of visual events.

6. Discussion

6.1 The fit of results to predictions

The results we report above fit, substantially if not absolutely, the predictions of both Hypotheses I and II. First, the significant sharp decline in the recall of visually-presented events in our experiment #I begins, as predicted, at speed 500 msecs. Second, with significant decline in the recall of base-line participants in experiment #II already evident at speed 500 msecs, the other sharp decline begins, as predicted, at speed 125 msecs.

6.2 Appropriateness of the methodology

Was our methodology appropriate for assessing the effect of presentation speed on the recall of visual events and participants ('objects')? Before answering, we had better recast this into two separate questions.

First, why did we propose to assess the speed of event processing by measuring the episodic recall of entire episodes rather than isolated visual events? Our answer to this is that in their valid adaptive context, events are seldom present in isolation, but rather as parts of coherent *episodes*. The significant task in creating a cognitive representation of an event is thus not merely a matter of representing the isolated event, but rather representing it in its natural adaptive context. In the same vein, in our experiment 2 we did not present visual objects in isolation, but rather as participants in integrated visual events. Again, we contend that this is the appropriate adaptive context within which visual objects are perceived, interpreted and stored.

The results of experiment 2 for the recall of full events — verb *and* participants — are yet to be analyzed. But the statistically significant difference between speeds 1.0sec and 500msec in the recall of participant-objects may suggest a possible interaction between event perception and participant-object perception.

Second, it is well known that the strictly-visual early phase of encoding both visual objects and events takes place much faster than the 1.0–.250 second range we have been concerned with, so that at presentation speeds as low as 10–15 msecs — far below the threshold of conscious recall — a lasting trace of a visually-presented information is retained and has visual priming effects (Bar and Biderman 1998; Subramaniam *et al.* 2000; Maunsell and Gibson 1992; Rolls *et al.* 1991; Ullman 1994). In the same vein, Treisman and Kanwisher (1998), in their review of subliminal vs. conscious visual object perception, make the following observation:

> . . . a consensus is developing that the critical processes involved in object recognition are remarkably fast, occurring within 100–200 ms of stimulus presentation. However, it may take another 100ms for subsequent processes to bring this information into awareness . . . (Treisman and Kanwisher 1998:219; see also DeSchepper and Treisman 1996; Treisman and DeSchepper 1996; Rock 1998)

The speed of *conscious* visual object recognition is, it appears, pegged firmly at the very same 250msec range we have predicted from the language processing evidence, and obtained in our experiment #II. This, rather than the much faster pre-conscious visual object processing, is the relevant processing speed for creating meaningful cognitive representations — semantic interpretations — of visually-perceived objects, the mental representation that can be accessed for the purpose of *conscious interpretation* and, in particular, appropriate *motor response.*

6.3 Why these particular processing rates?

Our results do not *prove* that the rates of processing of language-coded objects and events are an evolutionary outgrowth of the pre-linguistic rate of visual object and event processing. They only show that the two types of processing rates are substantially the same. Given the evolutionary precedence of visual information processing over language, an evolutionary interpretation of the amazing coincidence of the two types of rates is an eminently respectable theoretical gambit. Otherwise, the coincidence of four rhythms:

- verbal clause processing
- clausal-contexts effects on semantic priming
- ERP activation of Wernicke's area
- conscious/episodic recall of visual events

will remain just that — an inexplicable coincidence. Just as the coincidence of the four other rhythms:

- verbal word processing
- automatic activation of multiple word-senses
- ERP activation of the pre-frontal semantic area
- conscious/episodic recall of visual objects will remain another strange coincidence.

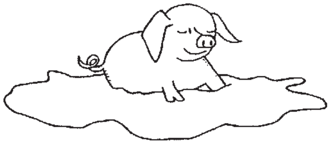
But why do event and object cognition proceed under *these* particular temporal constraints? Our tentative answer to this has already been hinted at in the preceding discussion. The rhythm of event and object cognition must have evolved in the ecological context of the need to respond — with appropriate action — to adaptively significant objects and events.

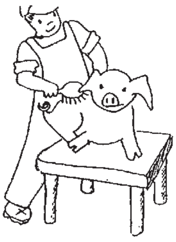
In the case of hominid primates, the relevant adaptive context must have been our species' ecological niche as *both* hunter and prey. Thus, the creation of a coherent cognitive representation of objects and events must have been timed, at the crucial upper limit, to the speed of motion, or perhaps reaction-time, of both relevant predators and relevant prey.

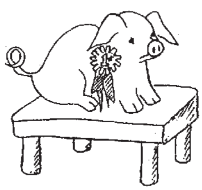
Appendix 1

Cartoon-strip stories, experiment 1

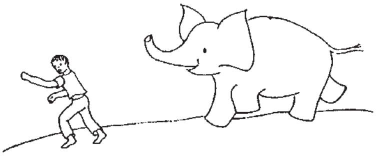
Trial story

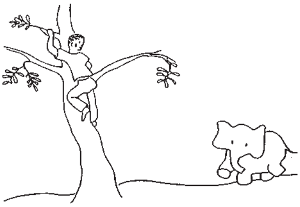


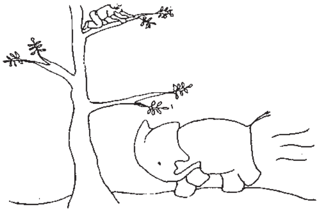


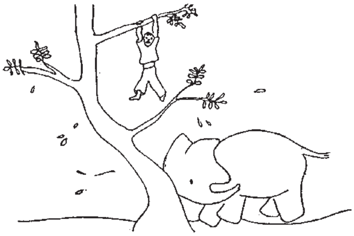


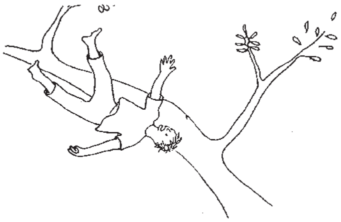
Story #1

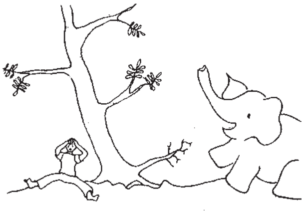




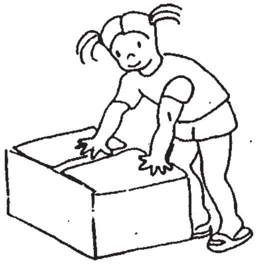


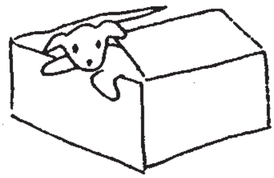


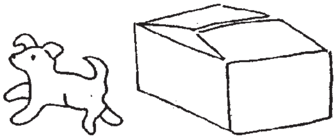


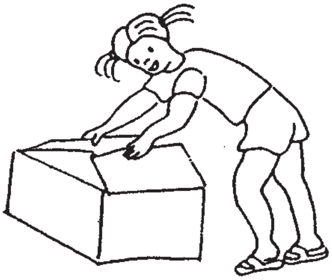


Story 2

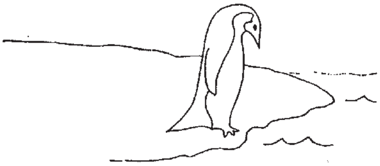


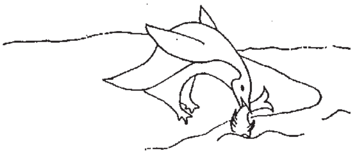


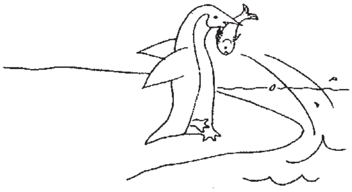


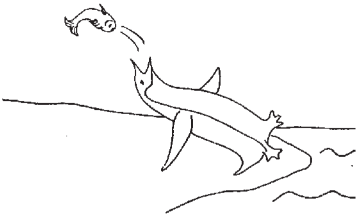


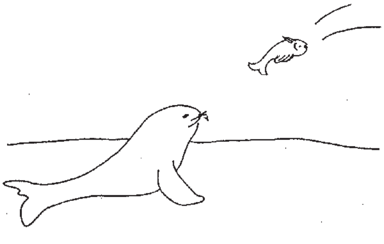
Story 3

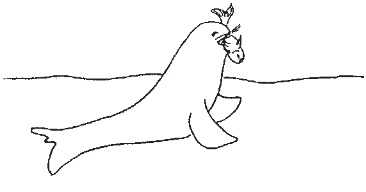




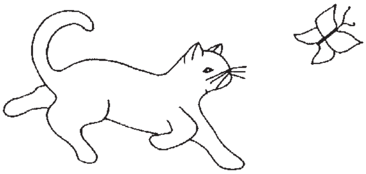


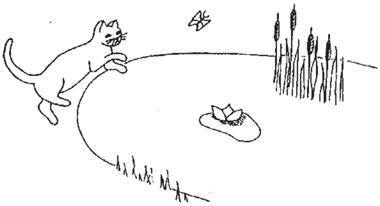


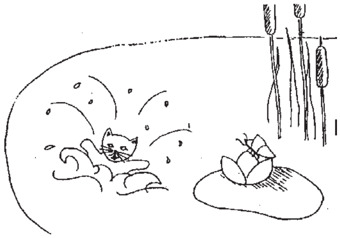




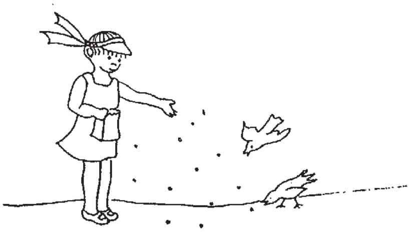
Story 4

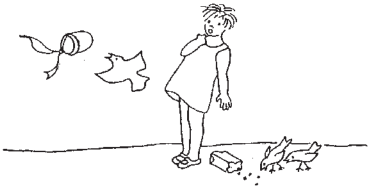


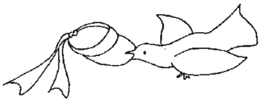




Story 5







Appendix 2

Schema of stimuli presentation for experiment 1: randomized presentation speeds for the five stories

Subject group 1 (subj. 1–25)

Story and speed combinations

Story	A	B	C	D	E
1	0.125	0.5	1.0	0.250	2.0
2	2.0	0.125	0.250	0.5	1.0
3	0.25	2.0	0.125	1.0	0.5
4	0.5	1.0	2.0	0.125	0.25
5	1.0	0.25	0.5	2.0	0.125

Story-speed Combinations A,B,C,D,E are assigned to five subjects in linear order, repeating five times in the first group of 25 subjects. (1–5, 6–10, 11–15, 16–20, 21–5)

Subject group 2 (subj. 26–50)

Story and speed combinations

Story	A	B	C	D	E
1	0.125	0.5	0.032	0.25	0.064
2	0.064	0.125	0.25	0.5	0.032
3	0.25	0.064	0.125	0.032	0.5
4	0.5	0.032	0.064	0.125	0.25
5	0.032	0.25	0.5	0.064	0.125

Story-speed Combinations A, B, C, D, E are assigned to five subjects in linear order, repeating five times in the second group of 25 subjects. (26–30, 31–5, 36–40, 41–5, 46–50)

APPENDIX 3

Instructions to subjects, experiment 1

You are going to see six stories on the computer screen. Each story is made out of six simple line drawings, like cartoons, that tell a story. After each story I will ask you to count backwards aloud; I will tell you when to stop counting by saying: “You can start telling me the story now”. At that point, I will turn on the tape-recorder and I’d like you to tell me as much as you can remember of the story you just saw, in as much detail as possible. The pictures in the different stories will be shown at different speeds; this means you’ll probably remember more of some stories than of others. That’s fine. The first story, the first set of pictures, is a trial run; so if you need to, you’ll be able to ask questions at that time. Do you have any questions now?

Appendix 4

Core clauses for the five stories (given in canonical forms), experiment 1

Story 1: a. Butterfly lands on cat
b. Cat chases butterfly
c. Butterfly goes over pond
d. Cat jumps after butterfly
e. Cat lands in water

Story 2: a. Penguin (stands) by water
b. Penguin catches fish
c. Fish flips out of penguin's mouth
d. Fish flips toward seal
e. Seal catches/eats fish

Story 3: a. Girl puts dog in box
b. Dog jumps out of box
c. Dog walks away
d. Girl looks in box
e. No dog in box

Story 4: a. Elephant chasing man
b. Man climbs tree
c. Elephant hits tree
d. Man falls from tree
e. Man is hurt

Story 5: a. Girl feeding birds
b. Winds blows
c. Wind snatches (girl's) hat
d. Bird grabs hat
e. Bird flies to tree (with hat)

APPENDIX 5

Example of transcribed recalled narratives for story 1 (one subject per speed), experiment 1

2.0 sec/frame: A cat is asleep, and it wakes up and sees a butterfly, ‘n the butterfly eventually lands, on, a, lily-pad, in a lake, and the cat, uh, is chasing the butterfly and falls into the lake, gets wet, and, is angry.

1.0 sec/frame: Uh, there’s a cat, laying on, a, a mat, a butterfly comes, the cat sees the butterfly . . . I believe, and the butterfly lands on the cat’s head, and the cat, begins to chase the butterfly, um, and the cat falls, in some water . . . the end.

0.5 sec/frame: OK, there was a cat that was sleeping, and a butterfly came and it started, chasing the butterfly, and it, fell into a big pond of water.

0.25 sec/frame: Uh, so, there’s a cat, running along sees a butterfly, jumps after it and falls into a lake.

0.125 sec/frame: Um, a cat was kind-of in a swamp, type thing, and, was chasing a bug? That’s all . . . I remember.

0.064 seconds/frame: OK, there was a, cat, and it was trying to catch a butterfly out in a field.

0.032 sec/frame: Well the cat, got all jumbled up and was chasing, a mouse and got, all tangled up in the strings and it was really chaotic, scene for the cat.

APPENDIX 6

Coding of recalled core clauses for all stories, presentation speeds and subjects, experiment I

	2 sec						total	mean	1 sec						total	mean	0.5 sec											total	mean	0.25 sec											total	mean	0.125 sec											total	mean	0.0625 sec					total	mean	0.03125 sec					total	mean
Story 1																																																																					
butterfly lands on cat	0	1	1	1	1	0	4		1	1	1	1	0	1	5		0	0	0	1	0	0	1	0	0	0	0	2		0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0		0	0	0	0	0	0	
cat chases butterfly	1	1	1	1	1	1	6		1	1	1	1	1	1	6		1	1	0	1	1	1	1	1	1	1	1	10		1	1	1	1	1	1	1	1	1	1	1	11		1	1	1	1	1	1	1	1	1	1	1	11		1	1	1	1	1	5		1	0	0	1	0	2	
butterfly goes over pond	0	0	1	1	1	1	4		1	1	0	1	1	1	5		0	0	0	1	0	1	0	0	0	1	0	3		0	0	1	1	0	0	0	0	0	0	0	2		1	1	0	0	0	0	0	0	0	0	0	2		0	0	0	0	0	0		0	0	0	0	0	0	
cat jumps after butterfly	1	1	1	1	1	0	5		1	0	0	0	1	1	3		1	1	0	0	1	1	0	0	1	1	1	7		0	1	0	0	0	0	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	1	0	1		0	0	0	0	0	0	
cat lands in water	1	1	1	1	1	1	6		1	4	1	1	1	1	6		1	1	0	1	1	1	0	1	1	1	1	9		1	1	1	0	1	1	1	0	1	0	1	8		0	1	1	0	0	0	1	1	0	0	0	4		1	0	1	1	0	3		0	0	0	1	0	1	
	3	4	5	5	5	3	25	4.2	5	4	3	4	4	5	25	4.2	3	3	0	4	3	4	2	2	3	4	3	31	3.0	2	3	3	2	2	2	2	1	2	1	2	22	1.7	2	3	2	1	1	1	2	2	1	1	1	17	1.3	2	1	2	3	1	9	1.7	1	0	0	2	0	3	0.8
Story 2																																																																					
penguin by some water	1	1	1	1	1	1	6		1	0	1	1	1	0	4		1	1	1	1	1	0	1	1	0	1	1	9		0	0	0	0	0	1	0	0	0	0	0	1		0	0	0	1	0	0	1	0	0	0	1	3		1	0	0	0	0	1		0	0	0	0	0	0	
penguin catches fish	1	1	1	1	1	1	6		1	1	1	1	1	0	5		1	1	1	1	1	1	1	1	1	1	1	11		1	1	1	1	0	1	1	1	1	1	1	10		1	1	0	0	1	1	1	0	0	0	1	6		0	0	0	0	0	0		0	0	0	0	0	0	
fish flips out of penguin's mouth	1	1	1	1	1	1	6		1	1	1	1	1	1	6		1	1	1	1	1	1	1	1	1	1	1	11		1	0	0	1	0	0	0	1	1	1	1	6		0	0	0	0	1	0	1	0	0	0	1	3		1	0	0	0	0	1		0	1	0	1	0	2	
fish flips to seal	0	1	1	1	1	1	5		0	0	1	1	1	0	3		1	0	0	1	1	1	0	1	1	1	0	7		0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	1	0	0	0	1	2		0	0	0	0	0	0		0	0	0	0	0	0	
seal catches/eats fish	1	1	1	0	1	1	5		1	1	1	1	1	1	6		0	1	1	1	1	0	1	1	0	1	1	8		1	1	1	1	0	0	0	1	1	1	1	8		0	0	0	1	1	0	1	1	0	1	1	6		1	0	0	0	1	2		0	0	0	1	0	1	
	4	5	5	4	5	5	28	4.7	4	3	5	5	5	2	24	4.0	4	4	4	5	5	3	4	5	3	5	4	46	4.0	3	2	2	3	0	2	1	3	3	3	3	25	2.5	1	1	0	2	3	1	5	1	0	1	5	20	2.2	3	0	0	0	1	4	1.0	0	1	0	2	0	3	0.7
Story 3																																																																					
girl puts dog in box	1	1	1	1	1	1	6		1	1	1	1	1	1	6		1	1	1	1	1	1	1	1	1	1	1	11		1	0	1	1	1	1	1	1	0	1	1	9		0	0	0	0	0	0	0	0	1	0	0	1		0	0	0	0	0	0		0	0	0	0	0	0	
dog jumps out	1	1	1	1	1	1	6		1	1	1	1	1	0	5		1	1	1	1	1	1	1	1	1	1	1	11		1	1	1	1	1	1	1	1	1	0	1	10		0	1	1	1	0	1	1	0	0	1	1	7		1	0	0	0	1	2		1	0	0	0	0	1	
dog walks away	1	1	1	1	1	1	6		1	1	1	0	1	1	5		0	0	1	1	1	0	0	0	0	0	0	3		0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	1	1		0	0	0	0	1	1		0	0	0	0	0	0	
girl looks in box	1	1	1	1	1	1	6		1	1	1	0	1	1	5		1	1	1	1	1	1	1	1	1	0	1	10		1	0	1	1	1	1	1	1	0	1	1	9		0	0	0	0	0	0	0	0	1	1	1	3		0	0	1	0	0	1		0	0	0	0	0	0	
dog gone/box empty	1	1	1	1	1	1	6		1	1	1	0	1	1	5		1	1	1	1	1	1	1	0	1	1	1	10		1	1	1	1	1	1	1	1	0	1	0	9		0	0	0	0	0	0	0	0	0	1	1	2		0	0	0	0	0	0		0	0	0	0	0	0	
	5	5	5	5	5	5	30	5.0	5	5	5	2	5	4	26	4.3	4	4	5	5	5	4	4	3	4	3	4	45	3.7	4	2	4	4	4	4	4	4	1	3	3	37	3.2	0	1	1	1	0	1	1	0	2	3	4	14	1.8	1	0	1	0	2	4	1.0	1	0	0	0	0	1	0.3
Story 4																																																																					
elephant chasing man	1	0	1	1	1	0	4		1	1	1	1	1	1	6		1	1	1	1	0	1	1	1	1	1	1	10		1	1	1	0	0	0	0	0	1	1	1	6		0	0	1	0	0	0	0	0	1	0	0	2		1	0	0	0	1	2		0	0	0	0	0	0	
man climbs tree	1	1	1	1	1	1	6		1	1	1	1	1	1	6		1	1	1	1	1	1	1	1	1	1	1	11		1	1	1	1	1	0	1	1	1	1	1	10		0	0	1	0	1	0	1	0	1	1	1	6		1	1	1	0	1	4		0	0	0	0	0	0	
elephant hits tree	1	1	1	1	1	1	6		1	1	0	1	1	1	5		1	1	0	1	1	1	1	1	0	1	1	9		0	0	1	1	1	1	1	1	1	1	1	9		0	0	0	0	0	0	1	1	1	0	0	3		0	0	0	0	0	0		0	0	0	0	0	0	
man falls	1	1	1	1	1	1	6		1	1	1	1	1	1	6		1	1	1	1	1	1	1	1	1	1	1	11		1	0	1	1	1	0	1	1	1	1	0	8		0	0	0	0	0	0	1	1	1	0	1	4		0	0	0	0	0	0		1	0	0	0	0	1	
man is hurt	1	1	1	0	1	1	5		1	1	1	1	1	0	5		1	0	0	0	0	0	0	0	0	1	0	2		0	0	0	0	0	0	0	1	1	0	0	2		0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0		0	0	0	0	0	0	
	5	4	5	4	5	4	27	4.5	5	5	4	5	5	4	28	4.7	5	4	3	4	3	4	4	4	3	5	4	43	4.0	3	2	4	3	3	1	3	4	5	4	3	35	3.3	0	0	2	0	1	0	3	2	4	1	2	15	2.0	2	1	1	0	2	6	1.3	1	0	0	0	0	1	0.4
Story 5																																																																					
girl feeding birds	1	1	1	1	1	1	6		1	1	1	1	1	1	6		1	1	0	1	1	1	0	1	1	1	0	8		0	0	0	1	1	0	1	0	1	1	1	6		0	0	0	0	1	0	0	0	0	0	1	2		0	0	0	0	0	0		0	0	0	0	0	0	
wind comes up	0	0	0	1	1	1	3		0	1	1	1	1	1	5		0	1	0	1	0	1	0	0	0	0	0	3		0	0	0	0	1	0	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0		0	0	0	0	0	0	
wind takes hat	1	1	1	1	1	1	6		1	1	1	1	1	1	6		0	1	1	1	1	1	1	0	0	1	1	8		0	0	0	0	1	0	1	0	0	0	0	2		0	0	0	0	0	0	0	0	0	0	0	0		0	0	1	0	0	1		0	0	0	0	0	0	
bird grabs hat	1	1	1	1	1	1	6		1	1	1	1	1	1	6		1	1	1	1	0	1	1	1	1	1	1	10		1	0	1	0	1	1	0	0	0	0	1	5		0	1	0	0	1	0	0	0	0	0	0	2		0	0	0	0	0	0		0	0	0	0	0	0	
bird flies to tree	1	1	1	1	1	1	6		1	1	1	1	1	1	6		1	1	0	0	0	1	1	0	1	1	1	7		1	0	1	0	1	0	1	0	1	0	0	5		0	1	0	1	1	1	0	0	0	0	1	5		0	0	0	0	0	0		0	0	0	0	0	0	
	4	4	4	5	5	5	27	4.5	4	5	5	5	5	5	29	4.8	3	5	2	4	2	5	3	2	3	4	3	38	3.3	2	0	2	1	5	1	3	0	2	1	2	19	1.5	0	2	0	1	3	1	0	0	0	0	2	9	0.5	0	0	1	0	0	1	0.3	0	0	0	0	0	0	0.0
							137	4.6							132	4.4												201	3.6												138	2.4												75	1.6						24	1.1						8	0.4

APPENDIX 7

Number of recalled core clauses for each story, presentation speed and subject, experiment I

Story 1

Subject	2.0	1.0	0.5	0.25	.125	.064	.032
1	3	5	3	2	2	2	1
2	4	4	3	3	3	1	0
3	5	3	0	3	1	2	0
4	5	4	4	2	1	3	2
5	5	4	3	2	1	1	0
6	3	5	4	2	1	/	/
7	/	/	2	2	2	/	/
8	/	/	2	1	2	/	/
9	/	/	3	2	1	/	/
10	/	/	4	1	1	/	/
11	/	/	3	2	1	/	/
Average:	4.2	4.2	2.8	2.0	1.5	1.8	0.6
Percent:	80	80	56	40	30	36	12

Story 2

Subject	2.0	1.0	0.5	0.25	.125	.064	.032
1	4	4	4	3	1	3	0
2	5	3	4	2	1	0	1
3	5	5	4	2	0	0	0
4	4	5	5	3	2	0	2
5	5	5	5	0	3	1	0
6	5	2	3	2	1	/	/
7	/	/	4	1	5	/	/
8	/	/	5	3	1	/	/
9	/	/	3	3	0	/	/
10	/	/	5	3	1	/	/
11	/	/	4	3	5	/	/
Average:	4.6	4.0	4.2	2.3	1.8	0.8	0.6
Percent:	92	80	84	46	36	16	12

Story 3

Subject	2.0	1.0	0.5	0.25	.125	.064	.032
1	5	5	4	4	0	1	1
2	5	5	4	2	1	0	0
3	5	5	5	4	1	1	0
4	5	2	5	4	1	0	0
5	5	5	5	4	0	2	0
6	5	4	4	4	1	/	/
7	/	/	4	4	1	/	/
8	/	/	3	4	0	/	/
9	/	/	4	1	2	/	/
10	/	/	3	3	3	/	/
11	/	/	4	3	4	/	/
Average:	5.0	4.3	4.7	3.4	1.3	0.8	0.2
Percent:	100	86	94	68	26	16	4

Story 4

Subject	2.0	1.0	0.5	0.25	.125	.064	.032
1	5	5	5	3	0	2	1
2	4	5	4	2	0	1	0
3	5	4	3	4	2	1	0
4	4	5	4	3	0	0	0
5	5	5	3	3	1	2	0
6	4	4	4	1	0	/	/
7	/	/	4	3	3	/	/
8	/	/	4	4	2	/	/
9	/	/	3	5	4	/	/
10	/	/	5	4	1	/	/
11	/	/	4	3	2	/	/
Average:	4.5	4.7	3.9	3.2	1.4	1.2	0.2
Percent:	90	94	78	64	28	24	4

Story 5

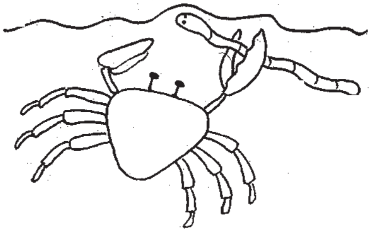
Subject	2.0	1.0	0.5	0.25	.125	.064	.032
1	4	4	3	2	0	0	0
2	4	5	5	0	2	0	0
3	4	5	2	2	0	1	0
4	5	5	4	1	1	0	0
5	5	5	2	5	3	0	0
6	5	5	5	1	1	/	/
7	/	/	3	3	0	/	/
8	/	/	2	0	0	/	/
9	/	/	3	2	0	/	/
10	/	/	4	1	0	/	/
11	/	/	3	2	2	/	/
Average:	4.6	4.8	3.3	1.7	0.8	0.2	0.0
Percent:	92	96	66	34	16	4	0

Appendix 8

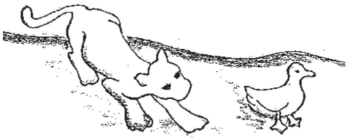
Stimuli event cartoons, experiment II

①

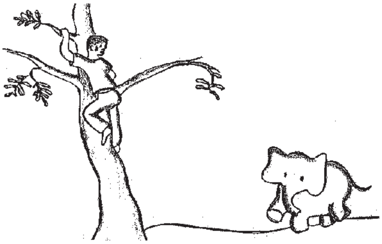
②



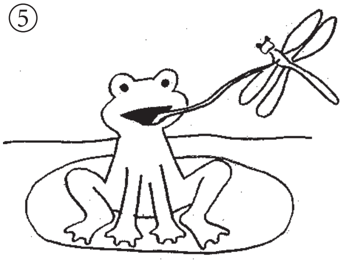
③



④

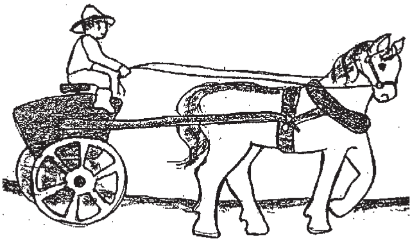


⑤

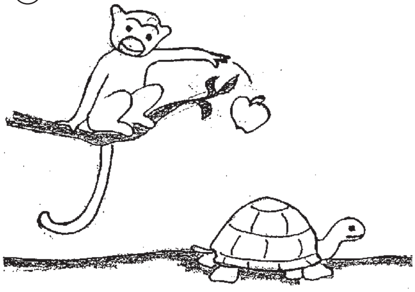


⑥

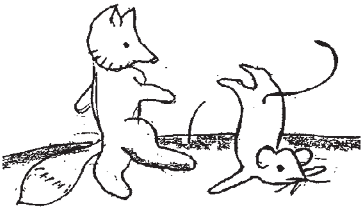
⑦



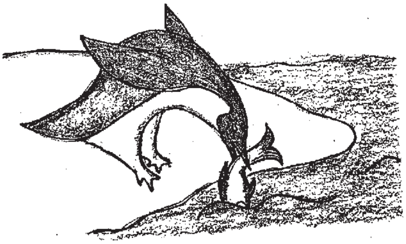
⑧



⑨

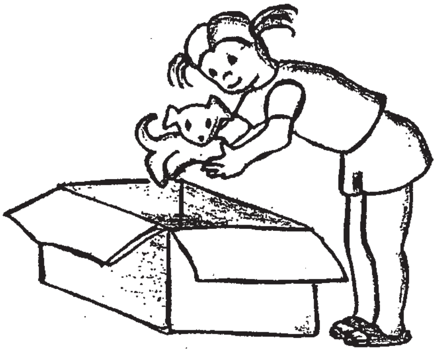


⑩

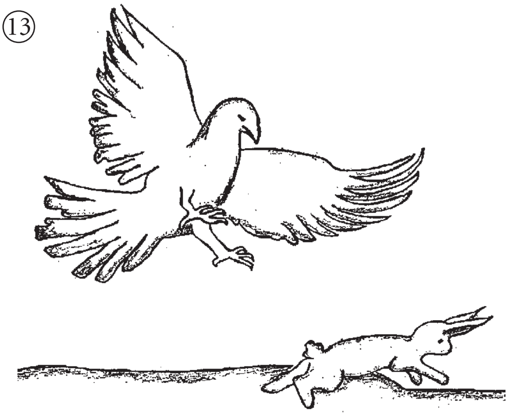


⑪

⑫



⑬



⑭

Appendix 9

Randomized presentation orders of the fourteen stimuli cartoons, experiment II

	Stimuli blocks						
	Speeds						
Picture	A	B	C	D	E	F	G
1	2000	1000	500	250	125	064	032
2	1000	500	250	125	064	032	2000
3	500	250	125	064	032	2000	1000
4	250	125	064	032	2000	1000	500
5	125	064	032	2000	1000	500	250
6	064	032	2000	1000	500	250	125
7	032	2000	1000	500	250	125	064
8	2000	1000	500	250	125	064	032
9	1000	500	250	125	064	032	2000
10	500	250	125	064	032	2000	1000
11	250	125	064	032	2000	1000	500
12	125	064	032	2000	1000	500	250
13	064	032	2000	1000	500	250	125
14	032	2000	1000	500	250	125	064

APPENDIX 10

The pairing of stimuli blocks (presentation orders) with subjects, experiment II

Subjects nos	Stimulus block
1,8,15,22,29	A
2,9,16,23,30	B
3,10,17,24,31	C
4,11,18,25,32	D
5,12,19,26,33	E
6,13,20,27,34	F
7,14,21,28,35	G

APPENDIX 11

List of base-line participants and accepted variants for the 14 cartoons, experiment II

Picture	Base-line participants	Total
1	rabbit/bunny, (foot)ball, (polar)bear	(3)
2	crab/lobster, snake/worm	(2)
3	cat/mountain-lion/cougar/puma/lion/ tiger/jaguar/leopard/panther/cheeta, duck(ling)	(2)
4	boy/man/guy, tree, elephant	(3)
5	frog, lily-pad/leaf, tongue, dragon-fly/fire-fly, moth, fly	(4)
6	girl, birds/ducks/geese/pigeons, food/seeds/feed/bread/crumbs/ popcorn/feeding	(3)
7	guy/boy/male/man/driver, mule/horse,carriage/cart/wagon/buggy/ coach/chariot	(3)
8	monkey, tree/branch/limb, apple,turtle/tortoise	(4)
9	fox/mouse/raccoon/hedgehog/mink, mouse/fox/raccoon/hedgehog/rodent	(2)
10	bird/penguin, fish, ocean/water/river/shore	(3)
11	boy/man/guy/farmer, blow-dryer/hair-drying, pig	(3)
12	girl, dog/puppy/cat/kitten, box	(3)
13	eagle/hawk/vulture/raptor, bunny/rabbit	(2)
14	cat/kitten, butterfly	(2)

Appendix 12

Sample recall texts of cartoon #1 at all presentation speeds, one subject per speed, experiment II

2000: I just saw a *rabbit*, like rabbit kicking a *football*, like, toward a *polar-bear*. (coded 3/3)

1000: There was two charaters, one looked like a *bunny*, and the other one looked like a *bear* and the bunny was kicking a *football* to the other character. (coded 3/3)

500: I think it was a *bunny* kicking a *football* at a *bear*. (coded 3/3)

250: I saw a *polar-bear*, and, a *bunny-rabbit* I think kicking a *football* at the polar-bear. (coded 3/3)

125: A *polar-bear* and a *rabbit* maybe a, could it have been a little, guy with a *balloon* in his hand and the, polar-bear, was jumping, at it or in it? (coded 2/3)

064: Uh, I saw a *bear*, and, *a little man* holding a *balloon*. (coded 1/3)

032: I saw *two polar bears* a large one on the left and a small one on the right, and they were stencil-drawings with sort of a light background an a, um, they were drawn in sort of a dark color. (coded 1/3)

Note

* We are indebted to Ed Awh, Mike Posner, Bertram Malle, Don Tucker and Eric Pederson for many helpful comments and suggestions, and to Linda Forrest for her generous help in the statistical analysis of experiment #II. The responsibility for the final product remains strictly our own.

References

Abdulaev, Y. G. and Posner, M. I. 1997. "Time course of activating brain areas in generating verbal associations". *Psychological Science* 8.

Abdulaev, Y. G. and Posner, M. I. 1998. "Event-related brain potential imaging of semantic encoding during processing single words". *Neuroimage* 7.

Atkinson, R. C. and Shiffrin, R. M. 1968. "Human memory: A proposed system and its control processes". In *The Psychology of Learning and Motivation*, Volume II, K. W. Spence and T. Spence (eds). New York: Academic Press.

Bar, M. and Biderman, I. 1998. "Subliminal visual priming". *Psychological Science* 9.

Barker, M. 2001. *An Evolutionary Perspective on Linguistic Processing Rates*. MA thesis, University of Oregon.

Chafe, W. 1994. *Discourse, Consciousness and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.

DeSchepper, B. and Treisman, A. M. 1996. "A visual memory for novel shapes: Implicit coding without attention". *Journal of Experimental Psychology* (Learning, Memory and Cognition) 22: 27–47.

Dickinson, C. and Givón, T. 1997. "Memory and conversation: Toward an experimental paradigm". In T. Givón (1997).

Givón, T. 1991. "Some substantive issues concerning verb serialization: Grammatical vs. cognitive packaging". In *Serial Verbs: Grammatical, Comparative and Cognitive Approaches*, C. Lefevre (ed). Amsterdam: John Benjamins.

Givón, T. (ed). 1997. *Conversation: Cognitive, Communicative and Social Perspectives*, Typological Studies in Language 34. Amsterdam: John Benjamins.

Givón, T. 1998. "On the co-evolution of language, mind and brain". *Evolution of Communication* 2.

Kaas, J. H. 1989. "Why does the brain have so many visual areas?" *Journal of Cognitive Neuroscience* 1.

Maunsell, J. H.R. and J. R. Gibson. 1992. "Visual response latencies in striate cortex of the macaque monkey". *Journal of Neurophysiology* 68.

Maunsell, J. H.R. and D. C. van Essen. 1983. "The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in macaque monkeys". *Journal of Neuroscience* 12.

Mishkin, M. 1978. "Memory in monkeys severely impaired by combined but not by separate removal of amygdala and hippocampus". *Nature* 273.

Mishkin, M. 1982. “A memory system in the monkey”. *Philosophical Society of London* [*Biol.*] 298.

Mishkin, M., Malamut, B., and Bachevalier, J. 1984. “Memories and habits: Two neural systems”. In *Neurobiology of Learning and Memory*, G. Lynch and J.L. McGaugh (eds). New York: Guilford Press.

Mishkin, M., Spiegler, B.J., Saunders, R.C., and Malamut, B.J. 1982. “An animal model of global amnesia”. In *Toward a Treatment of Alzheimer's Disease*, S. Corkin, K.L. Davis, H.Growdon, E.J. Usdin, and R.J. Wurtman (eds). New York: Raven Press.

Mishkin, M. and Petri, H.L. 1994. “Memories and habits: Some implications for the analysis of learning and retention”. In *Neuropsychology of Memory*, N. Butters and L.R. Squire (eds). New York: Guilford Press.

Neely, J.H. 1977. “Semantic priming and retrieval from lexical memory: Roles on inhibitionless spreading activation and limited-capacity attention”. *Journal of Experimental Psychology, General* 106.

Neely, J.H. 1990. “Semantic priming effects in visual word recognition: A selective review of current findings and theories”. In *Basic Processes in Reading: Visual Word Recognition*, D. Besner and G. Humphreys (eds). Hillsdale, NJ: Lawrence Erlbaum Associates.

Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M.A., and Raichle, M.E. 1988. “Positron emission tomographic studies of the cortical anatomy of single word processing”. *Nature* 331.

Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., and Raichle, M.E. 1989. “Positron emission tomographic studies of the processing of single words”. *Journal of Cognitive Neuroscience* 1.

Petersen, S.E., Fox, P.T., Snyder, A.Z., and Raichle, M.E. 1990. “Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli”. *Science* 249.

Posner, M. and Pavese., A. 1997. Anatomy of word and sentence meanings. Paper presented at the *Colloquium on Neuroimaging of Human Brain Functions.* Ms.

Posner, M.I., Petersen, S.E., Fox, P.T., and Raichle, M.E. 1988. “Localization of cognitive operations in the human brain”. *Science* 240.

Posner, M.I., Sandson, J., Dhawan, M., and Shulman, G.L. 1989. “Is word recognition automatic? A cognitive anatomical approach”. *Journal of Cognitive Neuroscience* 1.

Posner, M.I. and Tudela., P. 1997. “Imaging resources”. *Biological Psychology* 45.

Raichle, M.E., Fiez, J.A., Videen, T.O., MacLeod, M.K., Pardo, J., Fox, P.T., and Petersen, S.E. 1993. “Practice-related changes in human brain functional anatomy during nonmotor learning” *Cerebral Cortex.*

Rock, M.A. 1998. *Inattentional Blindness: Perception without attention.* Cambridge, Mass.: MIT Press.

Rolls, E.T., Tovee, M.J., and Lee, B. 1991. “Temporal response properties of neurons in the macaque inferior temporal cortex”. *European Journal of Neuroscience* 4.

Snyder, A.Z., Abdulaev, Y.G., Posner, M.I., and Raichle, M.E. 1995. “Scalp electrical potentials reflect regional cerebral blood flow responses during processing”. *Proceedings of the National Academy of Science* 92.

Squire, L.R 1987. *Memory and Brain.* Oxford: Oxford University Press.

Squire, L. R. and Zola-Morgan, S. 1991. "The medial temporal lobe memory system". *Science* 253.

Subramaniam, S., Biderman, I., and Madigan, S. 2000. "Accurate identification but no priming and chance recognition memory for pictures in RSVP sequences". *Visual Cognition* 7.

Swinney, D. 1979. "Lexical access during sentence comprehension: (Re)consideration of context effects". *Journal of Verbal Learning and Verbal Behavior* 18: 645–59.

Treisman, A. M. and DeSchepper, B. 1996. "Object tokens, attention and visual memory". In *Attention and Performance XVI: Information Integration in Perception and Communication*, T. Inui and J. McClelland (eds). Cambridge, Mass.: MIT Press.

Treisman, A. M. and Kanwisher, N. G. 1998. "Perceiving visually presented objects: Recognition, awareness, and modularity". *Current Opinion in Neurobiology* 8: 218–26.

Tucker, D. M 1991. "Developing emotions and cortical networks". In *Developmental Neuroscience*, *Minnesota Symposium on Child Psychology*, Volume 24, M. Gunnar and C. Nelson (eds). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ullman, S 1994. "Sequence seeking and counterstreams: A model for bidirectional information flow in the cortex". In *Large-Scale Neuronal Theories of the Brain*, C. Koch and J. L. Davis (eds). Cambridge, Mass.: MIT Press.

Ungerleider, L. A. 1985. "The cortical pathways for object recognition and spatial perception". In *Pattern Recognition Mechanisms*, *The Pontifical Academy of Sciences*, C. Chagas, R. Gattass, and C. Gross (eds), 21–37. Berlin/New York: Springer Verlag.

Ungerleider, L. A. and Mishkin, M. 1982. "Two cortical visual systems". In *Analysis of Visual Behavior*, D. G. Ingle, M. A. Goodale, and R. J. Q. Mansfield (eds). Cambridge: MIT Press.

CHAPTER 8

The clausal structure of linguistic and pre-linguistic behavior

Gertraud Fenk-Oczlon and August Fenk
University of Klagenfurt

1. Co-evolution of language and cognition: mutual stimulation and acceleration

The terms "evolution" and "co-evolution" imply pre-existing structures which undergo some developmental changes such as differentiation, specialization, and integration, and they imply that the scope of possible developmental directions is predisposed in these already existing structures.

Evolutionary processes are initiated or at least accelerated when the environment (e.g. the climate in a certain habitat) changes or when organisms change their environment, e.g. by conquering new habitats or niches, remigrating from land to water, etc. In the co-evolution of a system's (system 1) subsystems (1.1., 1.2. etc), the situation is similar in this respect. If we regard language (=subsystem 1.2.) as an environmental system of the cognitive system (=subsystem 1.1.), we may say that an increasing complexity of a community's language makes growing demands on relevant cognitive abilities of the "users" of this language. Individuals who are above average or extremely highly equipped with such abilities, gain higher, more "elaborated" levels in the use of this cognitive and communicative tool, and — therefore — higher chances to transmit this "equipment" to the next generation (generation $n+1$).[1] This selective mechanism is again effective in the transmission from generation $n+1$ to $n+2$, and so forth, and the more efficient, the higher the utility of the "organon" language for the individual user within his linguistic community and the higher its utility for the linguistic community as a whole.

The utility for the community (the family, the tribe, bigger societies) is immense. The activity of our cognitive apparatus is, first of all, concerned with the task of reducing uncertainty and of making events (e.g. the consequences of one own's actions) foreseeable by extracting invariants, rhythms, regularities, redundancies in the observations possible so far (Fenk and Vanoucek 1992). Anticipation allows "rational" decision-making under uncertainty in practical or even existential matters. To be able to communicate thoughts about indexical properties of events, about possibilities, conditionalities, contingencies, and probabilities, to communicate risks and chances — all that makes the social group not only fit for survival in a certain habitat, but makes it extremely fit for conquering new habitats and for adapting the habitat in an advantageous way. And the ability to succeed to meta-linguistic levels offers an extraordinarily high potential for completely new intellectual developments.

Thus, the phylogenetic and ontogenetic development of a complex language requires both: each step forward in this evolution has to allow for the actually reached "standard" of relevant cognitive capabilities. (It has to meet the demands of "cognitive ergonomics".) And each step in the complication of the "tool" language has to be accompanied (or proceeded, or followed) by an advance in the capabilities for acquisition and efficient use of this "organon". Developmental changes of language (subsystem 1.2.) stimulate and accelerate the development of cognition (subsystem 1.1.) and vice versa. This mechanism of mutual stimulation and mutual acceleration is the "motor" of the "self-acceleration" of the superordinate system 1. This view also holds, if — as in Wills (1993) or in Deacon (1997) — not cognition, but its biological substrate is taken as the language's partner in co-evolution: "The key to this is the co-evolutionary perspective which recognizes that the evolution of language took place neither inside nor outside brains, but at the interface where cultural evolutionary processes affect biological evolutionary processes". (Deacon 1997: 409).

Our recent and complicated languages most probably are traced back to less complicated rudimentary predecessors in vocal, mimic and gesture communication. New steps in evolution are grounded on pre-existing structures (see first paragraph of this section), and the introduction of a "new" referential system (like "picture language" and "diagrammaticity") will always refer to already existing referential systems (Fenk 1998). This is a matter of "pragmatics" and "cognitive economy", if not of necessity.

The aim of this paper is not to reconstruct the development of "pre-language" to "language", nor to define a borderline[2] between "pre-language" and "language", nor is it concerned with possible "ritual beginnings" (Deacon 1997: 401) of language and "symbolic reference", or with the changes from an initially visually-gesturally coding language to auditory coding (Givón 1998: 103, in this volume). Instead, it concentrates on some cognitive pre-conditions as a "matrix" allowing for, as well as constraining the development of a complex language system.

2. The pre-linguistic matrix of language

The impulses for typological differentiation and diachronic change of languages may come from the social-communicative environment and from language contact. But the evolution and possible forms of variation of language are determined by the pressure to improve, or maintain at least, the language's efficiency as a communicative and cognitive tool in a way that accounts for the constraints of the articulatory and the cognitive system.

These goals[3] and constraints can be seen as a matrix preforming language, and this matrix will be discussed in two respects: with regard to the goal of communicating thoughts about possibilities beyond the *hic et nunc* (section 2.1), and to the "clausal" structure of cognition and psychomotility (section 2.2) predisposing the clausal structure of language.

2.1 Mental propositions

Many psycholinguistic studies (already in the 1970s) view propositions as basic semantic entities of the organization of (sensory-motor and iconic) memory. This cognitive structure cannot be described in solely linguistic terms; but when it is verbalized it is likely to take the form of a predicate-argument-structure (Fenk-Oczlon1983:30). Givón (1990:896) states that something "like a mental proposition, under whatever guise, is the basic unit of mental information storage", and in Kintsch (1974:12) propositions are characterized as "n-tuples of word concepts, one of which serves as a predicate and the others as arguments".

According to Deacon (1997:334) "this operand-operator structure (and probably subject-predicate structure) /.../ is the minimum requirement to make the transition from indexical to symbolic reference". He states that already the earliest symbolic systems would necessarily have been combinatorial in this sense.

Considering the universal appearence of simple declarative sentences in all languages which are apt to transcend the *hic et nunc* and to communicate about assumptions etc. (see section 1), the crosslinguistic experiments described in section 3.2. used so called "kernel sentences". A "kernel sentence" can be described as a prototypical case of clauses: a simple declarative sentence encoding **one** proposition (in the sense of a pre-linguistic entity) within **one** intonation unit in the form of a predicative syntagma (Fenk-Oczlon and Fenk 1995:231).

In written communication, where the respective text also serves as an "external memory", long, complex, and embedding sentences are much more frequently used than in everyday oral communication — even in our literate society with its well trained skills in the reception of long and complex sentences.[4] In languages without a writing system or without a longer literate tradition — like Cayuga (Sasse 1991:204) or Eipo (Heeschen 1994) — a content that we would rather encode in a complex sentence is encoded in a series of minimal-predications, each of them included in a single intonation unit.

2.2 Segmentation of extralinguistic activities and the magical number seven, plus or minus two

In the psychology of information processing the number seven is a somewhat "magical" invariant: It manifests itself as a constraint of the span of absolute judgement, the span of immediate memory, and the span of attention. But Miller (1956) warned assuming "that all three spans are different aspects of a single underlying process". This limit of about seven (plus minus two) has since figured prominently in information processing theories.[5]

Linguistic information is a special type of information processed by our cognitive apparatus. If the number seven marks some general limits of this apparatus, it should also show in languages, because languages must have developed in adaptation to the general constraints of this apparatus.[6]

A main concern of Fraisse's meanwhile classical book on the psychology of time (Fraisse 1985/1957) is the "psychological present" or "immediate memory span". He states that this span cannot be understood as a window of consciousness, on which reality moves by, but rather as a beat in rhythmically organized attentional processes. One of his arguments (Fraisse 1985: 89 f.) says that we perceive the *tick-tack* of a watch always as this *tick-tack,* and not as a *tick-tack* followed by a *tack-tick* followed by a *tick-tack* etc., or that we perceive a waltz always as sequences of low-low-loud and not as a low-low-loud followed by low-loud-low followed by loud-low-low. In this context he already refers to the comprehension of sentences and states that the psychological present always corresponds to one sentence and never to the end of a sentence together with the first part of the following sentence with running elements (p 90). Regarding the **size** of the psychological present he mentions a series of strikes of a church bell: in the case of 3 or 4 strokes we know the indicated time immediately, without any counting. This is impossible with 12 strokes indicating the midnight hour. Children not yet able to count are able to correctly reproduce series of about 5 or 6 strokes.

In his experiments (Fraisse and Fraisse 1937, as discussed in Fraisse 1985: 93) on the retention of sound series, he had varied the between-sound intervals: 0.17sec, 0.37 sec, 0.63 sec, 1.2 sec, 1.8 sec. For these intervals the following mean number of sounds could be reproduced: 5.7, 5.7, 5.4, 4.0, 3.3. The overall duration of the retained series was 0.8 sec in the case of the 0.17 sec interval and 4.2 sec in the case of the 1.8 sec interval. From this he concluded

- that the optimal intervals for perception are in the area betwen 1.5 and 0.7 sec;
- that the psychological present is not a content-independent scope of time;
- that a number of 5 to 6 elements marks the limit of the perception of succession; and

- that the duration of the perceived present shows a higher variability than the number of perceived elements.

Such limits regarding the **number of elements** he assumes to be not only independent of sense modality — 6 to 7 elements was also a maximum in the retention of visual stimuli (Fraisse and Fraisse 1937) — but also of the "level of intelligence": Various birds such as daws can learn to look at first in that box for food on which a certain number (let us say 5) is represented by an arbitrary array of dots. (Only that box contains food. The position of this box among other boxes representing other numbers as well as the array of dots on all of the boxes is changed from trial to trial.) This works up to about a number of 7 dots.[7]

Corresponding results regarding the relevance of a fixed number of elements showed also in non-human primates. Rhesus Macaques were the "subjects" of the two following studies:

Swartz *et al.* (2000) report on experiments concerning the monkey's ability to memorize arbitrary sequences of arbitrary stimuli. Each of their subjects mastered 4 novel 4-item lists by trial and error, and the performance of some of the subjects suggested "that, with more list-learning experience, monkeys could master longer lists and determine the ordinal position of items on new lists with fewer errors" (Swartz *et al.* 2000: 284).

Experiments conducted by Brannon and Terrace (2000) "showed that monkeys trained on an ascending numerical rule spontaneously infer the ordinal values of novel numerosities when numerosity varied from 5 to 9 and "that neither language nor numerical symbols are necessary for discriminating and ordering visual stimuli on the basis of their numerosity." (Brannon and Terrace 2000: 48). This reminds us, in some respects at least, of the already mentioned findings of Köhler (1952) in his experiments with daws.

Results of brain research concerning the **duration** of perceptual and memory span are in line with Fraisse's findings as well. Elbert *et al.* (1991) report on experiments in which the subjects had to reproduce the duration (1, 2, 3, 4, 6, 8 sec intervals) of the illumination of a screen.

> Reproduction was accurate for standard intervals up to 3 s but deteriorated with increasing interval length. Brain potentials during reproduction intervals of 1–3 s differed from those recorded during the longer intervals. A CNV-like slow negative shift developed during the shorter reproduction intervals. Negativity was reduced or even absent, when subjects had to reproduce standard intervals of 4 s or longer. (Elbert *et al.* 1991: 648)

The authors concluded that short intervals with a maximum duration of 3–4 sec are processed primarily in left-hemispheric, frontal and temporal cortical networks and in a mode that differs qualitatively from processing events of longer duration.

Regarding the duration of the "perceived present", Fraisse (1985:96) had mentioned, among other examples, some visually ambiguous figures such as a cube with solid drawn edges throughout and with equal size of whatever one may perceive as the front plan or as the more distant back plan. The intervals in which the "foreground" subjectively becomes the "background" and then the other way round is assumed to reflect the periodicity of attentional shifts. According to more recent studies "the turnover time lies within a few seconds with a peak around 3 seconds" (Schleidt and Kien 1997:98). In these studies the respective cube representation is referred to as "Necker cube" (see also Pöppel 1986).

Rhythmical segmentation or "clausal" structure is, as we have seen, neither restricted to auditory perception nor to the species *Homo sapiens.* It is, moreover, not restricted to primarily perceptual/cognitive activities, but is a characteristic of motor activities as well. (Not surprisingly if one recalls the cognitive components of motor programming and the back-coupling of the effector-system to the cognitive system via the kinesthetic system.)

The aim of the above mentioned study of Schleidt and Kien (1997) and of Feldhütter *et al.*(1990) was an intercultural comparison of the duration and internal structure of movement patterns. A fundamental finding was that different sorts of behavior could be divided into small segments of functionally related movements. The authors' criteria for defining a sequence of movements as a separate action unit were:

> A sequence of movements was regarded as an action unit only if its beginning and end could be clearly seen, if it had an observable goal, and thus the individual movements are functionally related (e.g. wiping the nose with the finger). The consecutive movements do not belong to the same action unit if they are not functionally related (e.g. wiping the nose and then scratching it). (Schleidt and Kien 1997:79)

A further result was: from a total of 1542 action units analyzed, 93 % had a duration of 2–3 seconds.

Thus, segmentation in the range of a few seconds appears to be a universal phenomenon in human action. We find it in various cultures all over the world, not only in conscious and intentional acts like work activities, actions of communication like greeting behavior, and ritualized behavior such as dances, but also in less conscious and less culturally influenced activities like scratching one's body or playing with a baby (Schleidt and Kien 1997:81).

Such a segmentation (into somewhat shorter segments) was also observed in non-human primates, for instance in the hand movements of chimpanzees (Kien *et al.* 1991). And in Schleidt (1992) it is stated that in a higher number of movements per action unit the movements have to be shorter in order to fit into the restricted time span. (This gives the impression that the duration of the unit is rather the invariant dimension, and not the number of elements as suggested in Fraisse 1985: 93)

A central idea of the following section is that the **intonation units** in language are a special case of **action units**. To be more general: that the central nervous mechanisms underlying the segmentation of other activities are also effective in language behavior and are responsible for the clausal structure of both the perception and production of language (see Barker and Givón in this volume).

Language behavior is not only a matter of perceptual and a matter of productive activities, but a matter of intimate coordination between perception and articulation. In the role of the speaker, for instance, one is always also the hearer of one's own output. The possibility of perceiving and controlling one's own output, and the coordination and synchronization of the activities involved, must be considered as an essential characteristic of any communication system capable of development.(Similarly the possibility of the visual control of our own manual activities is often viewed as a prerequisite of tool-using and tool-manufacturing activities.)

Important indications for such a coordination between speaking and memory for the spoken come from experiments concerning the "articulatory loop hypothesis" (e.g. Baddeley 1986): One is able to recall as many words as one could pronounce in about 2 seconds. And the mapping between the psychologists' estimations of the duration of immediate memory (2 plus minus 1 sec) and the linguists' estimation of the duration of intonation units (see following section) will hardly be a mere accident.

3. Segmentation in language behavior

Spoken language is obviously segmented: "tone unit" (Quirk *et al.* 1964), "breath group" (Lieberman 1967), "tone group" (e.g. Halliday 1967) and "intonation unit" (e.g. Chafe 1994) — all these terms refer, though not always in identical way, to a prosodic unit which comprises a sequence of words spoken under a single intonation contour.

3.1 The duration of intonation units and clauses

Studies really measuring the duration of units seem to be rather rare. Määttä (1993: 109) studied the "portion of speech between two measurable pauses". In Finnish he found an average length of "breath groups" in the region of 2.1 to 2.2 sec. (The mean duration of breath groups inclusive pauses comes to 3.2 to 3.3 sec.) And according to Chafe (1987: 22), "new intonation units typically begin about two seconds apart. Evidently active information is replaced by other, partially different information at approximately two second intervals."

Most usually the length of intonation units is not measured in units of time but in number of words. Chafe (1994), for example, reports a mean length of 4.84 words per substantive intonation unit in English. Croft (1995), referring to Altenberg (1990) and Crystal (1969), reports similar numbers, ranging from 4 to 6 words per intonation unit. And when comparing English intonation units with intonation units in Seneca, Chafe (1994: 148) found that "with respect to number of words, then, Seneca intonation units are half as long as English ones." This shows that the average length of an intonation unit, when measured in words, is highly dependent on the language in question, and especially on its morphological type. In languages with a pronounced tendency to synthetic (agglutinative or fusional) morphology we have to expect a lower number of words per intonation unit (and in polysynthetic and incorporating languages even one long word that we would encode in a sentence comprising 5 or 6 words.) This is one of many reasons why the crosslinguistic study reported in the next section took the number of syllables as an appropriate measure for the size of a basic type of intonation unit.

3.2 Segmentation within clauses and again the magical number seven, plus or minus two

In spoken language there are only two entities corresponding to **rhythmic** processing — the **syllable** as the basic element and the **clause** or **intonation unit** at a higher order level. (The unit "in between" these two levels is the word. It is the most widely used material in memory experiments and is of course interesting because of its semiotic status. But it is not the appropriate candidate in the search for elements and components of rhythmic organization.) Thus, an experimental design was applied in order to study crosslinguistic regularities of within-clause segmentation. A central assumption of this crosslinguistic study (Fenk-Oczlon 1983) is that *the number of syllables per "clause" will vary within the range of the magical number seven plus minus two*. The clauses used were of a special quality: simple declarative sentences encoding **one proposition in one intonation unit**, such as *blood is red* or *the sun is shining*. 22 German sentences of this sort were presented to native speakers of 27 different languages. (Meanwhile, the sample was extended to 34 languages, 18 Indo-European, 16 Non-Indo-European (Fenk-Oczlon and Fenk 1999)). Native speakers were asked to translate the sentences into their mother tongue and to determine the length of their translations in syllables. The mean number of syllables per clause, computed for each one of these languages, was found to be located almost exactly within Miller's (1956) often quoted range of 7 plus minus 2 elements: The lowest size was 5.05 syllables (Dutch), and only Japanese with 10.2 syllables per clause was located outside the hypothesized range of 5–9 syllables. The overall length was 6.48 syllables per simple clause.[8]

A statistical reanalysis (Fenk-Oczlon and Fenk 1985) was intended to clarify the question about the relevant factor determining the position of single languages on the continuum "mean number of syllables per sentence". The factor which came under suspicion was syllable complexity, i.e. a language‘s mean number of phonemes per syllable. One cause of suspicion was that the articulation of more complex syllables takes up more time. Another cause was that Dutch is known for its complex syllables and Japanese for its simple CV-syllables. The result was a highly significant negative correlation ($r = -0.77$, $p < 0.1\%$) between mean number of syllables per sentence and mean number of phonemes per syllable. The interpretation therefore is that: languages with higher syllable complexity use, proportionate to the higher expenditure of time per syllable, fewer syllables for encoding a certain proposition.

In other words: for encoding a certain proposition, languages with simple syllable structure need the same time but more syllables than languages with high syllable complexity. The fact that in cross-linguistic comparison the number of syllables per simple sentence was found to be located in the area of 5–10 syllables (see Figure 1) agrees with our immediate memory span comprising about 5–9 units. And the location in this area corresponds to time-related limits, which might be operative at the level of syllable perception (and production): 200–300 milliseconds seems to be the duration necessary for auditory pattern recognition

	5 – 5.99	6 – 6.99	7 – 7.99	8 – 8.99	9 – 9.99	10 – 10.99
15	Dutch 5.05					
	French 5.32					
	Czech 5.36					
	Chin. 5.41					
	Hebr. 5.46					
10	Slov. 5.50					
	Germ. 5.50	Bamb. 6.46				
	Icel. 5.50	Turk. 6.46				
	Eston. 5.68	Alban. 6.55				
	Russ. 5.68	Yoruba 6.59				
5	Croat. 5.77	Port. 6.64	Hopi 7.12			
	Engl. 5.77	Pers. 6.64	Nava. 7.41			
	Ewon. 5.77	Hindi 6.77	Ital. 7.50	Korean 8.18		
	Hung. 5.91	Pan. 6.77	Greek 7.55	Anjang 8.23		
	Arab. 5.96	Mac. 6.96	Span. 7.96	Basque 8.27	Chiqu. 9.14	Japan. 10.23

syll./clause

Figure 1. The frequency distribution of languages over different classes of the parameter ''mean number of syllables per clause" (slightly changed from Fenk-Oczlon and Fenk 1999: 158)

(Massaro 1975) and for producing the right-ear advantage in dichotic-listening experiments.

If the duration of a simple sentence coincides with our "psychological present" (ca. 2 seconds) and if the minimum duration of a syllable is estimated at ca. 200 milliseconds, then the sentence comprises 10 syllables in a "pure CV-language" (see Japanese in Figure 1) and a lower number of syllables in the case of more complex syllables (CVC, CCVC, CCVCC, . . .), proportionate to the longer duration of these more complex syllables. In this respect, at least, there seems to be nothing magical in the "magical number seven".

The upper limit (2–3 sec per clause or simple sentence) and the lower limit (200–300 millisec per syllable) are operative in the rhythmic pattern organisation, and they might be operative like set points in the self-regulation of language systems, constraining for instance the typological differentiation of languages with regard to morphosyntactic structure and complexity of syllables.

Our negative crosslinguistic correlation between number and complexity of syllables points, first of all, to the efficency of such time related constraints, and so does the whole set of correlations — e.g. the more words per sentence, the fewer syllables per word — found in a later study (Fenk and Fenk-Oczlon 1993). More generally we may conclude:

- The central unit in the rhythmic organization of language is the clause, and the clausal structure of our language is apparently preformed (phylogenetically as well as ontogenetically) by the tonal utterances that we can isolate in the (prelinguistic?) communicative behavior of our infants and of recent non-human primates.
- The variability of clauses is, first of all, restricted in terms of duration. But this relatively "constant" time interval is also filled by a rather "constant" number of elements and a rather "constant" amount of information — probably due to economy principles providing "packages" with an optimal size for cognitive "handling".
- Optimal size of packages is achieved by trade-offs between sizes of within-clause elements (see our correlations) and between the size (complexity, duration) of these elements and the information carried by them: Higher token frequency goes hand in hand with lower size of elements (e.g. Zipf 1929) and, moreover, with lower information per element. (Higher relative frequency means higher probability and corresponds with lower information per mathematical definition. Similarly, higher "subjective" probability — or higher "familiarity" — results in lower "subjective" information and lower cognitive load.)
- Quantitative relations found between relevant dimensions in recent languages — be it within a single language (e.g. Menzerath's law)[9] or in crosslinguistic computation — reflect those dynamic and economic principles that govern

diachronic change and differentiation of language and probably the evolution of language out of pre-language.

4. Discussion

Dealing with the "roots" of recent human languages one probably should search for the — pre-linguistic or language-bound? — nature of propositional thinking, of man's propositional view of the world. Doing so, one might start with regularities of Gestalt-perception structuring our perceptual world by bringing about coherence between "elements" and separating groups of "elements".[10] From there one might proceed to our perceptual interpretation of — indexical, causal — relations and interactions between those objects separated by Gestalt-principles. Such "causal" interpretations on a **perceptual** level have been described by Michotte (1946). Further analyses might then be concerned with the size or duration of what can be kept within the conscious field (Mandler 1975) — be it a linguistically encoded proposition or an extralinguistic event which becomes the object of a "propositional" interpretation.

We did mention certain "positive" presuppositions of language evolution but concentrated on the "constraining" and "preforming" properties of these presuppositions. Such constraints can be observed on both levels, the **duration** of perceptual and action units and the level of the **number of elements**, and in both relevant systems, the afferent/perceptual and the efferent/motoric. Arguments were presented for an intimate perceptual/motoric coordination and back-coupling, especially in linguistic behavior.

Later studies on (a) linguistic and (b) pre-linguistic behavior correspond to the crosslinguistic and significant **negative correlation** found between number of syllables and complexity of syllables (Fenk-Oczlon and Fenk 1985):

(a) In the already mentioned linguistic study of Määttä (1993) a connection was found between syllable duration (in msec) and "length" of breath group, i.e. the number of syllables per breath group. He reports an inverse relationship between syllable duration and length of the breath group, and that the longest average syllable durations are restricted to short breath groups.

(b) Moreover, our negative correlation is in line with Feldhütter's *et al.* (1990) observation in extralinguistic and pre-linguistic action units: The higher the number of movements within the action unit, the shorter the duration of a single movement. In the case of spoken language these "single movements" are the articulatory movements forming a single syllable. Higher complexitiy of the single syllables results in longer duration of articulation and perception and in a lower number of syllables per clause. What remains relatively invariant and unaffected by this crosslinguistic within-clause variability is the duration of the clause and the information transmitted.

The underlying span of (about two seconds and) about seven syllables has, according to our findings, the appropriate size for encoding one proposition. It has the appropriate size for allowing the extraction of the meaning of a clause before moving to the following clause, or, to use Mandler's (1975: 236) words again, the size "that can be kept within the conscious field, the focal attention."

Notes

1. The selective advantage of tool-using abilities and their effect on cerebral growth is not restricted to the "tool" language, but is to be assumed for *tools* in the literal sense and for cultural techniques in general (Fenk-Oczlon and Fenk 1999). Rather specific neurophysiological and -anatomical arguments (of Lieberman 1991 and Greenfield 1991) aim, according to Beaken (1996: 17), at a close relationship between language and manual skills.

2. Such attempts are most commonly characterized by rather restrictive and arbitrary definitions of what is the constitutive property of "full" language. And "full" syntax which is most commonly suggested to be constitutive for "full" language, is again (e.g. Bickerton 1990) defined in a completely arbitrary way. Comprehensive criticism of Bickerton's attempt and of some other relevant approaches is presented in Sampson (1997). See also Givón's criticisms of any attempt to reduce language evolution "to a single initial cause" (Givón 1998: 105).

3. The term "goal" is just a substitute for circumstantial phrases like "advantages leading via selection mechanisms to . . .", and should in no way indicate a teleologic or vitalistic position of the authors.

4. Apart from the question of, whether or not a longer literate tradition is a prerequisite for a change from paratactic to hypotactic constructions (cf. the short discussion in Sampson 1997: 74), it has to be accepted as an empirical fact that in everyday oral communication hypotactical constructions are relatively rare. Impressive but unfortunately not yet published examples of an almost exclusive use of (fragments of) paratactic constructions are the dialogs recorded by Gigler (in preparation) in Carinthian.

5. One of the most recent and most interesting relevant findings is reported by Kareev (2000): small series of about 7 plus minus 2 data-pairs produce stronger correlations between the respective variables than the population. This would mean that a span of comprehension comprising about 7 elements or chunks of elements does not reflect a rather arbitrary perceptual/cognitive limit, but, furthermore, that there must have been a selective advantage and selective pressure to push up the limit to this region where minimal indications and minimal contingencies (see our section 1) can be detected with a minimum of "computational" work.

6. The present paper concentrates on the relevance of the magical number seven in the span of actual perceptual/cognitive activities. But it might also be relevant for the size of our languages' repertoire of e.g. categories (Fenk-Oczlon and Fenk 2000).

7. Fraisse's (1985: 94) report on these experiments should be supplemented by their original author Köhler (1952).

8. It is worth noticing that relevant data do not only show in our experimental setting with written clauses and translations but also in recordings of oral discourse: In 9 dialogues with a total number of 1055 intonation units Gigler (in preparation) found a mean size of 6.04 syllables per unit. The lowest mean value of these dialogues (Carinthian dialect) was 4.82 syllables, the highest was 7.12 syllables per intonation unit. The mean duration per intonation unit was 1.373 seconds.

9. Menzerath (1954: 100) found that German words composed of a high number of syllables tend to be composed of a "relatively" low number of phonemes.

10. Such structuring processes will be enhanced by the movements of the groups (cf. "Gesetz des gemeinsamen Schicksals") within the visual field and by eye-movements necessarily resulting in faster relative movements of those patterns which are nearer to the observer. These movements of patterns relative to the actually moving eye and/or head of the observer deliver indications regarding spatial depth as well as indications regarding coherence and separability of visual patterns.

References

Altenberg, B. 1990. *Prosodic Patterns in Spoken English: Studies in the Correlation Between Prosody and Grammar for Text-to-Speech Conversion.* [Lund Studies in English 76]. Lund: Lund University Press.

Baddeley, A. D. 1986. *Working Memory.* Oxford: Clarendon Press.

Barker, M. and Givón, T. This volume. "On the pre-linguistic origins of language processing rates".

Beaken, M. 1996. *The Making of Language.* Edinburgh: Edinburgh University Press.

Brannon, E. M. and Terrace, H. S. 2000. "Representation of the numerosities 1–9 by rhesus macaques (Macaca mulatta)". *Journal of Experimental Psychology: Animal Behavior Processes* 26: 31–49.

Bickerton, D. 1990. *Language and Species.* Chicago: University of Chicago Press.

Chafe, W. 1987. "Cognitive constraints on information flow". In *Coherence and grounding in discourse,* R. S. Tomlin (ed), 21–51. Amsterdam: Benjamins.

Chafe, W. 1994. *Discourse, Consciousness, and Time.* Chicago: University of Chicago Press.

Croft, W. 1995. "Intonation units and grammatical structure". *Linguistics* 33: 839–82.

Crystal, D. 1969. *Prosodic Systems and Intonation in English.* Cambridge: Cambridge University Press.

Deacon, T. W. 1997. *The Symbolic Species. The Co-evolution of Language and the Brain.* NewYork: W. W. Norton.

Elbert, T., Ulrich, R., Rockstroh, B., and Lutzenberger, W. 1991. "The Processing of Temporal Intervals Reflected by CNV-Like Brain Potentials". *Psychophysiology* 28: 648–55.

Feldhütter, I., Schleidt, M. and Eibl-Eibesfeldt. 1990. “Moving in the Beats of Seconds: Analysis of the Time Structure of Human Action”. *Ethology and Sociobiology* 11: 1–10.

Fenk, A. 1998. “Symbols and Icons in Diagrammatic Representation”. *Pragmatics and Cognition* 6: 301–34. (Special Issue on the *Concept of Reference in the Cognitive Sciences*).

Fenk, A. and Fenk-Oczlon, G. 1993. “Menzerath’s law and the constant flow of linguistic information”. In *Contributions to Quantitative Linguistics*, R. Köhler and B. Rieger (eds), 11–31. Dordrecht: Kluwer.

Fenk, A. and Vanoucek, J. 1992. “Zur Messung prognostischer Leistung”. *Zeitschrift für experimentelle und angewandte Psychologie* 39 (1): 18–55.

Fenk-Oczlon, G. 1983. *Bedeutungseinheiten und sprachliche Segmentierung. Eine sprachvergleichende Untersuchung über kognitive Determinanten der Kernsatzlänge.* Tübingen: Gunter Narr.

Fenk-Oczlon, G. and Fenk, A. 1985. “The mean length of propositions is 7 plus minus 2 syllables — but the position of languages within this range is not accidental.” In *Cognition, Information Processing, and Motivation*, G. d’Ydewalle (ed), 355–9. Amsterdam: North Holland.

Fenk-Oczlon, G. and Fenk, A. 1995. “Selbstorganisation und natürliche Typologie.” *Sprachtypologie und Universalienforschung* 48: 223–38.

Fenk-Oczlon, G. and Fenk, A. 1999. “Cognition, quantitative linguistics, and systemic typology”. *Linguistic Typology* 3: 151–77.

Fenk-Oczlon, G. and Fenk, A. 2000. “The magical number seven in language and cognition: empirical evidence and prospects of future research”. *Papiere zur Linguistik* 62–63: 3–14.

Fraisse, P. 1957. *Psychologie du temps.* Paris: Presses Universitaires de France.

Fraisse, P. 1985. *Psychologie der Zeit.* Munich/Basel: Ernst Reinhard Verlag.

Fraisse, P. and Fraisse, R. 1937. “Études sur la mémoire immédiate. I: L’appréhension des sons”. *Année psychologique* 38: 48–85.

Gigler, G. In preparation. *Intonationseinheiten im Kärntnerischen.* Dissertation, University of Klagenfurt.

Givón, T. 1990. *Syntax: a Functional-Typological Introduction,* Volume II. Amsterdam: John Benjamins.

Givón, T. 1998. “On the co-evolution of language, mind and brain”. *Evolution of Communication* 2: 45–116.

Givón, T. This volume. “The visual information-processing system as an evolutionary precursor of human language”.

Greenfield, P. M. 1991. “Language, tools and brain: The ontogeny and phylogeny of hierarchically organised sequential behaviour”. *Behavioral and Brain Sciences* 14: 531–92.

Halliday, M. A.K. 1967. *Intonation and Grammar in British English.* The Hague: Mouton.

Heeschen, V. 1994. “How long are clauses and sentences in a Papuan language like Eipo?” *Semaian* 10: 50–70.

Kareev, Y. 2000. “Seven (indeed, plus or minus two) and the detection of correlations”. *Psychological Review* 107: 397–402.

Kien, J., Schleidt, M., and Schöttner, B. 1991. "Temporal segmentation in hand movements of Chimpanzees (*Pan troglodytes*) and Comparison with Humans." *Ethology* 89: 297–304.

Kintsch, W. 1974. *The Representation of Meaning in Memory.* New York: J. Wiley.

Köhler, O. 1952. "Vom unbenannten Denken". *Verhandlungen der deutschen Zoologischen Gesellschaft* 16: 202–11.

Lieberman, P. 1967. *Intonation, Perception and Language.* Cambridge, Mass.: MIT Press.

Lieberman, P. 1991. *Uniquely Human: The Evolution of Speech, Thought and Selfless Behaviour.* Cambridge, Mass.: Harvard University Press.

Mandler, G. 1975. "Consciousness: Respectable, useful, and probably necessary". In *Information Processing and Cognition. The Loyola Symposium.* R. Solso (ed), 229–54. Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D. W. 1975. *Experimental psychology and information processing.* Chicago: Rand Mc Nally.

Määttä, T. 1993. "Prosodic structure indices in a folk tale". *Phonum* 2: 107–20.

Menzerath, P. 1954. "Die Architektonik des deutschen Wortschatzes". *Phonetische Studien* 3. Bonn: Dümmler.

Michotte, A. 1946. *La perception de la causalité.* Louvain and Paris: Fondation Universitaire de Belgique.

Miller, G. A. 1956. "The magical number seven, plus or minus two: Some limits on our capacity for processing information". *Psychological Review* 63, 81–97.

Pöppel, E. 1986. *Grenzen des Bewußtseins.* Stuttgart: Deutsche Verlags-Anstalt

Quirk, R., Duckworth, A., Svartvik, J. Rusiecki, J. P., and Colin, A. J. T. 1964. "Studies in the correspondence of prosodic to grammatical features in English". In *Proceedings of the IX International Congress of Linguists,* 679–91. The Hague: Mouton.

Sampson, G. 1997. *Educating Eve: The 'Language Instinct' Debate.* London and New York: Cassell.

Sasse, H.-J. 1991. "Predication and sentence constitution in universal perspective". In *Semantic Universals and Universal Semantics*, D. Zaefferer (ed), 187–221. Berlin: Mouton de Gruyter.

Schleidt, M. 1992. "Universeller Zeittakt im Wahrnehmen, Erleben und Verhalten". *Spektrum der Wissenschaft,* December: 111–15.

Schleidt, M. and Kien, J. 1997. "Segmentation in behavior and what it can tell us about brain function". *Human Nature* 8: 77–111.

Swartz, K. B., Chen, S., and Terrace, H. S. 2000. "Serial learning by Rhesus monkeys: II. Learning four-item lists by trial and error". *Journal of Experimental Psychology: Animal Behavior Processes* 26: 274–85.

Wills, C. 1993. *The Runaway Brain: The Evolution of Human Uniqueness.* New York: Basic Books.

Zipf, G. K. 1929. "Relative frequence as a determinant of phonetic change. *Harvard Studies in Classical Philology* 40: 1–95.

PART 3

Language and social cognition

CHAPTER 9

The gradual emergence of language*

Brian MacWhinney
Carnegie Mellon University

Language is a unique hallmark of the human species. Although many species can communicate about things that are physically present, only humans can use communication to construct a full narrative characterization of events occurring outside of the here and now. The religions of the world have interpreted this unique endowment as a Special Gift bestowed directly by the Creator. Scientists have also been influenced by this view of language, often attributing the emergence of language to some single, pivotal event in human evolution. The idea of linking language evolution to the emergence of a single Special Gift is certainly quite attractive, since it would account for so many facts in linguistics, psychology, and neuroscience through some single, simple mechanism.

The Special Gift view runs into a variety of problems, however, when we glance over the full landscape of six million years of human evolution. During this period, the hominid lineage has undergone a remarkable series of adaptations involving a straightening of upright posture, development of an opposing thumb, changes in the birth process (Hockett and Ascher 1964), loss of hair (Morgan 1997), adaptation of the gastrointestinal tract, increased innervation of the intercostal muscles (MacLarnon and Hewitt 1999), loss of pronounced canine teeth, bending of the vocal tract, refinement of the facial musculature, freeing of the vocal folds, and sharpening of the chin. These morphological changes have been accompanied by a gradual tripling of brain size (Holloway 1995) which has brought massive changes in the interconnectedness of the frontal lobes, changes in the linkage of vocal production to motor and emotional areas, linkages of the visual areas to motor areas, and expansion of many older areas, including the cerebellum, basal ganglion, and thalamus. Alongside these changes in morphology and neurology, human society has undergone a parallel process of development involving the expansion of social groups, the refinement of warfare, the development of tools, and the emergence of language.

Faced with evidence that evolution has effected so many detailed changes, adherents of the Special Gift approach will argue that one single change or process has been pivotal and the other changes are either unrelated to language emergence or else necessary consequences of the basic change. In most cases, the Special Gift analysis focuses on some morphological or cognitive change that has occurred in the last 100,000 years, after the exodus of the human race from Africa. The changes that were wrought during the previous six million years are considered irrelevant to the emergence of language in *Homo sapiens*.

The second major problem with Special Gift accounts is that they tend to treat language as if it were not grounded on cognition. In particular, they fail to account for the three-fold expansion of brain size that occurred primarily in the period between 2MYA (two million years ago) and .5MYA. The concept of coevolution (Deacon 1997; Givón 1998) serves as a useful way of redressing this imbalance. The theory of coevolution holds that changes in neurological structure have led to advances in planning and communication. These advances in cognitive and communicative capacity then set the stage for further neurological modifications. This notion of an evolutionary ratchet effect is fundamentally attractive. However, without further elaboration, it would seem to depict a species that is slowly and steadily moving toward fuller and fuller control over its environment. In this view, each new advance in ability should be accompanied by a greater control over the environment and a spread of habitat. Unfortunately, matters were not so simple. After a remarkable expansion of habitat by *Homo erectus* after 2MYA, the fortunes of our ancestors went into rapid decline. The population of our direct ancestors then bounced back a bit in the early Pleistocene to a population of perhaps 100,000, but declined to a mere 10,000 at a point some 70,000 years ago (Stringer and McKie 1996). If the story of the human race is one of the continual productive coevolution of language and the brain, we have to ask ourselves how it is that we so narrowly escaped extinction twice. The most recent near extinction is particularly problematic. At this point, *Homo* was a highly intelligent species with a well-configured vocal tract and fully articulated social relations. Yet, despite all of its gifts, this talented group of individuals nearly vanished from the planet.

In all fairness to our ancestors, we know that the Pleistocene glaciations put great stress on a human population that had adapted to warmer, lusher conditions. In addition, 70,000 years ago the massive explosion of the Toba Batak volcano in Indonesia brought a sudden drop in global temperatures that worsened an already bad situation. On top of this, one can always imagine that there may have been a plague or some other unrelated evolutionary pressure. My point here is not to criticize early humans for a lack of resourcefulness. My point is only that language, by itself, was not enough to save our ancestors from near extinction. More generally, I think that we have tended to overestimate the extent to which language, in itself, provided a clear evolutionary advantage for early men and women. Instead, the basic advantages that have accrued to individual hominid species are those that relate to their abilities to handle weapons, form tight social groups, navigate across territory, and adapt to new dietary pressures.

My basic thesis is that language evolution operated across the full six million years of human evolution, not just the last 100,000 years. The model of language functioning I am adapting here is one that I have developed elsewhere (MacWhinney 1999a). This model holds that language use involves the construction of a grounded cognitive simulation in which we enact the activities of others through processes of perspective-taking and perspective-shifting. I will not repeat those arguments in detail here. Instead, I will refer back to the core pieces of the analysis in terms of the ways that they help us to better understand the processes of co-evolution that have operated across the last six million years.

We can distinguish at least four major co-evolutionary periods. During the first period, which lasted from 8MYA to 4MYA, evolution focused on the linking of morphological and neurological changes to the attainment of bipedal gait and the elaboration of the changes brought about from that new posture. During the second period, from 4MYA to 2MYA, the major co-evolutionary pressures involved the development of methods for solidifying group structure during face-to-face vocal interactions (Dunbar 2000). This rather labile period is characterized by many failed evolutionary attempts (Wood and Collard 1999) and a shaky movement forward toward *Homo erectus*. During the third period, from 2MYA to 100,000 years ago, the focus of co-evolution was on the linkage of a broad set of symbolic processes to neural control mechanisms. These linkages involve the growth of prosody, chant, gesture, dance, and a variety of other largely social expressions of language functioning. As Donald (1991) has argued, during most of this period, the use of language was linked to mimetic forms that supported group processes and social structure. During the fourth period, which begins in earnest about 60,000 years ago, we see an increased evolutionary pressure for linking language to the production of a material culture. During this period, language assumed its current dual patterning with organization on the phonological and syntactic levels. However, many of these abilities involve precursors that were initially introduced during the third period for other functions. During each of these four periods, we see a different configuration of co-evolutionary pressures. To help us in thinking about these periods, let us refer to them by the challenges they face. The four challenges are: bipedalism, social cohesion, mimetics, and systematization.

Given this analysis, one might ask why we need to worry about the three earlier periods at all. If we just want to understand the evolution of language, could we just look for the major breakthrough adaptation or emergent process of the last 60,000 years and leave the rest of evolution to the physical anthropologists? The problem is that language itself is deeply rooted in the cognitive advances that were achieved during all four periods. One of the major achievements of cognitive linguistics (Givón 1979; Langacker 1989) has been the characterization of the extent which language is grounded on cognition. Much of this cognition is shared with our primate cousins, but much is uniquely human. If we focus only on recent evolution, we will have a hard time explaining the sudden emergence of human intelligence. By casting our net more broadly across the whole of the last six million years, we can derive a clearer view of the emergence of language from advances in cognition.

Consider the analysis proposed by Tomasello (1999). According to this analysis, the major cognitive difference between apes and humans is that only humans see the actions of conspecifics as involving intentionality. Tomasello holds that this "intentional stance" and its related theory of mind arose during the last 200,000 years, along with the advent of spoken language. Because Tomasello links this major leap forward in human cognition to the narrow time period of the last 200,000 years, he is forced to conclude that genetic changes alone were inadequate to support language evolution. In fact, he characterizes the brevity of the period available for adaptation as the major evolutionary puzzle that we have to solve when discussing language evolution. While it is certainly true that cultural organization has led to major advances in the last 40,000 years, there is no reason to argue that the human lineage has not had time to explore and consolidate a wide range of cognitive adaptations that support language use. If we view the relevant period as involving 6 million years and not 100,000 years, it is far more reasonable to imagine that there has been a series of fundamental cognitive changes providing support for the emergence of language. By expanding our timeframe, we can begin to analyze the specific conceptual supports that underly the attribution of intentionality to the actions of both objects and conspecifics postulated by Tomasello. These supports arise from abilities that we see in chimpanzees, but only as they have been refined by a further six million years of evolution.

To flesh out this analysis, let us take a closer look at how each of the four periods of co-evolution forced fundamental cognitive and physical adaptations in our predecessors.

1. Bipedalism

The advent of bipedal gait stands as a remarkably clear evolutionary watershed in the late Eocene. Between 10MYA and 7MYA, Africa experienced a major tectonic event that led to the formation of the Great Rift Valley. This valley runs down the center of the continent, dividing the flat jungle of the West from the more arid plain of the East (Coppens 1999). The rift valley includes high mountains such as Kilimanjaro and vast lakes such as Lake Tanganyika and Lake Nyasa. Connecting these lakes is the Nile River. The rain shadow created by these mountains also produces a major discontinuity in the two parts of sub-Saharan Africa. Before the emergence of the rift, the great apes of Africa thrived throughout the continent in the warm, lush forest of the Eocene that extended across the continent. With the advent of the rift during the Oligocene, the domain of the apes became restricted to the west of the continent, ending at the rift valley. Those apes that found themselves on the east of the rift were forced to adapt to the loss of the jungle. The major competitors of the apes were the monkeys who were better able to hide in the short bush and scramble for pieces of food, while still maintaining a clear arboreal presence. The challenge to the apes was to make use of the new drier savannah habitat in a way that did not compete with the monkeys. The major adaptation of this period was the shift to bipedalism. Instead of scampering about on four legs or even in a knuckle-walk posture, early hominids began to walk on their two hind legs. Many of the species of australopithecines, such as *Australopithecus afarensis* (Coppens 1999), maintained a full ability to climb trees, while still capable of walking on two legs on the ground. Other species, such as *Australopithecus anamensis*, adapted a more exclusively upright gait. The latter group included our direct ancestors.

1.1 Advantages of bipedalism

Although we know that bipedalism was favored in the sparser habitat of the east, we do not understand exactly why it was such a popular and successful adaptation for so many species of primates. One account focuses on the fact that upright posture allows the animal to expose less body surface to the sun. Having less skin exposed to the sun decreases the impact of radiation and the need for cooling. Another factor must be the ability of the primate to use the forelimbs for other purposes, ranging from tool use to flea picking. There is also the defensive effect of greater height. Hunters in the veldt emphasize the importance of always standing upright to convince lions, rhinos, and other animals of one's size to deter possible aggression.

When we compare the great apes to the monkeys, we can see some other pressures that might have favored the move to an upright posture. When chimps and gorillas are together in social groups, they move away from the four-legged knuckle walk position to a sitting position. This allows them to maintain better eye contact with the others in their group. In the sitting position, they do not have to bend their neck up to maintain eye contact. The shift away from knuckle walking to an upright posture continues this emphasis on body positions that allow full eye contact.

Upright posture and full eye contact also provided room for the emergence of the first gestural signals between early hominids. As many have argued, it is likely that hominids went through a period of relying on some forms of gestural communication. It is clear that upright posture provides room for such a development. However, the evolutionary advantage of early gestures may have been overestimated, since first bipedal primates had cognitive resources that were not yet greatly different from those of today's apes. Although we know that apes can learn and transmit a system of signs (Savage-Rumbaugh and Taglialatela 2001), there is little evidence that the level of sign use they display in natural contexts (Menzel 1975) would provide any major evolutionary advantage.

The major evolutionary advantage of bipedalism probably involved the fact that the arms could be used to hold sticks and other objects that could be used as tools and weapons (Coppens 1995). This ability to wield small weapons and tools would have given early hominids a clear advantage over monkeys in their search for food and defense against predators. Tools could be used to dig for roots, open up termite hills, and even catch fish. To take advantage of these opportunities, early hominids quickly developed an opposing thumb that allowed for a stronger and more precise grasp. Because the arms were no longer needed for tree-climbing, they could be committed in this way to object manipulation. In this account, the first co-evolutionary period focused on the development of new systems for control of the hands.

Bipedalism brought with it a series of evolutionary costs. It placed increased mechanical pressure on the neck, the spine, and all the joints of the legs. The reliance on the feet for walking made them less able to function for climbing. The lungs had to adapt to support the breathing needed for running. However, the most important of these costs involved the narrowing of the hips (Hockett and Ascher 1964). Because early hominids did not have a significantly enlarged cranium, this was not a problem during the period before 2MYA. However, after that time, brain expansion ran up against the earlier commitment to bipedalism, forcing a series of adaptations in female anatomy, parturition, and child-rearing (Li, this volume).

1.2 Neural modifications

The move to bipedalism opened up major cognitive challenges in terms of the control of the hands. Apes already have good control of reaching and basic object manipulation (Ingmanson 1996). However, with both hands now always free for motion, humans were able to explore still further uses of their hands. Rizzolatti (1996) has shown that monkeys (and presumably also primates) have "mirror" neurons in the supplementary eye fields of premotor cortex that respond with equal force when an action such as "grabbing" is carried out either by the self or by the other, including a human. This mechanism provides a way of equating actions performed by the self with actions or postures performed by the other. These neurons are part of the dorsal visual system (Goodale 1993) that allows us to represent the postural perspective of another in a way that is isomorphic with but non-identical to our own postural perspective. This dorsal system for motor matching also connects to systems in motor cortex, cerebellum, and hippocampus that represent various postures and movements of the body. These systems then connect to frontal mechanisms for storage and perspective-shifting that provide a way of using our own full body image to perceive the actions of others. At first this system was only partly elaborated. For early hominids, this system provided support for basic actions such as prying open shells, hitting things with clubs, and digging for roots. As hominids increased their ability to control hand motions and grasping actions, they could use the elaborations of this system to link specific actions to the affordances of different objects, as they are used for different purposes. Early hominids could use this ability to imagine various ways of picking up clubs, prying open termite mounds, and using stones to crack shells.

In the model developed by MacWhinney (1999a), this first level of perspective taking involves the shift from direct perception and action to stored mental representations of perceptions and actions. This is the first step in what I call the “ungrounding” of cognition. The perspective-taking model assumes that cognition is fundamentally grounded in the body. However, the evolution of perspective-taking involves the development of methods for escaping the directness of this grounding. Imagery is a clear example of this, since it is grounded on perception and action, but only through the mediation of additional cognitive storage systems. The additional processes that operate on the direct, grounded system achieve ungrounding. The first cognitive adaptations of the hominid line involved the ungrounding of direct actions and direct perceptions through sensori-motor imagery. When we are directly involved in an activity such as grasping a rock, the actions and perceptions form a direct, integrated whole. However, when we watch other human beings or animals grabbing a rock, the vicarious experience produces a secondary activation of the same pathways used for the direct perception. We can think of this as the social route for the ungrounding of direct perception. This social route is important in allowing us to understand the intentions of others and to learn by imitating their methods. This social route is particularly important during socialization, since it provides the infant with a way of learning skills such as hunting, carrying, building, and courtship.

The second route for the development of ungrounding involves the planning of motor sequences. Once we have developed a system of imagery to track the actions of others, we can use this same system to plan our own future actions. Unlike social imitation, planning must occur without support from the stimulus. To plan the actions involved in chipping an axe, we must be able to call up an image of the desired product, and we must be able to sequence a long series of specific motions that are needed to locate good stones and devise methods for chipping edges. Unassisted planning is relatively fragile in cognitive terms, whereas planning that occurs in a social context is relatively more stable. It is easier to imitate a plan than it is to build that plan without social support.

The movement to an upright posture and a bipedal gait was crucial in the evolution of motor imitation and planning. The ability to construct a planned sequence of actions appears to be a unique property of hominids, as opposed to monkeys and apes. Students of primate tool use (Anderson 1996; Visalberghi and Limongelli 1996) have shown that chimpanzees and capuchin monkeys can use tools in a productive and exploratory way. However, they do not appear to make planful use of mental imagery to limit their search through possible methods of tool use. Instead, they apply all directly perceptible methods in hopes that one may succeed.

The full development of this system for perspective-switching was achieved gradually across the full 6 million years of human evolution. During the first stage of this evolution, hominid abilities to imitate motor sequences were probably not much more advanced than the abilities we see in chimpanzees. However, because of the new centrality of the hands during this period, there was continual evolutionary pressure to improve abilities to imitate motor gestures. As we noted above, the first development were probably those that supported the direct social imitation of the manual gestures, body postures, and activities of others.

Early hominids did not simply reuse the neural areas that had previously been committed to manual activities in the arboreal environment. Because the move to a terrestrial environment was quite gradual (Corballis 1999), hominids needed to provide neural control for the use of the hands in both the arboreal and terrestrial environments. The arboreal environment favors the development of a specific type of motor imagery. Povinelli and Cant (1995) have noted that increases in body weight for larger apes such as orangutans make it important to be able to plan motions through the trees. To do this, the animal needs a map of the self as it executes possible motor actions. The reflexes of this penchant for postural adaptation are still evident in the human enjoyment of dance, dress, exercise, and sport. Donald (1999) makes a similar point in terms of his explication of earlier mimetic plans in hominids. The pressures in the arboreal environment that had favored some limited form of brain lateralization were then carried over to the terrestrial environment (McManus 1999). This ability to shift quickly between alternative environments required neural support for competing postural and affordance systems. This postural flexibility may also have allowed some early hominids to adapt partially to an aquatic environment (Morgan 1997).

Bipedalism also put some pressure on another set of neural mechanisms. Because hominids ceased relying on trees for refuge, and because they were now ranging over a wider territory, they needed to develop improved means of representing spaces and distances. All species must have some way of representing their territory. However, hominids faced the task of representing a rather large territory in which they were both the hunters and the hunted. To do this, they needed to further elaborate earlier mechanisms for spatial encoding. The basic neural mechanisms for this are already well-developed in many mammalian species, including primates (Menzel 1973), canines, and felines. By linking newly developed systems for body imagery to earlier systems for spatial localization, hominids could begin to plan for the motions of their own bodies through space and time. In addition, hominids could use the primate system of mirror neurons to track not only the posture of conspecifics, but also their movements through space and time. Given the structured nature of primate societies, and the increased mobility brought about from bipedalism, it is likely that some form of spatial perspective-taking developed in early hominids. In terms of the model of MacWhinney (1999a), the second level of degrounding from direct perception involved the development of systems for spatial imagery.

Holloway (1995) has presented evidence from endocasts indicating that there was, in fact, a major reorganization of parietal cortex after about 4MYA. This reorganization involved the reduction of primary visual striate cortex and the enlargement of extrastriate parietal cortex, angular gyrus, and supramarginal gyrus. Much of the evidence for Holloway's analysis comes from traces of the changing positions of the lunate sulcus and the intraparietal sulcus over time. According to Holloway, the areas that were expanded during these changes in the parietal support three basic cognitive functions:

1. Processing in the dorsal (parietal) stream of the visual field is important for representing actions of the other in terms of one's own body image.
2. The association areas of parietal maintain a map of the environment for navigation in the new bipedal mode.
3. The supramarginal gyrus is involved in face perception. Expansion of this area would facilitate the development of social patterns and memory for social relations.

The first two of these functions are directly related to the developments that resulted from the adoption of bipedal gait. The third function may have played a large role during the next period in which the focus of evolution was on social cohesion.

2. Social cohesion

The shift to bipedalism is clearly documented in the fossil record. However, fossils speak only indirectly about the evolution of primate social structures. Even facts about hominid group size are difficult to determine from the fossil record. We do know, from studies of primate groups, that increases in group size have positive consequences for protection, particularly of the young. We also know that maintaining larger groups also requires the refinement of methods for social communication and for gathering food.

2.1 Advantages of vocal support for social cohesion

As group size increases, there is increasing conflict for resources such as food, rank, control, and access to females. Dunbar (2000) argues that primates developed a larger neocortex to deal with these tensions. However, it is likely that there was marked pressure on early hominids to maintain large social groups well before we find evidence of a significant expansion in brain size. In particular, between 4.5MYA and 3.5MYA the hominids in East Africa went through an expansion of their range and a proliferation of species. This proliferation was then followed by a period of tight competition with range contraction (Foley 1999). During this period of range expansion and contraction, our ancestors did not improve their social organization simply by growing larger brains, although brain size did increase a bit, mostly allometrically (Holloway 1995). Instead, as Nettle and Dunbar (1997) have argued, it is likely that the hominids of the Pliocene consolidated their group structure by a set of targeted neural-behavioral adaptations. Chief among these, I would argue, is the subordination of the vocal system to cortical control.

It is likely that our Pliocene ancestors possessed the same set of vocal calls available to other primates. By linking communication to group solidarity and membership, our ancestors achieved a comparative advantage over their competitors. Cortical control of vocalization also allowed our ancestors to communicate more intentionally over a distance. More importantly, it allowed them to use chatter to gain the attention of their compatriots to negotiate the basics of group relations. The fact that the great apes did not go through a parallel evolutionary process in West Africa can be attributed to the different requirements on group size in their arboreal habitat.

It is likely that one of the side effects of an increase of cortical control over vocalization was the ability of hominid groups to lock in patterns of vocal behavior that characterize the local group, as opposed to the wider hominid community. At first, these local forms of communication would not be sharply defined. However, as the mechanics for vocalization come increasingly under cortical control, it would become easier for a group to differentiate itself from others by unique vocal features. Songbirds achieve this effect through species-specific processes of vocal learning (Konishi 1995). In birds, dialects allow individuals and groups to maintain their territory against competitors. Hominids could achieve the same effect through differentiation of local patterns for speech and gesture.

2.2 Neural modifications

Achieving neocortical control over the vocal-auditory channel required neuronal reorganization without a major increase in brain size. Ploog (1992) has shown that humans have more direct pyramidal connections between motor cortex and the speech and vocalization areas of the brain stem than do monkeys. Certain areas of the limbic system, such as the anterior thalamic limbic nuclei, have grown disproportionately large in humans. These nuclei serve the supplementary motor area and premotor and orbital frontal cortex. The expansion of these structures points to increased limbic input to the cortex as well as input from the cortex to the limbic structures. Tucker (this volume) shows that the basic adaptation here involved the absorption of the primate external striatum by the neocortex (Nauta and Karten 1970).

In macaques (Jürgens 1979), control of the vocal system relies on the periaqueductal gray matter of the lower midbrain. Additional midbrain regions can stimulate the periaqueductal gray, but the neocortex does not control or initiate primate vocalizations. In man, on the other hand, electrical stimulation of both the supplemental motor area and the anterior cingulate of the frontal cortex can reliably produce vocalization. Primates make few attempts to structure local dialects or otherwise structure their call system through learning (Seyfarth and Cheney 1999). Yerkes and Learned (1925) and others have tried to condition chimpanzee vocalizations in the laboratory and have failed. Human infants, on the other hand, rely at least in part on highly plastic cortical mechanisms to control vocalization. This allows them to pick up the sound patterns of their community through mere exposure. As a result, each hominid group can build up a local vocal accent that is passed on to the next generation through mere exposure. Other aspects of communication, such as conversational sequencing (Trevarthen 1984) may be more linked to modeling and imitation. However, this learning of conversational functions only occurs because the child is locked into the interaction by motivational forces that reward face-to-face vocalization (Locke 1995).

Although primate vocalization is not under cortical control, it has an extremely direct connection to midbrain motivational areas (Pandya, Seltzer, and Barbas 1988). Human language continues to rely on this underlying limbic architecture to provide emotional coloring to vocalization. As Tucker (this volume) argues, the linkage of the vocal system to limbic mechanisms provides grounding in terms of arousal (brainstem and amygdala), motivation (basal ganglion), patterning (striatal-thalamic ciruits), and memory (limbic circuits). Man also retains some direct links between audition and these limbic circuits, as evidenced in the directness of our responses to sounds such as infant cries or the growls of predators.

The linkage of vocalizations to cortical control allowed our ancestors to distinguish themselves from other hominids. It also allowed them to build up a system of face-to-face social interactions. MacNeilage (1998) has argued that the primate gesture of lip smacking is the source of the core CV syllabic structure of human language. The CV syllable has the same motoric structure as lip smacking. Moreover, it is produced in an area of inferior frontal cortex close to that used for lip smacking and other vocal gestures. Primates use lip smacks as one form of social interaction during face-to-face encounters. However, even bonobos, the most social of all primates, do not maintain face-to-face conversations for the long periods that we find in human interactions. Obviously, one must go beyond a boring repetition of lip smacking to maintain a reasonable level of sustained face-to-face vocal contact. Increased cortical control of vocalization allowed our ancestors to begin the process of developing these elaborations. By linking its members into tight affiliative relations through face-to-face interaction, our ancestors achieved a form of social organization that allowed them to maintain large social groups for defense against other hominid groups. Other primates have also responded to these pressures by developing a variety of social support mechanisms (de Waal and Aureli 1996). Other primates have also developed systems for attending to face-to-face interactions and pointing behavior (Gomez 1996). To maximize the effectiveness of face-to-face interactions, hominids then brought the production of facial gestures under cortical control. As in the case of the control of tool use through motor imagery, humans differ from monkeys in the extent to which the cortex can produce gestures upon demand (Myers 1976).

In considering the role of face-to-face vocalization in hominid groups, we must not forget the possible divisive role played by aggressive males (Anders 1994; Goodall 1979). Hominid groups relied on aggressive males for their skills as hunters and their ability to defend the group against attack. However, groups also needed to provide ways to avoid the direction of male aggression toward other members of the group, particularly other males. We know that primates had already developed various methods for handling these conflicts, including exile for problematic males, the formation of master-apprentice relations, and development of male social groups. Within this already established social framework, males could also benefit from ongoing reaffirmation of their social status through face-to-face chat. By socializing young males into this productive use of language for social cohesion, mothers could also contribute to the stability of the group. Breakdowns in these processes could threaten the survival of the group and even the species.

It is easy to assume that the acquisition of a local pattern of vocalization and gesture would lead to conflict between hominid groups. However, studies of kinship patterns (Fox 1976) show that societies often place a premium on the practice of outbreeding. In modern societies, this is formalized through rules of cross-cousin marriage or other forms of exogamy. It is possible that early humans developed similar patterns of controlled intermarriage as a way of solidifying relations between groups. However, the fuller articulation of this system probably occurred in later periods.

This account has emphasized the importance of cortical control over the vocal apparatus. I have made no proposals regarding the evolution of auditory abilities in man. The reason for this is that it appears that primates have already achieved a level of auditory processing ability sufficient to support analysis of all vocal communications (Hauser, Newport, and Aslin 2001; Kuhl and Miller 1978). There is currently no reason to believe that the human auditory system underwent any major adaptation in the last six million years. The linkage of vocal and facial expression to cortical control may seem like a fairly trivial neurological adaptation. However, it helped our ancestors through this period of intense competition between groups and set the stage for the major changes that were to come in the next period.

3. Mimesis

By 2MYA, *Homo erectus* emerged victorious from the period of intense competition. Recent analysis points to *Homo ergaster*, rather than *Homo erectus* as the direct ancestor of *Homo sapiens*. However, the details of the relations between *Homo erectus* and *Homo ergaster* are not yet clear. In this paper, I will refer to these two related populations as *Homo erectus*, understanding that the details of this particular lineage may soon be revised.

During this period, the species had achieved some level of group solidarity through the social use of vocalization. Beginning sometime before 2MYA, our ancestors were confronted with a third major evolutionary opportunity. Having committed themselves to face-to-face communication, and having elaborated their basic systems of social identification and imitation, the first groups of *Homo erectus* were then able to elaborate new forms of symbolic communication in both vocal and gestural modalities.[1] These new systems involved the conventionalization of vocalizations, postures, and gestures in specific social and pragmatic contexts. This conventionalization must be sufficient to allow for social transmission of newly-minted forms from one generation to the next.

Darwin (1877) thought it unlikely that a system such as language could have emerged from gesture. He believed that this would require the shifting of a function from one organ to another and then back again, as if flying had moved from the wing to the stomach and then back to the wing. However there is no reason to think that Darwin would have excluded the possibility that gesture and vocalization underwent a parallel co-evolution. Today, speech and gesture complement each other during communication (McNeill 1985). Some messages are conveyed through speech, but others are conveyed through the hands and the posture. Still other messages are conveyed by changes in the tone of voice. It is likely that each of these channels of communication were also available to *Homo* *erectus* and that they functioned in an interactive and complementary fashion, much as they do now.

3.1 Advantages of mimesis

Mimetic processes (Donald 1991) include gesture, signing, chant, song, drama, ritual, and basic forms of costume. Mimesis relies primarily on iconicity and aesthetics to establish communication. It is often difficult for us to imagine a species that relies on this type of communication. Our first intuition would be that these forms are created on-the-fly and vanish after each communication. However, a system of this sort would provide little evolutionary advantage. Instead, mimetic systems must rely on the construction of socially accepted forms. In this way, the early mimetic systems could support group solidarity, planning, and socialization in an increasingly structured way.

The operation of mimetic processes provides strong linkage of the individual to the group. Through mimetic elaborations of chants, songs, whistles, and yells, the group can contrast itself with its competitors. Through chant and dance, the male society can plan hunts and battles and reenact past struggles. At the same time, mothers can use song, sound play, and chant to hold the attention of their young children while socializing them into the practices of the community. Mimetic processes are linked to the emotional use of language that was elaborated during the previous evolutionary period. As a result, mimetic communications can be used to move people emotionally, preparing them for war, migration, or other major group activities.

It is important to understand the limits of mimesis. Mimesis is not a necessary condition for the execution of cooperative hunting. Felines (lions, tigers), canines (wolves, dogs), and chimpanzees (Goodall 1979) are able to organize sustained cooperative hunting without chant, dance, or other forms of mimesis. It is also misleading to equate mimesis in *Homo erectus* with the advanced forms of chant and dance found in modern-day tribal societies. These observations suggest that the major contribution of mimesis is its ability to elaborate the details of a shared social plan. It could remind group members of the ways in which animals would be trapped by herding into a closed canyon or even over a cliff. It could help organize the preparation and distribution of food. It could also be used to mark social relations of dominance and allegiance. In general, it would be open to much more concrete and immediate uses than we see in modern-day forms of mimesis.

Mimesis was enormously successful. It allowed *Homo erectus* to expand its territory to all of Africa, eventually leading to the extinction of all other hominid species. It also allowed *Homo erectus* to migrate successfully out of Africa to all parts of Asia, including the Middle East, China, and Indonesia. We should attribute this successful expansion to two basic processes. The first was an ability to eliminate competitors, particularly in Africa. To do this, erectus must have relied on group solidarity as a support for warfare. Second, to support the migration to new territories, erectus must have been extremely adaptive. This adaptivity could not have been grounded in some simple physical change. Instead, it must have resulted from a general improvement in cognitive capacity, particularly as reflected in group problem-solving and adaptation.

MacWhinney (1999a) and Tomasello (1999) have developed parallel accounts of these changes that emphasize the importance of being able to take the perspective of another human being as an intentional agent. In both accounts, this ability is a precondition to the full successful use of language. Tomasello locates the emergence of this ability in the period after 200,000 years ago. However, MacWhinney, like Donald (1991), sees the ability arising beginning nearly 2MYA. Malle (this volume) has asked whether it is possible that theory of mind predates language during evolution. Mimetic communication certainly satisfies Malle's criteria for a level 1 theory of mind, but so might many forms of primate communication. Given that mimesis implies capacities for imitation, joint attention, inferential sensitivity, perspective-taking, and the negotiation of unclear meanings, it would appear to satisfy much of Malle's requirements for a level 2 theory of mind. Thus, during this period, theory of mind and mimetic language underwent an intimate co-evolution. However, language as we currently know it was not yet invented.

3.2 Neuronal adaptations

The brain size of *Homo erectus* tripled in size during the period between 2MYA and 100,000 years ago. Some of this increase is allometrically related to the overall growth in stature during this period. Thus, it might be better to speak of a doubling in size, rather than a tripling in size. The brain expanded in size because the preconditions for a successful expansion were now all in place. In particular, previous evolution had already produced at least these five abilities:

1. The shift to bipedalism had freed the hands for gestural communication.
2. The freeing of the hands for tool use had led to further elaboration of primate mechanisms, such as "mirror" neurons, for representation of the actions of others.
3. During several millennia of migration and population movements across open ground and through the bush, our ancestors had developed systems for tracking their own spatial positions in the present, past, and future, as well as the spatial positions and perspectives of others.
4. Our ancestors could use their visual system to generate images of past actions and spatial configurations involving themselves and others.

5. During the period before 2MYA, our ancestors had developed a tight linkage of attention to vocalization processes during face-to-face interaction.

These are the preconditions for the evolution of mimesis. To support mimesis, the brain needed to provide complete episodic storage for combinations of gesture and vocalization. It had to store whole chants or gestures as they expressed particular events of importance to individuals and the group. Some of these chants and gestures might be rather extended. For example, there might be a dance that represented the time of ripening of the mongovo nut or one that outlined the components of the hunt for eland. At the same time, these larger mimetic sequences would be composed of smaller pieces that had achieved other cognitive-social grounding. For example, by pointing to a place in the area around the group's encampment, the dancer could signal the direction of a hunt. By holding his arm in a particular stance, the dancer could mime the attack on the game. These component gestures, movements of the body, and vocal chants could then be reused with other components in a relatively unsystematic fashion.

To fully process and interpret a mimetic communication such as a chant with dance, the onlookers would need to represent the perspective of the chanter, much as they now assume the perspective of the speaker of current human language (MacWhinney 1999a). However, even with the support of perspective-taking, communication still depends on the ability of group members to learn the meaning of new mimetic forms. Thus, mimesis confronts the neural system with a new set of cognitive challenges:

1. The production system must link up stored visual representations to the output processes of chant, gesture, and dance. This linkage of vision to gestural and vocal output requires not only the expansion of both central and peripheral (MacLarnon and Hewitt 1999) output control areas, but also the expansion of their connections to basic visual areas (Givón, this volume). More generally, control of this system requires the construction of a cognitive simulation of the human body (MacWhinney 1999a).
2. Mothers must be able to socialize their children into an understanding of the core mimetic sequences of their own social group.
3. The episodic memory system must store mimetic sequences and their components.
4. As mimetic sequences become elaborated, the brain will need to provide methods for storing whole perspectives, such as that of the hunter, to allow for a switching of perspective, as well as traditional reenactment of these shifts.

Of course, evolution does not respond directly to the changes in behavior that underly these new pressures. Instead, its response is indirect and operates in terms of the principles of population genetics. However, as Dobzhansky (1972) has noted, major evolutionary changes typically arise in response to opportunities opened up by behavioral shifts. In this case, the selective pressure would be for individuals who were able to store the multitude of representations made possible by mimetic communication.

Unlike the evolutionary pressures of earlier periods, the pressures imposed by mimesis cannot be solved simply by linking up older areas or by reusing earlier connections. Instead, the brain must add new computational space to store the multitude of new visual and auditory images (Li, this volume). In addition, the brain needs to expand the role of the frontal areas for storing and switching between perspectives. Because this system grew up in a haphazard way from earlier pieces of lip smacking, pointing, gesture, and rhythm, it would be difficult to extract a core set of elements from mimetic communications, thereby reducing requirements for storage space. Instead, many patterns and forms must be learned and stored as holistic unanalyzed sequences. This Gestalt-like shape of early mimetic patterns corresponds well with the Gestalt-like cognitions that we develop through our interactions with objects. For example, when we chop wood, there is a complete interpenetration of muscle actions, visual experiences, hand positions, and sounds. We can think of this as a single merged form such as I-hands-back-lift-axe-drop-split-chips-wood-cut. Mimetic forms have this same unanalyzed quality. This lack of analysis is not the result of chunking or automatization, since the Gestalt is not constructed by a system of combinatorial semantics. Instead, each chunk is a raw, unanalyzed whole that is fully grounded on direct action and perception. Because they are highly grounded on our direct perceptions and actions, they communicate in a basic way. However, they provide little support for cognitive organization.

The growth of the brain in response to these pressures was so rapid that it is typically assumed that it involves a single genetic mechanism. One such mechanism might involve regulatory genes (Allman 1999) that control the overall size of the cortex. Changes in the timing of the expression of these genes can lead to the observed across-the-board increase in size for the cortex and cerebellum that we see in *Homo erectus*. However, the expansion of the cortex placed additional adaptive pressures on *Homo erectus*. The bigger brain required a much greater level of caloric intake. This pressure could be met through changes in diet and modifications to the digestive system. A more fundamental pressure was the fact that increases in the size of the infant brain produce problems for the birth process. The width of the hips had narrowed in both men and women as a response to bipedalism. As long as the skull was not much larger than that found in the primates, this did not cause major problems for the birth process. However, the expansion of the skull in *Homo erectus* ran directly into this evolutionary barrier. To deal with this, the infant is born at a time when it is still fairly immature and the skull is relatively pliable. The increasingly organized shape of the society guarantees the survival of the child. In addition, women have had to sacrifice their ability to run quickly so that the hips could widen, permitting births with larger infant heads. The slowing of infant development not only helps in the birth process, but also helps the child maintain cortical plasticity (Elman, Bates, Plunkett, Johnson, and Karmiloff-Smith 1996; Julész and Kovács 1995) even into adolescence, thereby further enhancing the ability of the group to construct accepted mimetic patterns.

4. Systematization

By the end of the Pliocene, *Homo erectus* had achieved dominance over its hominid competitors. There were no remaining hominid species in Africa. Presumably *Homo erectus* had eliminated its competitors either through warfare or competition for resources. Homo erectus had also migrated to Eastern Europe, Asia, and Indonesia. However, with the onset of the glaciations of the Pleistocene, our ancestors came under increasing pressure to adapt to the colder, drier environment. It is remarkable to think that this smart primate with its large brain was unable to deal effectively with these new pressures. The contraction of territory back to Africa and then the near extermination of the species at about 70,000 years ago suggest three possibilities. First, it appears that much of the earlier success of *Homo erectus* was achieved at the expense of other hominids. In other words, it was not an expansion of an ecological niche as much as the seizure of a niche that been occupied by various species. Second, in addition to the evolutionary pressure of extreme climate change, *Homo erectus* may have had to deal with pressures from diseases that targeted this particular species. Third, the mere expansion of the brain by itself did not achieve a decisive evolutionary advantage.

The fact that the brain expansion was not enough by itself helps us to understand the shape of recent processes in human evolution. The analysis of mitochondrial DNA (Cann 1995; Cavalli-Sforza 1991; Templeton 1992) allows us to trace the lineage of all current humans back to a single population and perhaps a single mother that lived in Africa about 200,000 years ago. By this time, the brain size of *Homo sapiens* had nearly reached its full proportions. Mere increases in brain size were unlikely to achieve any further evolutionary advantage. Instead, humans needed some way to systematize the profuse growth in mimetic processes that had occurred during the Pliocene. We now have a set of fairly well-articulated proposals regarding the nature of this change. The core of the new system involved the introduction of a set of phonological contrasts (Hockett and Altmann 1973). To achieve accurate articulation of these contrasts, a further set of adaptions were needed for the serial ordering of actions and the precise articulation of sounds. This phonetic revolution occurred gradually over the period from 200,000 to 50,000 years ago.

4.1 Advantages of systematization

Archeological and linguistic research has provided a clear picture of the advantages of phonological and lexical systematization. By coding words into a compact set of contrastive features, *Homo sapiens* was able to build up a lexicon that can name and encode all of the important objects, properties, and actions in our environment. Having a full inventory of the physical world allowed early humans to use different animal and plant species for increasingly refined purposes. Language was used to pinpoint the location of quarry sites for flints and other materials for stone weapons and tools. Master craftsmen used language to teach apprentices how to chip points, prepare hides for tanning, carve out wooden bowls, bind axes to poles with sinew, sew hides into tents, and tame pets. As toolmaking progressed, language was used to explain how, where, and when to plant and water seeds. By codifying these times and practices in verse and religion, the first agricultural systems were developed and shortly thereafter the first settled villages. From these roots, came the civilizations of the ancient Middle East, Egypt, and China with their invention of writing, organized warfare, chariots, metalmaking, government, and increasingly formalized religion. Modern society has elaborated on this foundation with the creation of cities, books, bombs, law, medicine, and the Internet. All of these developments are consequences of the introduction of systematization for phonology and lexicon. However, this recent expansion would not have been possible without the major cognitive modifications of the full 6 million years of human evolution.

4.2 Neuronal adaptations

Some of the adaptations required for smooth vocal production are quite peripheral. (Lieberman 1973), involving changes to the vocal tract, the structure of the larynx, muscle innervation, tongue support, and facial musculature. Some of these changes were underway before the Pleistocene; others have been more recent. To control this additional external hardware, the brain has needed to fine-tune its mechanisms for motor control. This fine-tuning does not require the type of brain expansion that occurred in *Homo erectus*. Instead, it involves the linking of inferior frontal areas for motor control to temporal areas (Gabrieli, Brewer, Desmond, and Glover 1997) for sequence storage. These linkages (Damasio and Damasio 1988) involve pathways that lie under the central sulcus. They constitute a functional neural circuit that implements a phonological loop for learning new words (Gupta and MacWhinney 1997). The auditory shapes of words are stored in topological maps (Miikkulainen 1990) in superior temporal auditory cortex and can be associated to visual images in inferior temporal areas. This linkage of the vocal-auditory channel to the visual channel further develops binding the entrainment of the vocal-auditory channel by the visual channel (Givón, this volume).

Once *Homo sapiens* had achieved an ability to produce, store, and learn a large vocabulary of phonologically organized forms, the remaining steps in the evolution of language were comparatively easy. Humans had already achieved a mimetic system for perspective-taking and perspective-switching. This system allowed listeners to mentally reenact the motions, rhythms, and chants of the speaker as they depicted movement between places and actions on objects. Once words became available, speakers and listeners could parse these single-package gestalt-like communications into their components. With words to name specific objects and participants, it was possible to separate out nouns from verbs. This adaptation to grammar required no particular new cognitive skill for nouns. However, for predicates such as verbs, it was important to store linkages between the overall configuration of the action and the specific uses with participants. In other words, children had to learn how to manage language in terms of item-based syntactic constructions (MacWhinney 1975, 1982) and phrasal collocations (Li, this volume), including "verb islands" (Tomasello 2000). Neuronal processes for this level of control involve little in the way of new evolution. However, they place storage demands on the pathways between the temporal lexical areas and the frontal planning and sequencing areas.

As speakers build up longer and longer strings of propositions, they rely increasingly on frontal areas, such as dorsolateral prefrontal cortex (DLPFC) for the storage of one perspective that allows shifting to a secondary perspective. Shifts of this type are central in the processing of anaphors and gaps in argument structure. As MacWhinney (2000) has shown, these various syntactic processes are grounded not on the construction of abstract syntactic trees, but on the direct processing of embodied perspectives of the type that were also important during the period of mimetic communication.

4.3 Cultural evolution

Given the minimal nature of the additional adaptations needed to support human language, why did the human species suddenly become so successful? The reason for this great success is that these minimal changes were crucial in triggering a reuse of previous adaptations. With the onset of phonological systematization, humans were able make a much fuller use of the massive expansion in brain size. They did this by constructing a linguistic system that uses the entire brain to represent experience. This was a social construction of the last 60,000 years, not a process grounded on the slow timetable of biological evolution.

Language builds systematically on each of the achievements of the six million years of evolution.

1. It uses the basic primate sensorimotor systems of posterior cortex to encode objects in terms of direct perceptions and properties.
2. It uses the dorsal systems developed in the late Miocene for storing mental representations of action sequences on objects. This system supports the representation of concrete nouns, adjectives, and intransitive verbs.
3. It uses the hippocampal and parietal systems developed in the late Miocene for encoding navigation paths and the spatial location of others. This system supports the representation of deictic terms, prepositions, and locative adverbs.
4. It uses the system developed during the Pliocene for the planning of mimetic sequences. MacWhinney (2000) views this ability as forming the core of the grammatical system of transitivity.
5. It uses the perspective-taking involved in mimetic communication as the core of perspective-taking in language. Systems for expressing these shifts include complementation, embedding, binding and anaphora, factivity, evidentiality, and marking for politeness and kinship.

Language is grounded entirely on these achievements of the past 6 million years. However, once language is available as a method of cognitive representation, cognition itself is fundamentally altered. The most remarkable property of human language is that it has smooth and nearly immediate access to the entire brain. Through this ability to integrate across modalities and modules, language is able to overcome modularity and open up the mind to fully human consciousness. Language relies on the entire brain to achieve its complete cognitive simulation of experience in terms of objects, space, action, and social relations. Because it integrates these separate modules so thoroughly, it allows us to fully escape the modularity that is present in primates (Russon and Bard 1996) and young children (Hermer-Vazquez, Moffet, and Munkholm 2001). Without language, it may be possible to focus directly on the position of an object without regard to earlier orientations or the orientations of others. Without language, we can focus on an action without breaking it apart into its component participants. In order to achieve lexicalization, language forces us to analyze experience into its components. Although language forces us to break our experiences into pieces, it provides ways of then recombining these pieces into much larger edifices. Moreover, narrative and discourse allow us to integrate our own experiences more fully with those of others. In this way, language bundles the whole of mental life into a single, more fully conscious, but relatively less grounded whole.

Because language is a human invention, the brain provides us with a great deal of latitude in the way it can be represented. As Wittgenstein (1953) observed, language is a like a well-pruned hedge. Externally, each hedge must look like the others. However, internally each hedge can have a very different shape. This means that we should not be surprised to find large individual differences in the neuronal basis of higher-level dynamic control for language. For example, Children with large focal lesions to the left hemisphere areas that typically control language are able to achieve normal language functioning by using parallel areas in the right hemisphere (Booth *et al.* 2000). Several patterns of reorganization to damage have been identified for young children (MacWhinney, Feldman, Sacco, and Valdes-Perez 2000) and similar reorganization certainly occurs in adults (Holland *et al.* 1985). In addition to the flexibility found in patients with brain lesions, there are important individual differences in the way that the brain organizes for language in normally-developing speakers. In particular, it is likely that the process of cerebral lateralization operates in very different ways in different children, with some making a sharp separation between the functions of the two hemispheres (Beeman *et al.* 1994) and others allowing for more redundancy.

5. Social-developmental consequences

Although the emergence of language in humans places only modest demands on additional neuronal evolution, it had major consequences for the way in which we use our brains. Vygotsky (1962) showed how language begins as an external social phenomenon and then is internalized to provide the backbone of human cognition. Vygotsky attributed this internalization to the emergence of "inner speech" and compressed forms of reference and topicalization, but provided little additional detail regarding his proposal. Tomasello (1999) extended Vygotsky's analysis by linking language learning to the child's ability to treat others as intentional agents. This assumption of an intentional stance helps the child follow the meaning of the parent when learning new words. Tomasello also shows how intentionality allows the child to construct the representations underlying causal predications. Infants will extend intentionality even to inanimate objects, when they have eyes that represent a face (Gergely, Nádasdy, Csibra, and Bíró 1997) or even when they do not but appear to have intention (Csibra *et al.* 1999). Later, through fictive extensional processes (Lakoff 1987; Talmy 1988), children treat even inanimate objects as acting intentionally and causally. This system of causal representation forms the basis for grammatical processes such as case marking, conflation, passivization, inverse, and causitivization.

MacWhinney (1999a) shows how the construction of the intentional perspective is a preliminary to a wider expansion of language and perspective-taking to incorporate the specific linguistic propositions needed for an articulated theory of mind. In computing the details of social relations, we often need to take the perspective of the other and imagine how that person will react to a variety of possible actions that we might perform. Typically, we select the exact shape of a given utterance or action out of a much larger field of potential actions based largely on this act of social perspective-taking. Would our utterances offend some particular individual? Would they violate some moral precept? Would we fail to properly articulate our position and therefore leave ourselves open to later misinterpretation? As we compute these various possible scenarios, we use the three lower levels of perspective-taking (direct perception, space-time, causal action) to move about in a conceptual space that was constructed through linguistic interactions embedded in a social context. These processes allow us to construct enormously complex systematic (Levi-Strauss 1963) views of social structure. Yet the cognitive resources we are using are little more extensive than those available to *Homo erectus*. Instead, through the construction of a method for creating new words, we have been able to reuse our mimetic resources to support the full articulation of human culture. In this sense, the appearance of modern language is a paradigm case of an emergent behavior (MacWhinney 1999b).

6. Disorders of communication

A good understanding of language evolution should be able to illuminate issues in communicative disorders. In particular, we might expect that the most recent adaptations would be the ones that are least fully integrated into the human genome. One clear prediction of the current analysis is that the most recent adaptations for language involve motor control for phonology. This view is highly compatible with the fact that we see a series of language disorders such as:

1. stuttering: a process that involves problems in the activation and coordination of output processes based on individual lexical items,
2. dysphagia: problems with chewing and swallowing that may reflect some instability in the rewiring of innervation of the vocal tract,
3. articulation disorders: between the ages of 4 and 8, many children have problems articulating the exact shapes of words. Disorders such as lisping sometimes continue into adulthood.
4. tongue curling: about 30% of the adult population cannot perform the type of tongue curling or the type of tongue bending needed to properly articulate sounds such as the retroflex stops of Hindi. There are similar problems with the articulation of one of the three Czech r-sounds.

There is good reason to believe that these disabilities represent incomplete consolidation of recent evolutionary changes. If we then further parcel out cases of mental retardation[2], autism[3], fetal damage, and chromosomal abnormalities[4], we are left with a group of children who are said to have specific language impairment (SLI). The incidence of some form of specific language impairment in the population is often estimated to be about 7%. In the clinic, language disorders are nearly four times more frequent in males than in females (Bishop 1997). However, epidemiological studies have shown that the actual balance in the population is nearly equally distributed between the sexes (Leonard 1998). We can divide the overall category of specific language impairment into three subtypes:

1. expressive language disorders,
2. pragmatic disorders,
3. grammatical disorders

The category of expressive language disorders is probably just a further project of the group of motor speech disorders we mentioned above. However, this group has even deeper problems with motor output and formulation. The instability of these skills is, again, evidence for the recency of these evolutionary developments. The existence of a class of children with pragmatic disorders indicates that some aspects of perspective-taking are still variable in the human species. The account offered above suggests that these abilities were consolidated long ago during evolution. It is likely that children with pragmatic disorders are actually reflecting incomplete consolidation of the more recent evolutionary changes in frontal lobe memory processes used to store and switch perspectives.

Finally, we turn to a consideration of the evolutionary status of children with both expressive and receptive disorders, including grammatical disorders[5]. Recent studies (Franks and Connell 1996; Van der Lely and Stollwerk 1996, 1997) have suggested that children with grammatical disorders have a specific problem with argument chains. However, it may be that these problems are based on representational deficits, rather than any specifically grammatical problem. Some recent accounts (Merzenich *et al.* 1996; Tallal *et al.* 1996) have attempted to link specific language impairment to an underlying auditory deficit. Given the fact that it is difficult to find any clear structural or functional differences between human and primate auditory processing or learning (Hauser *et al.* 2001), it would seem that raw auditory processing is an unlikely candidate for a weak link in the evolutionary chain. Grammatical SLI may reflect difficulties in the representation of lexical items (Chiat 2001; Gupta and MacWhinney 1997) or problems in the storage of chains of lexical items for syntactic combination.

The patterns of population dispersion that occurred during the last 100,000 years might well have an impact on the relative frequency of language disorders in different human populations. For example, there is now evidence that Cro-Magnon groups intermarried with Neanderthal groups in both Europe and the Middle East. It is generally accepted that Neanderthal people, although derivative from the basic human stock, had a relatively low level of cultural innovation and perhaps a low level of vocal control (Lieberman 1973). Any intermarriage between groups with greater linguistic ability and groups with less linguistic ability could lead to a population that included individuals with some form of language impairment or limitation. Currently, our information on the actual distribution of language disorders in different parts of the world is so incomplete that we cannot evaluate these patterns. Moreover, languages with different structures would put different strains on children's ability to learn and produce language (Leonard 1998).

If language is such an important aspect of human existence, then how can it be that such a sizeable portion of the population has language impairment? We know that adults with language impairments are less successful in finding mates (Leonard 1998). Given this, it is likely that the genes selecting for specific language impairment are linked chromosomally to other genes that provide compensatory adaptive advantages. However, we have currently no ideas about what these additional characteristics might be.

7. Conclusion

The study of language evolution has advanced enormously in recent years. New evidence from the fossil record, paleoclimatology, genetic analysis, neuroscience, infancy research, and cognitive grammar has fueled these advances. As the database of evidence regarding man's last six million years continues to grow, we will be able to articulate increasingly precise ideas about the co-evolution of language, brain, and social processes. In this paper, I have presented one fairly detailed account that takes into account the complete set of adaptations that occurred across the full range of six million years. By casting a broad net across this whole period, we can see more clearly how modern language relies on a set of adaptations of which some are shared with our primate cousins, some are ancient hominid inventions, and some have emerged in our recent past.

Notes

* My thanks to Paul Fletcher and the University of Hong Kong for providing me with time, support, and encouragement to work out the ideas summarized in this paper. My thanks to the editors for their detailed comments on the first draft.

1. As Hewes and many others have argued, there is reason to believe that early humans went through a period of heavy use of gestural communication. Some have suggested that the freeing of the hands by the shift to bipedalism should lead directly to the use of gesture. However, in cognitive terms, our ancestors from 6MYA were still close to the ancestors of the great apes. Their systems of natural gestural communication might well have been

parallel to those observed in today's chimpanzees . Thus, it seems to me that the real flowering of gestural communication probably did not occur until this third period, when systems for perspective-taking, sustained attention, imitation, and planning could make better use of conventionalized gestures.

2. Many children with language impairments also have some organic brain problem. These difficulties can include smaller brain sizes, lower neural activation, early brain lesions, traumatic brain injury, hydrocephalus, and epilepsy.

3. Some authors have claimed that autism and Landau-Kleffner's syndrome represent disorders in the biology of a cognitive module supporting theory of mind . Although some linkage of this type seems likely and would fit in well with the current theory, the embryological and genetic facts about the origin of autism cast some doubt on any theory that views it as a unitary disorder with a single consequence.

4. Another set of developmental problems is based not on specific motor disabilities, but on malformations that occur during fetal development from various imbalances. These include cleft palate and cerebral palsy. We also know that the processes of meiosis and mitosis are vulnerable to basic errors in translocation and nondisjunction in all species. These processes might also impact the use of language, but would not represent any pattern with a clear status in recent evolution. The most important syndromes here are Down Syndrome, Williams Syndrome, Turner's syndrome, and Kleinfelter's syndrome. These genetic abnormalities have important impacts on language development, but they cannot be linked in any direct way to the evolution of language. For example, Bates and Goodman have shown how we can compare normal language development with language development in children with Down and Williams syndrome. This comparison underscores the linkage of grammar to the lexicon in the first stages of language learning. However, this type of comparison does not tell us whether grammar and the lexicon developed in parallel across evolution.

5. One study of a family in London has shown a pattern of disorders consistent with autosomal dominant inheritance. Affected family members have problems forming regular past tense forms such as *jumped*, but can produce irregulars such as *fell.* However, this disorder involves not only language, but also chewing, swallowing, and a variety of other behaviors .

References

Alcock, K. J., Passingham, R., Watkins, K. E., and Vargha-Khadem, F. 2000. "Oral dyspraxia in inherited speech and language impairment and acquired dysphasia". *Brain and Language* 75: 17–33.

Anders, T. 1994. *The Origins of Evil: An Inquiry into the Ultimate Origins of Human Suffering.* Chicago: Open Court.

Anderson, J. 1996. "Chimpanzees and capuchin monkeys: Comparative cognition". In *Reaching into Thought: The Minds of the Great Apes*, A. E. Russon, K. A. Bard, and S. T. Parker (eds), 23–56. Cambridge: Cambridge University Press.

Baron-Cohen, S., Baldwin, D.A., and Crowson, M. 1997. "Do children with autism use the speaker's direction of gaze strategy to crack the code of language?" *Child Development* 68: 48–57.

Bates, E., and Goodman, J.C. 1999. "On the emergence of grammar from the lexicon". In *The emergence of language*, B. MacWhinney (ed), 29–80. Mahwah, NJ: Lawrence Erlbaum Associates.

Beeman, M., Friedman, R.B., Grafman, J., Perez, E., Diamond, S., and Lindsay, M.B. 1994. "Summation priming and coarse semantic coding in the right hemisphere". *Journal of Cognitive Neuroscience* 6: 26–45.

Bishop, D. 1997. *Uncommon Understanding*. Hove: Psychology Press.

Booth, J.R., MacWhinney, B., Thulborn, K.R., Sacco, K., Voyvodic, J.T., and Feldman, H.M. 2000. "Developmental and lesion effects during brain activation for sentence comprehension and mental rotation". *Developmental Neuropsychology* 18: 139–69.

Cann, R., L. 1995. "Mitochondrial DNA and human evolution". In *Origins of the Human Brain*, J.-P. Changeux and J. Chavaillon (eds). Oxford: Clarendon Press.

Cavalli-Sforza, L. 1991. "Genes, people, and languages". *Scientific American* 265: 72–8.

Chiat, S. 2001. "Mapping theories of developmental language impairment: Premises, predictions, and evidence". *Language and Cognitive Processes* 16: 113–42.

Coppens, Y. 1995. Brain, locomotion, diet, and culture: how a primate, by chance, became a man. In *Origins of the Human Brain*, J.-P. Changeux and J. Chavaillon (eds). Oxford: Clarendon Press.

Coppens, Y. 1999. "Introduction". In *African Biogeography, Climate Change, and Human Evolution*, T.G. Bromage and F. Schrenk (eds). New York: Oxford University Press.

Corballis, M.C. 1999. "Phylogeny from apes to humans". In *The Descent of Mind: Psychological Perspectives on Hominid Evolution*, M.C. Corballis and S.E.G. Lea (eds), 40–70. Oxford: Oxford University Press.

Csibra, G., Gergely, G., Bíró, S., Koos, O., and Brockbank, M. 1999. "Goal attribution without agency cues: The perception of pure reason in infancy". *Cognition* 72: 237–67.

Damasio, A.R., and Damasio, H. 1988. "The neuroanatomical correlates of aphasia and the understanding of the neural substrates of language". In *Language, Speech and Mind: Studies in Honor of Victoria A. Fromkin*, L.M. Hyman and C.N. Li (eds). London: Routledge and Kegan Paul.

Darwin, C. 1877. "A biographical sketch of an infant". *Mind* 2: 292–4.

de Waal, F.B.M., and Aureli, F. 1996. "Consolation, reconciliation, and a possible cognitive difference between macaques and chimpanzees". In *Reaching into thought: The Minds of the Great Apes*, A.E. Russon and K.A. Bard and S.T. Parker (eds), 80–110. Cambridge: Cambridge University Press.

Deacon, T. 1997. *The Symbolic Species: The Co-evolution of Language and the Brain*. New York: Norton.

Dobzhansky, T. 1972. "Genetics and the diversity of behavior". *American Psychologist* 27: 523–30.

Donald, M. 1991. *Origins of the Modern Mind*. Cambridge, Mass.: Harvard University Press.

Donald, M. 1999. "Preconditons for the evolution of protolanguages". In *The Descent of Mind: Psychological Perspectives on Hominid Evolution*, M.C. Corballis and S.E.G. Lea (eds), 138–54. Oxford: Oxford University Press.

Dunbar, R. 2000. "Causal reasoning, mental rehearsal, and the evolution of primate cognition". In *The Evolution of Cognition*, C. Heyes and L. Huber (eds). Cambridge, Mass.: MIT Press.

Elman, J., Bates, E., Plunkett, K., Johnson, M., and Karmiloff-Smith, A. 1996. *Rethinking Innateness.* Cambridge, Mass.: MIT Press.

Foley, R. 1999. "Evolutionary geography of Pliocene African hominids". In *African Biogeography, Climate Change, and Human Evolution*, T.G. Bromage and F. Schrenk (eds). New York: Oxford University Press.

Fox, R. 1976. *Kinship and Marriage: An Anthropological Perspective.* Cambridge: Cambridge University Press.

Franks, S.L., and Connell, P.J. 1996. "Knowledge of binding in normal and SLI children". *Journal of Child Language* 23: 431–64.

Gabrieli, J.D.E., Brewer, J.B., Desmond, J.E., and Glover, G.H. 1997. "Separate neural bases of two fundamental memory processes in the human medial temporal lobe". *Science* 276: 264–6.

Gauger, L.M., Lombardino, L.J., and Leonard, C.M. 1997. "Brain morphology in children with specific language impairment". *Journal of Speech, Language, and Hearing Research* 40: 1272–84.

Gergely, G., Nádasdy, Z., Csibra, G., and Bíró, S. 1997. "Taking the intentional stance at 12 months of age". *Cognition* 56: 165–93.

Givón, T. 1979. *On Understanding Grammar.* New York: Academic Press.

Givón, T. 1998. "On the co-evolution of language, mind and brain". *Evolution of Communication* 2: 45–116.

Gomez, J.C. 1996. "Ostensive behavior in great apes: The role of eye contact". In *Reaching into Thought: The Minds of the Great Apes*, A.E. Russon, K.A. Bard, and S.T. Parker (eds), 131–51. Cambridge: Cambridge University Press.

Goodale, M.A. 1993. "Visual pathways supporting perception and action in the primate cerebral cortex". *Current Opinion in Neurobiology* 3: 578–85.

Goodall, J. 1979. "Life and death at Gombe". *National Geographic* 155: 592–620.

Gopnik, M. 1990. "Feature blindness: A case study". *Language Acquisition* 1: 139–64.

Gupta, P., and MacWhinney, B. 1997. "Vocabulary acquisition and verbal short-term memory: Computational and neural bases". *Brain and Language* 59: 267–333.

Hauser, M., Newport, E., and Aslin, R. 2001. "Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins". *Cognition* 78: B53-B64.

Hermer-Vazquez, L., Moffet, A., and Munkholm, P. 2001. "Language, space, and the development of cognitive flexibility in humans: The case of two spatial memory tasks". *Cognition* 79: 263–99.

Hewes, G.W. 1973. "Primate communication and the gestural origin of language". *Current Anthropology* 14.

Hockett, C., and Ascher, R. 1964. "The human revolution". *Current Anthropology* 5: 135–67.

Hockett, C.F., and Altmann, S.A. 1973. “A note on design features”. In *Animal Communication*, T.A. Sebeok (ed). Bloomington, Ind.: Indiana University Press.

Holland, A., Miller, J., Reinmuth, O., Bartlett, C., Fromm, D., Pashek, G., Stein, D., and Swindell, C. 1985. “Rapid recovery from aphasia: A detailed language analysis”. *Brain and Language* 24: 156–73.

Holloway, R. 1995. “Toward a synthetic theory of human brain evolution”. In *Origins of the Human Brain*, J.-P. Changeux and J. Chavaillon (eds). Oxford: Clarendon Press.

Ingmanson, E.J. 1996. “Tool-using behavior in wold Pan paniscus: Social and ecological considerations”. In *Reaching into thought: The minds of the great apes*, A.E. Russon, K.A. Bard, and S.T. Parker (eds). Cambridge: Cambridge University Press.

Julesz, B., and Kovacs, I. (eds). 1995. *Maturational Windows and Adult Cortical Plasticity*. New York: Addison-Wesley.

Jürgens, U. 1979. “Neural control of vocalization in nonhuman primates”. In *Neurobiology of Social Communication in Primates*, H.D. Steklis and M.J. Raleigh (eds). New York: Academic Press.

Konishi, M. 1995. “A sensitive period for birdsong learning”. In *Maturational Windows and Adult Cortical Plasticity*, B. Julesz and I. Kovacs (eds), 87–92. New York: Addison-Wesley.

Kuhl, P.K., and Miller, J.D. 1978. “Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli”. *Journal of the Acoustical Society of America* 63: 905–17.

Lakoff, G. 1987. *Women, Fire, and Dangerous Things*. Chicago: Chicago University Press.

Langacker, R. 1989. *Foundations of Cognitive Grammar*, Volume II: *Applications*. Stanford, Calif.: Stanford University Press.

Leonard, L. 1998. *Children with Specific Language Impairment*. Cambridge, Mass.: MIT Press.

Levi-Strauss, C. 1963. *Structural Anthropology*. New York: Basic Books.

Lieberman, P. 1973. “On the evolution of language: A unified view”. *Cognition* 2: 59–94.

Locke, J.L. 1995. “Development of the capacity for spoken language”. In *The Handbook of Child Language*, P. Fletcher and B. MacWhinney (eds). Oxford: Blackwell.

MacLarnon, A., and Hewitt, G. 1999. “The evolution of human speech”. *American Journal of Physical Anthropology* 109: 341–63.

MacNeilage, P. 1998. “The frame/content theory of evolution of speech production”. *Behavioral and Brain Sciences* 21: 499–546.

MacWhinney, B. 1975. “Pragmatic patterns in child syntax”. *Stanford Papers And Reports on Child Language Development* 10: 153–65.

MacWhinney, B. 1982. “Basic syntactic processes”. In *Language acquisition*, Volume I. *Syntax and semantics*, S. Kuczaj (ed), 73–136. Hillsdale, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. 1999a. “The emergence of language from embodiment”. In MacWhinney (1999b: 213–56).

MacWhinney, B. (ed). 1999b. *The Emergence of Language*. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. 2000. “Perspective-taking and grammar”. *Japanese Society for the Language Sciences* 1: 1–25.

MacWhinney, B., Feldman, H. M., Sacco, K., and Valdes-Perez, R. 2000. “Online measures of basic language skills in children with early focal brain lesions”. *Brain and Language* 71: 400–31.

McManus, I. C. 1999. “Handedness, cerebral lateralization, and the evolution of language”. In *The Descent of Mind: Psychological Perspectives on Hominid Evolution*, M. C. Corballis and S. E. G. Lea (eds), 194–217. Oxford: Oxford University Press.

McNeill, D. 1985. “So you think gestures are nonverbal?” *Psychological Review* 92: 350–71.

Menzel, E. 1973. “Chimpanzee spatial memory organization”. *Science* 182: 943–5.

Menzel, E. 1975. “Purposive behavior as a basis for objective communication between chimpanzee”. *Science* 189: 652–4.

Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., and Tallal, P. 1996. “Temporal processing deficits of language-learning impaired children ameliorated by training”. *Science* 271: 77–81.

Miikkulainen, R. 1990. “A distributed feature map model of the lexicon, *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, 447–54. Hillsdale, NJ: Lawrence Erlbaum Associates.

Morgan, E. 1997. *The Aquatic Ape Hypothesis*. London: Souvenir Press.

Myers, R. E. 1976. “Origins and evolution of language and speech”. *Annals of the New York Academy of Sciences* 280: 745–57.

Nauta, W. J. H., and Karten, H. J. 1970. “A general profile of the vertebrate brain, with sidelights on the ancestry of cerebral cortex”. In *The Neurosciences*, G. C. Quarton and T. Melnechuck and G. Adelman (eds), 7–26. New York: Rockefeller University Press.

Nettle, D., and Dunbar, R. 1997. “Social markers and the evolution of reciprocal exchange”. *Current Anthropology* 38: 93–9.

Pandya, D. P., Seltzer, B., and Barbas, H. 1988. “Input-out organization of the primate cerebral cortex”. In *Comparative primate biology: Neurosciences*, H. Steklis and J. Irwin (eds). New York: Liss.

Plante, E. 1996. “Phenotypic variability in brain-behavior studies of specific language impairment”. In *Toward a Genetics of Language*, M. L. Rice (ed). Mahwah, NJ: Lawrence Erlbaum Associates.

Ploog, D. W. 1992. “Neuroethological perspectives on the human brain: From the expression of emotions to intentional signing and speech”. In *So Human a Brain: Knowledge and Values in the Neurosciences*, A. Harrington (ed), 3–13. Boston: Birkhauser.

Povinelli, D. J., and Cant, J. G. H. 1995. “Arboreal clambering and the evolution of self-conception”. *Quarterly Journal of Biology* 70: 393–421.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. 1996. “Premotor cortex and the recognition of motor actions”. *Cognitive Brain Research* 3: 131–41.

Russon, A. E., and Bard, K. A. 1996. “Exploring the minds of the great apes: Issues and controversies”. In *Reaching into Thought: The Minds of the Great Apes*, A. E. Russon and K. A. Bard (eds). Cambridge: Cambridge University Press.

Seyfarth, R., and Cheney, D. 1999. "Production, usage, and resonse in honhuman primate vocal development". In *The Design of Animal Communication*, M. Hauser and M. Konishi (eds). Cambridge, Mass.: MIT Press.

Stringer, C., and McKie, R. 1996. *African Exodus*. London: Pimlico.

Tallal, P., Miller, S.L., Bedi, G., Byma, G., Wang, X., Nagarajan, S.S., Schreiner, C., Jenkins, W.M., and Merzenich, M.M. 1996. "Language comprehension in language-learning impaired children improved with acoustically modified speech". *Science* 271:81–4.

Talmy, L. 1988. "Force dynamics in language and cognition". *Cognitive Science* 12:59–100.

Templeton, A.R. 1992. "Human origins and analysis of mitochondrial DNA sequences". *Science* 255:737.

Tomasello, M. 1999. *The Cultural Origins of Human Communication*. Cambridge: Cambridge University Press.

Tomasello, M. 2000. "Do young children have adult syntactic competence?" *Cognition* 74: 209–53.

Trevarthen, C. 1984. "Biodynamic structures, cognitive correlates of motive sets and the development of motives in infants". In *Cognition and Motor Processes*, W. Prinz and A.F. Sanders (eds). Berlin: Springer.

van der Lely, H., and Stollwerk, L. 1997. "Binding theory and grammatical specific language impairment in children". *Cognition* 62:245–90.

van der Lely, H.K.J., and Stollwerk, L. 1996. "A grammatical specific language impairment in children: An autosomal dominant inheritance?" *Brain and Language* 52:484–504.

Visalberghi, E., and Limongelli, L. 1996. "Acting and undertanding: Tool use revisited through the minds of capuchin monkeys". In *Reaching into thought: The minds of the great apes*, A.E. Russon, K.A. Bard and S.T. Parker (eds), 57–79. Cambridge: Cambridge University Press.

Vygotsky, L. 1962. *Thought and Language*. Cambridge, Mass.: MIT Press.

Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Blackwell.

Wood, B., and Collard, M. 1999. "The human genus". *Science* 284:65–71.

Yerkes, R.M., and Learned, B.W. 1925. *Chimpanzee Intelligence and its Vocal Expressions*. Baltimore: Williams and Wilkins.

Chapter 10

The relation between language and theory of mind in development and evolution*

Bertram F. Malle
University of Oregon

When Tom Givón asked me a while ago what my chapter would be about, I said, "Roughly, about the relation between language and theory of mind." His laconic response was, "Well, they are inseparable." So, I thought, there goes my chapter.

But not really. There is reason to believe that language and theory of mind have co-evolved, given their close relation in development and their tight connection in social behavior. However, they are clearly not inseparable — neurologically, cognitively, or functionally. So the question becomes, "What is the *exact* relation between language and theory of mind, in evolution, development, and social behavior?" To answer this question is a daunting task; I will try merely to clear a path toward an answer. I will consider several possible relations between the two faculties, bring conceptual arguments and empirical evidence to bear on them, and end up arguing for an escalation process in which language and theory of mind have fueled each other's evolution.

1. Some stipulations about language and theory of mind

1.1 Language

What constitutes genuine language? Modern language is typically described on several distinct levels: phonology, morphology, syntax, and semantics. To define language, the phonological level should be unimportant, because sign systems can be full-blown languages (Morford, this volume; Goldin-Meadow, this volume). Also, even though phonetic advances may have played an important role in the evolution of human speech (Studdert-Kennedy 2000b), very early forms of language may not have been spoken but signed (e.g., Corballis 2000; Donald 1998; Givón, this volume).

The next two levels, morphology and syntax, can vary extensively within systems that should count as language. I am not just referring to the difference between analytic and synthetic languages but to the degree that a symbolic system includes (in morphology, syntax, or lexicon) central cognitive distinctions, such as perspective (I—you), modality (is—ought), spatial and temporal specifications, causality, intentionality, and so on. If a representational system contains none of these distinctions (e.g., a purely depictive system such as photography), it should not be called a language. A key component of language is that it offers choices in its representational repertoire. (Drawing may thus be called a language if it is understood to offer choices and distinctions). But the number of distinctions will vary, and so will the specific format used to mark these distinctions (such as morphology, syntax, or lexicon).

That a language must be a symbolic system (i.e., some entities within the system are understood to represent some entities outside the system) is uncontroversial. However, we must allow wide variation in semantic complexity. Just one notch outside language I would position expressive systems, such as unintentional screams of pain and satisfaction, facial features of emotion, and so on. Because they allow no choice in expression, they are not language (Li, this volume). They still "represent," but they do so by their very nature, not because a language user decides to use them this way.[1] (In Paul Grice's 1957, terms, they have *natural* meaning.) When symbols are iconic but under the control of the user, we have a primitive form of language. Ullin T. Place (2000) writes:

> Before human beings developed the ability to communicate vocally, they communicated with one another by means of some form of sign language whose primary function was to organize the cooperative social activities involved in hunting and foraging. In such a language the signs wold be entirely iconic, that is to say they would function as signs entirely by virtue of a resemblance between the sign and the object or movement it stands for.

Michael Corballis (2000) concurs:

> Language evolved from manual gestures, since gestures have at least the potential to represent concepts iconically rather than in abstract form. Once a set of iconic representations is established, increasing usage can then lead to more stylized and ultimately abstract representation, as has occurred in the evolution of writing systems.

As long as the iconic signs are understood to represent something else (an object, action, or relation), they are symbolic, and such a system of symbols, when used deliberately, should count as a primitive language. More elaborate systems will, for reasons of generativity, make use of arbitrary signs that represent a variety of entities, concrete or abstract, existent or nonexistent, sensory or conceptual.

Implicit in what I have said so far is an assumed feature of language that seems obvious once mentioned: Genuine language is shared among a community of speakers, including at least two (such as in the "secret" language of close childhood friends), but more likely involving a whole group (Wittgenstein 1953). Sometimes it is argued that the function of language in a social group is “for communication.” However, this position pushes us only one step further back: What is communication for? I assume that communication–or better, language in communication–has as its major adaptive advantage *improved social coordination.* Such coordination has adaptive advantages in response to environmental challenges, in warfare, organized hunting (Whiten 1999), predatory defense, and social foraging (Bickerton 2000). Of course, it also enables teaching and learning (Tomasello 1996) and the strengthening of social bonds (Dunbar 1993). Whether these adaptive advantages of social coordination were in fact the driving forces for the emergence and refinement of language is a question we cannot answer until we consider other possible adaptive advantages, such as being a generative format for thinking and planning (Vaas 2000) or for reasoning about other minds (de Villiers 2000). I will have more to say about adaptive functions later.

1.2 Theory of mind

Theory of mind refers to the ability to represent, conceptualize, and reason about mental states. In its fully mature stage, theory of mind is a domain-specific conceptual framework that treats certain perceptual input as an agent, an intentional action, a belief, and so forth. This framework can be activated very easily, as Heider and Simmel (1944) have shown with stimuli as simple as triangles that move about in space. Theory of mind arguably underlies all conscious and unconscious cognition of human behavior (Malle 2001a), thus resembling a system of Kantian categories of social perception — i.e., the fundamental concepts by which people grasp social reality (Kant 1998/1787). But the framework not only classifies perceptual stimuli; it also directs further processing of the classified input, including inference, prediction, and explanation (Malle in press).

Of course, humans are not born with a fully mature theory of mind, so the developmental precursors of this theory are revealing components of the cognitive system that is at issue here. In particular, the conceptual framework of agency, intentionality, and mind most likely grows out of perceptual discrimination in infancy, between birth and about 18 months. All we see at birth is a capacity to imitate (Meltzoff and Moore 1977, 1989). But by about 9 months we find reliable evidence of children’s perceptual sensitivity to self-propelled movements and to goal-directed action (Gergely, Nádasdy, Csibra, and Bíró 1995; Wellman and Philips 2001; Woodward 2001). By 14 months we see an ability to parse human action streams into meaningful (intention-relevant) units (Baldwin, Baird, Saylor, and Clark 2001; Sharon and Wynn 1998). And by 18 months we see an ability to infer intentions even from unsuccessful surface behavior. Out of this stage of perceptual discrimination and early concept formation grows the conceptual understanding of desire (by age 2) and belief (by age 3), culminating in an understanding of false belief (by age 4), which many consider the watershed of theory of mind development (Perner 1991; Wellman 1990). At this point children understand that different people represent the world in different ways and therefore can believe or know different things. The concepts of this emerging framework are tightly interconnected, with the intentionality concept being an organizing node (Malle, Moses, and Baldwin 2001), and children show increasing versatility in their reasoning with these concepts (Baird 2000; Malle and Knobe 1997a).

The emergence of this conceptual and processing framework can be seen as an attempt to strike a balance between two counteracting forces: increasing self-other differentiation on the one hand and increasing self-other coordination on the other hand. After almost complete dependence in the early months of life, the child's autonomy increases over the next few years, and with it a potential for separation from the caregiver. Attachment researchers speak about the dangers of physical and emotional separation and postulate an evolutionary response in the form of an affective bond (Bowlby 1969). However, there is also the threat of mental separation, which can be illustrated most powerfully with the autistic child's disconnection from the caregiver's attention and emotion (Sigman, Kasari, Kwon, and Yirmiya 1992). As a result of such disconnection, observational and instructional learning, interpersonal coordination, and social information processing are all seriously hampered (Baron-Cohen 1995). With little sensitivity to others' mental life, one's own mental life may remain mysterious; and with little self-insight, mental interpretations of others' behaviors may be impossible (see Goldman 1989; Gordon 1986).

Among the functions of a theory of mind, the achievement of social coordination (of both behavior and mind) seems critical. In the course of development, the coordination of minds may become more important precisely when the coordination of behaviors becomes difficult, that is, when behavioral responses between self and other begin to diverge. Such divergence may come about, for example, because of differential motives and affect, multiple opportunities to act, and multiple interaction partners.

After these sketches of both language and theory of mind, I begin with my explorations of possible relations between the two.

2. Does one faculty precede the other?

2.1 Language precedes theory of mind

In adult social behavior, language is an important vehicle by which theory of mind skills are expressed and put to use. For example, behavior explanations, fundamental to making sense of other minds, are to a significant extent embedded in conversation (Kidd and Amabile 1981; Hilton 1990; Malle and Knobe 1997b) and exhibit complex linguistic regularities (Malle 1999; Malle, Knobe, O'Laughlin, Pearce, and Nelson 2000). Even though some causal judgments may be pre-verbal (Cheng and Novick 1990; van Overwalle 1998), those that explicitly take into account other minds seem to rely on a sophisticated conceptual framework that is unlikely to be pre-verbal (Malle 1999). Similarly, explicit mental state inferences (such as attempts to empathize with or take the perspective of another person) rely on verbal reasoning, representation, and eventually expression (e.g., Ickes 1993). Thus, to the extent that we understand theory of mind as a high-level skill of making sense of (others') mental states, language is almost invariably enmeshed in it (Smith 1996).

As argued earlier, however, restricting theory of mind capacities solely to high-level (and explicit/conscious) processes would be misleading. The developmental literature has identified a number of theory of mind skills that predate early language competence, such as classification of self-propelled movement (Premack 1990), eye-gaze tracking (Farroni, Johnson, Brockbank, and Simion 2000), imitation (Meltzoff and Moore 1977), recognizing goal-directed action (Gergely *et al.* 1995; Wellman and Philips 2001; Woodward 2001), and parsing the behavior stream into intention-relevant units (Baldwin, this volume; Baldwin *et al.* 2001). These skills do not incontestably demonstrate reasoning about *mind* (Povinelli 2001), but they are at least precursors or early forms of mental-state inference without which later, more explicit forms may well be impossible.

Perhaps the most striking evidence against the notion that language precedes theory of mind (at least in development) is that language acquisition itself appears to rely on theory of mind skills. Research demonstrates that the capacity to engage in joint attention (i.e., registering that self and other are both attending to the same object) is critical in early word learning and referential communication (Baldwin 1991, 1993; Tomasello 1998). Autistic children, by contrast, who have great difficulties with joint attention, show difficulties in language learning (Carpenter and Tomasello 2000), especially compared to children with other mental handicaps, who have general cognitive deficits but no deficiency in joint attention (Baron-Cohen, Baldwin, and Crowson 1997).

Another recent suggestion about theory of mind capacities involved in language comes from Abry and Labossiere (2000), who claim that vocalizations in communication require a monitoring system to distinguish own thought, own vocalization, other's vocalization, and other's thought. However, even though communication may indeed presuppose a self-other distinction, it does not require a thought-action distinction; for otherwise every organism's goal-directed action would presuppose this distinction and therefore every organism capable of goal-directed action would automatically have a theory of mind, which is obviously not the case. Early or primitive communication may well exist without consideration of the other's *mind*; it would merely expand the organism's attempts to influence the other's behavior by using certain (iconic) expressions.

2.2 Theory of mind precedes language

Some of the arguments against language preceding theory of mind (e.g., joint attention as a prerequisite of language acquisition) suggest that theory of mind may precede language (e.g., Baron-Cohen 1999). And indeed, several authors maintain that having the ability to infer others' communicative intentions is a precondition for language (e.g., Origgi and Sperber 2000; Sperber 2000). Suppose a social perceiver infers that an agent's expressive gesture *G* indicates a certain mental state *M*. Once the agent infers that the social perceiver made such a $G \rightarrow M$ inference, the agent can use *G* to intentionally communicate to the perceiver that *M* obtains. This simple scheme could be expanded from natural, iconic symbols to increasingly arbitrary symbols; and it can include a variety of referent mental states, such as bodily states, feelings, desires, plans, perceptions and perhaps even beliefs. Note that such a communicative system would have to be used primarily for coordination, because deception (the agent communicates *G* without *M* obtaining) would, over time, weaken the perceiver's willingness to make the $G \rightarrow M$ inference and thus weaken the semantic bond that people perceive between *G* and *M*.

Some scholars have gone one step further and argued that, in principle, once an organism uses a *symbol*, a convention is involved, requiring that the symbol user assume that others interpret the symbol the same way as the symbol user does (e.g., Hobson 2000). If one defines *symbol* in this way, and if language uses such symbols, then it follows that theory of mind precedes language. However, could an organism not use a symbol without an appreciation of minds and meanings? Suppose Franzy, a chimpanzee, displays a red card whenever he wants to play ball. Does he have to understand that the card signifies "I want to play ball" and understand that the caretaker knows that the card signifies 'Franzy wants to play ball'? Or could Franzy just have acquired a habit of prediction such that, whenever he has the urge to play ball, he shows the red card to his caretaker, because he has learned that under these circumstances his wish is likely to be fulfilled? In that case, would we say that Franzy communicates with his caretaker? If communication is defined in its functional role as facilitating social coordination, we should seriously consider calling Franzy's behavior communication. If, however, we define communication by means of symbol use, symbol use by means of appreciation of semiotic conventions, and the latter by means of representations of others' mental states, we cannot escape the conclusion that theory of mind precedes language.

What would weaken the position that theory of mind precedes language? For one, there is evidence that metarepresentation itself may be dependent on language. For example, Harris (1996) argues that the child's well-known transition from appreciating desires to appreciating both desires and beliefs comes with engagement in conversation. However, Bartsch and Wellman (1995) have evidence showing that children's conversations first feature desire verbs and then, after about a year, begin to feature belief verbs alongside desire verbs. This pattern renders conversation per se a less likely candidate for proffering a belief concept–for why would this concept not emerge early in the conversational practice but rather a full year later?

De Villiers (2000) offers another line of evidence for the notion that advancements in language engender advancements in theory of mind. She reports data showing that 3- to 4-year-old children's mastery, in comprehension and production, of oblique complement structures (e.g., "He said it was raining"; "She thought the bowl was empty")[2] predicts children's success at false-belief tasks, a conservative index of metarepresentational capacity (Perner 1991). One problem with these data is that language mastery was also presupposed in the false-belief tasks, which were all verbal. The predictive relation may therefore rely on the common linguistic demands in the two types of measures (oblique complements and false beliefs) rather than on a causal influence of linguistic advances on theory of mind capacities. It is conceivable that children acquire some false-belief understanding before they can couch it in linguistic structures of the sort needed in standard, verbal false-belief tasks. De Villiers' (2000) data would therefore have to be replicated with non-linguistic false-belief assessments to shore up the argument for linguistic influence on theory of mind development.

But even if we interpret De Villiers's (2000) results as suggesting a leading influence of language competence on theory of mind competence, we are considering here fairly mature levels of language as well as theory of mind. The findings on earlier stages of development remain, which showed that theory of mind (ToM) skills are more likely to precede than to follow language skills. Thus, at best we can conclude that the direction of influence shifts from ToM → Language in early stages of development to Language → ToM in later stages. Pushing this claim further we would expect to see an increasing influence of linguistic mastery and verbal intelligence on the refinement of theory of mind skills in later childhood and early adolescence. Unfortunately, the literature on developing theory of mind past the age of 7 is very limited, so data of the sort needed have yet to be collected. However, what supports the suggestion that language becomes more autonomous from theory of mind in later stages of development is the fact that at least some autistic children are able to catch up on their language skills in later years without overcoming their deficit in theory of mind (Baron-Cohen 1995). This linguistic competence in turn allows them to compensate for their inability to infer other people's mental states by using complex linguistic representations (multi-layered if-then rules) when interacting with others (Sacks 1995).

3. Did language and theory of mind co-evolve?

In light of the current developmental evidence, neither the position that language precedes theory of mind nor the position that theory of mind precedes language is compelling, so the claim that the two faculties co-evolved becomes increasingly plausible. However, claims of co-development and co-evolution are easily made; the challenge is to flesh out the details of such a claim and say exactly how and why the two faculties co-emerged. Moreover, treating language and theory of mind as monolithic entities will not do. A plausible model of co-evolution will have to include the specification of elementary components of both language and theory of mind, because it is likely that those components influenced one another in a gradual escalation.

But before I discuss such escalation, I would like to address the reader's possible discomfort with my quick transition from developmental to evolutionary conclusions about the possible ordering of language and theory of mind.

There are two questions to consider, of which the first one is this: Even if language does not precede theory of mind in development, could it nevertheless precede it in evolution? I would claim that this is highly unlikely, because children in the evolutionary past had to learn their language too, and we know of no other plausible mechanisms for acquiring language than through social-interactive learning, imitation, and practice, relying heavily on early theory of mind skills. What still seems feasible is that some sort of protolanguage could have been acquired without a theory of mind (see Franzy's symbolic use under section 1.2), whereas full-fledged language became possible precisely because of the evolution of a theory of mind. I will return to this chain model of *protolanguage → theory of mind → language* in the next section.

The second question is this: Even if theory of mind did not precede language in development, could it nevertheless precede it in evolution? This seems quite plausible. First, as we have seen, any documented influence of language on theory of mind occurs later in development and is therefore compatible with an evolutionary primacy of theory of mind. Second, to the extent that only advanced language skills can exert a causal influence on theory of mind progression, we still have to explain the evolutionary emergence of those advanced language skills, which seem difficult to acquire without some form of theory of mind.

These deliberations underscore the need for a fine-grained analysis of the multi-stage evolution of language and the possible multi-stage co-evolution of theory of mind. I organize my thoughts on such an analysis into three sections: Co-evolution with coincidental parallelism; co-evolution due to the emergence of a third factor that enabled both language and theory of mind; and co-evolution in the form of mutual escalation toward a shared, or at least partially shared, adaptive advantage.

3.1 Coincidental concurrent evolution

The weakest model of co-evolution would assume no shared adaptive pressures and no mutual influence during evolution. The two faculties would then have evolved for distinct reasons, though roughly during the same time period. According to this model, humans just happen to have both a linguistic capacity and a refined theory of mind, and their respective architectures would have been established independently, leading to no more than coincidental functional parallels between the two.

Several facts speak against this position. First, we find intimate ties between language and theory of mind both in development (e.g., language learning relies on theory of mind) and in adult social behavior (mental-state inferences are required and practiced in conversation). Second, there is likely a common adaptive advantage to both faculties in the form of vastly improved social coordination (which does not rule out additional, unshared adaptive advantages, such as art, planning, counterfactuals, etc.). Third, the striking parallels between fundamental linguistic structures (e.g., thematic roles, verb classes, and perspective) and their corresponding elements in theory of mind (the concepts of agency, intentionality, self vs. other; see Malle 2001b) would remain unexplained if we did not postulate an interdependence between the two faculties during evolution. Such an interdependence can be treated like a correlation between variables–it may be due to a causal influence of one on the other or due to a third factor that exerts a causal influence on both. Below I explore the possibility of a third factor and then turn to the scenario of mutual causal influence.

3.2 In search of a third factor

Two types of factors might have fostered the co-evolution of language and theory of mind: a powerful adaptive demand or another adaptation that made both faculties possible. As candidates for the latter we may think of adaptations such as consciousness, executive control, increased memory capacity, representational ability, or propositional thought. The problem with all of these factors is that they are at least as likely to be consequences of language and theory of mind as they are antecedents of them (see Bogdan 2000, and Carruthers 1998). More important, neither theory of mind nor language necessarily requires the operation of these factors. For example, theory of mind is not particularly memory-intensive and many of its functions do not require consciousness or executive control (Malle 2001a). Similarly, phonology and syntax do not require consciousness or executive control (although syntax may require fairly sophisticated memory). In addition, all of the mentioned third factors leave most of the specific features of language unaccounted for (e.g., Hurford 2000). The same may hold true for features of theory of mind. The concept of intentionality, for example, can in no way be derived from any cognitive mechanisms in consciousness, memory, or the like. Concepts such as intentionality must be adaptations to a reality (or a conceived reality) that is better managed and mastered with than without these concepts. Of course, advances in cognitive processes may improve the resolution, efficiency, and manipulability of a conceptual framework, but the concepts themselves must be there in the first place, as an adaptation to the social world around.

This leads us to the second potential driving force for the co-evolution of language and theory of mind: a consistent external demand that had an impact on both faculties. The most likely candidate for such a demand is the need for improved social coordination, with at least two things speaking for it. First, a number of different environmental challenges that early hominids encountered (e.g., climate shifts, wildlife disruptions, novel germs, increased group size, migration, hostile encounters with other hominid groups) likely resulted in a need for greater social coordination. When food is tight, life is novel or dangerous, and groups become large, social coordination and organization are called for. Second, among the contemporary functions of language and theory of mind, improved social coordination is the most salient function they share. (I am not aware of any other shared functions; but I would be happy to be corrected.)

By itself, a powerful demand such as social coordination would not explain the emergence of two adaptations. However, a plausible explanation can be constructed if we presume that one of the faculties emerged first in a primitive form and facilitated a primitive form of the other, and soon the two plunged into a race in which advances in one faculty repeatedly enabled or demanded advances in the other. What we have here is a version of the arms-race argument, but I am thinking more of an arms race between psychological faculties than between different genetic pools, and so I prefer the term *escalation*.

4. Evolutionary escalation

The balance of evidence reviewed earlier favors the emergence of a primitive theory of mind (TOM-1) as a first adaptive step, because language learning is rather inconceivable without it. What might such a primitive theory look like? Povinelli (2001) reviewed the literature to date on primate theory of mind and concluded that chimpanzees have a refined behavior reading and detection system but no clear understanding of the mind as the underlying source of observed behavioral patterns. Unenculturated chimpanzees also show no reliable skills of joint attention (Tomasello 1996), and the evidence on imitation is mixed (Smith 1996; Whiten 1999). Thus I postulate that TOM-1 includes:

a. a capacity for imitation (a mechanism that matches others' behavior to self's internal states and provides a necessary condition for inferences about those internal states in others; Meltzoff and Brooks 2001);
b. a capacity for joint attention (representing oneself and another person attending to an object); and
c. an inferential sensitivity to some mental states associated with action (approximating a desire or goal concept).

Most likely, the three elements emerged in the listed order and built on each other until reaching a stage that is roughly comparable to a modern 9–12 month-old infant, which is also the stage when language learning begins. In human evolution, however, the emergence of these three components may well have been separated by a few million years (MacWhinney, this volume).

Imitation and joint attention may be easy postulates to accept, but how could a sensitivity to desires have emerged? The early hominid brain may have fed data from behavioral pattern analysis into a reasoning device that could make inferences about unobservable forces. Some have argued that these inferences are enabled by a sort of analogical reasoning from one's own case (Goldman 2001; Meltzoff and Brooks 2001). This system would match up patterns in the other's behavior with patterns in one's own behavior (Blakemore and Decety 2001), then patterns in one's own behavior with (a vague awareness of) one's own mental state, finally leading to a projective inference of one's own mental state to the other person (Gallese and Goldman 1998). Such an analogical process has few rivals in explaining what pushed the primate perceiver from sophisticated behavior analysis to the crude beginnings of a mental analysis. If this notion of an analogical process hits the mark, however, then a mature TOM-1 would presuppose advances in introspective access (as vague as it may be), and such introspective access may qualify as a "driving third variable." How such introspective access might have evolved is unclear. Some form of brain reorganization seems inevitable (Humphrey 2000), perhaps with increases in self-other differentiation playing a crucial triggering role. For example, encounters with variants of one's genus or species would highlight how the self is different from the other, and living in larger-sized groups might expand the gene pool to allow for more individual differences, leading to more self-other distinctions. In both cases, global awareness of self as different from others could set the stage for more specific awareness of self's distinct preferences, emotions, and so on.

If we equip TOM-1 with the three elements of imitation, joint attention, and inferential sensitivity to desires, these elements would likely reinforce one another: Joint attention facilitates sensitivity to others' object-directed desires, and desire sensitivity improves imitation of behaviors as well as of goal states (Meltzoff 1995). Operating together, the three elements also open new possibilities. Imitation and joint attention may initiate cooperative learning, especially if the learner also recognizes the teacher's goal in the process. Joint attention with desire sensitivity should improve behavior prediction and thus the coordination of shared activities (e.g., hunting, combat, tool making). And, most important, the three elements would enable expressive gestures or vocalizations to become predictably associated with certain objects (especially objects of goal-directed actions), laying the groundwork for a simple protolanguage (Bickerton 1990). Specifically, imitation would enable cross-member reliability in vocalization (De Boer 2000; Studdert-Kennedy 2000a); joint attention would enable cross-member reliability for acts of vocal reference (Baldwin 1993); and desire sensitivity would launch the interpretation of vocalized reference as a form of communicative intention.

At the next stage of the hypothesized escalation process, the achieved communication system becomes a spring board for further advances in mental-state inference. We can assume that hominid agents' mental attitudes were more complex than could be represented by contemporary social perceivers' mental inferences. The expanding vocabulary of protolanguage may then have taken on an intermediary role by differentiating between various relations that agents might have had toward objects (e.g., like vs. dislike; weak vs. strong desire; desire vs. intention) and by differentiating between types of behaviors–primarily intentional vs. unintentional, but also among specific speech acts such as command, request, offer, decline, or warning. If used reliably, and practiced over numerous generations, these communicative tools must have refined early hominids' conceptual framework of mind and behavior.

But there was likely another process at work that helped refine theory of mind. The increasing vocabulary and differentiation of this early communication system introduced a great deal of ambiguity and inferential challenge. For one thing, the linguistic forms included in protolanguage were probably not names for objects (Wray 2000). Instead, they were speech acts (commands, threats, warnings etc.) that holistically expressed person-object relations and perhaps included interjections to clarify such person-object relations (e.g., a prehistoric "oops" to indicate unintentionality or an ancient "there" to indicate intentionality; see Carpenter, Akhtar, and Tomasello 1998). Add to that the fact that even in modern times communicators overestimate the clarity of their expressions and the transparency of their communicative intentions (Gilovich, Savitsky, and Medvec 1998; Keysar 1994). We can conclude, then, that protolanguage posed serious difficulties for the ancient social perceiver and may have caused a fair amount of misunderstanding and conflict within and between groups.

In response to both the opportunities and challenges posed by protolanguage, theory of mind must have advanced to TOM-2, through an extension of introspective access and/or an improvement of cognitive capacities, such as inference, conceptual distinctions, contextual awareness, and memory. The change in memory capacity may be a particularly important link, because protolanguage should have increased interactions among group members (in learning, teaching, and joint action), which in turn demanded better memory for one's relations with each group member, especially if the groups also increased in size. One may think here of cheater detection and tracking the balance of one's reciprocal exchanges (Cosmides and Tooby 1992). But the tracking must have gone further, to improved representations of people's habitual desires, skills, and behavior trends. Reliable memory representations of interactions with others would also allow the social agent to better recognize individual differences among people in their desires, skills, and attitudes toward the agent. These emerging differentiations may then be the starting point for the realization that people represent the world in different ways and act on those representations (Perner 1991).

Once mental-state inferences in TOM-2 were sharpened and memory for people and interaction contexts was recruited to disambiguate communication, language itself progressed toward greater clarity. This may seem paradoxical, because it was precisely the ambiguity of protolanguage that demanded the more refined TOM-2, and one might expect that this refinement would have allowed protolanguage to remain ambiguous. But I think otherwise. Consider a conversation between two people in which one tries to express a complex internal state and is rather vague and incoherent. A listener who genuinely wants to understand the speaker will pose questions to clarify what the speaker means, will propose and "test" various formulations, until the speaker feels that a certain formulation adequately depicts his or her internal state. Thus, the two negotiate and agree upon a set of expressions that the speaker will use henceforth to describe the relevant internal state (see Krauss and Fussell 1991). Thus, modern speakers become clearer and more precise in their expressions when listeners use their mental-state inference skills to help disambiguate expressions, and I would assume that something similar occurred in the evolutionary past.

These processes must have given rise to a more advanced protolanguage (PL-2), which was less holistic and more precise and incorporated a greater number of linguistic references to internal states. Some basic grammatical distinctions may have been added, such as between names for objects (nouns) and names for activities (verbs). Increasing intragroup interactions may have also demanded names for people, and combinations between these three classes of words would have permitted depiction of important social events, such as a certain person manipulating a certain object.

If PL-2 was spoken by adults around young children, its representational uses should have helped children differentiate among their own emerging behavior patterns, their preference structures, and even their social emotions, which would in turn sharpen their introspection from an earlier age on. Thus we can derive the modern-day prediction that the degree to which children are exposed to representational language should accelerate their introspective abilities. At the opposite end would be deaf children with parents whose sign language is primitive and lacks representational elaboration, resulting in slowed maturation of the children's introspective abilities.

Sharpened introspection within any new generation may then lead to increasingly subtle and differentiated first-person statements about preferences or emotional responses, combining the functions of mere expression in protolanguage-1 with those of representation in protolanguage-2. Successive generations would thus refine the shared linguistic tools, not unlike successive generations of signers have been shown to refine their language (Senghas 2000). Once more, these advances have implications for theory of mind, because if agents make their internal states explicit to others (at least under conditions of intimacy and trust), social perceivers receive more validating information for their mental state inferences, which would become increasingly differentiated and spiral up toward a level of TOM-3.

How and why fully syntactic language emerged out of PL-2 and how TOM-3 differs from a full-fledged modern theory of mind would have to be the topic of another paper. The same holds for the intriguing idea that contemporary scientific psychology may be the latest step in this escalation between language and theory of mind, because it tries to offer explicit verbal representations and extensions of folk psychology (theory of mind). That then raises the question whether, on the next level of the spiral, theory of mind will change under the weight of scientific findings, as some have confidently predicted (Churchland 1988). Whatever future evolution may bring, the evidence at hand suggests that language and theory of mind evolved in grades (MacWhinney, this volume) and in constant interaction, serving one primary adaptive goal: to improve social coordination.

5. Coda

Dan Slobin, skeptical about the scientific study of language evolution (Slobin, this volume) characterized my preceding proposal as an excellent science-fiction story. Being rather skeptical myself of much of evolutionary psychology, I tend to agree with this characterization. But an open-minded view of science (finally shedding the shackles of neopositivism) allows for speculative, even fictional models to serve a heuristic or hypothesis-generating function. A cognitive-evolutionary model of what happened, say, 100,000 years ago will never be empirically testable in the strict sense; but neither will an astrophysical model of what happened 100 Million years ago. In both cases we build models that best fit the facts we have available, and we often begin to search for new facts that would have to be true if the model were correct. It may take a while to find these facts (or their refutation), and several competing models may be "best-fitting." But no philosophy of science should prevent us from discussing, comparing, and challenging these models. In this spirit, the skeptical reader as well as the aficionado of science fiction are invited to propose a better model than the one here offered, point to existing facts overlooked, and search for new facts not yet considered. It is this method of scrutinizing claims and models in a rational, systematic manner that counts as science.

Notes

* Preparation of this chapter was supported by NSF CAREER award SBR-9703315. I am grateful to Katie Burns, Tom Givón, Lara London, Sarah Nelson, John Orbell, and Brian MacWhinney for inspirations and comments on an earlier draft of this chapter.

1. If an organism has control over expressive features and begins to conceal or feign them for particular purposes, I would still not call it a language but rather a sophisticated manipulation of the other organism's perception or (more likely) behavior. Whether or not one believes the arguments that some animals are capable of full-blown deception (manipulating the other's mind, not just behavior) is irrelevant here, because language requires at least two people who both understand that a symbolic system is being used.

2. What is important about these structures is that the complement can be false. That is, the speaker understands that, even though the agent may have said or believed or wanted that it was raining outside, it may in reality not have been raining outside. This linguistic structure thus enables the representation of another person's subjective representation of reality, or "metarepresentation."

References

Abry, and Laboissiere, R. 2000. Who's afraid of the co-evolution of medial and lateral cortices for speech? Paper presented at the 3rd Conference on the evolution of language. Ecole Nationale Supérieure des Télécommunications, Paris, France. http://www.infres.enst.fr/confs/evolang/actes/_actes01.html.

Baird, J.A. 2000. *Young Children's Understanding of the Relation Between Actions and Intentions*. Dissertation, University of Oregon.

Baldwin, D.A. 1991. "Infants' contribution to the achievement of joint reference". *Child Development* 62:875–90.

Baldwin, D.A. 1993. "Early referential understanding: Infants' ability to recognize referential acts for what they are". *Developmental Psychology* 29:832–43.

Baldwin, D. A, and Tomasello, M. 1998. "Word learning: A window on early pragmatic understanding". In *Proceedings of the 29th Annual Child Language Research Forum*, E.V. Clark (ed), 3–23. Stanford, Calif.: Center for the Study of Language and Information.

Baldwin, D.A., Baird, J.A., Saylor, M.M., and Clark, M.A. 2001. "Infants parse dynamic action". *Child Development* 72:708–17.

Baron-Cohen, S. 1999. "The evolution of a theory of mind". In *The Descent of Mind: Psychological Perspectives on Hominid Evolution*, M.C. Corballis and S.E.G. Lea (eds), 261–77. New York: Oxford University Press.

Baron-Cohen, S. 1995. *Mindblindness: An Essay on Autism and Theory of Mind.* Cambridge, Mass.: MIT Press.

Baron-Cohen, S., Baldwin, D. A, and Crowson, M. 1997. "Do children with autism use the speaker's direction of gaze strategy to crack the code of language?" *Child Development* 68:48–57.

Bartsch, K., and Wellman, H.M. 1995. *Children Talk about the Mind.* New York: Oxford University Press.

Bickerton, D. 1990. *Language and Species.* Chicago: University of Chicago Press

Bickerton, D. 2000. Foraging versus social intelligence in the evolution of protolanguage. Paper presented at the 3rd Conference on the evolution of language. Ecole Nationale Supérieure des Télécommunications. Paris, France. http://www.infres.enst.fr/confs/evolang/actes/_actes05.html.

Blakemore, S. and Decety, J. 2001. "From the perception of action to the understanding of intention". *Nature Reviews Neuroscience* 2:561–7.

Bogdan, R.J. 2000. *Minding Minds: Evolving a Reflexive Mind by Interpreting Others.* Cambridge, Mass.: MIT Press.

Bowlby, J. 1969. *Attachment and Loss.* New York: Basic Books.

Carpenter, M., Akhtar, N., and Tomasello, M. 1998. "Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions". *Infant Behavior and Development* 21:315–30.

Carpenter, M., and Tomasello, M. 2000. "Joint attention, cultural learning, and language acquisition: Implications for children with autism". In *Autism Spectrum Disorders: A Transactional Developmental Perspective*, A.M. Wetherby and B.M. Prizant (eds), 31–54. Baltimore: Paul H. Brookes.

Carruthers, P. 1998. "Thinking in language? Evolution and a modularist possibility". In *Language and Thought: Interdisciplinary Themes*, P. Carruthers and J. Boucher (eds), 94–119. Cambridge: Cambridge University Press.

Cheng, P.W., and Novick, L.R. 1990. "A probabilistic contrast model of causal induction". *Journal of Personality and Social Psychology* 58:545–67.

Churchland, P.M. 1988. *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind* (Rev. edition). Cambridge, Mass.: MIT Press.

Cosmides, L., and Tooby, J. 1992. "Cognitive adaptations for social exchange". In *The adapted mind: Evolutionary psychology and the generation of culture*, J.H. Barkow, L. Cosmides, and J. Tooby (eds), 163–228. New York: Oxford University Press.

Deacon, T. 1997. *The Symbolic Species: The Co-evolution of Language and the Brain.* New York: W.W. Norton.

de Boer, B. 2000. "Emergence of sound systems through self-organization". In *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*,

C. Knight, M. Studdert-Kennedy, and J.R. Hurford (eds), 177–98. Cambridge: Cambridge University Press.

De Villiers, J. 2000. “Language and theory of mind: What are the developmental relationships?” In *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*, S. Baron-Cohen, H. Tager-Flusberg, and D.J. Cohen (eds), 83–123. 2nd edition. New York: Oxford University Press.

Donald, M. 1998. “Mimesis and the excessive suite: Missing links in language evolution”. In *Approaches to the Evolution of Language: Social and Cognitive Bases*, J.R. Hurford, M. Studdert-Kennedy, and C. Knight (eds), 44–67. Cambridge: Cambridge University Press.

Farroni, T., Johnson, M. H, Brockbank, M., and Simion, F. 2000. “Infants’ use of gaze direction to cue attention: The importance of perceived motion”. *Visual Cognition* 7: 705–18.

Gallese, V., and Goldman, A. 1998. “Mirror neurons and the simulation theory of mind-reading”. *Trends in Cognitive Sciences* 2: 493–501.

Gergely, G., Nádasdy, Z., Csibra, G., and Bíró, S. 1995. “Taking the intentional stance at 12 months of age”. *Cognition* 56: 165–93.

Gilovich, T., Savitsky, K., and Medvec, V.H. 1998. “The illusion of transparency: Biased assessments of others’ ability to read one’s emotional states”. *Journal of Personality and Social Psychology* 75: 332–46.

Goldman, A.I. 1989. “Interpretation psychologized”. *Mind and Language* 4: 161–85.

Goldman, A.I. 2001. “Desire, intention, and the simulation theory”. In *Intentions and Intentionality: Foundations of Social Cognition*, B.F. Malle, L.J. Moses, and D.A. Baldwin (eds), 207–25. Cambridge, Mass.: MIT Press.

Gordon, R.M. 1986. “Folk psychology as simulation”. *Mind and Language* 1: 158–71.

Grice, H.P. 1957. “Meaning”. *Philosophical Review* 64: 377–88.

Harris, P. 1996. “Desires, beliefs, and language”. In *Theories of Theories of Mind*, P. Carruthers and P.K. Smith (eds), 200–20. Cambridge: Cambridge University Press.

Heider, F., and Simmel, M. 1944. “An experimental study of apparent behavior”. *American Journal of Psychology* 57: 243–59.

Hilton, D.J. 1990. “Conversational processes and causal explanation”. *Psychological Bulletin* 107: 65–81

Hobson, R.P. 2000. “The grounding of symbols: A social-developmental account”. In *Children’s Reasoning and the Mind*, P. Mitchell and K.J. Riggs (eds), 11–35. Hove: Psychology Press.

Humphrey, N. 2000. “The privatization of sensation”. In *The Evolution of Cognition*, C. Heyes and L. Huber (eds), 241–52. Cambridge, Mass.: MIT Press.

Hurford, J.R. 2000. The roles of communication and representation in language evolution. Paper presented at the 3rd Conference on the evolution of language. Ecole Nationale Supérieure des Télécommunications. Paris, France. http://www.infres.enst.fr/confs/evolang/actes/_actes37.html.

Ickes, W. 1993. “Empathic accuracy”. *Journal of Personality* 61: 587–610.

Kant, I. 1998/1787. *Critique of Pure Reason.* (Translated and edited by Paul Guyer, Allen W. Wood). Cambridge: Cambridge University Press.

Keysar, B. 1994. “The illusory transparency of intention: Linguistic perspective taking in text”. *Cognitive Psychology* 26: 165–208.

Kidd, R.F., Amabile, T.M. 1981. “Causal explanations in social interaction: Some dialogues on dialogue”. In *New directions in attribution research*, Volume III, J.H. Harvey, W.J. Ickes, and R.F. Kidd (eds), 307–28. Hillsdale, NJ: Lawrence Erlbaum Associates.

Krauss, R.M., and Fussell, S.R. 1991. “Constructing shared communicative environments”. In *Perspectives on Socially Shared Cognition*, L.B. Resnick, J.M. Levine, and S.D. Teasley (eds), 172–200. Washington: American Psychological Association.

Malle, B.F. 1999. “How people explain behavior: A new theoretical framework”. *Personality and Social Psychology Review* 3:23–48.

Malle, B.F. In press. “The folk theory of mind: Conceptual foundations for social cognition”. In *The New Unconscious*, R. Hassin, J.S. Uleman, and J.A. Bargh (eds). New York: Oxford University Press.

Malle, B.F. 2002. “Verbs of interpersonal causality and the folk theory of mind and behavior”. In *The Grammar of Causation and Interpersonal Manipulation*, M. Shibatani (ed), 57–83. Amsterdam: Benjamins.

Malle, B.F., and Knobe, J. 1997a. “The folk concept of intentionality”. *Journal of Experimental Social Psychology* 33:101–21.

Malle, B.F., and Knobe, J. 1997b. “Which behaviors do people explain? A basic actor-observer asymmetry”. *Journal of Personality and Social Psychology* 72:288–304.

Malle, B.F., Knobe, J., O’Laughlin, M., Pearce, G.E., and Nelson, S.E. 2000. “Conceptual structure and social functions of behavior explanations: Beyond person–situation attributions”. *Journal of Personality and Social Psychology* 79: 309–26.

Meltzoff, A.N. 1995. “Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children”. *Developmental Psychology* 31:838–50.

Meltzoff, A.N., and Brooks, R. 2001. “‘Like me’ as a building block for understanding other minds: Bodily acts, attention, and intention”. In *Intentions and Intentionality: Foundations of Social Cognition*, B.F. Malle, L.J. Moses, and D.A. Baldwin (eds), 171–91. Cambridge, Mass.: MIT Press.

Meltzoff, A.N., and Moore, M.K. 1977. “Imitation of facial and manual gestures by human neonates”. *Science* 198:75–8.

Meltzoff, A.N., and Moore, M.K. 1989. “Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms”. *Developmental Psychology* 25:954–62.

Mithen, S. 1996. *The Prehistory of the Mind.* London: Thames and Hudson.

Origgi, G. and Sperber, D. 2000. “Evolution, communication and the proper function of language”. In *Evolution and the Human Mind: Modularity, Language and Metacognition*, P. Carruthers and A. Chamberlain (eds), 140–69. Cambridge: Cambridge University Press.

Perner, J. 1991. *Understanding the Representational Mind.* Cambridge, Mass.: MIT Press.

Povinelli, D.M. 2001. “On the possibilities of detecting intentions prior to understanding them”. In *Intentions and intentionality: Foundations of social cognition*, B.F. Malle, L.J. Moses, and D.A. Baldwin (eds), 225–48. Cambridge, Mass.: MIT Press.

Premack, D. 1990. “The infant’s theory of self-propelled objects”. *Cognition* 36:1–16.

Sacks, O. 1995. "An anthropologist on Mars". In *An anthropologist on Mars*, O. Sacks, 244–96. New York: Vintage Books.

Senghas, A. 2000 Evolution of Grammar: The Grammaticization of Nicaraguan Sign Language by sequential iterations of native acquisition. Paper presented at the 3rd Conference on the evolution of language. Ecole Nationale Supérieure des Télécommunications. Paris, France. http://www.infres.enst.fr/confs/evolang/actes/_actes64.html.

Sharon, T., and Wynn, K. 1998. "Individuation of actions from continuous motion". *Psychological Science* 9: 357–62.

Sigman, M. D, Kasari, C., Kwon, J., and Yirmiya, N. 1992. "Responses to the negative emotions of others by autistic, mentally retarded, and normal children". *Child Development* 63: 796–807.

Smith, P.K. 1996. "Language and the evolution of mind-reading". In *Theories of Theories of Mind*, P. Carruthers and P.K. Smith (eds), 344–54. Cambridge: Cambridge University Press.

Sperber, D. 2000. "Metarepresentations in an evolutionary perspective". In *Metarepresentations: A Multidisciplinary Perspective*, D. Sperber (ed), 117–37. New York: Oxford University Press.

Studdert-Kennedy, M. 2000a. "Evolutionary implications of the particulate principle: Imitation and the dissociation of phonetic form from semantic function". In *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, C. Knight, M. Studdert-Kennedy, and J.R. Hurford (eds), 161–76. Cambridge: Cambridge University Press.

Studdert-Kennedy, M. 2000b. "Introduction: The emergence of phonetic structure". In *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, C. Knight, M. Studdert-Kennedy, and J.R. Hurford (eds), 123–9. Cambridge: Cambridge University Press.

Tomasello, M. 1996. "The cultural roots of language". In *Communicating Meaning: The Evolution and Development of Language*, B.M. Velichkovsky and D.M. Rumbaugh (eds), 275–307. Mahwah, NJ: Lawrence Erlbaum Associates.

Tomasello, M. 1998. "Reference: Intending that Others Jointly Attend". *Pragmatics and Cognition* 6: 229–43.

Vaas, R. 2000. Evolving language, I-consciousness and free will. Paper presented at the 3rd Conference on the evolution of language. Ecole Nationale Supérieure des Télécommuniycations. Paris, France. http://www.infres.enst.fr/confs/evolang/actes/_actes73.html.

van Overwalle, F. 1998. "Causal explanation as constraint satisfaction: A critique and a feed-forward connectionist alternative". *Journal of Personality and Social Psychology* 74: 312–28.

Wellman, H. 1990. *The Child's Theory of Mind.* Cambridge, Mass.: MIT Press.

Wellman, H.W., and Phillips, A.T. 2001. "Developing intentional understandings". In *Intentions and Intentionality: Foundations of Social Cognition*, B.F. Malle, L.J. Moses, and D.A. Baldwin (eds), 125–48. Cambridge, Mass.: MIT Press.

Whiten, A. 1999. "The evolution of deep social mind in humans". In *The descent of mind: Psychological Perspectives on Hominid Evolution*, M.C. Corballis and S.E.G. Lea (eds), 173–93. New York: Oxford University Press.

Wittgenstein, L. 1953. *Philosophical Investigations.* (Transl. G.E.M. Anscombe). Malden, Mass.: Blackwell.

Woodward, A.L., Sommervile, J.A., and Guajardo, J.J. 2001. "How infants make sense of intentional action". In *Intentions and Intentionality: Foundations of Social Cognition*, B.F. Malle, L.J. Moses, and D.A. Baldwin (eds), 149–70. Cambridge, Mass.: MIT Press.

Wray, A. 2000. "Holistic utterances in protolanguage: The link from primates to humans". In *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, C. Knight, M. Studdert-Kennedy, and J.R. Hurford (eds), 285–302. Cambridge: Cambridge University Press.

CHAPTER 11

The rise of intentional understanding in human development

Analogies to the ontogenesis of language

Dare A. Baldwin
University of Oregon

Constructing sound hypotheses about the origins of language — ontogenetic or phylogenetic — can only gain from having a good analysis of the systems with which language interacts. Intentional understanding is one such closely-allied system. Intentional understanding is what enables us, among other things, to discern or "recover" peoples' intentions as we observe the dynamic patterns of motion they exhibit in the course of everyday intentional action. Our ability to do this is currently little understood. I approach this chapter as an opportunity to speculate about mechanisms that enable humans to discern intentions within the dynamic motion stream, and about the ontogenetic development of this skill. Some of the hypotheses that arise in the course of such speculation have possible implications for our understanding of language evolution.

1. Intentional understanding pervades language

First, I should clarify at least some of the ways in which language is closely allied with intentional understanding. For one thing, our everyday language is chock-full of intentionally laden terms. Some of this is obvious. Languages include terms that explicitly encode intention concepts, such as words like intention, motivation, goal, purpose, meaning, to intend, to attempt, and to try.

Some of this is a little less obvious. When we describe our own or others' everyday actions — their graspings, kickings, ticklings, and huggings, the terms we use to do so typically capture the content of their intentions, not just the nature of their motions themselves. For example, let's say I witnessed Tom reaching for a coffee cup. As an English speaker I might, quite naturally, describe this as "Tom reached for the coffee cup." In contrast, it would be decidedly odd for me to describe this action strictly in terms of motion parameters. Consider, for example, how peculiar (and technically challenging) the following description seems:

> Tom executed a motion of the right arm in which (a) the arm originated in a position of nearly full extension beside the body when the major axis of the body was perpendicular to the earth's surface, (b) the arm moved in a ballistic fashion with initial acceleration of blah-blah, at angle blah-blah with respect to a coffee cup, with deceleration blah-blah beginning as the fingers of the left arm achieved roughly blah-blah distance from the coffee-cup, and (c) the hand and fingers of the arm simultaneously underwent position changes predicted by the coffee cup's height (blah-blah), shape (blah-blah), and position relative to immediately adjacent objects (blah-blah-blah-blah-blah).

If the term *reach* in the context of Tom executing a reach were fundamentally equivalent to this kind of direct motion description, why should such a passage strike us as so outrageously ridiculous and unnatural? Perhaps one could argue that we react this way because long-winded contributions of this sort would drastically violate the convention, and the practical need, to be efficient communicators. Obviously, attempts to describe motion patterns in their own right are impossibly unwieldy and would overwhelm listeners with potentially irrelevant verbiage. Could it be, then, that the term *reach* is simply a short-hand for all the unpleasant detail, and hence still equivalent to a (summarized) motion description? This seems unlikely. If the meanings of descriptive terms such as *reach* are equivalent to motion descriptions, then multiple descriptive terms that apply to one and the same motion pattern ought to be equivalent in meaning, and clearly they are not. In the *reach* instance, for example, other terms such as "Tom *grasped* the coffee cup," "Tom *touched* the coffee cup," "Tom *prepared to pick up* his coffee cup", "Tom *prepared to caffeinate* his system" are all appropriate descriptors of the relevant motion pattern, but they don't communicate the same thing. They each communicate a different perspective on what Tom's intentions are in a given situation at a given point in time.

The upshot of all this? What usually matters for communicating with each other about action is to convey the *meaning* of action — the goals or intentions motivating it — not the nature of the motion pattern itself. Of course, this is not to say that motion patterns play no role whatsoever in the meaning of verbs of action; I'm simply arguing that motion patterns don't exhaust the meaning (or reference or use) of such terms.

That terms like reach, pull, grasp, kick, and tickle represent concepts imbued with attribution of intention is also clear from everyday arguments people engage in, such as

A: Stop tickling me!
B: I'm not tickling you!
A: Yes you are! Cut it out!

B: I am not. I'm just moving my foot. It's not my fault if you're *feeling* tickled.

Aggravator B won't agree to the charge of tickling because he had no such intent. We might customarily think of the term *tickle* as merely a marker for a certain category of motion patterns, but this just isn't so. It's more complicated than that, and one of the complexities is the role of intention. An enormous range of terms in any language is imbued with intentional concepts in just this way.

Clearly, using language, acquiring language, and evolving language to the level that enables discussion of the ethics of tickling all require an apparatus for interpreting others' motions in intentional terms. Individuals who can't do so simply couldn't understand or produce an enormous range of very ordinary, everyday utterances in their native language.

Another way in which language makes contact with intentional understanding is in its use as a medium for communication, as of course, the whole notion of communication is grounded in mentalistic understanding, and intentional understanding is a core component of our everyday thinking about the mind [see also Malle (this volume)]. The point is simply that to be a *bona fide* language user, one must be a communicator. And to be a communicator, one must appreciate that language typically influences others not because it alters people's behavior directly, but by virtue of its impact on mental life — its impact on others' intentions, beliefs, desires, knowledge, and the like. For present purposes I will focus just on the intentions component of folk psychology as, for reasons I won't explain here, I'm guessing it has some ontogenetic primacy, and others (e.g., Call and Tomasello 1998) argue for phylogenetic primacy as well.

2. Language and action amalgamated

I've argued thus far that everyday language use both explicitly and implicitly trades on the operation of intentional understanding. A powerful case can also be made about a related, but more specific issue: engaging in ordinary linguistic activity hinges on parallel processing of intentions across modalities (e.g., auditory information such as speech, visual information such as motion). An obvious demonstration of such parallel processing is the consternation one experiences on seeing people behave in ways seemingly inconsistent with their verbally-stated intentions. We expect the intentions we infer based on motion information to cohere with our linguistically-based intentional analysis. Susan Goldin-Meadow and colleague's research provides another and more striking example of the parallel interplay of multi-modal information in the processing of intentions (e.g., Alibali, Bassok, Solomon, Syc, and Goldin-Meadow 1999; Goldin-Meadow 1998;

Iverson and Goldin-Meadow 1998). They provide a compelling case that gestures (motion information) accompanying speech make independent contributions to a speaker's message that go beyond what is available in the linguistic stream alone, and listeners capitalize on the additional motion information. The broader point such work makes is that our analysis of others' intentions can be rampantly incomplete when we lack multi-modal information about their behavior.

Finally, there is reason to believe that skill at interpreting others' motions in intentional terms plays a crucial and quite broad role in facilitating language learning. In a limited way I have demonstrated this in previous research: Sensitivity to intentions revealed in a speaker's motions (e.g., changes in gaze direction, head orientation, voice direction, and the like) helps infants to home in on what's being talked about in the linguistic stream, enabling them to construct sound hypotheses about word meaning (e.g., Baldwin 1991, 1993a, 1993b, 2000). Tomasello and colleagues' ingenious research has broadened the case considerably, showing how these skills play a role not only in word learning, but in grammar acquisition as well (e.g., Tomasello 1999). Tomasello's chapter in this volume makes a compelling case, for example, that grammar acquisition originates in children noting correspondences between whole sentences and their interpretation of what the speaker intends to communicate in the particular context. Intentional understanding, then, is the nexus around which language learning revolves. And this can only be the case if children, like adults, process intentions in parallel across modalities, and expect cross-modal coherence in the intentional analyses they derive.

3. Processing motion to discern intentions

If infants' ability to discern intentions within a dynamic motion stream facilitates language learning, as I've just argued, then processing motion to recover intentions must have some primacy, developmentally speaking. Yet at present we know only a little about the nature of the early-emerging system that "detects" intentions in the flow of motion. At least one thing is obvious, however: it seems to be another generative knowledge system. Jodie Baird and I have articulated this elsewhere (Baldwin and Baird 2001) so I will only summarize here. Some of the points I raise make contact with important issues elaborated by Fenk-Oczlon and Fenk (this volume).

Generative systems enable us to interpret novel sequences systematically, and shared knowledge of the principled regularities of the system enable multiple individuals to agree on an interpretation for one and the same novel sequence (e.g., Bloom 1994). Our system for human motion processing displays this generative quality. Novel sequences of human motion typically pose little obstacle to processing, and multiple witnesses of a given motion sequence show a high degree of convergence on a similar intentional account even when motion is novel.

Generative systems also possess the general quality of representing dynamic and continuous stimulus streams in discrete terms. Our system for processing human motion reflects this property as well: we interpret dynamic, continuous motion in terms of discrete intention concepts. These concepts seem to be organized hierarchically. For example, when we observe others pursuing tasks such as kitchen clean-up, we not only note this task-oriented intention, but also discern subsidiary sub-goals, such as placing a dish in a dishwasher or putting ice cream away in the freezer, which themselves are each composed of intentions to carry out specific actions, such as grasping a dish or grasping the dishwasher handle.

The hierarchical structure implicit in the encoding of intentions organizes our understanding of relations among individual intentions. This hierarchical organization often diverges from the relations that we experience in the motion stream itself. That this is so is obvious when one realizes that people frequently interrupt their pursuit of one intention to enact another, as when one "pauses" for a sip of wine in between loading dishes into the dishwasher. Wine-sipping and acts of dishwasher loading may be adjacent to each other within the stream of motion, yet we understand the non-adjacent acts of dishwasher loading as conceptually related to a much greater degree. As observers, we would encode the motion stream itself (represented in Figure 1) in terms of a structure something like that represented schematically in Figure 2. Note that "drink wine", while immediately adjacent in the *surface stream* to the motions "glass into dishwasher" and "cup into dishwasher", is *not* adjacent to either in the encoded intentional structure.

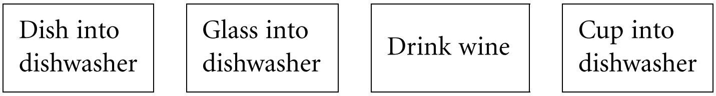


Figure 1. Actions adjacent in the surface motion stream

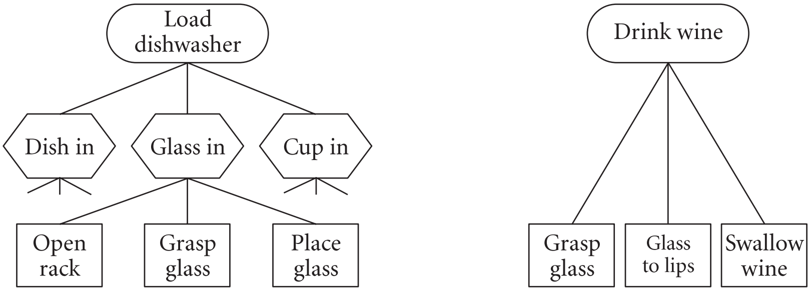


Figure 2. Encoded representation of the motion stream

On the more general level, this example illustrates in yet another way that our encoding of others' intentions goes beyond registering the surface properties of the motions themselves. Intentions thus represent the meaning or "gist" of motion, and it seems almost certain that our analysis of intentions is what we typically retain when recalling others' motions, rather than the "verbatim" motion stream. This kind of complex interplay between different levels of analysis seems typical of generative systems.

4. Acquisition of intentional understanding

If we buy that discerning intentions rides on the operation of a generative knowledge system, how is this system acquired? The analogies I've been drawing to language, and the history of debate on the topic of language acquisition suggest we should consider whether intentional understanding (IU), or at least some universal core of IU, might be innate, and thus not really acquired at all. We could launch a poverty of the stimulus argument that invokes Gold's theorem (applicable to IU given the involvement of hierarchical, embedded structures) that provides impetus for such a radical nativist account. On the other hand, contrary to what is traditionally argued regarding language, intentional understanding might be a domain in which negative evidence is available to learners. That is, children might have access to evidence that clarifies which actions are structurally ill-formed for fulfilling intentions. Afterall, it is not uncommon for intentional action to fail in achieving its goals, and infants might be able to discriminate when failures are due to disrupted structure as opposed to disrupted content. If negative evidence is indeed available to infants, IU might be readily "learnable" (in Gold's and Chomsky's sense of that term). And there are other reasons why we can safely dismiss a *radical* nativist account. For one, people's analysis of others' intentions is so obviously colored by experience and world knowledge, to such a very great degree, that a learning account is obviously imperative.

If we then approach the genesis of IU as a learning problem, what kind of acquisition account might do the work? Here I move into the realm of rampant speculation, although these speculations aren't entirely unconstrained by data. In broad outline, I suspect that the mechanisms enabling acquisition of IU overlap heavily with the mechanisms enabling the acquisition of language. More specifically, I have two proposals:

Structure detection. First, I propose that the acquisition of IU (like the acquisition of language) originates in a relatively general propensity to extract patterns — structure that is configurational and/or stochastic in nature — from the sensory array. This structure extraction or pattern recognition capacity jump-starts analysis of the stimulus stream. In particular, statistical analysis of structures within the sensory array enables children to identify relevant structural units and segment them from the continuous stream. These structural units can then be encoded as discrete entities. The nature of the structure across domains such as speech versus motion differs considerably, and hence the same structure detection mechanisms brought to bear in analyzing these different stimuli rapidly give rise to different kinds of structural representations for these domains. This is an important point. In arguing that the same structure detection mechanisms jump-start knowledge acquisition across language and intentional action, I am not arguing for iconicity across these domains in the knowledge children acquire. They are separate knowledge systems.

Induction. Second, I propose that the acquisition of IU (like language) involves the operation of inductive mechanisms that not only seek to systematize representations in terms of shared properties, but also can — and are possibly even predisposed to — group representations with respect to commonalities *not* available in surface sensory experience. This is how "content" or "meaning" emerges within the system. This kind of capacity for organizing knowledge with respect to non-obvious commonalities is at the heart of theory formation of any kind. It enables knowledge to be encoded in ways that diverge from surface experience, and it enables new structural relations to be discovered that organize across non-obvious or abstracted commonalities. Computational models of just such inductive processes — mechanisms that achieve progressively abstract representations from structured surface representations — are provided by Gentner and colleagues (e.g., Gentner and Medina 1998) as well as Holyoak and his colleagues (e.g., Holyoak and Thagard 1995). The genesis of intentional understanding represents a new and potentially illuminating application for these models.

5. Structure detection and induction in language learning

First, I'll illustrate these two proposals in the context of language, because language is the generative system we currently understand best. Regarding structure detection mechanisms, we have abundant reason to believe they are crucially involved in language learning. Recently, Saffran, Aslin, and Newport (e.g., 1996) directly demonstrated that infants as young as 8 months can learn to segment a novel, continuous stream (bidakupadotigolabubidakutupirogolabu) into "words" based on purely statistical evidence. These "words" (e.g., bidaku) are not semantic entities, but purely structural entities — they have no content or meaning. Skill at extracting them, on structural grounds alone, however, would be of great value to infants: once extracted as discrete units, infants are nicely positioned to note how they are used in relation to other stuff in their experience (such as objects, actions, emotions, etc.). Processing of relations between structurally-derived units and things in the world can also begin on statistical grounds (e.g., infants might be primed to note the degree to which a given unit of speech covaries with something else in their experience), but there is reason to believe that processing or world-world relations quickly goes beyond the statistical level and begins to take on a character indicative of the inductive mechanism I outlined in the second proposal. Recall that the idea here is that the human mind is inclined to abstract beyond surface relations toward registering non-obvious commonalities. One source of evidence that such a mechanism operates in word learning comes from recent work by Graham (in press) as well from my own research lab (Desjardins and Baldwin, 1995). Both of these research programs found that 16- to 18-month-old infants expect objects given the same label to share non-obvious properties. For example, if infants encounter a novel device that emits a "moo" when tipped and it is called a *modi* (a word infants haven't encountered before), infants readily expect a rather dissimilar object to likewise "moo" when tipped as long as it, also, is called a *modi.* When labels are *not* provided (instead, the objects are talked about in a generally enthusiastic manner), infants are disinclined to predict the presence of the hidden property. It seems that, without the shared label, infants simply aren't sure whether to relate the two rather dissimilar objects. Yet infants accept a novel label as providing such guidance, even with respect to properties that must be inferred. Infants could display this pattern only if their inductive mechanisms are open, and perhaps even predisposed, to representing commonalities that go beyond what is directly accessible in the sensory array.

My proposal is that similar processes are operative in infants' processing of dynamic human motion, and these processes play a crucial role in the acquisition of IU. I offer two illustrations. One illustration concerns segmentation in action processing, the other concerns categorical judgments in action processing.

6. Structure detection and induction in the acquisition of intentional understanding

Segmenting the continuous motion stream. The idea that structure detection mechanisms play an important role in infants' segmentation of the continuous motion stream seems plausible, in part, because intentional action exhibits considerable structure. To illustrate, when we act intentionally on objects in the world, as in kitchen clean-up, grocery shopping, or tooth-brushing, we display a sequence of maneuvers that is highly predictable across many diverse action contexts. These sequences are predictable in terms of how body parts move in relation to one another and the timing relations binding these sequential movements. For example, to act on most objects we must first locate the relevant object (necessitating gaze and correlated head direction changes), gain proximity to the object (necessitating predictable changes in body position and trajectory, leg motion, etc.), contact the object (necessitating predictable changes in arm position and trajectory as well as hand position and finger configuration), manipulate the object (involving predictable correspondences between gaze direction, hand and finger motions, and object position), and release the object (involving predictable changes in relative position of body to object, etc.). As well, a predictable temporal contour — patterns of acceleration, steady state along a trajectory, deceleration — accompanies the sequence.

If infants are sensitive to these structural properties, they might be able to identify such an "intention-relevant" segment on purely structural grounds, through the operation of statistical learning or other skills that support the extraction of configurational structure. Structure detection processes would thus enable them to extract discrete units from within a largely continuous and dynamic stream of motion. Of course, their representation of such discrete segments would be solely structural in nature — they wouldn't have content or meaning. Put another way, statistical learning and pattern extraction might enable infants to extract and represent "intention relevant" segments within human motion without any actual understanding of the intentions at play. Such structure detection processes could thus help to explain how infants break into processing a complex motion stimulus — human intentional action — without yet understanding the intentional content underlying the motion stream. By positing such structure detection capacities, we might be able to account for how learning could begin in this domain in the absence of content-rich innate knowledge structures.

Attributing content to structurally-based units of motion. If the above ideas about structure-detection skills are on the correct track, then how do infants go on to infuse their structurally-derived "intention-relevant" units with content? How do they enrich their initial, dry-bones structural segments to encompass an appreciation for the intentions motivating these motion patterns? An inductive mechanism oriented toward representing commonalities that go beyond the surface would need to be involved. Infants would need to be inclined to note commonalities across different kinds of segments to build action categories, and they would need to be inclined to posit non-obvious commonalities (such as intentions) that are shared by category members.

What reason do we have to believe this structural segmentation/induced content story? Our reasons for belief are few, but not nonexistent. First, research led by Jodie Baird, that Bertram Malle and I joined, provides evidence that adults, at least, indeed engage in segmentation as part of their processing of continuous intentional action (Baird, Baldwin, and Malle, under revision). We showed adults digitized videos of continuous intentional action, such as kitchen clean-up, in which tones were superimposed at various points within the motion stream.

Adults showed substantially greater accuracy in recalling the locations of tones that coincided with the completion of intentions (endpoint tones) relative to the location of tones that coincided with points prior to the completion of intentions (midpoint tones). In particular, adults misremembered midpoint tones as occurring much closer to the completion of an intentional unit than they actually did (see Figure 3). Moreover, a control study demonstrated that these accuracy differences emerged only when correspondence between the tones and the motion stream was at issue. These findings support the idea that adults segment continuous intentional action into units coinciding with intentional boundaries, and these units resist interruption.

Second, in other research my collegues and I found that infants 10–11 months of age — by many accounts too young to yet be interpreting human motion in intentional terms — similarly segment a continuous motion stream along intention boundaries (Baldwin, Baird, Saylor, and Clark 2000). I'll describe one study in some detail. On repeated trials infants watched a digitized video depicting a woman in a kitchen turning, noticing a towel on the floor, moving toward and bending down to grasp the towel, picking up the towel, straightening, and moving across the room to hang the towel on a towel rack. Infants then saw two kinds of test videos: *completing* test videos, in which the continuous motion stream was interrupted by a 1.5 second still-frame pause just as grasping of the towel was completed, and *interrupting* test videos, in which a 1.5 second still-frame pause interrupted midstream the motion to reach toward the towel. Infants looked longer at the interrupting test videos than the completing test videos, suggesting that the interrupting test videos violated the structure they had extracted from the motion stream in their prior processing (see Figure 4). A control study revealed that the interrupting test videos were no more salient to infants, in a superficial

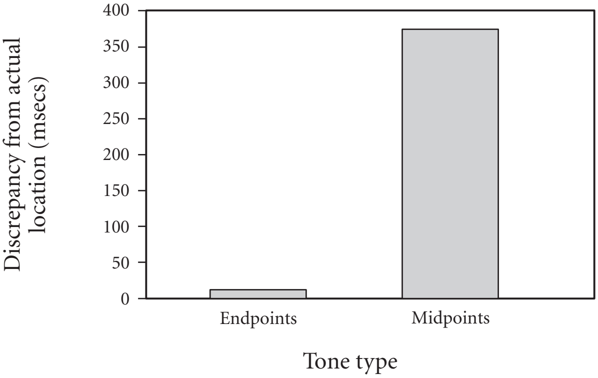


Figure 3. Adults' judgments of tone location (N = 32)

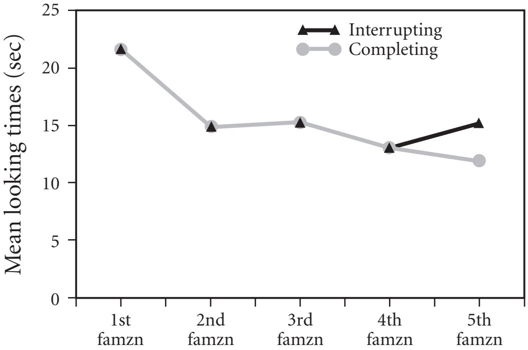


Figure 4. Infants' looking to familiarization and test videos (Age: 10–11 months, N = 16)

sense, than were the completing test videos (see Figure 5). Thus the findings clarify that 10- to 11-month-olds spontaneously process continuously flowing intentional action with respect to intention-relevant units. A more recent study (Saylor, Baird, Killingsworth, and Baldwin, in preparation) replicated the basic finding — 10- to 11-month-olds segment continuous intentional action along intention boundaries — with a different methodology in which infants viewed live action rather than videos, received much less exposure to any given action, and viewed an expanded range of actions. Thus infants' ability to extract "intention-relevant" units from within continuous motion seems fairly robust.

We suspect that infants' skill at segmenting continuous motion along intention boundaries derives, at least in part, from their skill at detection of statistical

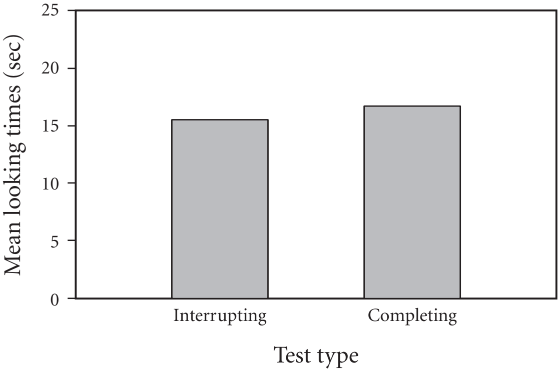


Figure 5. Infants' looking to test sequences without prior familiarization (Age: 10–11 months, N = 16)

and/or configurational structure, rather than depending solely on a conceptual-level understanding of the content of the intentions. Several studies currently underway directly investigate this issue. I'll describe just one. It is a direct analog of the Saffran, Aslin, and Newport (1996) research I described earlier that demonstrates the role of statistical learning in infants' extraction of words from continuously flowing speech. In this study, infants are introduced to a series of continuous novel motion sequences in which transitional probabilities between adjacent movements vary. Because the motion sequences are novel, statistical information about transitional probabilities will be the only basis on which infants might come to identify units — that is, "actions" — within the continuous motion stream. In a subsequent test phase, infants view two kinds of motion sequences (in isolation): "actions" (motion sequences in which adjacent movements had high transitional probabilities during training) and "non-actions" (motion sequences in which adjacent movements were low in transitional probability during training). If infants can use statistical information about transitional probabilities between adjacent movements to drive segmentation of the continuous motion stream, they should look longer at non-actions (which violate their prior statistical experience) than actions (which match their prior statistical experience). This pattern would indicate that infants can use purely statistical information to identify "actions" within a continuous motion stream.

Finally, what reason do we have to believe that inductive mechanisms oriented toward representing non-obvious commonalities play a role in infants' discovery of the intentional content of "actions" (intention-relevant segments) they extract from the motion stream? None yet, but I view this as a crucial issue for understanding knowledge acquisition in this domain. Of course the issues Bertram Malle (this volume) discusses — mutual escalation processes that promote inductive progress in knowledge about language on the one hand and knowledge about the mind on the other — will likely be a part of any fully elaborated story here. That is, language seems to facilitate the ability to register or encode non-obvious commonalities, and thus learning terms for actions may well promote the acquisition of quintessentially theoretical concepts such as intention.

Categorical judgments in action processing. A second illustration of my two-fold proposal (the acquisition of IU depends heavily on both structure-detection and inductive mechanisms) involves judgments that we readily make about the content and quality of others' intentional actions. Here I'm concerned with judgments as to whether intentional action is hostile or benign for actions such as caressing versus slapping. I suspect that such judgments display categorical discrimination in just the way that native speaker's judgments about consonant contrasts (e.g., /ba/ vs. /pa/) do. For the sake of establishing common ground, I'll briefly review the notion of categorical discrimination in the consonant-contrast context (Whitney 1998).

Described in purely acoustical terms, the distinction between /ba/ and /pa/ can be carried solely through changes in voice-onset time. Voice-onset time for /ba/ is shorter (vocal cords resonate simultaneously or shortly after air is released from between closed lips) than for /pa/ (vocal cords resonate considerably later than the time at which air is released from between the lips). Voice-onset time is of course a continuous temporal dimension. However, when we vary voice-onset time in continuous intervals, we find that people's judgments of which sound they hear (/ba/ vs. /pa/) is not continuous at all. Significant changes in voice-onset time can occur without concomitant change in people's judgment of the syllable they hear, while relatively small changes in voice-onset time at one specific region within the continuum can produce abrupt changes in judgment. In other words, discriminability is enhanced at phonetic boundaries. Speech-sound discrimination thus reflects the imposition of categories on what is, physically speaking, a continuous dimension.

Given that speech sounds are created by motions of the articulatory apparatus, it doesn't seem far-fetched to think that motions produced on a grander scale — by the body as a whole — might be processed in ways similar to those involved for speech sounds. And if you consider minimally contrasting actions, such as caressing the face with the hand as opposed to slapping the face with the hand, it is clear that these motions primarily contrast along a particular dimension. In the case of *caress* versus *slap*, the contrast is temporal in nature; these actions display different rates of acceleration and deceleration. It seems plausible that people's identification of the action (*caress* vs. *slap*) reflects the imposition of categories on continuous changes along the temporal dimension. In other words, if we create a continuum of stimuli that incrementally vary between a *slap* and a *caress*, we may find that observers are inclined to gloss over substantial change in one region of the continuum and count all as instances of *caress*, while changes in one particular, quite narrow, region of the continuum may yield abrupt changes in judgment, to *slap*.

Why would categorical discrimination in the domain of intentional action be useful for processing? For the same reasons that categorical discrimination is touted as essential in the language domain. That is, categorical discrimination enables the observer to transform continuously varying parameters into discrete judgments, which opens the door to building a combinatorial (and potentially generative) system.

Second, categorical discrimination enables the observer to key on the information-rich aspects of a motion display, saving the processing and memory system from being overwhelmed by the vast amount of continuously varying information. Considering the case of *caress* versus *slap* helps to make this intuition concrete: as an observer of action it is important to me to know whether someone else is about to slap me or caress me — this judgment will radically change my plans for how to prepare and react. Encoding *precisely* how vicious the slap or how gentle the caress seems decidedly less important. Likely I *can* make such fine-grained distinctions if the need arises (for instance, if my language "lures" me to do so (Slobin 2001), but I might not do so on a categorical basis).

What kind of mechanism might enable infants to begin discriminating minimally contrasting actions categorically? If the answer is anything like what has emerged in the language domain (e.g., Werker 1989), then categorical discrimination of certain action contrasts might well be available very early in life. At the same time, we know that categorical discrimination abilities in the language domain are affected by language input. That is, by the end of the first year of life infants retain categorical discrimination only for native consonant contrasts. Kuhl and colleagues (Kuhl, Williams, Lacerda, Stevens, and Lindblom 1992) suggest that this "tuning" or calibration process might well occur solely via perceptual analysis of the sounds heard in the ambient language environment. If this is correct, it is another example of the key role that structural analysis capabilities play in the acquisition of a generative system.

It seems at least possible that categorical discrimination within the action domain might display, developmentally, a structurally-based calibration to the ambient action input analogous to that observed in speech perception. This calibration process could take a similar form to that occurring in speech perception (i.e., infants initially discriminate a greater range of distinctions than is supported by the input), but might well take a different form (i.e., initial categorical distinctions are few, and expand with experience) or a more complex form (i.e., initial categorical distinctions are to some degree readjusted to different regions within the relevant psychophysical dimensions). The currently impoverished state of our knowledge about the psychophysical structure of intentional action is what makes it difficult to opt for any one particular hypothesis here. Hopefully this will soon change. Obviously, these ideas are radically speculative, but also highly investigable.

Finally, a point of connection to language: Languages of course differ considerably in the lexical machinery they devote to marking subtle differences in action categories (Bybee 1985; Slobin 2001; Talmy 1980). English emphasizes manner differences in its verb lexicon, so English speakers have at their disposal a very large range of "manual contact with hostile intent" verbs, such as hit, punch, whack, pummel, smack, slap, spank, clout, cuff, and whallop, and similarly a range of verbs communicating "manual contact with benign/caring intent", such as caress, fondle, pet, pat, stroke, tap, and touch. Other languages, such as Spanish, provide less conventional lexical machinery for remarking on differences in manner, and without knowing Spanish, I'm hazarding that it displays reduced lexicalization than English with respect to verbs encoding manual contact. In a recent talk, Slobin (2001) pointed out that languages tend to display propensities either for or against the marking of such subtleties. That is, if a language tends to elaborate distinctions in one arena, then subtle distinctions in other arenas are likely. He suggests, following Bowerman's evidence (e.g., Bowerman 1989) that children very early become sensitive to the propensities of their native language. If so, the language children are learning may influence the extent to which they engage in detailed analysis of subtle differences in motion patterns and the intentions that give rise to them. Differences in marking of subtle distinctions between similar motion patterns across languages/cultures clearly points to the involvement of learning processes, and the developmental patterns here may be analogous, at least in some respects, to those uncovered in the case of speech perception.

Understanding action in intentional terms. If infants are inclined, on a purely structural basis, to distinguish categorically between actions varying incrementally along a continuous dimension, then how do they come to imbue these action categories with content or meaning? What would lead them to regard a slap as originating in hostile *intent*, and a caress as having its source in benign intent, for instance? Again, I suspect that inductive mechanisms promoting attention to non-obvious commonalities is at play. Slaps feel yucky, and one seeks to avoid them. Others on the receiving end appear to find slapping yucky, too. People who slap often inflict other actions that make one feel yucky. The notion of hostile intent is a non-obvious organizing principle that makes sense of some of the disparate experiences that give rise to this aversive quality; it enables one to predict (and hopefully avoid or combat) such experiences in the future.

Of course, one can experience pain or discomfort at the hands of another without hostile intent being involved — accidental infliction happens as well. If inductive mechanisms are involved in children's ability to sort all this out, we would expect the sorting out process — at the conceptual, but not necessarily structural, level — to take time and experience. And indeed it seems to. I recall making little headway, for example, when trying to convince my 3-year-old that a nurse who administered a painful injection was not meaning to hurt him (he was pretty mad at her). Children seem to have little difficulty identifying which actions will hurt and which will not. The attribution of intent poses greater obstacles for them, however. A fundamental issue children need to sort out, here, is that action and intent are in many-to-many relation: a given action can be consistent with many intentions (e.g., poking a needle into an arm could be done to fulfill an intention to heal or an intention to harm), and a given intention can give rise to many actions (e.g., an intention to heal can lead to actions as diverse as administering an injection, advice, referral, or surgery). Recent research by Baird and Moses (in press) indicates that even four-year-olds do not yet fully understand these issues; they are inclined, for example, to think that a given motion pattern arises from just one intention.

7. Structure-analysis and induction: can they be decoupled?

If there is anything to the two-part proposal I'm putting forward, then we might expect that the parts would be, to some degree, dissociable. Are there instances in which individuals seem to possess the relevant structural analysis abilities, but lack the inductive mechanism to invest structured units of motion with the attribution of intention? Just above I suggested that we might see such a dissociation in children's development. Children at a given age may be able to discriminate between actions (an ability perhaps grounded in structural analysis), while at the same time displaying deficits (relative to adults) in understanding the motivations engendering those distinct actions.

A different case for a possible dissociation between structural and conceptual analyses of action — chimpanzees — has recently been put forward by Povinelli and colleagues (e.g., Povinelli 2001; Povinelli and Giambrone, in press). They have carried out an extended series of carefully designed probes of mentalistic understanding, and have repeatedly found chimpanzees' understanding of folk psychological notions such as intention, belief, knowledge, and seeing to be impoverished at best. Clearly, however, chimpanzees can process others' behavior at some level — they show appropriate and planful responses to others' actions across diverse contexts. Povinelli and colleagues have suggested what they call the "reinterpretation hypothesis" to make sense of this apparent paradox. They propose that chimpanzees share with humans a basic, and phylogenetically ancient, set of skills for analyzing behavior in a structural sense (Hauser 2001, likewise suspects this), but lack the ability to reinterpret their "surfacy" representations of behavior in terms of higher-level, more abstract, folk psychological notions such as intention. On this account, then, it is the powerful inductive mechanism for reinterpretation of surface behavior — in terms of intentions, beliefs, and the like — that is special to humans.

In possible contrast, recent findings from Tomasello and his colleagues (e.g., Call and Tomasello 1998; Hare, *et al.* 2000) as well as Savage-Rumbaugh and her colleagues (e.g., Savage-Rumbaugh, Fields, and Taglialatela 2001) raise the possibility that chimpanzees possess genuine mentalistic understanding. They, like Hauser (2001), point to other factors (e.g., enculturation, ability to represent recursive structures, etc.) as the source of cross-species differences in cognition. Clearly, much more work is needed before the dust can settle on these fascinating issues.

8. Summary and developmental/evolutionary speculation

What I've proposed thus far can be readily summarized: Attribution of intentions is a theory-driven exercise, and to build a sound theory one must first be able to isolate the relevant phenomena. Structure-analysis capabilities help to make this possible. Two examples I considered are structure-based extraction of "intention-relevant" units within the continuous flow of motion, and categorical discrimination that enables one to key on information-rich regions within continuous motion parameters to aid in establishing relevant action categories. Building the theory also requires that one go beyond surface patterns of motion to infer underlying cause, which, in the case of intentional action, is, at least in part, intention. The particular theory that is built — Intentional Understanding — is a powerful one: it enables us to make sense of continuous motion streams even when these are novel, and it enables us to assemble new classes of intentions.

Regarding evolutionary speculation, I began with the suggestion that language development and language evolution will be better understood if we also understand the development and evolution of other cognitive systems that are closely allied with language. Our system for discerning others' intentions within their bodily motion is closely allied with language in large part because action processing and language processing are directed toward the same overarching goal: to figure out the other's purpose in moving or speaking. And the purpose behind people's movements is frequently concordant with the purpose behind their speech. This is why, in development, being good at discerning intentions within others' motion patterns helps one to figure out the meaning of their speech.

I've suggested in this chapter that human skill at intention detection in the action arena may be closely allied with language for an entirely different reason as well: the cognitive systems that make possible the acquisition of these two generative knowledge systems may hinge on some of the same learning mechanisms. In particular, I've suggested that knowledge in both domains requires a combination of structure-detection skills and inductive propensities.

I've gone further to suggest that some of the particular structure-detection skills thought to support language acquisition (e.g., statistical learning) might likewise support the acquisition of IU. If this is correct, we can transfer some of what we suspect about language evolution to the case of the evolution of IU, and vice versa.

Take the case of categorical discrimination of consonant contrasts, for example. In research by Kuhl (e.g. 1981), chinchillas demonstrate categorical discrimination for consonant contrasts. Also, others (e.g., Cutting and Rosner 1974) have shown that people discriminate some non-linguistic sounds categorically. These findings lead to the conclusion that categorical discrimination arises from a domain-general processing propensity, and that this propensity is not unique to humans, nor even to primates. We also have reason to believe, from a long tradition of research on animal behavior, that structure-detection skills of other sorts — such as statistical learning — are evolutionarily ancient (e.g., Holland, Holyoak, Nisbett, and Thagard 1986; Rescorla 1972). Along these same lines, Ramus, Hauser, Miller, Morris, and Mehler (2000) recently reported that cotton-top tamarin monkeys share human infants' ability to discriminate between different languages, indicating that such discrimination hinges on perceptual and neurobiological mechanisms that are general properties of the primate auditory system. This set of findings regarding structure-detection skills for language suggests that structure-detection in the domain of intentional action might also be evolutionarily ancient. And, as I described earlier, this is precisely what Povinelli and colleagues have explicitly argued in a series of recent papers.

Perhaps this isn't a surprising idea. Likely, basic structure-detection skills for discrimination and segmentation are prerequisite for language and IU to evolve at all. Still, it remains possible that humans have evolved specialized forms of structure-detection. This possibility is currently under investigation by researchers such as Newport, Hauser, and their colleagues (e.g., Newport 2000; Hauser 2001). For example, they are probing cotton-top tamarins' ability to detect statistical structure of various kinds, and initial findings suggest that the tamarins versus human infants display differing strengths and weaknesses. It may be, then, that the stories we can ultimately tell regarding language evolution and IU evolution will include some chapters about specialized structure-recovery mechanisms that arose for the first time with *Homo*.

As you recall, I've argued that structure detection alone won't give us either language or IU in children's development. An inductive system that is highly flexible and open to inferring non-obvious commonalities across a broad range of domains and contexts seems also prerequisite. At this point, we have every reason to believe that induction is something humans do far more openly and flexibly than any other species (e.g., James, 1890; Povinelli 2000). This is not to say that we don't see genuine induction in other species. Quite the contrary, I believe. There are many striking examples of non-human inductive skill, such as stellar orientation in migratory songbirds and dead-reckoning by ants (e.g., Gallistel, Brown, Carey, Gelman, and Keil 1991). It is simply that songbirds and ants are enormously constrained in the range of domains in which they can execute such skills, whereas humans seem able to readily establish new representations and derive inferences from them across a wide range of domains, modalities, and circumstances. The sheer power of human induction, in my view, is what stands out most in relation to other species. I suspect that it is a key element in accounting for the evolution of language and IU. Likewise, I, like others (e.g., Premack 1984; Tomasello 1999) suspect that language and IU have, in turn, enabled us to exploit our special inductive capacity to an unprecedented degree, opening up, among other things, new realms of cultural learning. If this is right, then what we have is an evolutionary and developmental *quid pro quo* (inductive skill makes possible the acquisition of language and IU; language and IU in turn expand inductive potential) that has been all to human advantage.

References

Alibali, M.W., Bassok, M., Solomon, K.O., Syc, S.E., Goldin-Meadow, S., 1999. “Illuminating mental representations through speech and gesture”. *Psychological Science*: 327–33.

Baird, J.A. and Moses, L.J. In press. “Young children’s understanding that identical actions may be motivated by different intentions”. *Journal of Cognition and Development.*

Baird, J.A., Baldwin, D.A., and Malle, B.F. Under revision. “Adults’ parsing of continuous action: The psychological reality of units corresponding to actors’ initiation and completion of intentions”. *Cognitive Psychology.*

Baldwin, D.A. 1991. “Infants’ contribution to the achievement of joint reference”. *Child Development* 62: 875–90.

Baldwin, D.A. 1993a. “Infants’ consult the speaker for clues to word reference”. Journal of Child Language 20: 395–418.

Baldwin: D.A. 1993b. “Early referential understanding: Infants’ ability to recognize referential acts for what they are”. *Developmental Psychology* 29: 832–43.

Baldwin, D.A. 2000. “Interpersonal understanding fuels knowledge acquisition”. *Current Directions in Psychological Science* 9: 40–5.

Baldwin, D.A., and Baird, J.A. 2001. “Discerning intentions in dynamic human action”. *Trends in Cognitive Sciences* 5: 171–8.

Baldwin, D.A., Baird, J.A., Saylor, M.M., and Clark, M.A. 2001. “Infants parse dynamic action”. *Child Development* 72: 708–17.

Bloom, P. 1994. “Generativity within language and other cognitive domains”. *Cognition* 51: 177–89.

Bowerman, M. 1989. “Learning a semantic system: What role do cognitive predispositions play?” In *The Teachability of Language*, Schiefelbusch, R.L., and Rice, M.L. (eds). Paul H. Brookes Publishing Co.

Bybee, J.L. 1985. *Morphology: A Study of the Relation Between Meaning and Form.* Amsterdam: John Benjamins.

Call, J., and Tomasello, M. 1998. “Distinguishing intentional from accidental actions in orangutans (*Pongo pygmaeus*), chimpanzees (*Pan troglodytes*), and human children (*Homo sapiens*)”. *Journal of Comparative Psychology* 112: 192–206.

Cutting, J.E., and Rosner, B.S. 1974. “Categories and boundaries in speech and music”. *Perception and Psychophysics* 16: 564–70.

Desjardins, R., and Baldwin, D.A. 1992. Infants’ use of language to guide reasoning about the object world. Paper presented at the biennial meeting of the International Society for Infant Studies. Miami, Fla.

Fenk-Oczlon, G., and Fenk, A. 2001. The clausal structure of linguistic and pre-linguistic behavior. Paper presented at the symposium on the evolution of language out of pre-language. Eugene, Or.

Gallistel, C.R., Brown, A.L., Carey, S., Gelman, R., and Keil, F.C. 1991. “Lessons from animal learning for the study of cognitive development”. In *The Epigenesis of Mind: Essays on Biology and Cognition*, S. Carey and R. Gelman (eds), 3–36. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gentner, D. and Medina, 1998. "Similarity and the development of rules". *Cognition* 65: 263–97.

Goldin-Meadow, S. 1998. "The development of gesture and speech as an integrated system". In *The Nature and Functions of Gesture in Children's Communication: New Directions for Child Development,* J.M. Iverson *et al.* (eds), 29–42. San Francisco: Jossey-Bass.

Hare, B., Call, J., Agneta, B., and Tomasello, M. 2000. "Chimpanzees know what conspecifics do and do not see". *Animal Behaviour* 59: 771–85.

Hauser, M. 2001. "To understand the evolution of language and number in humans, don't study humans!" Paper presented at the Institute of Neuroscience, University of Oregon. Eugene, Or.

Holland, J.H., Holyoak, K.J., Nisbett, R.E., and Thagard, P.R. 1986. *Induction: Processes of Inference, Learning, and Discovery*. Cambridge, Mass.: MIT Press.

Holyoak, K. and Thagard, P. 1995. *Mental Leaps: Analogy in Creative Thought.* Cambridge, Mass.: MIT Press.

Iverson, J.M., Goldin-Meadow, S. 1998. "Why people gesture when they speak". *Nature* 228.

James, W. 1980. *The Principles of Psychology*. New York: Dover.

Kuhl, P.K. 1981. "Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech sound categories". *Journal of the Acoustical Society of America* 70: 340–9.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., and Lindblom, B. 1992. "Linguistic experiences alter phonetic perception in infants by 6 months of age". *Science* 255: 606–8.

Newport, E. 2000. "Innately constrained learning: The place of statistical learning in the acquisition of language". Paper presented at a conference on "Finding the Words" Stanford, Calif.: Stanford University.

Premack, D. 1984. "Possible general effects of language training on the chimpanzee". *Human Development* 27: 268–81.

Povinelli, D.J. 2000. *Folk Physics for Apes: The Chimpanzee's Theory of How the World Works.* Oxford: Oxford University Press.

Povinelli, D.J. 2001. "On the possibilities of detecting intentions prior to understanding them". In *Intentions and Intentionality: Foundations of Social Cognition*, B.F. Malle, J.L. Moses, and D.A. Baldwin (eds), 225–50, Cambridge, Mass.: MIT Press.

Povinelli, D.J., and Giambrone, S. In press. "Reasoning about beliefs: A human specialization?" *Child Development.*

Ramus, F., Hauser, M.D., Miller, C., Morris, D., and Mehler, J. 2000. "Language discrimination by human newborns and by cotton-top tamarin monkeys". *Science* 88: 349–51.

Rescorla, R.A. 1972. "Informational variables in Pavlovian conditioning". In *The Psychology of Learning and Motivation*, G.H. Bower (ed). New York: Academic Press.

Saffran, J.R., Aslin, R.N., and Newport, E.L. 1996. "Statistical learning by 8-month-old infants". *Science* 274: 1926–8.

Saylor, M., Baird, J.A., Killingsworth, J., and Baldwin, D. In preparation. "Scope and robustness of infants' action parsing skill".

Savage-Rumbaugh, S., Fields, W. M., and Taglialatela. 2001. "Language, speech, tools, and writing: A cultural imperative". *Journal of Conciousness Studies* 8: 273–92.

Slobin, D. I. 2001. The many sources of structure in child language. Master lecture presented at the biennial meeting of the Society for Research in Child Development, Minneapolis, Minn.

Talmy, L. 1985. "Lexicalization patterns: Semantic structure in lexical form description". In *Language Typology and Syntactic Fieldwork*, T. Shopen (ed). Cambridge: Cambridge University Press.

Tomasello, M. 1999. *Cultural Origins of Human Cognition*. Cambridge, Mass.: Harvard University Press.

Welder, A. N., and Graham, S. A. 2001. "The influence of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties". *Child Development* 72: 1653–73.

Werker, J. 1989. "Becoming a native listener". *American Scientist* 77: 54–9

PART 4

Language development

CHAPTER 12

The emergence of grammar in early child language

Michael Tomasello
Max Planck Institute for Evolutionary Anthropology

It is presumably the case that the first humans to communicate with one another symbolically did so with single symbols. They then began to use multiple symbols in particular communicative contexts, and in some cases particular strings of symbols came to be used as multi-symbol constructions. Historical processes of grammaticalization and syntacticization led to the creation of a whole inventory of such constructions, some based on particular symbols and some based on more abstract patterns of symbols.

An important part of this process was, of course, ontogeny. Each new generation of speakers was exposed to whatever constructions were currently in use, with no access to the historical development. Children thus did not need to create constructions from discourse sequences of symbols. It is also important that children were not exposed to abstract constructions directly, in the sense that one cannot be directly exposed to abstract categories or schemas per se. Learners experience only individual utterances, and they must create for themselves patterns or abstractions across those utterances.

In this chapter I would like to outline some of the basic processes by means of which young children comprehend the utterances of other people and from this basic experience construct an inventory of grammatical constructions of various shapes, sizes, and degrees of abstraction. This process played a key role, and continues to play a key role, in the evolution and historical development of languages. In the process of doing this, I will also relate several lines of research on child language acquisition in which my colleagues and I are currently engaged. The model with which I will be working is a usage-based model broadly consistent with the approaches of Langacker (1987, 1988, 2000) and Bybee (1985, 1995, 1998), with a few bits from Givón (1995), Croft (2000), and Fillmore/Goldberg (1995). It consists of three main propositions:

1. Children hear and attempt to learn utterances, i.e., phonological forms for expressing communicative intentions; in doing so they assign a communica-

tive function both to the utterance as a whole and to certain of its constituents (sub-functions of the utterance's overall function).[1]

2. Children store comprehended and produced utterances and constituents, along with their functional analyses; token and type frequency of stored pieces of language lead to the construction of (a) utterance schemas and (b) constituent schemas, both of which may contain relatively abstract 'slots' (and it is an empirical fact that children construct these abstractions only gradually and in piecemeal, item-based fashion; see below).
3. Children produce utterances by "cutting and pasting" from their inventory of stored utterance schemas and constituent schemas.

I treat each of these in turn, followed by an account of some of our recent research that is relevant to these and related issues.

1. Functional analysis of utterances

Focusing on linguistic communication, not LANGUAGE, the most basic unit of linguistic experience is the **utterance**. It is most basic because it is the smallest unit in which a person expresses a complete communicative intention — that is, an intention that another person share attention with her to something (see Clark 1996, on how this analysis applies even to requests and performatives). Accordingly (and iconically), utterances have a coherent phonological form — typically, pauses on either side and one of a delimited number of distinct intonation contours. This is perhaps especially clear in early adult-child discourse since typically each participant's conversational turn is exactly one utterance long.

When an adult speaks to her, what the child is attempting to do most urgently is to comprehend the overall communicative intention behind the utterance; what does the adult intend for me to attend to? At the same time, she is also attempting to determine the communicative intentions behind particular constituents within the utterance. This is a kind of "blame assignment" procedure in which the child attempts to determine the functional role of a constituent in the communicative intention of the utterance as a whole. (NB: the child cannot determine the function of a novel constituent without knowing at least something about the communicative intention of the whole utterance — one cannot determine a novel sub-function without knowing something about the overall function.) Presumably, particular utterance constituents are most easily identified — and emerge as independent units — when the same phonological form appears in different utterance-schema slots over time with some functional consistency. Thus, if the child hears "There's the ball", "Gimme my ball", "The ball's rolling", "The ball's bouncing", "I want a ball", "Throw the ball", "That ball's Jeffery's", "Where's your ball?", etc., the word *ball* comes to exist as a potential utterance constituent for future use when the child needs to indicate one of a certain class of objects as one sub-function of an utterance. (NB: if the child hears "We are going to the ball tonight" the completely different function of the phonological form *ball* in this instance means that it will not be assimilated to the more familiar instances of *ball* as toy.) Children are aided in this process by various kinds of cue validity and cue reliability in experienced exemplars over time, as outlined by Slobin (1985), Bates and MacWhinney (1989), and others.

As a nonlinguistic example, a child may see an adult use a stapler and understand that his goal is to staple together two pieces of paper. In some cases, the child may understand also that the sub-goal/function of placing the papers inside the stapler's jaws is to align them with the stapling mechanism inside the stapler, and that the sub-goal/function of pressing down on the stapler is to eject the staple through the two papers — with both of these sub-functions being in the service of the overall goal of attaching the two sheets of paper. The child does not need to understand all of this to mimic an adult stapling papers with the same stapler over and over again. (Analogy: the child can say *There-ya-go* over and over again without understanding its internal constituents). But to the extent that the child does not understand these sub-functions, she will be lost when she encounters some new stapler in which the sub-functions are effected by a different means, for example, one whose stapling mechanism does not require pressing down but rather squeezing. Only to the extent that the child understands the relevant sub-functions will she be able to adapt to new situations creatively by, for example, adjusting her behavior to effect the same outcome with the new stapling mechanism.

The comparable linguistic example is that the child hears an adult say "I stapled your papers" and comprehends not only the utterance and its overall communicative intention, but also, for example, the words *I* and *stapled* and their communicative sub-functions in the utterance (the contributions they are making to the utterance as a whole), along with the phrase *your papers* and its communicative sub-function in the utterance (and the sub-sub-functions of *your* and *papers*). As in the case of the stapler, it is only if the child performs some hierarchically organized 'functionally based distributional analysis' of this type that she will be able to comprehend the constituent linguistic elements in a deep enough way to enable her in the future to use them creatively in novel utterances.

By the way, this is the way children learn words. That is, children do not try to learn words directly; they try to comprehend utterances and in doing so they often must comprehend a word in the sense of determining the functional role it is playing in the utterance (i.e., its particular sub-function in the utterance). The lexicon — if there is such a thing — is thus only an emergent phenomenon in the sense of Bybee (1998). This is true despite the fact that the process is sometimes obscured in Western middle-class culture because parents and children often establish highly frequent utterance schemas for naming objects (e.g., "That's a ___". "It's a ___", "Here's the ___", etc.). Children understand quite well the overall function of these utterances as well as the function of the open slot, with the new word in the slot always serving to name the new object in the situation. But these kinds of schemas do not exist for many other word types — for instance, most verbs and many function words, which can only be learned through efforts to isolate their functional contribution in some larger and less predictable set of phrases (Tomasello 1987, 1995).

2. Creating schemas

Children hear and learn ("store") individual utterances that they experience in particular usage events (including their function). Stored utterances that are similar in phonological form and communicative function "congeal" into **utterance schemas**. The greater the similarity of form and function of stored utterances the more readily they congeal together into a single schema. Utterance schemas are more or less entrenched on the basis of their token frequency. Highly entrenched schemas resist internal disruptions and resist influence from other schemas (e.g., regularization). Acquiring skill with a stock of utterance schemas — more or less automated procedures for fluently expressing recurrent communicative intentions (perhaps with slight variations from occasion to occasion) — is the major early way station on the child's road to mature linguistic competence.

In one limiting case an utterance schema may be phonologically and functionally invariant, as in utterance types such as *There-ya-go* and *I-dunno*. On the other extreme, utterance schemas may theoretically be highly abstract in the sense that there is possible phonological variation of all possible schema constituents, with functional consistency only (e.g., the English ditransitive construction as in NP+VP+NP+NP to indicate instances of 'transfer', broadly defined). The more normal case — and the one most important for language acquisition — is that the utterance schema has an invariant phonological core and also, in addition, one or more **slots** that embody some possible variation (as in many adult constructions identified by Fillmore and other construction grammarians — e.g., "What's NP doing VERB-ing?"). A slot is defined by the phonological form and communicative function of the pieces of language that conventionally go into it (although, unlike much morphology, in syntax phonological form is seldom restrictive, and so mostly syntactic slots are defined purely by their functions). The productivity of a slot is a function of: (1) the amount of type variation that has occurred in it previously, and (2) the nature of this type variation, especially in terms of function.

Children's first utterance schemas consist of a consistent phonological form paired with a consistent communicative function (no slots). In many cases, the form corresponds to a single adult word — thus, the "single word period" — and this is sometimes called a holophrase. Selection of the particular linguistic form from adult models is due to a number of principles of salience as described, for example, in Slobin's (1973) operating principles (e.g., at the beginning children often store the last word of an utterance only — as when an English-speaking child learns to request "Up!" from such adult utterances as "Shall I pick you *up*?"). In addition, most children begin language acquisition by learning some unparsed adult expressions as holophrases — such things as "I-wanna-do-it", "Lemme-see", and "Where's-the-bottle"? A bit later, children learn to say things with some internal complexity, but these first complex utterance schemas are still consituted quite narrowly by particular phonological forms — typically with at most one open slot. These utterance schemas are not tightly interrelated or organized into A GRAMMAR which interrelated them in terms of abstract categories, parameters, or deep structures; they simply constitute an inventory of things the child can say in particular communicative circumstances ("islands of organization in an otherwise unorganized grammatical system"). In the account of Tomasello (1992, 1999) these individual "constructional islands" (holophrases, verb islands, and other item-based constructions) constitute the totality of the child's early linguistic competence. Early utterance schemas in English that children both comprehend and produce are such things as: *Where's the X?, I wanna X, Hi, Bye-bye, More X, It's a X, I'm X-ing it, Hello, Thank-you, Put X here, Mommy's X-ing it, Let's X it, Throw X, X gone, I X-ed it, Sit on the X, Open X, X here, There's a X, There-ya-go, No, Yes, X broken,* and so forth and so on.

Like utterances, utterance constituents also may form schemas, in this case called **constituent schemas** — which are also subject to entrenchment effects and which may also have slots subject to the same productivity principles as utterance schemas (e.g., *a* __, *the* __, *my/your* __, *must* __, *in the* __). (For current purposes, a partcular word in the lexicon, e.g., *ball*, is also considered a consituent schema; substitute the word construction for schema, if you like.)

To summarize, children are consistently and persistently:

- Storing utterances and their functions.
- Extracting functionally coherent constituents from those utterances
- Abstracting across stored utterances and stored constituents to form utterance schemas and constituent schemas with open slots.

Empirical research has established that initially these are narrowly defined schemas based on specific phonological content, and the child does not relate these to one another structurally. They are thus constructional islands, and the child's early grammatical competence is best thought of as a relatively unstructured inventory of these (see below). Obviously, there are major cross-linguistic differences in the nature of the process. One major factor may be the role of "local cues" (mainly, morphology on single words) as opposed to more dispersed structural cues (word order, some forms of agreement, etc.) to signal grammatical relations in these constructions (Slobin 1982, 1985).

3. Cutting and pasting schemas into utterances

So, on the basis of the utterances they hear from adults (which embody in a large percentage of cases utterance schemas with a fair amount of phonological consistency used repeatedly; see below), children are acquiring both utterance schemas (with and without slots) and constituent schemas (with and without slots). Once they have acquired some of these, children have three basic options for producing an utterance.

(1) First, they might retrieve a functionally appropriate utterance schema containing no slots and just say it as they have heard it said (and probably said it before themselves). For example, they say "Up!" or "There-ya-go".

(2) Second, they might retrieve an utterance schema and simultaneously "tweak" it to fit the current communicative situation. The basic ways they can do this are as follows (see below for specifics):

– filling a new constituent into a slot in the utterance schema [e.g., *Where's-the __* and *ball* combine to make "Where's-the ball?"];
– adding a new constituent onto the beginning or end of the utterance schema[e.g., *Throw it* and *here* combine to make "Throw it here"]; and
– inserting a new constituent into the middle of an utterance schema (e.g. the way a German child might insert *auch* [too] into a schema position where nothing had ever before appeared).[2]

(3) The third way a child might produce an utterance is by combining constituent schemas without using an utterance schema. In the absence of an utterance schema (and obviously without "rules" — since, on this account, rules as traditionally conceived do not exist), they presumably do this on the basis of various kinds of pragmatic principles governing the ordering of old and new information, etc. — although there is very little relevant research here. The idea is very close to Braine's (1976) "groping patterns".

These processes of utterance production may be called **usage-based syntactic operations**, because the child does not begin with words and morphemes and glue them together with "rules". Rather, the child starts with already constructed pieces of language of various shapes, sizes, and degrees of abstraction (and whose internal complexities she may control to varying degrees), and then cuts and pastes these together in a way appropriate to the current communicative situation.

It is important to note that to cut and paste effectively a speaker is always making sure that the functions of the various pieces fit together functionally in the intended manner — this is not the mindless process this metaphor might suggest to some. These processes may also work at the level of utterance constituents and their internal structure. Evidence for these processes — and their combinations — is presented below.

There may be differences in the way individual children use these three strategies, but it is likely that all children early in development learn at least some utterance schemas as production units that allow them to fluently and efficiently express entire communicative intentions (at the very least such things as *Thank-you* and *Bye-bye*, but also many frequent utterance schemas like *I wanna X*, *Where's-the X?*, or similar things.). Utterance schemas are thus, to repeat, a major, if not the major, target of children's early language-learning efforts — a major way-station on the way to more adult-like linguistic competence. Constituent schemas are also major targets of children's early language-learning efforts as they enable the fluent expression of recurrent utterance sub-functions flexibly in novel utterances.

There are clearly important cross-linguistic differences in how all of this works. But in some cases it is simply a case of children working with bound morphemes rather than words, in many of the same ways outlined above. (That is to say, there is no difference in principle between an English-speaking child learning an utterance schema with 3 adult words and an open slot, and an Inuktitut child learning an utterance schema with one free morpheme, two bound morphemes, and an open morpheme slot.) However, one major issue is that the role of constituent order (e.g., word order) in utterance schemas may be different in typologically different languages, and so beginning with English (as we do empirically) biases us towards emphasizing the role of order. Ongoing cross-linguistic work should help to alleviate this concern (see below).

After children have mastered a number of utterance schemas as local constructional islands, and a number of constituent schemas as potential utterance constituents that can be used flexibly in different utterances schemas and in combination among themselves, the major processes leading to more adult-like grammatical competence are:

- Abstracting across utterance schemas to create more abstract linguistic constructions such as, in English, the simple transitive, the simple intransitive, the resultative, the ditransitive, the passive, etc., (or perhaps more narrowly defined constructions such as, e.g., the *get* and *be* passive) — which go along with plenty of more item-based constructional islands that persist into adulthood (naturally, the more high frequency utterance and constituent schemas);

- Abstracting across constituent schemas to create more abstract phrasal constructions and lexical categories such as NPs, Ns, VPs, Vs, PPs, Ps, APs, As — and perhaps some "functional categories" such as determiners and TAM marking as well, although many high frequency words and morphemes are probably learned and used as individual items (not as members of clases); and
- Coordinating utterance schemas and constituent schemas in creative ways in complex constructions and discourse (e.g., sentential complements, adverbials, coordinations, and various narrative—discourse structures).

4. Studies

I will now relate some of our recent research relevant to this proposed usage-based theory (and a few relevant studies of some other researchers).

4.1 The utterances children hear

Cameron-Faulkner, Lieven, and Tomasello (submitted) examined the child-directed speech (CDS) of twelve English-speaking mothers and analyzed it in two ways, both of them from the perspective of the utterance-level constructions mothers used. In the first analysis, mothers' utterances were categorized in terms of general constructional categories such as Wh- Questions, Copula Constructions, Transitive Constructions, etc. In the second analysis, maternal utterances within these categories were further specified in terms of the initial word or words that framed the utterance, such item-based frames as *Are you* . . ., *I'll* . . ., *It's* . . ., *Let's* . . ., *What did* . . ., etc. The main findings were:

- these 12 English-speaking two-year-olds heard about 5–7,000 utterances per day;
- overall, only about 15% of all maternal utterances had SVO form (most utterances were questions (32%), imperatives (9%), copulas (15%), and sentence fragments(20%));
- more than half of all maternal utterances began with one of 52 highly frequent item-based frames (more than an estimated 40 times per day for more than half the children), mostly consisting of 2 words or morphemes [45% began with one of just 17 words: *What* (8.6%), *That* (5.3%), *It* (4.2%), *You* (3.1%), *Are/Aren't* (3.0%), *I* (2.9%), *Do/Does/Did/Don't* (2.9%), *Is* (2.3%), *Shall* (2.1%), *A* (1.7%), *Can/Can't* (1.7%), *Where* (1.6%), *There* (1.5%), *Who* (1.4%), *Come* (1.0%), *Look* (1.0%), and *Let's* (1.0%)]; and
- children used many of these same item-based frames, in some cases at a rate that correlated highly with their own mother's frequency of use.

These results clearly demonstrate that many of, indeed the majority of, the utterances children hear are grounded in highly repetitive lexical frames that they hear dozens, in some cases hundreds, of times every day.

4.2 The item-based nature of early language

It has now been demonstrated beyond a reasonable doubt that young children's early syntactic constructions are highly concrete, that is to say, organized around individual lexical items or phrases. Following in the pioneering work of Braine (1976) in particular (see also MacWhinney 1978), Tomasello (1992), Lieven, Pine, and Baldwin (1997), and others have shown that English-speaking children use their early words and phrases in highly restricted syntactic contexts with development to more adult-like usage proceeding only gradually and in peicemeal fashion — suggesting that early grammatical organization takes the form of an inventory of item-based constructional islands ('limited scope formulae').

A number of systematic studies of children learning languages other than English have also found early item-based organization. For example, in a study of young Italian-speaking children Pizzuto and Caselli (1992, 1994) found that of the six possible person-number forms for each verb in the present tense, about half of all verbs were used in one form only, and an additional 40% were used with 2 or 3 forms. Of the 10% of verbs that appeared in 4 or more forms, approximately half were highly frequent, highly irregular forms that could only have been learned by rote — not by application of an abstract schema or rule. In a similar study of one child learning to speak Brazilian Portugese, Rubino and Pine (1998) found adult-like subject-verb agreement patterns only for the parts of the verb paradigm that appeared with high frequency in adult language (e.g., first-person singular), not for low frequency parts of the paradigm (e.g., third-person plural). It is thus clear that Romance-speaking children do not master the whole verb paradigm for all their verbs at once, but rather they only master some endings with some verbs — and often different ones with different verbs. (For additional findings of this same type, see Serrat 1997, for Catalan; Behrens 1998, for Dutch; Allen 1996, for Inuktitut; Gathercole, Sebastián, and Soto 1999, for Spanish; Kauppinin 1999, for Finnish; Stoll 1998, for Russian; and Berman and Armon-Lotem 1995, for Hebrew.)

Experiments using novel verbs have also found that young children's early productivity with syntactic constructions is highly limited. For example, Tomasello and Brooks (1998) exposed 2- to 3-year-old children to a novel verb used to refer to a highly transitive and novel action in which an agent was doing something to a patient. In the key condition the novel verb was used in an intransitive sentence frame such as *The sock is tamming* (to refer to a situation in which, for example, a bear was doing something that caused a sock to "tam" — similar to the verb *roll* or *spin*). Then, with novel characters performing the target action, the adult asked children the question: *What is the doggie doing?* (when the dog was causing some new character to tam). Agent questions of this type encourage a transitive reply such as *He's tamming the car* — which would be creative since the child has heard this verb only in an intransitive sentence frame. The outcome was that very few children produced a transitive utterance with the novel verb, and in another study they were quite poor at two tests of comprehension as well (Akhtar and Tomasello 1997). As a control, children also heard another novel verb introduced in a transitive sentence frame, and in this case virtually all of them produced a transitive utterance — demonstrating that they could use novel verbs in the transitive construction when they heard them used in that way. Moreover, 4- to 5-year-old children are quite good at using novel verbs in transitive utterances creatively, demonstrating that once they have acquired more abstract linguistic skills children are perfectly competent in these tasks (Pinker *et al.* 1987; Maratsos *et al.* 1987; see Tomasello 2000, for a review).

In an interesting variation on this theme, Akhtar (1999) found that if 2.5- to 3.5-year-old children heard such things as *The bird the bus meeked*, when given new toys they quite often repeated the pattern and said such things as *The bear the cow meeked* — only consistently correcting to canonical English word order at 4.5 years of age. This behavior is consistent with the view that when 2- to 3-year-olds are learning about *meeking* they are just learning about *meeking*; they do not assimilate this newly learned verb to some more abstract, verb-general linguistic category or construction that would license a canonical English transitive utterance (see Smith, Lieven, and Tomasello, submitted, for further support). It should also be noted that syntactic overgeneralization errors such as *Don't fall me down* — which might be seen as evidence of more general and categorical syntactic knowledge — are almost never produced before about 2.5 to 3 years of age (see Pinker 1989).

Interestingly, and perhaps surprisingly, this same item-based approach is quite revealing in the case of many of children's more complex constructions as well. For example, Dabrowska (in press) looked in detail at one child's earliest uses of Wh-questions in English. Her most general finding was that 83 percent of this child's questions during her third year of life came from one of just 20 formulas such as *Where's THING? Where THING go? Can I PROCESS? Is it PROPERTY?* and so forth. Relatedly, Rowland and Pine (2000) attempted to explain why English-speaking children sometimes invert the subject and auxiliary in Wh- questions and sometimes not — leading to errors such as *Why they're not going?* What they found was that the child they studied from age 2 to 4 consistently inverted or failed to invert particular auxiliaries occurring with particular Wh-words. She thus consistently said such incorrect things as *Why I can . . .? What she will . . .? What you* *can . . .?*, but at the same time she also said such correct things as *How did . . .? How do . . .? What do . . .?* In all, of the 50 particular Wh-word–auxiliary pairs this child produced, 47 of them were produced either 100% correctly or 100% incorrectly. Both of these studies of children's questions thus show again the item-based nature of children's early constructions, in this case for a set of constructions that develop well into the preschool years.

The general conclusion is clear. In the early stages, children mostly use language the way they have heard adults using it; they learn via imitation, where imitation is not characterized as blind mimicry but as reproducing the same behavior for the same purpose (one form of "cultural learning"; Tomasello, Kruger, and Ratner 1993). This leads to an inventory of item-based utterance schemas, with perhaps some slots in them built up through observed type variation in that utterance position (see below). The reason that children do not operate with more abstract linguistic categories and schemas is quite simply because they have not yet had sufficient linguistic experience in particular usage events (enough of the right kind of type variation) to construct these adult-like linguistic abstractions.

4.3 The process of abstraction

The only experimental study of children's construction of an abstract constructional schema is by Childers and Tomasello (2001). They investigated the linguistic skills and representations underlying English-speaking 2 1/2-year-olds' production of transitive utterances such as *He's kicking it.* The main study was a training study in which children heard several hundred transitive utterances in three separate sessions. Half the children learned new English verbs (and so increased their transitive verb vocabularies during training) whereas the other half heard only verbs they already knew. Within these groups, some children heard all of the utterances with full nouns as agent and patient, whereas others heard utterances with both pronouns (i.e., *He's VERB-ing it*) and also full nouns as agent and patient. They were then tested to see if they could creatively produce a transitive utterance with a nonce verb.

Children were best at generalizing the transitive construction to the nonce verb if they had been trained with pronouns and nouns, regardless of the familiarity of the trained verbs. That is, the consistent pronoun frame *He's VERB-ing it* seemed to facilitate children's formation of a verb-general transitive schema to a greater degree than the learning of additional transitive verbs alone, in the absence of such a stabilzing frame. This suggests that children construct their early abstract constructions out of both (i) particular lexical or morphological items and patterns, and (ii) observed type variation (with some functional consistency) in particular utterance constituents. A posible graphic depiction of

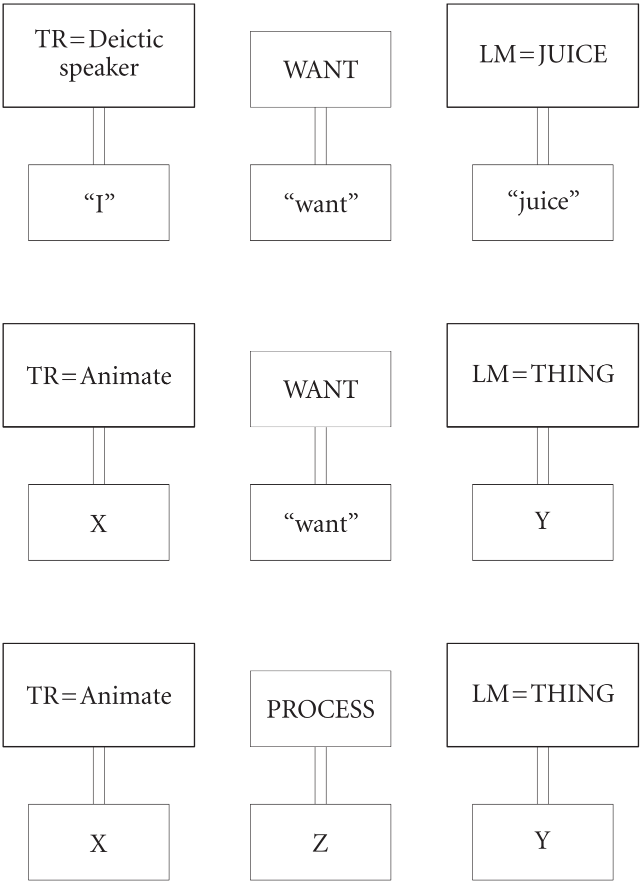


Figure 1. Three levels of schematicity in the child’s constructional schemas. (TR=trajector; LM=landmark)

the process may be seen in Figure 1 (based on Dabrowsa, in press, who based hers on Langacker 1987).

4.4 The process of constraint

As constructions become more abstract, their generalizing tendencies must also be constrained; all verbs cannot be used in all constructions (see Pinker 1989). One hypothesis is that they do this by becoming more entrenched and resitant to flexible use (see Bybee 1995). In the only experimental investigation of this process in the domain of early syntax, Brooks, Tomasello, Lewis, and Dodson (1999) modeled the use of a number of fixed-transitivity English verbs for children from

3;5 to 8;0 years — verbs such as *disappear* that are exclusively intransitive and verbs such as *hit* that are exclusively transitive. There were four pairs of verbs, one member of each pair typically learned early by children and used often by adults (and so presumably more entrenched) and one member of each pair typically learned later by children and used less frequently by adults (less entrenched). The four pairs were: *come-arrive, take-remove, hit-strike, disappear-vanish* (the first member of each pair being more entrenched). The finding was that, in the face of adult questions attempting to induce them to overgeneralize, children of all ages were less likely to overgeneralize the strongly entrenched verbs than the weakly entrenched verbs; that is, they were more likely to produce *I vanished it* than *I disappeared it.* This finding suggests not only that children say what they hear but that the more they hear it the more it seems to them that this is the only way it can be said.

Brooks and Tomasello (1999) demonstrated that the alternative constructions children have available to meet discourse demands (e.g., in the above example, the availability of *I made it disappear*) also play an important role in constraining abstract constructions. Children are less likely to overgeneralize if they have a good alternative. In addition, after four years of age, children also rely on generalizations about which semantic classes of verbs (as identified by Pinker 1989) participate in which abstract constructions.

4.5 Usage-based syntactic operations

Given that children are creating linguistic schemas of various shapes and sizes and degrees of abstraction throughout early development (i.e., building their linguistic inventories), we may now ask about their ability to put these schemas together creatively in order to adapt to the exigencies of particular usage events. Lieven, Tomasello, Behrens, and Speares (in press) addressed this issue in a naturalistic study of one 2-year-old child learning English. The novelty was that this child's language was recorded using extremely dense taping intervals — 5 hours per week for 6 weeks, roughly 5 to 10 times denser than most existing databases of child language and accounting for approximately 8 to 10% of all of the child's utterances during this period. In order to investigate this child's constructional creativity, all of her 500+ utterances produced during the last one-hour taping session at the end of the 6 week period were designated as target utterances (300+ multi-word utterances). Then, for each target utterance, there was a search for 'similar' utterances produced by the child (not the mother) in the previous 6 weeks of taping. The main goal was thus to determine for each utterance recorded on the final day of the study what kinds of syntactic operations were necessary for its production, that is to say, in what ways did the child have to modify things she had previously said (her 'stored linguistic experience') to produce the thing she was now saying. We may call these operations ‘usage-based syntactic operations’ since they explicitly take into account that the child does not put together each of her utterances from scratch, morpheme by morpheme, but rather, she puts together her utterances from a motley assortment of different kinds of pre-existing psycholinguistic units.

And so, following the usage-based models of Bybee (1995), Langacker (2000), and Croft (2000), the question was how this child was able to “cut and paste” together her previously constructed linguistic schemas (mostly item-based in one way or another) in order to create a novel utterance in a specific usage event. What was found by this procedure was that:

- the majority of the 300+ multi-word utterances the child produced on this day were things she had previosuly said before during the six week taping period (approximately two-thirds);
- of the remaining third of the multi-word utterances, the majority consisted of repetitions of an established utterance schema plus other linguistic material ‘filled in’ to a slot or ‘added on’ to the beginning or end. For example, the child had said many hundreds of times previously *Where’s the* __?, and on the target tape she produced the novel utterance *Where’s the butter?* Utterances schemas that the child used on the last day of the study had been used by the child, on average, an estimated 200–400 times during the previous six weeks.
- a small minority of multi-word utterances on the last tape (3%) differed from things this child had said before in more than one way. These mostly involved the combination of ‘filling in’ and ‘adding on’ to an established utterance

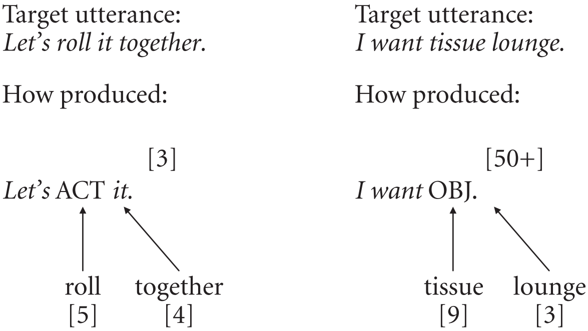


Figure 2. Two examples of cut-and-paste operations. Utterances schemas are in italics, constituent schemas are in regular type, previous frequencies are in brackets, functional characterizations are in CAPS, and the arrows represent ‘filling in’ (vertical) and ‘adding on’ (slanted) usage-based syntactic operations

> schema, but there were 3 utterances that seemed to be novel combinations of utterance constituents only.

It is important to note that there was also very high functional consistency across different uses of this child's utterance schemas, that is, the child filled a given slot with basically the same kind or kinds of linguistic items or phrases across the entire six week period of the study. Figure 2 displays some examples of creative utterances of different types and their hypothesized prescedents from the child's already mastered utterance schemas and consituents.

The overall picture, then, is that young two-year-old English-speaking children most often produce multi-word utterances that they have previously produced before (two-thirds of the time), but they quite often produce creative utterances as well, mostly (97% of the time) on the basis of their already well-established item-based utterance and constituent schemas. To try to expand our picture of the process, we are currently collecting similar dense data bases in German (with Behrens), Estonian (with Vihman), and Polish (with Dabrowska).

4.6 Complex utterances

Some version of this 'cut-and-paste' approach is also used by young children to construct some of their more complex utterances a bit later in their development. For example, Diessel and Tomasello (in press) looked at young English-speaking children's earliest utterances with sentential complements from 2 to 5 years of age. They found that virtually all of them were composed of a simple sentence schema that the child had already mastered combined with one of a delimited set of complement-taking matrix verbs (see also Bloom 1992). These matrix verbs were of two types. First were epistemic verbs such as *think* and *know*. As one example, in almost all cases children used *I think* to indicate their own uncertainty about something, and they basically never used the verb *think* in anything but this first-person, present tense form; that is, there were virtually no examples of *he thinks, she thinks* . . ., etc., virtually no examples of *I don't think* . . ., *I can't think* . . ., etc., and virtually no examples of *I thought* . . ., *I didn't think* . . ., etc. And there were almost no uses with a complementizer (virtually no examples of *I think that*). It thus appears that for many young children *I think* is a relatively fixed phrase meaning something like *maybe*. The child then pieces together this fixed phrase (or one of the other similar phrases like *I hope* . . . *I bet* . . ., etc.) with a full proposition, with its function being as a sort of evidential marker (not as a matrix clause that embeds another as in traditional analyses). The second kind of matrix verbs were attention-getting verbs like *look* and *see*, used in conjunction with full finite clauses. In this case, children used these matrix verbs almost exclusively in imperative form (again almost no negations, no non-present tenses, no complementizers), suggesting again an item-based approach not involving syntactic embedding. Thus, when examined closely, children's earliest complex sentences look much less like adult sentential complements (which are used most often in written discourse) and much more like various kinds of 'pastiches' of various kinds of established item-based constructions (see Hopper 1998).

Similarly, Diessel and Tomasello (in press) investigated young English-speaking children's earliest relative clauses. Surprisingly, these did not turn out to be what are often thought of as prototypical relative clauses used to restrict referents with all kinds of nominals. Instead, virtually all of children's earliest relative clauses had as the main clause a presentational construction with a copular verb, for example, *it's a . . .*, *here are the . . .*, *there's a . . .*, *where's the . . .?*, etc. The relative clause then served to provide new information about the predicate nominal (object). (See Fox and Thompson 1990, and Lambrecht 1988, for some similar analyses of many relative clauses in informal adult conversation.) This led to such utterances as "That's the toy I found", "Here's the cup that's empty", and so forth. Unlike in adult written discourse, and most experiments with young children, there were virtually no relative clauses modifying subjects. The main point in the current context is that even this very complex construction is firmly based in a set of simpler constructions (copular presentationals) that children have mastered as item-based constructions some time before relative clauses are first acquired and produced.

4.7 Summary

One way of thinking about all of this (following Dabrowska, in press) is to think of children's early utterance schemas as "Big Words". This means that many of the same hypotheses and theoretical options that exist in the literature on the acquisition of morphology also apply in the acquisition of syntax. In this context, the current proposal is fully consonant with Single Process Theory, as espoused by many different linguists and psycholinguists (e.g., Bybee 1995; Bates and MacWhinney 1989).

The alternative option is Dual Process Theory, which divides the linguistic world into two indpendent parts: (1) all of the idiosyncratic aspects of language that must be learned by normal learning methods (the linguistic periphery), and (2) all of the regular aspects of language that don't need to be learned in the normal way because they are assimilated to a rule (the linguistic core, which hooks up in some mysterious way with an hypothesized innate universal grammar; e.g., Chomsky 1980; Pinker 1999; Clahsen 2000). Dual Process theorists typically do not believe that the kind of bottom-up approach I have outlined here can ever account for the rule-based aspects of human linguistic competence — the skepticism being based on various specious philosophical arguments about the inadequacies of induction. My view, however, is that Dual Process theorists have a false view of the nature of human linguistic competence that serves as the endpoint of child language acquisition. Recent research has revealed that much of adult language is based in specific linguistic items and schemas as well, and many of these item-based constructions are fully productive — which makes the line between core and periphery very difficult to draw (e.g., see Fillmore *et al.* 1988 on the *let alone* construction; Michaelis and Lambrecht 1999, on the nominal extraposition construction; Kay and Fillmore 1999 on the WXDY construction; and many others). Indeed, a number of prominent formal theorists have recently come to the conclusion, based on detailed analyses of particular linguistic constructions, that it is basically impossible to make a principled distinction between the linguistic 'core' and 'periphery' (e.g., Jackendoff 1996; Cullicover 1998).

5. Conclusion

I am not sure precisely what should be called pre-language and what should be called language. If reliance on single symbol utterances — holophrases — is pre-language, then this would seem to be a common initial step in the evolution of language and the ontogeny of language. But children do not have to create more complex constructions *de novo*; they hear thousands of multi-word utterances every day and so 'only' must learn them. Our holophrastic ancestors did not have the benefit of this input, and so had to create from discourse sequences multi-word utterances and constructional patterns. Children have to make abstractions. The process of abstraction is not part of the external collectivity of linguistic communication in which children hear utterances, however, but rather it is an individual psychological process and so part of cognitive ontogeny — which is a crucial component in the historical emergence of new linguistic structures.

The actual events that led to the emergence of language historically are gone and can never be observed. But some of the processes that were at work in this evolutionary event are still at work in the ontogeny of language in young children. Much can be gained by looking at these processes in detail and by comparing them to those that might have occurred when language began.

Notes

1. The terms communicative intention and communicative function are thus corelative. A person expresses a communicative intention in a phonological form (or other physical

medium), and that phonological form then possesses a communicative function , viz., of expressing the underlying communicative intention.

2. Another possibility is rearranging elements, but that occurs very infrequently in English.

References

Akhtar, N. and Tomasello, M. 1997. "Young children's productivity with word order and verb morphology". *Developmental Psychology* 33: 952–65.

Akhtar, N. 1999. "Acquiring basic word order: Evidence for data-driven learning of syntactic structure". *Journal of Child Language* 26: 339–56.

Allen, S. 1996. *Aspects of Argument Structure Acquisition in Inuktitut.* Amsterdam: John Benjamins.

Bates, E., and MacWhinney, B. 1982. "Functionalist approaches to grammar". In *Language Acquisition: State of the Art*, E. Wanner and L. Gleitman (eds), Cambridge: Cambridge University Press.

Bates, E., and MacWhinney, B. 1989. "Functionalism and the competition model". *The Crosslinguistic Study of Sentence Processing*, In B. MacWhinney and E. Bates (eds). Cambridge: Cambridge University Press.

Behrens, Heike 1998. Where does the information go? Paper presented at MPI workshop on argument structure. Nijmegen.

Berman, R. A., and Armon-Lotem, S. 1995. How grammatical are early verbs? Paper presented at the *Colloque International de Besançon sur l'Acquisition de la Syntaxe.* Besançon.

Bloom, L. 1992. *Language Development from Two to Three.* Cambridge: Cambridge University Press.

Braine, M. 1976. "Children's first word combinations". *Monographs of the Society for Research in Child Development* 41: no. 1.

Brooks, P., Tomasello, M. 1999. "Young children learn to produce passives with nonce verbs". *Developmental Psychology* 35: 29–44.

Brooks, P., Tomasello, M., Lewis, L., and Dodson, K. 1999. "Young children's overgeneralizations with fixed transitivity verbs". *Child Development* 70: 1325–37.

Bybee, J. 1998. "The emergent lexicon". *Proceedings of the Chicago Linguistics Society* 34: 421–35.

Bybee, J. 1995. "Regular morphology and the lexicon". *Language and Cognitive Processes* 10: 425–55.

Bybee, J. 1985. *Morphology: A Study of the Relation Between Meaning and Form.* Amsterdam: John Benjamins.

Cameron-Faulkner, T. , Lieven, E. and Tomasello, M. Submitted. "A construction based analysis of child directed speech".

Childers, J. B. and Tomasello, M. 2001. "The role of pronouns in young children's acquisition of the English transitive construction". *Developmental Psychology* 37: 739–48.

Chomsky, N. 1980. "Rules and representations". *Behavioral and Brain Sciences* 3: 1–61.

Clahsen, H. 2000. "Lexical entries and rules of language: A multidisciplinary study of German inflection". *Behavioral and Brain Sciences* 22 991–1060.

Clark, E. 1996. “Early verbs, event types, and inflections”. In *Children’s Language*, Volume IX, C. Johnson and J. Gilbert (eds). Mahwah, NJ: Lawrence Erlbaum Associates.

Croft, W. 2000. *Explaining Language Change: An Evolutionary Approach.* London: Longman.

Cullicover, P. 1998. *Syntactic Nuts.* Oxford: Oxford University Press.

Dabrowska, E. In press. “Formulaicity and children’s early questions”. *Cognitive Linguistics.*

Diessel, H. and Tomasello, M. In press a. “The acquisition of relative constructions in spontaneous child speech”. *Cognitive Linguistics*

Diessel, H. and Tomasello, M. in press b. “The acquisition of finite complement clauses in English: A corpus-based analysis”. *Cognitive Linguistics.*

Goldberg, A. 1995. *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago: University of Chicago Press.

Fillmore, C., Kaye, P., and O’Conner, M. 1988. “Regularity and idiomaticity in grammatical constructions: The case of let alone”. *Language* 64: 501–38.

Fox, B. and Thompson, S. 1990. “A discourse explanation of ‘The Grammar’ of relative clauses in English conversation”. *Language* 66: 297–316.

Gathercole, V., Sebastián, E., and Soto, P. 1999. “The early acquisition of Spanish verbal morphology: Across-the-board or piecemeal knowledge?” *International Journal of Bilingualism* 3: 133–82.

Givón, T. 1995. *Functionalism and Grammar.* Amsterdam: John Benjamins.

Hopper, P. 1998. “Emergent grammar”. In *The New Psychology of Language: Cognitive and Functional Approaches to Lanuage Structure*, M. Tomasello (ed). Hillsdale, NJ: Lawrence Erlbaum Associates.

Jackendoff, R. 1996. “Twistin’ the night away”. *Language* 73: 534–59.

Kay, P. and Fillmore, C. 1999. “Grammatical constructions and linguistic generalizations”. *Language* 75: 1–33.

Lambrecht, K. 1988. “There was a farmer had a dog: Syntactic amalgams revisited”. *Berkeley Linguistics Society* 14: 319–39.

Langacker, R. 1987a. *Foundations of Cognitive Grammar,* Volume I. Stanford, Calif: Stanford University Press.

Langacker, R. 1987b. “Nouns and verbs”. *Language* 63: 53–94.

Langacker, R. 2000. “A dynamic usage-based model”. In *Usage-Based Models of Language*, M. Barlow and S. Kemmerer (eds). Stanford, Calif.: SLI Publications.

Lieven, E., Tomasello, M. , Behrens, H., and Speares, J. In press. “Item-based creativity”. *Journal of Child Language.*

Lieven, E., Pine, J., and Baldwin, G. 1997. “Lexically-based learning and early grammatical development”. *Journal of Child Language* 24: 187–220.

MacWhinney, B. 1978. “The acquisition of morphophonology”. *Monographs of the Society for Research in Child Development* 43.

Maratsos, M., Gudeman, R., Gerard-Ngo, P., and DeHart, G. 1987. “A study in novel word learning: The productivity of the causative”. In *Mechanisms of Language Acquisition*, B. MacWhinney (ed). Hillsdale, NJ: Lawrence Erlbaum Associates.

Michaelis, L. and Lambrecht, K. 1996. “Toward a construction-based theory of language function: The case of nominal extraposition”. *Language* 72: 215–47.

Pinker, S. 1989. *Learnability and Cognition: The Acquisition of Verb–Argument Structure.* Cambridge, Mass.: Harvard University Press.

Pinker, S. 1999. *Words and Rules.* New York: Basic Books.

Pinker, S., Lebeaux, D. S., Frost, L. A. 1987. "Productivity and constraints in the acquisition of the passive". *Cognition* 26: 195–267.

Pizutto, E. and Caselli, C. 1994. "The acquisition of Italian verb morphology in a cross-linguistic perspective". In *Other Children, Other Languages*, Y. Levy (ed). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pizutto, E. and Caselli, C. 1992. "The acquisition of Italian morphology". *Journal of Child Language* 19: 491–557.

Rowland, C. and Pine, J. 2000. "Subject-auxiliary inversion errors and *wh*-question acquisition: What do children know?" *Journal of Child Language* 27: 157–82.

Rubino, R. and Pine, J. 1998. "Subject-verb agreement in Brazilian Portuguese: What low error rates hide". *Journal of Child Language* 25: 35–60.

Serrat, E. 1997. *Acquisition of verb category in Catalan.* Dissertation, University of Barcelona.

Slobin, D. 1985. *The Cross-Linguistic Study of Language Acquisition.* Hillsdale, N J: Lawrence Erlbaum Associates.

Slobin, D. 1982. "Universal and particular in the acquisition of language". In *Language Acquisition: The State of the Art*, L. Gleitman and E. Wanner (eds). Cambridge: Cambridge University Press.

Slobin, D. 1973. "Cognitive prerequisites for the development of grammar". In *Studies of Child Language Development*, D. Slobin (ed). New York: Holt, Rinehart, Winston.

Smith, K., Lieven, E., and Tomasello, M. Submitted. "What children do and do not do with ungrammatical word orders".

Stoll, S. 1998. "The acquisition of Russian aspect". *First Language* 18: 351–78.

Tomasello, Michael 2000. "Do young children have adult syntactic competence?" *Cognition* 74: 209–53.

Tomasello, M. 1987. "Learning to use prepositions: A case study". *Journal of Child Language* 14: 79–98.

Tomasello, M. 1992. *First Verbs: A Case Study in Early Grammatical Development.* Cambridge: Cambridge University Press.

Tomasello, M. 1995. "Pragmatic contexts for early verb learning". In *Beyond Names for Things: Young Children's Acquisition of Verbs*, M. Tomasello and W. Merriman (eds). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tomasello, M. 1999. *The Cultural Origins of Human Cognition.* Cambridge, Mass.: Harvard University Press.

Tomasello, M., Kruger, A. C., and Ratner, H. H. 1993. "Cultural learning". *Behavioral and Brain Sciences* 16: 495–552.

CHAPTER 13

Why does exposure to language matter?

Jill P. Morford
University of New Mexico

The goal of this chapter is to explore why exposure to language is important for language learners, first from an ontogenetic perspective, and subsequently with respect to language origins. Over the past 40 years the idea that some type of language structure may be innate has played a dominant role in much thinking about language acquisition. Consequently, research on the role of language exposure in the acquisition process has focused primarily on how language input affects (or doesn't affect) language structure. On the one hand, exposure is argued to influence only minor details of language structure that is essentially innate — details such as setting parameters, triggering default switches, and selecting between structural options (e.g., Radford 1994; Roeper 1998; Roeper and Williams 1987). Alternatively, it has been argued that characteristics of language input such as salience and frequency can account for the order and type of language structure in children's speech (e.g., Bates and MacWhinney 1987; Rumelhart and McClelland 1986; Tardif, Shatz and Naigles 1997). While this debate has been fruitful, it has focused our questions on one issue — the mastery of structure. An issue that has received less attention is how children learn to process language, and what the role of language exposure is in that process.

This chapter explores the notion that the consequences of language exposure for language processing may be ultimately more significant to language outcomes, both ontogenetically and historically, than the consequences of language exposure for the mastery of language structure. This is not to say that input does not influence language structure. It surely does. However, as the studies reviewed here will demonstrate, children structure their communication whether or not they are exposed to language input. By contrast, they do not exhibit rapid and efficient language processing abilities without exposure from an early age. This distinction has important implications for language origins. Certain types of grammatical structures arise historically as a result of the automation of language processing. This set of facts suggests that we can and should distinguish between two types of grammatical properties. One set of grammatical properties, referred to as *system-internal* grammatical properties, concern the systematicity of structures within a single grammar, and can emerge through innovation in an individual who has not been exposed to language. A second set of grammatical properties, referred to as *processing-dependent* grammatical properties, emerge through language use and are the direct result of the automation of language processing. Processing-dependent grammatical properties cannot emerge without language exposure from an early age, and thus, could only have emerged after a community of hominids shared a structured communication system to which their offspring were exposed.

The chapter is divided into four sections. The first section describes cases of language innovation by deaf children of hearing parents and demonstrates that language structure emerges even when there is very little structure in the input. The second section describes the grammaticization of innovated structure when a community of former linguistic isolates was formed in Nicaragua. The third section of the chapter outlines studies of adolescent first language learners showing that language processing skills are not efficient when language exposure is delayed. In the final section, implications for theories of language origins are addressed.

1. The forbidden experiment: children deprived of exposure to language

Even in ancient and medieval times, a number of curious monarchs were interested in questions about original languages and undertook what today is considered "The Forbidden Experiment", that is, they purposely deprived infants of exposure to language. Despite some disastrous results — at least from the perspective of the unfortunate research subjects — these experiments persisted across the ages (Campbell and Grieve 1982). Consider just one such experiment, purportedly carried out in 1500 A.D. by Akbar, King of Hindustan. In his version of the experiment, the research subjects were sent to live in an asylum, and were cared for by deaf wetnurses, who obviously, from Akbar's perspective, could not use language with the infants.[1] Curiously enough, although these infants did not learn to speak, they did thrive, and became quite skilled at communicating their needs by using their hands (see Bonvillian *et al.* 1997 for a fuller account of this experiment). Although it is nearly impossible to determine whether the account is true or not, it is exactly what we would expect from our current understanding of language as a modality-free phenomenon. Children will not innovate a spoken language without any exposure to speech, particularly when they are exposed to a gestural model of language.

An opportunity to address these issues using more modern experimental methods arose in the 1970s with the discovery of an "experiment of nature", namely, deaf children of hearing parents who were not being exposed to a signed language, but who were communicating with their hands (Goldin-Meadow and

Feldman 1977). These cases are valuable because the experimental subjects are individuals who are loved and cared for by their parents, and because they do not constitute isolated cases. Deaf individuals who grow up without exposure to a signed language typically depend on the use of gesture to communicate. The gesture use of these individuals is noticeably more prolific than in hearing individuals of the same culture, and is commonly referred to as HOME SIGN in American Sign Language. There are studies documenting the use of homesign by deaf individuals living in 13 different countries and studied by as many different investigators (see Morford 1996 for a list). Homesign is used by children under 2 years old and adults as old as 60. Thus, the phenomenon is widespread and easily identified.

Susan Goldin-Meadow and colleagues have produced the most sustained research effort investigating the structural properties of homesign systems used by young children. They have documented structural properties of individual signs, including stability of the lexicon over time, systematicity of the form-meaning relationships of the handshapes and movements used across a sign lexicon, and form-based distinctions of signs playing different grammatical roles. Analyses of gesture strings also provide evidence of structural properties, such that thematic roles can be predicted on the basis of (1) sign order, (2) sign production probabilities, (3) the movement of signs in space, and (4) the way that complex propositions are produced (Goldin-Meadow, this volume). Thus, both within signs and across signs, there is evidence that the gesture production of deaf individuals who have not been exposed to language is not haphazard, but is characterized by organizational properties. Importantly, these properties do not appear in the gesture use of the caretakers of these deaf individuals (Goldin-Meadow and Mylander 1984, 1990).[2] It appears then, that one characteristic of the communication systems of individuals who are not exposed to language is *innovation* of organizational properties that are not in the input.

Signed languages that evolved over multiple generations also exhibit the types of organizational properties that have been documented in homesign systems. They have stable lexicons, and stable form-meaning relationships of handshapes and movements across signs. The arguments of a verb can be determined by the order of signs or by the spatial arrangement of signs. In addition to these properties, they have some grammatical elements that can be traced historically to the lexicon. For example, Janzen (1995) provides an analysis of the completive marker in American Sign Language (ASL), arguing that it entered the language after repeated and frequent use of the verb FINISH was reduced to a marker conjoined to frequent verbs. Likewise, MacFarlane (1998) has demonstrated that an adverbial marker for unexpected events has its source in the adjective WRONG, and Shaffer (2000) has outlined the development of a future gram in ASL from the verb PARTIR in Old French Sign Language. In homesign systems, the organizational properties appear to emerge either on-the-spot, or through an internal analysis of the relationship of signs to each other (Goldin-Meadow, this volume). The gradual movement of lexical items into grammatical roles through repeated use and automation does not characterize the organizational properties of homesign.

2. The emergence of a signed language in Nicaragua: gesture vs. homesign as input

A rare opportunity to observe the emergence of a new signed language was made possible by the social, political and demographic circumstances in Nicaragua in the 1980s. According to Polich (2000), the widespread view that facility in oral language is a prerequisite to societal roles for adults prevented the formation of a deaf community in Nicaragua prior to the 1980s. Deaf persons were viewed as "eternal children". When the Sandinista government was elected in 1979, the first broad efforts to provide special education were undertaken (Kegl and McWhorter 1997). These efforts affected deaf individuals living primarily in or near Managua. There was little support for the government or from the government to establish special education on the Atlantic Coast. Education for the deaf on the Atlantic Coast has primarily come about as the result of efforts by private and religious individuals and foundations, such as the Nicaraguan Sign Language Projects, Inc. based in the U.S., and a Catholic order known as "Las Monhas de Santa Agnes". These private efforts followed the government efforts by about a decade. In the intervening decade, Judy Kegl and her associates were able to document the homesign in usage on the Atlantic Coast, while at the same time investigating the form of communal signing being used in the educational facilities in Managua. Signs were not taught in the Managua schools. These schools merely provided a social context for deaf individuals to interact, and in some cases, to reject the notion that societal roles require facility in oral language.

To describe the range of language use by deaf individuals in Nicaragua, we need to distinguish three groups of deaf individuals. Starting in 1994, Kegl and colleagues began to document the homesign of deaf individuals living in Bluefields, on the Atlantic Coast. These individuals had some prior contact to one another, but without educational support of signed language, and without the social support that comes in the form of the idea that deaf persons can participate in society. Gesture use by this group has been assumed to be most similar to previously documented cases of homesign. A second group includes deaf individuals attending the educational facilities in Managua, but who, according to Senghas (1995), were too old or entered the schools at a time when there was too much variability in the language in use to undergo the typical language acquisition process, and thus, only used a form of signing referred to as a pidgin, or a homesign contact form. A third group, consisting of deaf individuals who were exposed to the contact form of homesign by age seven, are argued to have acquired, indeed, to have generated, a fully grammatical language, Nicaraguan Sign Language (Senghas 1995).

Morford and Kegl (2000) provide a comparison of the organizational properties observed in the gesture use of these three groups. Several features observed in the homesign systems are closely related in form but differ slightly in function relative to the contact form and to Nicaraguan Sign Language. For example, the tendency of homesigners to encode manipulable objects with a gesture showing how the object is used, eventually surfaces in the sign language as a system of what have traditionally been called "classifiers", but are currently being re-categorized as polymorphemic predicates (see, e.g., Schembri 2000a, b). In the homesign used in Nicaragua, a distinction between references to objects themselves vs. references to the manipulation of objects has not been identified.[3] By contrast, this distinction is made in the signed language. A second example is the use of role shift to express the signer's perspective on an event and to encode grammatical relations in the signed language, which appears to have its roots in the tendency of homesigners to act out events from a first-person perspective. In other words, the homesigner often takes on the role first of the agent, and then of the patient, to describe a transitive action. In the signed language, serial verb constructions encode similar sequences of actions, but role shift is used in a consistent manner to distinguish the identity of the participants across a stretch of discourse. As a final example, lexical items that consist of a string of descriptors in homesign undergo a process of compounding that is sensitive to articulatory constraints that can actually interfere with the conceptual basis of the sign. Homesigns for fruits are invariably a sequence in which preparation of the fruit for eating was followed by the sign to eat (e.g., SLICE-OFF-TOP-WITH-MACHETE EAT=PINEAPPLE, PEEL EAT=BANANA, RUB-ON-SHIRT EAT=APPLE). In Nicaraguan Sign Language, the order of these components is reversed for the sign APPLE, even though it makes no sense to polish your apple on your shirt after you've already eaten it. The point of these examples is to suggest that we see a change in the structural properties of the signed language relative to the homesign system, but that these properties are very much linked to the fact that the children who produced the signed language started out with homesign as input rather than gesture. Thus, an important prerequisite to the emergence of Nicaraguan Sign Language was the presence of regular patterns in the signing they were exposed to, even if that signing did not constitute a full language. The process of language emergence in this community is better described as a process of *grammaticization* than of innovation. Homesigners started with inconsistent gesture as input, and innovated structure. Nicaraguan signers started with structured input (i.e., homesign) and grammaticized elements of the input.

What is so different about gesture vs. homesign as input? For one, there are perceptual regularities in homesign that may not be available in gesture. Homesigners draw their handshapes from a small inventory of relatively unmarked shapes (Morford 1999). For example, Goldin-Meadow and colleagues (1995) found that 4 homesigners used a total of 15 different handshapes in a corpus of 3,042 signs (cf. ASL signers draw from an inventory of about 40–45 handshapes). Since these same handshapes are used in the production of many different gestures, an observer has the opportunity to see each handshape produced at a high frequency, but in combination with a variety of different movement and location parameters (Goldin-Meadow *et al.* 1995).

Homesigners also use longer strings of signs than speakers who gesture. McNeill (1992) finds that the typical rate of gesture during speech is one gesture per clause. Thus, it appears that speakers rarely combine gestures into strings. By contrast, homesigners combine gestures in strings, and they combine them in regular orders. Thus, an important difference in having homesign vs. gesture as input in a language learning context is that the statistical regularities of the input are much greater in homesign.[4] Based on recent research showing that infants are highly sensitive to both phonological (Saffran *et al.* 1996) and lexical (Gomez and Gerkin 1999) patterns in language, we can surmise that the developing processing systems of Nicaraguan signers had a basis for drawing generalizations, generating categories, and computing probabilities. In other words, they were not merely learning individual gestures in isolation, focusing on the form-meaning relationship of individual signs. Rather, they were identifying patterns of phonemes within signs, and patterns of signs within utterances, and developing strong expectations about the frequency and co-occurrence of different elements in the input.

3. First language acquisition in adolescence: Why is it so hard?

There is widespread agreement that delaying exposure to language until adolescence has detrimental effects on the acquisition process. Susan Curtiss (1977) has provided the most detailed account of the only well-documented case in which a *hearing* individual was not exposed to language until adolescence. She describes the linguistic development of a girl referred to as "Genie", after 13 years and 7 months of isolation from language and other basic needs. Several investigators have also studied the use of ASL by deaf adults who were first exposed to language in adolescence (e.g., Emmorey, Bellugi, *et al.* 1995; Mayberry and Fischer 1989; Newport 1990). Unlike Genie, the deaf adults who participated in these studies were not the victims of abuse and neglect. They were exposed to spoken English throughout childhood, but did not learn it due to their deafness. Eventually, they had the opportunity to learn ASL, and continued to use it as their primary language in adulthood.

These studies are often construed as providing evidence that grammatical structure cannot be mastered if exposure to language is delayed. However, two considerations about these datasets are important to note. First, these individuals do acquire a few grammatical elements of their respective languages, even though they demonstrate difficulties with many others. For example, Genie learned to use the progressive, and basic word order in English (Curtiss 1977), and signers apparently have no trouble identifying the antecedents of overt and null pronouns (Emmorey and Lillo-Martin 1995) or mastering sign order in three-sign transitive sentences (Newport 1990) in ASL despite delayed exposure to the language. A second consideration is that the majority of data documenting poor grammatical skills have been collected through comprehension tasks. Comprehension tasks require that subjects perceive and decode the grammatical elements of the stimuli prior to making a response about the meaning or grammaticality. In other words, comprehension tasks confound measures of the subjects' grammatical knowledge with measures of the subjects' processing abilities.

An alternative explanation for the performance of adolescent first language learners is that these individuals are not able to process English and ASL as efficiently as native speakers and signers, and that their difficulties with the grammar are a consequence of their processing difficulties. The strongest support for this explanation comes from two studies showing that adolescent first language learners of ASL display atypical performance in real-time, but improve dramatically when the processing demands are reduced. In one study, Emmorey, Bellugi and colleagues (1995) found that native signers but not delayed signers were slower to process sentences with errors in agreement during an on-line sign monitoring task. However, when shown the same sentences in an off-line grammaticality judgment task, the delayed signers were just as accurate as the native signers in identifying the agreement errors. A second study found poor comprehension of utterances with polymorphemic predicates and verb agreement by two adolescent first language learners when the utterances were viewed in real-time. Subjects' performance improved considerably when they were allowed to view the same utterances several times before responding. Interestingly, the same subjects spontaneously produced utterances with polymorphemic predicates and verb agreement even though they had difficulties comprehending utterances with these structures (Morford 2003).

With these results in mind, it is important ask what the processing demands were of tasks used to document the grammatical deficiencies of this population. Newport's (1990) study included both tests of comprehension and of production. The comprehension measures required processing sentences in real-time, so poor performance could reflect a failure of the participants to comprehend the stimuli.

The production measures varied in the task requirements. There is little detail about the actual performance of subjects on these tasks since results are reported in terms of z-scores relative to the native signers' performance. Newport does report that the late learners of ASL use the target structures well above chance but with much more variability than native signers.

Mayberry and colleagues (Mayberry 1993; Mayberry and Fischer 1989; Mayberry and Eichen 1991) have used sentence recall and shadowing tasks. Subjects view a sentence and reproduce it simultaneously (shadowing) or immediately after viewing the sentence (recall). They have found that adolescent first language learners produce a number of substitutions that are phonologically similar to the target signs, but are semantically and grammatically inappropriate to their contexts.[5] Mayberry (1993: 1268) has interpreted this result as evidence that late first language learners are "intermittently stuck at the surface level of language structure." Interestingly, second language learners do not make these errors (Mayberry 1993).

Emmorey and colleagues' work is consistent with these results since they have found that delayed signers need more visual information relative to native signers to identify signs (Emmorey and Corina 1990), suggesting that lexical access in this population is indeed slower although not necessarily different than in native signers. Additional evidence that phonological information may "hang around" in working memory for non-native signers comes from a study reported by Emmorey, Corina and Bellugi (1995). They describe a probe recognition study in which subjects saw a sentence, followed by a probe sign. Subjects merely respond "yes" if the probe appeared in the sentence, or "no" if the probe did not appear in the sentence. There were no differences between native and non-native signers on accuracy or reaction time for "yes" responses, but non-native signers were much slower to reject distracter probes that were phonologically-related to signs that appeared in the target sentence.

Do problems in language processing merely reflect slower phonological processing? According to Givón and Barker (this volume), single lexical items in spoken languages are processed at the predictable rate of 250 msec per word. Emmorey and Corina (1990) report that single signs are isolated about 25 msec later by signers who learned ASL at age 11 on average relative to native signers. For single signs, this delay may have few consequences, but across a sequence of 10 signs, the delay would increase to 250 msec — enough to miss an entire sign. Based on these calculations, delayed signers would understand only 90% of the signs in a stretch of discourse. Actual analyses of immediate recall of ASL sentences by delayed signers suggest that comprehension is even poorer than this. Mayberry (1993) found that delayed signers were able to recall 56% of the lexical stems in target sentences, while Morford (2003) found that delayed signers recalled fewer than 25% of signs verbatim. Thus, in addition to the speed of processing, it is likely that other factors are disrupting language processing in this population.

A common characteristic of Genie's use of English and of adolescent first language learners' use of ASL is the variability in their production. This variability may be an important key to understanding how exposure to language affects language outcomes since it suggests that language use in this population is not as routinized or automated as in individuals who learn language from birth. As Bybee (e.g., Bybee and Scheibman 1999) and Haiman (e.g. 1994) have argued, degrees of constituency and grammatical elements arise out of the repeated processing of co-occurring words in language. However, this argument assumes that the language user will chunk the co-occurring elements, as is the case for native speakers or signers. Variability in delayed language learners may be an indication that this chunking process does not occur when exposure does not begin from birth.

In sum, the range of studies that have been carried out on non-native signers provides a great deal of evidence that processing is slower and less efficient in this population relative to native signers. Non-native signers appear to dwell on the surface structure of language, and are slow to deactivate phonological information after lexical access has taken place. This pattern of processing interferes with comprehension, which in turn, most likely affects the pattern of language production in this population as well. One way in which we may be underestimating the effects of delaying exposure to language is by assuming that delayed language learners are able to perceive the language input once they are exposed to it (Morford 2003). Poor processing skills may function as a continued form of linguistic isolation for delayed language learners during the course of acquisition.

4. The role of language exposure in language origins

The research described so far is motivated by questions about language development over the ontogenetic timespan. This collection of studies can nevertheless inform our thinking about language origins. Specifically, the results of these studies suggest that we can and should distinguish between two types of grammatical properties. Some grammatical elements emerge through the innovation of structure by individuals who have had no exposure to language, as is documented in the case of homesign. We could call these *system-internal* grammatical properties because they concern the systematicity of structures within a single grammar. Other grammatical properties emerge as a result of grammaticization (cf. Bybee, this volume), and depend crucially on the pre-existence of regular patterns in the input, as was described in the case of Nicaraguan Sign Language. These *processing-dependent* grammatical properties emerge through language use and are the direct result of characteristics of the way humans automate the perception, storage and retrieval of linguistic signals. Moreover, since we have seen that language processing does not become automated when language exposure is delayed until adolescence, we can surmise that these properties could not emerge within a single generation. In other words, the same individual who innovates structure in a communication system will not have the opportunity to automate processing of that structure. Automation, and subsequently grammaticization, require exposure in childhood to a structured communication system, in which case the innovation of structure would have already occurred. In sum, processing-dependent grammatical properties could only have emerged after a community of hominids shared a structured communication system to which their offspring were exposed.

Many investigators have proposed that a lexicon must have emerged prior to grammatical properties in the origin of human language (e.g., Givón 1998; Nowak, Plotkin and Jansen 2000), and these data are certainly consistent with this view. Further, these data indicate that not all grammatical properties emerged simultaneously. Some properties could have emerged without exposure to input, as in the case of homesign, and some properties could only have emerged after hominids had automated their language processing skills through exposure to language from an early age. Thus, the implication of this work is that there may be three distinctive stages in the emergence of language: (1) the emergence of the lexicon, (2) the emergence of system-internal grammatical properties, and (3) the emergence of processing-dependent grammatical properties.

Notes

1. We might think differently today, since it is likely that these deaf women communicated at least with spontaneously generated gestures, and more likely with a regional variety of a signed language.

2. Goldin-Meadow's subjects were all raised in homes where the parents' goal was to teach their children to speak. The parents' continued use of speech while communicating with their children may account for the lack of organizational properties in their gestures when analyzed alone. Gesture and speech typically form a unified organizational system (McNeill 1992). In other words, there may have been organizational properties of the gesture use of the parents that was inaccessible to the children because it required integration with properties expressed in speech.

3. But see Goldin-Meadow *et al.* 1995, for the description of how an American homesigner can distinguish between nouns and verbs.

4. It is important to consider the possibility that structural properties of gesture may be difficult to discern when analyzing it in isolation of the speech it accompanied. It is possible

that to the deaf child, there is little structure, even though to a hearing interlocutor there would be considerably more structure. See note 2.

5. A comparable error in English would be hearing the phrase “a *sweet* little girl” and responding “a *sweep* little girl.”

References

Bates, E. and MacWhinney, B. 1987. “Competition, variation and language learning”. In *Mechanisms of Language Acquisition*, B. MacWhinney (ed), 157–93. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bonvillian, J., Garber, A.M., and Dell, S.B. 1997. “Language origin accounts: Was the gesture in the beginning?” *First Language* 17:219–39.

Bybee, J. and Scheibman, J. 1999. “The effect of usage on degrees of constituency: The reduction of don’t in English”. *Linguistics* 37:575–96.

Campbell, R.N. and Grieve, R. 1982. Royal“ investigations of the origin of language”. *Historiographia Linguistica* 9:43–74.

Curtiss, S. 1977. *Genie: A Psycholinguistic Study of a Modern-Day ‘Wild-Child’*. New York: Academic Press.

Emmorey, K., Bellugi, U. Frederici, A., and Horn, P. 1995. “Effects of age of acquisition on grammatical sensitivity: Evidence from on-line and off-line tasks”. *Applied Psycholinguistics* 16:1–23.

Emmorey, K. and Corina, D. 1992. Differential sensitivity to classifier morphology in ASL signers. Paper presented at the Linguistic Society of America, Chicago, Ill.

Emmorey, K. and Corina, D. 1990. “Lexical recognition in sign language: Effects of phonetic structure and morphology”. *Perceptual and Motor Skills* 71:1227–52.

Emmorey, K. Corina, D. and Bellugi, U. 1995. “Differential processing of topographic and referential functions of space”. In *Language, Gesture and Space*, K. Emmorey and J. Reilly (eds), 43–62. Hillsdale, NJ: Lawrence Erlbaum Associates.

Emmorey, K. and Lillo-Martin, D. 1995. “Processing spatial anaphora: Referent reactivation with overt and null pronouns in American Sign Language”. *Language and Cognitive Processes* 10:631–53.

Givón, T. 1998. “On the co-evolution of language, mind and brain”. *Evolution of Communication* 2:45–116.

Goldin-Meadow, S. and Feldman, H. 1977. “The development of language-like communication without a language model”. *Science* 197:401–3.

Goldin-Meadow, S. and Mylander, C. 1984. “Gestural communication in deaf children: The effects and non-effects of parental input on early language development”. *Monographs of the Society for Research in Child Development* 49:1–121.

Goldin-Meadow, S. and Mylander, C. 1990. “The role of a language model in the development of a morphological system”. *Journal of Child Language* 17:527–63.

Goldin-Meadow, S., Mylander, C., and Butcher, C. 1995. “The resilience of combinatorial structure at the word level: Morphology in self-styled gesture systems”. *Cognition* 56: 195–262.

Gomez, R.L. and Gerken, L. 1999. "Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge". *Cognition* 70: 109–35.

Haiman, J. 1994. "Ritualization and the development of language". In *Perspectives on Grammaticalization*, W. Pagliuca (ed), 3–28. Amsterdam: John Benjamins.

Janzen, T. 1995. *The polygrammaticalization of FINISH in ASL.* Ms., University of Manitoba.

Kegl, J.A. and McWhorter, J. 1997. "Perspectives on an emerging language". In *Proceedings of the 28th Annual Child Language Research Forum*, E.V. Clark (ed), 15–38. Stanford, Calif.: Center for the Study of Language and Information.

MacFarlane, J. 1998. "The grammaticization of WRONG in American Sign Language". In *Proceedings of the 1st Annual High Desert Linguistics Conference*, C. Berkenfield, D. Nordquist and A. Grieve-Smith (eds), 17–23. Albuquerque, N. Mex.: High Desert Linguistics Society.

Mayberry, R.I. 1993. "First-language acquisition after childhood differs from second-language acquisition: The case of American Sign Language". *Journal of Speech and Hearing Research* 36: 1258–70.

Mayberry, R.I. and Eichen, E.B. 1991. "The long-lasting advantage of learning sign language in childhood: Another look at the critical period for language acquisition". *Journal of Memory and Language* 30: 486–512.

Mayberry, R.I. and Fischer, S.D. 1989. "Looking through phonological shape to sentence meaning: The bottleneck of non-native sign language processing". *Memory and Cognition* 17: 740–54.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal About Thought.* Chicago: University of Chicago Press.

Morford, J.P. 2003. "Grammatical development in adolescent first language learners". *Linguistics* 41.

Morford, J.P. 1999. On the origins of phonology: From homesign to ASL in two adolescent first language learners. Paper presented at the 15th Annual Language Origins Society Meeting. Naples, Italy.

Morford, J.P. 1996. "Insights to language from the study of gesture: A review of research on the gestural communication of non-signing deaf people". *Language and Communication* 16: 165–78.

Morford, J.P. and Kegl, J.A. 2000. "Gestural precursors of linguistic constructs: How input shapes the form of language". In *Language and Gesture*, D. McNeill (ed), 358–87. Cambridge: Cambridge University Press.

Morford, J.P. and Mayberry, R.I. 2000. "A reexamination of 'early exposure' and its implications for language acquisition by eye". In *Language Acquisition by Eye*, C.D. Chamberlain, J.P. Morford, and R.I. Mayberry (eds), 111–27. Mawtaw, NJ: Lawrence Erlbaum Associates.

Newport, E.L. 1990. "Maturational constraints on language learning". *Cognitive Science* 14: 11–28.

Nowak, M.A., Plotkin, J.B., and Jansen, V.A.A. 2000. "The evolution of syntactic communication". *Nature* 404, 495–8.

Polich, L.G. 2000. *Social Agency and Deaf Communities: A Nicaraguan Case Study*. Ph.D. dissertation, University of Texas.

Radford, A. 1994. “The syntax of questions in child English”. *Journal of Child Language* 21: 211–36.

Roeper, T. 1998. “Finding fundamental operations in language acquisition: Formal features as triggers”. *University of Massachusetts Occasional Papers in Linguistics* 22: 197–214.

Roeper, T. and Williams, E. (eds) 1987. *Parameter Setting*. Dordrecht: D. Reidel.

Rumelhart, D.E. and McClelland, J.L. 1986. “On learning the past tenses of English verbs”. In *Parallel distributed processing: Explorations in the Microstructure of Cognition*, Volume II: *Psychological and Biological Models*, J.L. McClelland and D.E. Rumelhart (eds). Cambridge, Mass.: MIT Press.

Saffran, J.R., Aslin, R.N., and Newport, E.L. 1996. “Statistical learning by 8-month-old infants”. *Science* 274: 1926–8.

Schembri, A. 2000a. “Nachdenken über Klassifikatoren in Gebärdensprachen (Teil I)”. *Das Zeichen* 14: 470–83.

Schembri, A. 2000b. “Nachdenken über Klassifikatoren in Gebärdensprachen (Teil II)”. *Das Zeichen* 14: 629–39.

Senghas, A. 1995. “The development of Nicaraguan Sign Language via the language acquisition process”. *Proceedings of the Annual Boston University Conference on Language Development* 19: 543–52.

Shaffer, B. 2000. *A Syntactic, Pragmatic Analysis of the Expression of Necessity and Possibility in American Sign Language*. Ph.D. dissertation, University of New Mexico.

Tardif, T., Schatz, M., and Naigles, L. 1997. “Caregiver speech and children’s use of nouns versus verbs”. *Journal of Child Language* 24: 535–66.

Chapter 14

Getting a handle on language creation*

Susan Goldin-Meadow
University of Chicago

In 1866, the Linguistic Society of Paris declared a moratorium on research into the origins of language, asserting that the crucial studies could not be done. Although it is certainly true that we cannot go back in time to observe the birth of language, we can watch language as it changes over historical and ontogenetic timespans. And getting a sense of where language comes from over the short-term may, in the end, offer insight into where it came from originally.

We can watch language as it changes over historical time. Languages that already have grammatical structure create new structure and often do so in similar ways. For example, grammatical structure can grow out of lexical items, and the same lexical items often give birth to the same grammatical structures across various unrelated languages (Bybee 1998). Grammatical structure can even grow out of very limited structure. Pidgin languages, which typically arise in colonial situations and initially are used as trade languages, are simple systems with little structure. Creole languages develop out of these simple pidgins, and each creole has a more expanded vocabulary and a more complex grammatical structure than the pidgin out of which it emerged. Indeed, not only do all creoles have grammatical structure, but they may all have the same grammatical structure, despite having evolved out of very different pidgins (Bickerton 1990, 1999).[1]

In addition to observing language change across generations, we can also observe language change within a single individual over ontogenetic time. When children are exposed to a model for language (as most children are), they learn that language, though not all at once. The steps children follow in progressing from a one-word stage to a multi-word and multi-proposition stage have the potential to offer insight into mechanisms of language change in general (cf. Givón 1998: 101). But, of course, the nature and direction of the changes that children experience as they learn language are determined in large part by the language to which they are exposed. Although children's creative tendencies are at times visible in the act of language-learning, those creative acts are greatly constrained by the to-be-learned language and can easily be masked.

A potentially more revealing way to explore language creation in individual children is to observe children who are exposed to language models that are not intact. For example, Singleton and Newport (2002) describe language-learning in a deaf child whose deaf parents were late-learners of sign and thus produced signs that provided an incomplete model of the morphological structure in American Sign Language (ASL). The child, exposed only to this imperfect model of ASL, nevertheless developed a sign language with morphological structure more complex than that of his parents, and comparable in many respects to the morphological structure developed by other deaf children exposed to complete models of ASL. He changed the language as he learned it.

We find comparable creativity when deaf children are not exposed to ASL but rather to Manually Coded English (MCE). MCE is the name for a set of sign systems which map English surface structure onto the visual/gestural modality. MCE was invented by educators to teach English to deaf children and, as such, is not a "natural" language system spontaneously developed by language users. Unlike ASL, which uses simultaneous spatial devices to mark morphology, MCE uses invented signs that are designed to map onto the morphologic structure of English. English-like sequential structure within a sign is apparently very difficult to process. Indeed, deaf children exposed only to MCE alter the input they receive, innovating forms that systematically use space to convey meaning, as do many of the grammatical devices of ASL (Gee and Goodhart 1985; Goodhart 1984; Livingston 1983; S. Supalla 1991; Suty and Friel-Patti 1982). Thus, when provided with input that is difficult to process, children change that input, creating new structures out of old.

Children have perhaps the greatest opportunity to create language when they are exposed to no language model whatsoever. It is rare that children find themselves in such situations, but it does happen. For example, many children who are congenitally deaf have hearing losses so severe that they are unable to acquire spoken language, even with intensive instruction. If these deaf children are not exposed to sign language input until adolescence, they will be for all intents and purposes deprived of a usable model for language during early childhood — although, importantly, they are not deprived of other aspects of human social interaction. Despite their lack of linguistic input, deaf children in this situation use gestures to communicate. These gestures, called "homesigns," assume the form of a rudimentary linguistic system, one that displays structure at both word and sentences levels and is used for many of the functions served by conventional language.

My goal here is to describe the steps that deaf children follow in fashioning their homesign gesture systems — in other words, to observe language-creation over the ontogenetic timespan. I begin by providing the background on deafness and language-learning that is necessary to understand the unusual language-learning circumstances in which homesigners find themselves. I then briefly summarize the properties of language that we have found in the deaf children's gesture systems. These are linguistic properties that children can invent *de novo* without explicit guidance from a language model — the "resilient" properties of language. In the next two sections, I examine the developmental steps deaf children take when creating their homesign gesture systems. I focus first on the parts of the system that do change over time, and then on the parts of the system that don't. I then explore what we can learn about language-creation from a controlled experimental study conducted with adults.

Deaf children generating their own gesture systems are in a unique language-creation situation in that they are doing their creating without a communication partner who is willing to participate with them in the process of language invention. As a result, their gesture systems cannot achieve the level of arbitrariness or automaticity found in other natural language situations. In the final section, I consider the implications of this difference for language creation.

1. Background on deafness and language-learning

Deaf children born to deaf parents and exposed from birth to a conventional sign language such as American Sign Language (ASL) acquire that language naturally; that is, these children progress through stages in acquiring sign language similar to those of hearing children acquiring a spoken language (Newport and Meier 1985). However, 90% of deaf children are not born to deaf parents who could provide early exposure to a conventional sign language. Rather, they are born to hearing parents who, quite naturally, tend to expose their children to speech (Hoffmeister and Wilbur 1980). Unfortunately, it is extremely uncommon for deaf children with severe to profound hearing losses to acquire the spoken language of their hearing parents naturally, that is, without intensive and specialized instruction. Even with instruction, deaf children's acquisition of speech is markedly delayed when compared either to the acquisition of speech by hearing children of hearing parents, or to the acquisition of sign by deaf children of deaf parents. By age 5 or 6, and despite intensive early training programs, the average profoundly deaf child has limited linguistic skills in speech (Conrad 1979; Mayberry 1992; K. Meadow 1968). Moreover, although many hearing parents of deaf children send their children to schools in which one of the manually coded systems of English is taught, some hearing parents send their deaf children to "oral" schools in which sign systems are neither taught nor encouraged; thus, these deaf children are not likely to receive input in a conventional sign system.

The ten children in our studies are severely (70–90 dB bilateral hearing loss) to profoundly (>90dB bilateral hearing loss) deaf, and their hearing parents chose to educate them using an oral method. At the time of our observations, the children ranged in age from 14 months to 4 years, 10 months and had made little progress in oral language, occasionally producing single words but never combining those words into sentences. In addition, at the time of our observations, the children had not been exposed to ASL or to a manual code of English. As preschoolers in oral schools for the deaf, the children spent very little time with the older deaf children in the school who might have had some knowledge of a conventional sign system (i.e., the preschoolers only attended school a few hours a day and were not on the playground at the same time as the older children). In addition, the children's families knew no deaf adults socially and interacted only with other hearing families, typically those with hearing children. One of the primary reasons we were convinced that the children in our studies had had no exposure to a conventional sign system at the time of our observations was that they did not know even the most common lexical items of ASL or Signed English (i.e., when a native deaf signer reviewed our tapes, she found no evidence of any conventional signs; moreover, when we informally presented to the children common signs such as those for mother, father, boy, girl, dog, we found that they neither recognized nor understood any of these signs).

Under such inopportune circumstances, these deaf children might be expected to fail to communicate at all, or perhaps to communicate only in non-symbolic ways. This turns out not to be the case. Studies of deaf children of hearing parents in general have shown that these children spontaneously use gestures — "homesigns" — to communicate even if they are not exposed to a conventional sign language model (Fant 1972; Lenneberg 1964; Moores 1974; Tervoort 1961). Given a home environment in which family members communicate with each other through many different channels, one might expect that the deaf child would exploit the accessible modality (the manual modality) for the purposes of communication. The question is whether the gestures the deaf child uses to communicate are structured in language-like ways. In the next section, I describe the properties of the deaf child's gestures that can be considered language-like.

2. The resilient properties of language

The linguistic properties that appear in the deaf children's gesture systems are resilient — likely to crop up in a child's communications whether or not that child is exposed to a conventional language model. Table 1 lists the properties of language that we have found in the deaf children's gesture systems. There may, of course, be many others — just because we haven't found a particular property in the deaf children's homesign gesture system doesn't mean it's not there. The table lists properties at the word- and sentence-levels, as well as properties of language

Table 1. The resilient properties of language

The resilient property	As instantiated in the deaf children's gesture systems
Words	
Stability	Gesture forms are stable and do not change capriciously with changing situations
Paradigms	Gestures consist of smaller parts that can be recombined to produce new gestures with different meanings
Categories	The parts of gestures are composed of a limited set of forms, each associated with a particular meaning
Arbitrariness	Pairings between gesture forms and meanings can have arbitrary aspects, albeit within an iconic framework
Grammatical Functions	Gestures are differentiated by the noun, verb, and adjective grammatical functions they serve
Sentences	
Underlying Frames	Predicate frames underlie gesture sentences
Deletion	Consistent production and deletion of gestures within a sentence mark particular thematic roles
Word Order	Consistent orderings of gestures within a sentence mark particular thematic roles
Inflections	Consistent inflections on gestures mark particular thematic roles
Recursion	Complex gesture sentences are created by recursion
Redundancy Reduction	Redundancy is systematically reduced in the surface of complex gesture sentences
Language use	
Here-and-Now Talk	Gesturing is used to make requests, comments, and queries about the present
Displaced Talk	Gesturing is used to communicate about the past, future, and hypothetical
Narrative	Gesturing is used to tell stories about self and others
Self-Talk	Gesturing is used to communicate with oneself
Meta-language	Gesturing is used to refer to one's own and others' gestures

use, and details how each property is instantiated in the deaf children's gesture systems.

2.1 Words

The deaf children's gesture words have five properties that are found in all natural languages. The gestures are *stable* in form, although they needn't be. It would be easy for the children to make up a new gesture to fit every new situation (and, indeed, that appears to be what hearing speakers do when they gesture along with their speech, cf. McNeill 1992). But that's not what the deaf children do. They develop a stable store of forms which they use in a range of situations — they develop a lexicon, an essential component of all languages (Goldin-Meadow, Butcher, Mylander and Dodge 1994).

Moreover, the gestures they develop are composed of parts that form *paradigms*, or systems of contrasts. When the children invent a gesture form, they do so with two goals in mind — the form must not only capture the meaning they intend (a gesture-world relation), but it must also contrast in a systematic way with other forms in their repertoire (a gesture-gesture relation). In addition, the parts that form these paradigms are *categorical*. For example, one child used a *Fist* handshape to represent grasping a balloon string, a drumstick, and handlebars — grasping actions requiring considerable variety in diameter in the real world. The child did not distinguish objects of varying diameters within the *Fist* category, but did use his handshapes to distinguish objects with small diameters as a set from objects with large diameters (e.g., a cup, a guitar neck, the length of a straw) which were represented by a *CLarge* hand. The manual modality can easily support a system of analog representation, with hands and motions reflecting precisely the positions and trajectories used to act on objects in the real world. But, again, the children don't choose this route. They develop categories of meanings that, although essentially iconic, have hints of *arbitrariness* about them (the children don't, for example, all have the same form-meaning pairings for handshapes, Goldin-Meadow, Mylander and Butcher 1995).

Finally, the gestures the children develop are differentiated by *grammatical function*. Some serve as nouns, some as verbs, some as adjectives. As in natural languages, when the same gesture is used for more than one grammatical function, that gesture is marked (morphologically and syntactically) according to the function it plays in the particular sentence (Goldin-Meadow *et al.* 1994).

2.2 Sentences

The deaf children's gesture sentences have six properties found in all natural languages. Underlying each sentence is a *predicate frame* that determines how many arguments can appear along with the verb in the surface structure of that sentence. Moreover, the arguments of each sentence are marked according to the thematic role they play. There are three types of markings that are resilient: (1) *deletion* — the children consistently produce and delete gestures for arguments as a function of thematic role; (2) *word order* — the children consistently order gestures for arguments as a function of thematic role; and (3) *inflection*[2] — the children mark with inflections gestures for arguments as a function of thematic role (Goldin-Meadow and Mylander 1984; Goldin-Meadow *et al.* 1994).

In addition, *recursion*, which gives natural languages their generative capacity, is a resilient property of language. The children form complex gesture sentences out of simple ones. They systematically combine the predicate frames underlying each simple sentence, following principles of sentential and phrasal conjunction. When there are semantic elements that appear in both propositions of a complex sentence, the children have a systematic way of *reducing redundancy*, as do all natural languages (Goldin-Meadow 1982, 1987).

2.3 Language use

The deaf children use their gestures for five central functions that all natural languages serve. They use gesture to make requests, comments, and queries about things and events that are happening in the situation — that is, to communicate about the *here-and-now*. Importantly, however, they also use their gestures to communicate about the non-present — *displaced* objects and events that take place in the past, the future, or in a hypothetical world (Butcher, Mylander and Goldin-Meadow 1991; Morford and Goldin-Meadow 1997).

In addition to these rather obvious functions that language serves, the children use their gestures to tell stories about themselves and others — to *narrate* (Phillips, Goldin-Meadow and Miller 2001). They use their gestures to communicate with themselves — to *self-talk* (Goldin-Meadow 1997). And finally, they use their gestures to refer to their own or to others' gestures — for *metalinguistic* purposes (Singleton, Morford and Goldin-Meadow 1993).

All of the properties listed in the table are resilient in the sense that they can be developed by children without guidance from a language model. Interestingly, however, some of these properties take time to develop, others do not. If we are interested in observing changes in a linguistic system — changes that are *not* brought about by a conventional language model — an excellent place to look is at those aspects of the deaf children's gesture systems that develop over time.

3. Properties of the homesign gesture system that change over time

3.1 An utterance grows in size and scope

The most obvious developmental change in the deaf children's homesign gesture systems is that the children's utterances grow in size and scope. We first observed some of the children in what might be called a "one-gesture" period akin to the one-word stage found in children learning conventional languages (e.g., Bloom 1973). Although two of the children in our studies, Kathy and Abe, were gesturing during their first observation session, neither one produced more than one gesture at a time. Some of their gestures were pointing gestures and some were iconic gestures, but at no time during these early sessions did the children combine two gestures into a single sentence.

Why might a *hearing* child be limited to producing one word at a time? One possibility is that young children find it hard to remember two different words at one time. Another is that it is motorically difficult for young children to produce two different words as a single unit. But these explanations won't work for gesture. It seems easy enough to combine one pointing gesture with another pointing gesture — there is little strain on either memory or motor function. The one-unit constraint found in all young communicators may therefore stem from deeper causes, ones tied to neither the modality nor the arbitrariness of the language.

The fact that the deaf children experience a one-gesture period comparable to the one-word period of children learning conventional languages suggests that the deaf children are following a *language* trajectory — their early gestures have the same constraints as early words. Moreover, their later steps are also in synchrony with children learning conventional languages. Kathy and Abe first began producing two-gesture sentences at ages 1;6 and 2;5, respectively — around the same time as English-learning children first produce two-word sentences (Brown 1973), and only slightly later than ASL-learning children first produce two-sign sentences (Bonvillian, Orlansky, and Novack 1983). These early two-gesture sentences, for the most part, convey only one proposition, and do so using structures found in early child language learned from conventional models.

In addition, four of the deaf children in our sample began producing more than one proposition within the bounds of a single sentence — complex sentences, having the important property of recursion — during our observation sessions (the other six produced complex sentences when we first observed them). Mildred and Kathy first produced complex sentences at 2;2, Abe began at 2;5, and Donald began sometime between 3;1 and 3;11 (we did not observe Donald during this time period; when we resumed observations, he was already producing complex sentences). Children learning spoken (Brown 1973) and signed (Hoffmeister 1978) languages first produce sentences conveying more than one proposition around these same ages.

Thus, the deaf children's sentences grow in the same way as sentences do when children are exposed to conventional language models. They first experience a one-word period during which they are limited to one gesture at a time. They then combine those gestures into two-word sentences characterized by simple structural properties. Finally, they produce longer sentences which convey two or more propositions and thus exhibit recursion. The deaf children's gesture sentences never get as consistently long as the sentences English-learners produce —

they are closer in length to sentences produced by children whose language permits a great deal of ellipsis (e.g., Japanese or Sinhalese). Yet they do grow in patterned ways.

3.2 The onset of morphological structure

3.2.1 *The morphological system*

The deaf children's gestures not only formed parts of longer sentence-units but they themselves were made up of smaller parts. For example, to request the experimenter to lay a penny down flat on a toy, one of the deaf children, David, produced a short downward arcing motion with his hand shaped like an O (with thumb and fingers touching). In itself this could be a global gesture presenting the shape and trajectory as an unanalyzed whole. The experimenter pretended not to understand and, after several repetitions, the child factored the gesture into its components: first he statically held up the gesture for a round object (the *OTouch* handshape) and then, quite deliberately and with his hand no longer in the *OTouch* shape but exhibiting a flat *Palm,* made the *Short Arc* trajectory for downward movement. The original gesture was thus decomposed into two elements. This example hints at the presence of a system of linguistic segments in which the complex meaning of "round-thing-moving-downward" is broken into components which are then combined into a gesture. Although the experimenter's feigned lack of understanding was undoubtedly important in getting David to decompose his gesture at that particular moment, the point I want to stress here is that when David did break his gesture into parts, those parts were elements of a wider system — one that accounted for virtually all of the gestures that David produced (Goldin-Meadow and Mylander 1990).

Thus, David had devised a morphological system in which each gesture was a complex of simpler gesture elements (Goldin-Meadow and Mylander 1990). Systematic compositionality of gestures within a system of contrasts can only come about by segmentation and combination. As an example of how this child's gestures formed a system of contrasts, a *CMedium* handshape (the hand shaped in a C with the fingers 1–3 inches from the thumb) meant 'handle an object 2–3 inches wide,' and a *Revolve* motion meant 'rotate around an axis.' When combined, these two components created a gesture whose meaning was a composite of the two meanings — 'rotate an object 2–3 inches wide' (e.g., twist a jar lid). When the same *CMedium* handshape was combined with a different motion, a *Short Arc* (meaning 'reposition'), the resulting combination had a predictably different meaning — 'change the position of an object 2–3 inches wide' (e.g., tilt a cup). As a result, the child's gestures can be said to conform to a framework or system of contrasts.

We have analyzed the gesture systems of four children (David, Marvin, Kathy, Abe) at this level (Goldin-Meadow *et al.* 1995), and have found that the gestures of all four children could be characterized by paradigms of handshape and motion combinations. Each child (1) used a limited set of discrete handshape and motion forms; (2) consistently associated each handshape or motion form with a particular meaning (or set of meanings) throughout the corpus; and (3) produced most of the handshapes with more than one motion, and most of the motions with more than one handshape. Each handshape and motion was an independent and meaningful morpheme that could combine with other morphemes in the system to create larger meaningful units — the system was combinatorial.

Although similar in many respects, the gesture systems produced by these four children were sufficiently different to suggest that the children had introduced relatively arbitrary — albeit still iconic — distinctions into their systems. For example, in contrast to David and Marvin who used the *CMedium* handshape to represent objects 2–3 inches in width (e.g., a cup or a box), Kathy and Abe used the same *CMedium* handshape to represent objects that were slightly smaller, 1–2 inches in width (e.g., a banana or a toy soldier, Goldin-Meadow *et al.* 1995). The fact that there were differences in the ways the children defined a particular morpheme suggests that there were choices to be made (although all of the choices still were transparent with respect to their referents). Moreover, the choices that a given child made could not be determined without knowing that child's individual system. In other words, one cannot predict the precise boundaries of a child's morphemes without knowing that child's individual system. It is in this sense that the deaf children's gesture systems can be said to be *arbitrary.*

3.2.2 *The parts grow out of wholes*

At the very earliest stages of development, children acquiring conventional languages initially learn words as rote wholes (MacWhinney 1978). They then realize — relatively quickly in some languages, e.g., K'iche' Maya (Pye 1992), Turkish (Aksu-Koch and Slobin 1985), West Greenlandic (Fortescue and Olsen 1992) and more slowly in other languages, e.g., English (Bowerman 1982), ASL (Newport 1984) — that those wholes are composed of meaningful parts and begin to use those parts as productive morphemes. Since the deaf children in our study are not learning their gestures from adult models, we might expect them to show a different developmental pattern — that is, to use the sub-gesture hand and motion components that we have just described productively even at the earliest stages of development. If so, we would then conclude that children begin by learning words as wholes rather than as combinations of parts *only* when they learn their words from a conventional language model.

On the other hand, it is possible that, even without a conventional language model, the child's first representation of an event is not in terms of parts, but rather in terms of the event as a whole. If so, the deaf child's first lexical items would not be composed of component parts but would instead be unanalyzed wholes which map (as wholes) onto an event. For example, the gesture *OTouch+*

Revolve may, early in development, be used in the context of twisting a small key and for no other objects or actions, and thus may function as an unanalyzed label for key-twisting. Later, perhaps when the child has accumulated a sufficient number of gestures in the lexicon, the child may begin to consider his or her gestures in relation to one another and organize the gestures around any regularities that happen to appear in the lexicon (i.e., the child treats his or her own gestures as a "problem space" that needs systematization, cf. Karmiloff-Smith 1992). For example, over time the child may add an *OTouch+Short Arc* combination used exclusively for hat-putting-on, and a *CLarge+Revolve* combination used exclusively for jar-twisting, to his or her repertoire. At some point, the child then pulls back and considers the relation between these three gestures: *OTouch+ Revolve, OTouch+Short Arc* and *CLarge+Revolve.* The child notices that the *OTouch* handshape recurs across the gestures, as does the *Revolve* motion. These recurring forms are then, for the first time, separated out from the wholes and treated as component parts, allowing the construction of new combinations (e.g., *Clarge+Short Arc*, meaning reorient a large object, as in putting a lid on a box). The transition is from a state in which the child considers a gesture only in relation to the situation conveyed — that is, a gesture-*world* relation — to a state in which the child begins to consider gestures in relation to other gestures in the system — a gesture-*gesture* relation (see Table 2).

If the deaf children were to follow this developmental path, we would expect that a particular handshape/motion combination, when still an undecomposed whole, might be used exclusively for a single object/action pairing. Later, when the parts of the gesture have been isolated, that same combination would be used for a variety of related objects and a variety of related actions. This is precisely the pattern we find. The *OTouch+Revolve* gesture is used first to refer uniquely to "key-twisting," and only later to refer to "knob-rotating" and "dial-spinning" as well. That is, the *OTouch* handshape in this and in other gestures is now used in

Table 2. Changes in how handshapes and motions map onto objects and actions over time

(i) Whole gestures mapped onto events in their entirety
OTouch + Revolve = Key/Twist
OTouch + Short Arc = Hat/Put-on
CLarge + Revolve = Jar/Twist

(ii) Handshapes mapped onto classes of objects, and motions mapped onto classes of actions
OTouch = Small Object
CLarge = Large Object
Revolve = Rotate
Short Arc = Reorient

relation to a variety of related objects (keys, knobs, dials — all of which are narrow in diameter), and the *Revolve* motion in this and in other gestures is used in relation to a variety of related actions (twisting, rotating, spinning — all of which involve revolving around a point). We found this developmental pattern in three of the four children — David, Kathy, and Marvin (Goldin-Meadow *et al.* 1995). The fourth child, Abe, was already using a large number of handshape/motion combinations for a variety of object and actions from the first of his observations sessions. We may not have caught Abe early enough to observe the first steps he took in fashioning a morphological system — or Abe may have begun his gesture system, not with representations of events as wholes, but with representations of parts of events.

What I am suggesting is that the deaf children induce their morphological systems from the earliest gestures they themselves create. Indeed, the first holistic gestures that the children used seemed to set the stage for the system each child eventually generated. For example, in session I, David used the *OTouch+No Motion* combination to describe holding a bubble wand, a narrow *long* object. In addition, he also used the *OTouch+Circular* combination to describe twisting a small key, a narrow *short* object. If these examples are representative of the gestures David used at the time, he would infer that the *OTouch* handshape is used for objects that have relatively narrow diameters but that can be either long (like the wand) or short (like the key). Thus, on the basis of *his own* gestures, David would infer a form/meaning pairing in which the *OTouch* form is associated with the meaning "handle an object <2 inches in width and any length".

In contrast, the first time David produced the *Fist* handshape, he used it in session II combined with *No Motion* to describe holding a bubble wand; that is, the *Fist+No Motion* combination was used for the same event as the *OTouch+No Motion* combination. However, the *Fist* was *not* used to describe any other objects during the early sessions. On the basis of these gestures, David ought to infer that the *Fist* handshape is used for objects that have narrow diameters and *long* lengths. In fact, he did. When he began to consistently use gestures in relation to a variety of objects and actions in session IV, David used the *Fist* (combined with the *Arc To and Fro* and the *Short Arc* motions) to describe a set of objects, all having narrow diameters (<2 inches) and long lengths (>3 inches), e.g., the handle of a hammer, the handlebars of a bike, a newspaper, and the brim of a hat — precisely the range of objects eventually seen for this form in his gesture system.

The earliest gestures that Kathy and Marvin created also set the stage for the categories they eventually settled on in their gesture systems (Goldin-Meadow *et al.* 1995: 241–2) — and thus set the stage for the similarities and differences seen across the children's systems. Before each child began to consistently use a handshape/motion combination in relation to a variety of objects and actions, the child had already used that handshape in different gestures in relation to precisely the range of objects that would eventually fall within a given morpheme type in that child's system. Thus, when the child was ready to survey his or her gestures and analyze them to extract handshape and motion components, the outlines of the system were already present.

Just as children provided with a conventional language model induce rules and categories from the input they receive, the deaf children in this study induce the structure of their categories from their input — the difference is that the deaf children are forced by their circumstances to provide, and reflect upon, their own gestures as input.

3.3 The onset of grammatical categories

3.3.1 *Identifying nouns, verbs, and adjectives in gesture*

The children combined handshapes and motions to create gestures which they then used to refer to objects, actions, and attributes. The next question we ask is whether there is any reason to believe that these gestures functioned like the grammatical categories nouns, verb, and adjectives. We addressed this question by examining the gestures produced by David, our most prolific gesturer, over a two-year time period, from ages 2;10 to 4;10 (Goldin-Meadow *et al.* 1994). Our challenge was to figure out how to break into David's system of grammatical categories without knowing the properties of that system (or even whether the system existed).

Languages vary in the way nouns and verbs manifest themselves — the syntactic positions they occupy, and the morphological inflections they assume. What is common across languages is the functional roles nouns and verbs play in discourse, and the semantic characteristics that have evolved as a function of these discourse roles (Hopper and Thompson 1984, 1988; Sapir 1921). Sapir (1921) grounds the universality of the noun-verb distinction in the basic fact that language consists of a series of propositions. On intuitive grounds, there must be something to talk about and something to be said (or to predicate) of this subject once it is brought into focus. According to Sapir, this particular distinction is of such fundamental importance that languages emphasize it by creating a formal barrier between the two terms of the proposition — the subject of discourse, that is, the noun, and the predicate of the discourse, the verb.

We followed Sapir (1921) in considering a noun to be the focus or subject of the discourse (i.e., the something that is talked about), and verbs and adjectives to be the predicates of the discourse (i.e., what is said of this something). Thus, if David uses an iconic gesture to focus attention on an object, it is coded as a noun, but if he uses the gesture to say something about that object (i.e., to predicate something of the object), it is coded as either a verb or an adjective, depending upon whether the gesture depicts an action or an attribute.

Table 3. Progression of techniques used to distinguish noun gestures from verb gestures over time

		Noun gestures	Verb gestures
I.	Distinct forms	Point at jar	TWIST
II.	Distinct lexicons	ROUND	TWIST
III.	One lexicon with grammatical distinctions	TWIST [+abbreviation] [1st position]	TWIST [+inflection] [2nd position]

For example, if David uses the "flap" gesture (two palms, each held at a shoulder, arced to and fro as though flapping wings) to comment on a picture of a bird riding a bicycle with its wings on the handlebars (i.e., to focus attention on the bird rather than to comment on wing-flapping), the gesture is considered a *noun.* In contrast, if the "flap" gesture is used to describe a toy penguin that is at that moment flapping its wings, the gesture is considered a *verb* (although we do recognize that David could be commenting on the presence of the bird itself). As a second example, if David uses the "high" gesture (a flat palm held horizontally in the air) to comment on the fact that a cardboard chimney typically stands in the corner at Christmas time (i.e., to focus attention on the chimney rather than to comment on the chimney's height), the gesture is considered a *noun.* In contrast, if the "high" gesture is used to describe the temporary height of the tower before urging his mother to hit it with a hammer and topple it, the gesture is considered an *adjective.* Not surprisingly, material entities (cf. Bloom 1990) turn out to be the most common subjects of the discourse — the nouns — and relations (actions and attributes) turn out to be the most common predicates — the verbs and adjectives.

We found that David distinguished nouns from verbs throughout the two-year period. Thus, like all natural languages, David's gestures have a noun-verb distinction. Interestingly, however, the way in which David maintained the distinction between nouns and verbs changed over time, becoming more and more linguistically sophisticated with age (see Table 3). I describe the three stages David traversed in the next section.

3.3.2 *From separate forms, to separate lexicons, to separate grammatical markings and positions*

Initially, David used pointing gestures to fill noun roles and iconic gestures to fill verb and adjective roles. Thus, at the earliest stage, David distinguished nouns from verbs and adjectives through a gross distinction in gesture form: The stationary and directed index finger signaled nominal functions; iconic gesture forms signaled predicate functions.

Sometime after age 2;10, David started to use iconic forms as nouns, while continuing to use these forms as verbs and adjectives. The question is whether he found some other way to distinguish between nouns, verbs, and adjectives now that gesture *form* no longer served the purpose. English-learners, at the earliest stages, maintain inter-category boundaries by having distinct lexicons for nouns, verbs, and adjectives. And they do so despite the fact that the English language does *not* have distinct lexicons. In other words, there are words in English that cross noun-verb boundaries, but young learners don't seem to notice. For example, "comb" can be both a noun and a verb. Although young English-learners do use words like "comb" at the earliest stages of language-learning, they use these words in only one role (Macnamara 1982). A child might use "comb" to describe what she does to her doll's hair, but then would *not* also use "comb" to refer to the instrument involved in this activity. It is not until later in development that the child begins to use the same word in two different roles.

David showed this same constraint. He restricted his use of a particular iconic gesture to a single role. For example, David used his "laugh" gesture as a noun (to refer to Santa Claus) and never as a verb. He violated these boundaries for the first time at age 3;3, when he used the same gesture as a noun and a verb. He did not use the same gesture as a noun and an adjective until age 3;11. Thus, like young children learning conventional languages, David did *not* violate inter-category boundaries at the earliest stages of development. His first inclination was to respect these boundaries as do all young communicators. When gesture form (i.e., pointing vs. iconic forms) no longer served to distinguish nouns from verbs and adjectives in David's gesture system, he maintained the distinction lexically, that is, by using separate lexical items as nouns, verbs, and adjectives.

Eventually, however, English-learning children do learn words that cross the noun-verb boundary, that is, they use words like "comb" for both their noun and verb meanings. But when they do, they (like adult English-speakers) treat the noun uses of the word differently from the verb uses: (1) Noun uses appear in different positions within a sentence than do verb uses; that is, they are marked differently with respect to syntax, e.g., "I *comb* my hair" vs. "The *comb* is lovely." (2) Noun uses are marked with different inflections than are verb uses; that is, they are marked differently with respect to morphology, e.g, "I *combed* my hair" vs. "The *combs* are lovely."

David began to use the same iconic gesture as both noun and verb at age 3;3. And, like children learning conventional languages, when he did, he used morphological and syntactic techniques to distinguish the different uses. Nouns were more likely to be abbreviated and less likely to be inflected (morphological distinctions) than verbs in David's system. In addition, nouns were more likely to precede pointing gestures and verbs were more likely to follow them (a syntactic distinction) in David's system. For example, if using a "twist" gesture as a noun to mean "jar," David produced the gesture with only one rotation rather than several (with abbreviation), produced it in neutral space (without inflection), and produced it *before* a pointing gesture at the jar (pre-point). In contrast, if using the gesture as a verb to mean "twist," he produced the gesture with several rotations (without abbreviation), produced it near the jar (with inflection), and produced it *after* the pointing gesture at the jar (post-point).

Interestingly, adjectives were a mixed category in David's system, as they are in many languages (Thompson 1988; Dixon 1994). In David's gestures, adjectives resembled nouns in terms of morphological markings, but verbs in terms of syntactic position. For example, when David used the adjective "broken," he produced only one breaking-apart motion rather than several or the motion with only one hand rather than two (with abbreviation), and produced the gesture in neutral space (without inflection) — that is, he treated it like a noun. However, when positioning it in a sentence, David produced "broken" *after* a pointing gesture at the broken object (post-point) — that is, he treated it like a verb. David thus maintained a distinction between nouns, verbs, and adjectives, but he did so *grammatically* rather than lexically.

3.3.3 *Are the nouns and verbs grammatical categories or names for objects and actions?*

We began our search for a noun-verb distinction in David's gesture system with an intuitive guess as to which of his iconic gestures are nouns and which are verbs. Using these noun and verb categories, we found both morphological (i.e., variations within the gesture itself) and syntactic (i.e., variations across a string of gestures) patterns that distinguished between nouns and verbs in David's system. We take these formal patterns to be evidence for the noun and verb categories we code in David's gestures since the former (the patterns) are formulated in terms of the latter (the categories). The question then arises — what are these categories that we call *nouns* and *verbs* in David's gestures? Are they truly grammatical categories that are part of a linguistic system, or are they semantic categories naming objects and actions? After all, as is the case in all natural languages and particularly child language (Brown 1958), most of David's nouns turned out to refer to objects, and most of his verbs turned out to refer to actions.

To pursue this question, we recoded David's gestures, this time assessing whether they referred to objects or actions (Goldin-Meadow *et al.* 1994). We then asked whether David's noun-verb categories could be reduced to these object-action categories. We found that they could *not*, particularly after age 3;3. Before age 3;3, coding David's gestures in terms of nouns and verbs resulted in precisely the same categories as coding them in terms of objects and actions. However, after age 3;3, the two sets of categories were distinguishable and, impressively, David's morphological and syntactic devices patterned according to noun-verb grammatical categories rather than object-action semantic categories (Goldin-Meadow 2002a; Goldin-Meadow *et al.* 1994:300–1).

3.4 Reorganization across the system

To recap David's trajectory — after abandoning a distinction between nouns, verbs, and adjectives based purely on gesture form, David used separate sets of lexical items as nouns, verbs, and adjectives. Thus, he persistently respected inter-category boundaries in his early lexicon, as do children learning conventional languages, be they spoken (Huttenlocher and Smiley 1987) or signed (Petitto 1992). At 3;3, David changed once again. He began to use some of his lexical items for more than one function, most commonly using the same gestural form as a noun and a verb. However, he continued to maintain a distinction between categories by abbreviating nouns but not verbs (akin to derivational morphology), by producing verbs but not nouns in marked locations (akin to inflectional morphology), and by producing verbs and nouns in distinct positions in gesture sentences (akin to syntax). Thus, at 3;3, David began to use grammatical devices to maintain inter-category boundaries.

There are other changes that coincided with what appears to be the onset of a grammatical system in David's gestures. Prior to age 3;3, David produced all of his iconic gestures in contexts consistent with their grammatical function. During the first three observation sessions, he produced verbs *only* in contexts in which the relevant action and object were both present, and nouns *only* in contexts in which the object and the action were both absent — prototypical contexts for identifying actions and objects (cf. Huttenlocher and Smiley 1987, 1989). Thus, just as children learning English initially distinguish between nouns and verbs on the basis of a semantic rather than a grammatical distinction (Macnamara 1982), David may be basing his first categories on a semantic (object-action) rather than a grammatical (noun-verb) distinction. However, at age 3;3, David began using noun and verb gestures in intermediate contexts where the action was absent and the object was present — precisely the moment when he introduced grammatical devices to keep noun and verb categories distinct. Thus, David began to use gestures in what amounts to ambiguous action and object contexts at just the moment that he secured his grammatical system for distinguishing nouns from verbs.

At this same point in development, David's gestures could, for the first time, be characterized as having two levels of structure — structure across gestures within a sentence (akin to syntactic structure), and structure within each gesture (akin to morphological structure). Before this age, there was evidence for structure across gestures in David's gesture system, but no evidence that David had broken his gestures into component parts. At age 3;3, however, David began to systematize his lexicon, changing it from a collection of gestures, each treated as a whole, into a system in which the component parts of each gesture contrasted in a meaningful way with the component parts of the other gestures in the lexicon

(Goldin-Meadow and Mylander 1990; Goldin-Meadow *et al.* 1995). Handshape and motion components combined to form word-stems. These stems were then fed to the next level of the system — they were abbreviated when used as nouns or inflected when used as verbs, and placed into distinctive positions in sentences. Thus, we see sets of units corresponding to the different levels found in conventional languages (word-stem morphology, derivational and inflectional morphology, syntax) which come together at the same moment in development and constitute the building blocks of David's gesture system.

The transformation of David's lexicon from an unorganized collection of gestures into a system of contrasting morphemes that work within a grammatical system co-occurred with yet one other change. At age 3;3, David began to refer, either via pointing gestures or the placement of iconic gestures, to objects that were not in the here-and-now (Butcher, Mylander and Goldin-Meadow 1991). For example, David pointed at the head chair at the dining room table, followed by a "sleep" gesture. No one was sleeping in the chair at the time, but David's father was sleeping in his bedroom down the hall. David used the dining room chair, which is where his father typically sat, as a "stand-in" for dad — the sentence thus meant "dad sleeps." David had pointed at an object in the room to refer to an object that was not in the room. At the same age, David also began inflecting his verb gestures toward objects in the room to indicate objects that were not in the room (Butcher *et al.* 1991). For example, David displaced a "twist" gesture toward an open jar in order to request his mother to twist the lid off a closed jar in the kitchen. Thus, David began to systematize his lexicon in the context of a grammatical system at the same time as he began to use his gestures in an increasingly symbolic fashion.

The impetus for a massive reorganization of the sort David experienced at age 3;3 might have been the child's maturational state — that is, the fact that he had reached a certain age. Or perhaps the impetus came from the state of the gesture system itself — that is, the fact that the system had become sufficiently cumbersome to require reorganization. One approach to pulling apart these possibilities is to conduct developmental analyses on the gesture systems of the remaining deaf children in our study. Whatever the outcome of such analyses, however, it is clear that the reorganization that occurred resulted in a *system* that looks very much like natural language. Moreover, the development of this system did *not* depend on guidance from a conventional language model.

4. Properties of the homesign gesture system that do not change over time

It seems to require time for the deaf children to construct morphological structure and a set of grammatical categories within their gesture systems. In contrast, as soon as the children began producing gestures together within a single sentence, they developed devices for marking semantic roles in those sentences. All ten of the children we have studied used two different sentence-level devices to indicate "who does what to whom."

4.1 Marking roles by producing them at a particular rate in a sentence

As described above, the deaf children went through a stage during which they produced no more than two gestures in a sentence. As a result, children at this stage cannot produce gestures for all of the roles they might like to express in a single sentence. One solution to this problem is for the children to produce gestures for the different roles haphazardly. However, this is not the solution the children chose — they systematically produced gestures for some roles and omitted gestures for others, thereby distinguishing one role from another.

Production probability is the likelihood that a particular thematic role will be gestured in a sentence of a given length. All ten of the deaf children we have observed were more likely to produce a gesture for the patient (e.g., the eaten cheese) in a sentence about eating than to produce a gesture for the actor (e.g., the eating mouse). In other words, production probability was significantly lower for transitive actors than for patients for each of the ten children (Goldin-Meadow and Mylander 1984). Two points are worth noting.

First, the children's production probability patterns convey information about who is the doer and the done-to in a two-gesture sentence. If, for example, a deaf child produces the gesture sentence "boy hit," we would guess from this utterance that the boy is the hittee (patient) in the scene rather than the hitter (actor) simply because the deaf children tend to produce gestures for patients at the expense of transitive actors.

Second, note that the deaf children's particular production probability pattern tends to result in two-gesture sentences that preserve the unity of the predicate — that is, patient + act transitive sentences (akin to OV in conventional systems) are more frequent in the deaf children's gestures than actor + act transitive sentences (akin to SV in conventional systems).

We also calculated production probability for intranstive sentences. Nine of the ten children produced gestures for the intransitive actor (e.g., the mouse in a sentence describing a mouse going to a hole) as often as they produced gestures for the patient (e.g., the cheese in a sentence describing a mouse eating cheese), and far more often than they produced gestures for the transitive actor (e.g., the mouse in a sentence describing a mouse eating cheese). This production probability pattern (see Figure 1, top graph) is reminiscent of case-marking patterns found in ergative languages — gesture production is high and equal for intransitive actors and patients, and low for transitive actors (cf., Dixon 1979; Silverstein 1976).

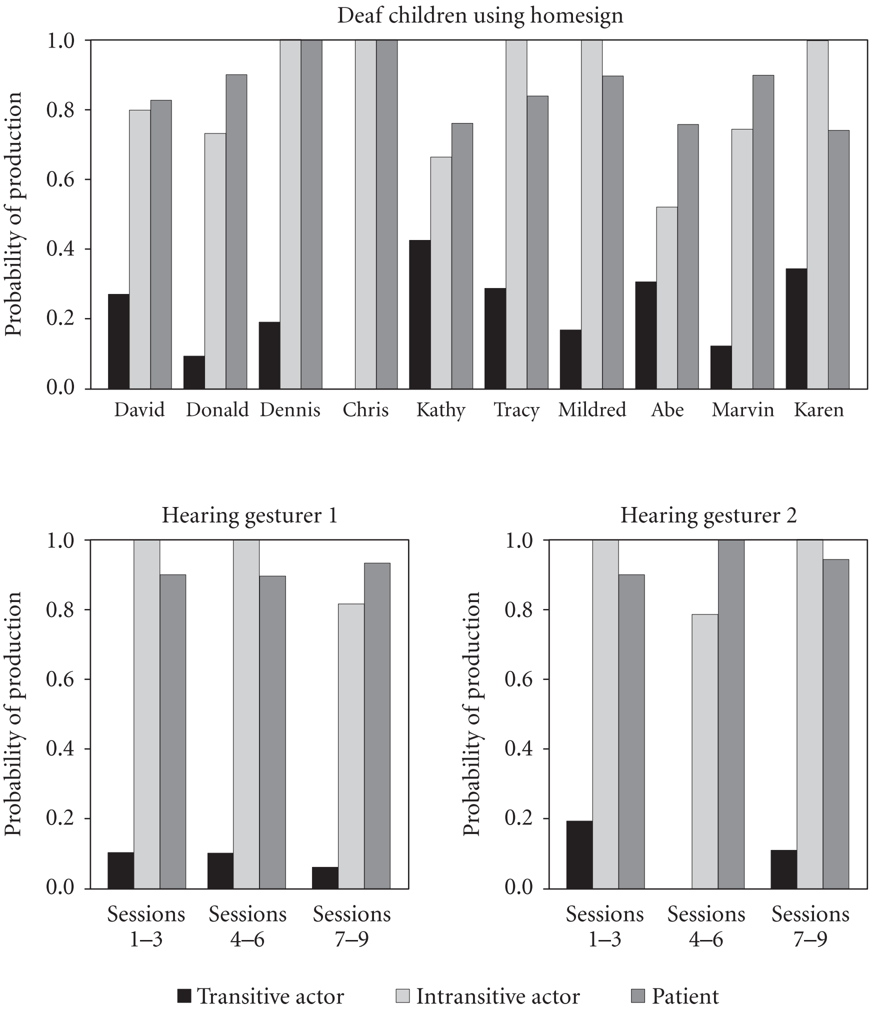


Figure 1. The likelihood that a gesture will be produced for a transitive actor (black bars), an intransitive actor (light grey bars), or a patient (dark grey bars) in a two-gesture sentence that permits all of these elements. The top graph displays the production probability patterns in the homesigns of ten deaf children of hearing parents. The bottom graphs display the production probability patterns in the gestures invented by two adult English-speakers who were asked to gesture without speaking. The data for the adults are divided into three parts: the first 3 sessions in which the two gesturers participated, the second 3, and the last 3. In both the deaf children and the hearing adults, the gesturer was more likely to produce gestures for intransitive actors and patients than for transitive actors (an ergative pattern)

It is important to note that the deaf children really were marking thematic role, and not just producing gestures for the most salient or most informative element in the context. One very sensible (albeit, in this case, wrong) possibility is that the deaf children produced gestures for intransitive actors and patients more often than for transitive actors because intransitive actors and patients tend to be new to the discourse more often than transitive actors (cf. DuBois 1987). In other words, the deaf children's production probability patterns could be an outgrowth of a semantic element's status as "new" or "old" in the discourse. If the novelty of a semantic element is responsible for how often that element is gestured, we would expect production probability to be high for all "new" elements (regardless of role) and low for all "old" elements (again, regardless of role). We found no evidence to support this hypothesis (Schulman, Mylander and Goldin-Meadow 2001; see also Goldin-Meadow and Mylander 1984: 49). Rather, we found an ergative production probablity pattern for "new" elements when analyzed on their own, as well as for "old" elements when analyzed on their own, as we would expect if thematic role, rather than novelty, determined how often an element was gestured.

4.2 Marking roles by placing them in a particular position in a sentence

In addition to reliably producing gestures for some thematic roles at the expense of others, the children were also consistent in where the gestures they did produce were positioned in their two-gesture sentences. Many, but not all, of the children ordered gestures for patients, acts, and recipients in a consistent way in their two-gesture sentences (Goldin-Meadow and Mylander 1984: 35–6): (1) gestures for *patients* (e.g., cheese) preceded gestures for *acts* (e.g., eat), (2) gestures for *patients* (e.g., hat) preceded gestures for *recipients* (e.g., cowboy's head), and (3) gestures for *acts* (e.g., move-to) preceded gestures for *recipients* (e.g., table). Importantly, these ordering patterns were not reducible to the discourse status of the semantic elements — if we reanalyze the sentences in terms of whether an element is "new" or "old" to the discourse, we find that most of the children's gesture sentences were "old-old" or "new-new," and that the "old-new" sentences were approximately as frequent "new-old" sentences. In other words, "new" elements did not consistently occupy the initial position in the deaf children's gestures sentences, nor did "old" elements (Goldin-Meadow and Mylander 1984: 51).

Nine of the 10 children produced gestures for patients before gestures for acts. Moreover, 7 children also produced gestures for intransitive actors before gestures for acts (Goldin-Meadow 2002b). Thus, many of the children treated intransitive actors like patients with respect to gesture order as well as production probability. David was the only child who produced a sufficient number of sentences with transitive actors to allow us to discern an ordering pattern for this thematic role. David not only treated patients and intransitive actors alike with respect to gesture order, but he ordered them differently from transitive actors — he produced gestures for patients and intransitive actors *before* gestures for acts, but gestures for transitive actors *after* gestures for acts (Goldin-Meadow and Mylander 1984: 39). For example, David pointed first at the snack he was offering me, then produced an "eat" gesture, and finally pointed at me, roughly translated as "snacks eat Susan" with the patient preceding the act and the transitive actor following it. In contrast, when David asked me to move to a particular spot — that is, when I played the role of intransitive actor rather than transitive actor — David pointed first at me and then produced a "move-over" gesture. David thus treated patients and intransitive actors alike and distinct from transitive actors — he displayed an ergative pattern — not only with respect to production probability, but also with respect to gesture order.

5. Language creation in adults

The deaf children we study are not exposed to a conventional sign language and thus cannot be fashioning their gestures after such a system. They are, however, exposed to the gestures that their hearing parents use when they speak — and these gestures could serve as a model for the deaf children's system. To explore this hypothesis, we examined the gestures that the hearing mothers of six of the deaf children spontaneously produced as they talked to their children. We found that the mothers' gestures displayed little systematicity either within or across individuals, and that the consistent patterns that the mothers did display did not match their children's patterns (Goldin-Meadow and Mylander 1983, 1984, 1998). Thus, the mothers' gestures could not have served as a straightforward model for the gesture systems their children created.

The hearing mothers and their deaf children interacted with one another on a daily basis. We might therefore have expected mother and child to develop a shared gesture system. But they didn't. The question is — what leads gesture to assume language-like structure in the deaf children, but not their hearing parents? I suggest that it is only when gesture is called upon to carry the full burden of communication that it assumes a language-like form. When produced along with speech, gestures do not and, I would argue, because of the constraints imposed by speech, *cannot* assume the segmented form of the words they accompany. One might suspect that if the deaf children's hearing mothers had merely refrained from speaking as they gestured, their gestures would have become more language-like in structure. In other words, the mothers might have been more likely to use gestures that mirrored their children's if they kept their mouths shut.

To test this hypothesis, we conducted a series of experiments in which hearing adults were asked to describe scenes with and without speech. We predicted that the gestures the adults would produce without speech would be distinct from the gestures these same adults produced when they described the scenes verbally. We thus attempted to simulate the deaf child's language-creating situation but with hearing adults as creators. There are two, very obvious differences between hearing adults and the deaf children in our studies. First, the adults already know a conventional language (English) and thus their created gestures could be heavily influenced by the particular language that they know. Second, the adults are not children and thus are well beyond whatever critical period there is for language-learning (and perhaps language-creating). To the extent that we find differences between the gestures that the adults and the deaf children create, age and language-knowledge become likely candidates for causing those differences. But to the extent that the gestures created by the adults and deaf children resemble one another, we have evidence that the created structures do *not* reflect a child-like way of organizing the world. Adults, even those who already have a language, may organize their communications in precisely the same ways as the deaf children, raising the possibility that the language-like properties found in the deaf children's systems result from trying to get information from one human mind to another in real time.

5.1 Gestures with and without speech

We asked English-speakers who had no knowledge of sign language to participate in the study (Goldin-Meadow, McNeill and Singleton 1996). We showed the adults videotaped vignettes of objects and people moving in space from the test battery designed by Supalla, Newport and their colleagues (2002) to assess knowledge of ASL. Half the scenes contained only one moving object (e.g., a porcupine wandering across the screen) while the other half contained one moving object and one stationary object (e.g., a girl jumping into a hoop). The adults were asked to describe each event depicted on the videotape twice, first using speech and a second time using only their hands. We examined whatever gestures the adults produced in their first pass through the events (the Gesture+Speech condition) and compared them to the gestures they produced in their second pass (the Gesture condition). As predicted, we found that the adults' gestures resembled the deaf children's in the Gesture condition but not the Gesture+Speech condition.

Specifically, in the Gesture condition, the adults frequently combined their gestures into strings and those strings were reliably ordered, with gestures for certain semantic elements occurring in particular positions in the string. Interestingly, the gesture order that the adults used did *not* follow canonical English word order. For example, to describe a girl jumping into a hoop, the adults gestured "hoop girl jump," rather than the more typical English order "girl jump hoop". In contrast, although the adults did produce gestures that stood for actions and objects in the Gesture+Speech condition, they rarely combined those gestures into strings. In this condition, the adults used their gestures as representations for global wholes rather than as building blocks for larger units.

5.2 Ergative structure emerges again

All of the vignettes in our initial study involved objects and people moving about in space, events that elicit intransitive sentences. With only these stimuli, we were unable to determine whether the adults used ergative constructions in their gestures, as did the deaf children. We therefore conducted a second study with vignettes involving some events that elicit intransitive sentences and others that elicit transitive sentences (Goldin-Meadow, Yalabik and Gershkoff-Stowe 2000). We included only the Gesture condition in this study — the adults were asked to describe each scene using their hands and not their mouths. Because we were interested in whether there would be changes in the gestures over time, we arranged for the two adults to meet twice a week for several weeks.

We used the same system of analysis for the adults as we did for the deaf children. We looked at gesture strings that could have contained three semantic elements but, in fact, only contained two (e.g., transitive sentences with an underlying structure of actor-act-patient, and intransitive sentences with an underlying structure of actor-act-recipient). Both adults produced gestures for intransitive actors as often as they produced gestures for patients, and far more often than they produced gestures for transitive actors. In other words, they displayed the same ergative pattern seen in the deaf children's gestures. And they did so immediately — the ergative pattern was evident in the adults' initial sessions and did not change over time (see Figure 1, bottom graphs; Goldin-Meadow 2002b).

In terms of gesture order, both adults produced gestures for intransitive actors in first position of their two-gestures sentences (e.g., 'mouse runs'). This result is hardly surprising as the pattern parallels typical word order for intransitive actors in English. Neither adult produced many gestures for transitive actors, which made it impossible to determine an order preference for this semantic element. More interestingly, both gesturers also produced gestures for patients in first position of their two-gesture sentences ('cheese eat'). Not only is this pattern identical to the deaf children's gesture order for patients, but it is also different from the pattern typically found in English (i.e., "eat cheese").

The patient-first pattern is particularly interesting in the adults. The deaf children often (although not always) used pointing gestures to convey patients. The adults were unable to take advantage of this strategy simply because there were no objects in the room to point at. The adults were forced to invent an iconic gesture for their patients — for example, a smoking movement at the mouth to refer to an ashtray, which was then followed by a gesture representing the action that was done on that ashtray (e.g., a throwing action). Despite the fact that they used iconic rather than pointing gestures to refer to patients, the adults followed the same (non-English) ordering patterns as the deaf children.

Thus, when asked to describe a series of action vignettes using their hands rather than words, English-speaking adults invent an ergative structure identical to the one developed by the deaf children, rather than the accusative pattern found in their own spoken language. Ergative structure is *not* unique to child language-creators and therefore cannot reflect a child-like way of organizing information for communication. Rather, the ergative pattern may reflect a robust solution to the problem of communicating information from one mind to another, be it an adult or child mind.

5.3 Morphological structure does not emerge

The emergence of production probability and gesture ordering patterns immediately in the experimental paradigms that we used with adults underscores the resilience of these grammatical properties in symbolic human communication. With no time for reflection, the adults in our studies constructed a set of gesture sentences in which thematic roles were marked by gesture production probability and gesture ordering. Interestingly, however, our simple experimental paradigm was not sufficient to support the emergence of all of the grammatical properties found in the deaf children's gesture systems. The adults' gestures were not systematically organized into a system of internal contrasts, that is, into a morphology (Goldin-Meadow *et al.* 1996).

When the hearing adults generated a gesture, their goal was to produce a handshape that adequately represented the object, and their choice of handshapes appeared to be constrained only by their imaginations and the physical limitations imposed by their hands (gesture-to-world relations). For example, a hearing adult might produce a different handshape for each of the five airplanes on the test, with each handshape capturing an idiosyncratic property of the airplane pictured in that event. In contrast, when the deaf children in our studies generate a gesture, their choice of handshapes is guided not only by how well the handshape captures the features of the object, but also by how well that handshape fits into the set of handshapes allowed in their individual gesture systems (gesture-to-gesture relations). Thus, they use the same handshape for all airplanes (indeed, for all vehicles), regardless of their individual idiosyncracies, and this handshape contrasts with the handshape used to represent, say, curved objects. The fact that adults instantly invent a gesture system with devices for marking thematic roles but without a system of internal contrasts suggests that some properties of language may be more resilient than others.

6. Language-creation without a willing partner

The deaf children are creating their gesture systems without the benefit of a language model and, in this sense, are like the original creators of language. However, their language-creation situation is clearly not a simulation of the situation in which language was created for the first time. The deaf children are developing their communication systems in a world in which language and its consequences are pervasive. It may not be necessary for a child to be exposed to a language model in order to create a communication system with language-like structure. However, it may well be necessary for that child to experience the human cultural world. It is very likely that, as language evolved, the cultural artifacts that characterize our world evolved along with it. Indeed, Hockett (1977: 149) argues that the ability to carry artifacts (in particular, tools) and the ability to refer to objects that are not visible (communication beyond the here-and-now) developed side-by-side, each developing in small increments furthered by the already-achieved increments of itself and of the other. The deaf children in our studies, while lacking conventional language, nevertheless had access to the artifacts that evolved along with language and that could have served as supports for the child's invention of a language-like system for communicating both within and beyond the here-and-now.

Although not lacking modern-day artifacts, the children did lack a communication partner who was willing to create a system with them. The families in our study chose to educate their deaf children through an oral method and their focus was on their children's verbal abilities. The families did not treat the children's gesture as though it were a language and, in this sense, they were not partners in their children's gestural communication. As a result, the children were forced to keep their gestures iconic — if they had invented relatively arbitrary symbols, no one would have been able to understand them. This aspect of the deaf children's gesture-creation situation may, in fact, be comparable to the original language-creation scene. The first language-creator is very likely to have been communicating with others who (like the deaf children's hearing parents) were willing, even eager, to interact with the creator, but did not share the creator's emerging language.[3]

As another symptom of their unusual communication situation, the deaf children were producers, but not receivers, of their gestures. The gestures that the deaf children saw were the gestures that their hearing parents produced as they talked, gestures that were qualitatively different in both form and function from the gestures that the children themselves produced (cf. Goldin-Meadow *et al.* 1996). The give-and-take between parent and child thus never became seamless and automatic. The interesting point is how far these deaf children could go in constructing a grammatical system despite the fact that they experienced no pressure to automatize their gestures.

If grammar is truly “an instrument of speeded-up, more automated language processing” (Givón 1998: 92), we might expect that the deaf children would have no need for it. The fact that the deaf children did invent gesture systems with both morphological and syntactic structure suggests that at least a rudimentary grammatical system can evolve without the pressure to speed up processing. There must be other forces propelling the deaf children (and perhaps all language-creators and -learners) toward grammatical structure.

In fact, I suggest that there may be two types of forces at work. Some forces lead to grammatical structure *immediately* whenever two individuals (children or adults) attempt to communicate with one another, whether those individuals know a conventional language or not. Freyd (1983) has suggested that certain knowledge structures emerge because of their social utility at the group level rather than at the individual level. For example, scenes are perceived seamlessly along continuous dimensions, yet are made accessible to the minds of others by fitting individual experiences into discrete, conventional categories. Perhaps the production and ordering regularities that we observed in the deaf children’s and adults’ gestural communications emerged because the task required imparting information to another person. An alternative and equally likely possibility, however, is that the regularities grew out of the way the children and adults conceptualized and parsed the scenes (Gershkoff-Stowe and Goldin-Meadow 2002). Whatever the cause, the effect is that sentence-level structure is introduced immediately into communication that is created *de novo*.

Another set of forces seem to work *over time* and lead to the evolution of a system. These forces transform the initial communication into a system characterized by word-level structure and grammatical categories, as well as sentence-level structure. The push to systematize may come from the state of the communication itself — that is, the communication may become sufficiently cumbersome that reorganization becomes a necessity. Children may begin to have trouble recalling their gestures and, in response, may find it necessary to organize their gestures in relation to one another (and not just in relation to the world); that is, they may find it necessary to introduce morphological structure into their gestures. Similarly, children may begin to have trouble creating different gestures for nouns and verbs and, in response, may introduce grammatical devices that mark nouns differently from verbs and thus allow the same gesture form to be used for both functions. Finally, the fact that, in David, the changes in morphological structure and grammatical categories occurred at the same moment in development raises the intriguing possibility that these changes are synergistic and must both occur in order for either to occur.

Our challenge is to discover the forces that shape gesture creation in deaf children — for these are the forces that are likely to play a role in language creation every time it takes place, perhaps even the very first time.

Notes

* This research was supported by BNS 8810769 from the National Science Foundation, and RO1 DC00491 from the National Institute of Deafness and other Communicative Disorders.

1. The newly developing Nicaraguan Sign Language is a creole-like situation which is taking place right now. Opening the first school for the deaf in Managua in the late 1970s created an opportunity for deaf children to interact with one another for the very first time. The result was a sign language. This language, which resembles pidgin languages in many respects, is now being learned by a second generation of young signers who are introducing new grammatical structure into the language (Kegl, Senghas and Coppola 1999; Senghas 1995, 2000; Senghas and Coppola 2001; Senghas, Coppola, Newport and Supalla 1997).

2. We follow sign language researchers in using the term "inflection" for the displacement of gestures away from neutral space (the chest-level area). The directionality of an inflecting verb reflects agreement of the verb with its subject or object, just as a verb in English agrees with its subject in number. Verbs in ASL agree with the person (I, you, he/she/it) of its subject or object. The 1st person affix places the sign near the signer's body; the 2nd person affix places the sign in the direction of the addressee; and the 3rd person affix places the sign at the locus assigned to that entity (Padden 1983).

3. I thank Eric Pederson for this insight.

References

Aksu-Koc, A.A., and Slobin, D.I. 1985. "The acquisition of Turkish". In *A Cross-Linguistic Study of Language Acquisition*, Volume I, D.I. Slobin (ed), 839–78. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bickerton, D. 1990. *Language and Species.* Chicago: University of Chicago Press.

Bickerton, D. 1999. "Creole languages, the language bioprogram hypothesis, and language acquisition". In *Handbook of Child Language Acquisition*, W.C. Ritchie and T.K. Bhatia (eds), 195–220. New York: Academic Press.

Bloom, L. 1973. *One Word at a Time.* The Hague: Mouton.

Bloom, P. 1990. *Semantic Structure and Language Development.* Doctoral dissertation, Massachusetts Institute of Technology.

Bonvillian, J.D., Orlansky, M.O., and Novack, L.L. 1983. "Developmental milestones: Sign language acquisition and motor development". *Child Development* 54: 1435–45.

Bowerman, M. 1982. "Reorganizational processes in lexical and syntactic development". In *Language Acquisition: The State of the Art*, E. Wanner and L.R. Gleitman (eds). Cambridge: Cambridge University Press.

Brown, R. 1958. *Words and Things.* New York: The Free Press.

Brown, R. 1973. *A First Language.* Cambridge, Mass.: Harvard University Press.

Butcher, C., Mylander, C., and Goldin-Meadow, S. 1991. "Displaced communication in a self-styled gesture system: Pointing at the non-present". *Cognitive Development* 6: 315–42.

Bybee, J. 1998. "A functionalist approach to grammar and its evolution". *Evolution of Communication* 2: 249–78.

Conrad, R. 1979. *The Deaf Child*. London: Harper and Row.

Dixon, R.M.W. 1979. "Ergativity". *Language* 55: 59–138.

Dixon, R.M.W. 1994. "Adjectives". In *The Encyclopedia of Language and Linguistics*, R.E. Asher (ed). New York: Pergamon Press.

DuBois, J.W. 1987. "The discourse basis of ergativity". *Language* 63: 805–55.

Fant, L.J. 1972. *Ameslan: An Introduction to American Sign Language*. Silver Springs, Md.: National Association of the Deaf.

Fortescue, M., and Olsen, L.L. 1992. "The acquisition of West Greenlandic". In *A Cross-Linguistic Study of Language Acquisition*, Volume III, D.I. Slobin (ed), 111–219. Hillsdale, NJ: Lawrence Erlbaum Associates.

Freyd, J.J. 1983. "Shareability: The social psychology of epistemology". *Cognitive Science* 7: 191–210.

Gee, J.P. and Goodhart, W. 1985. "Nativization, linguistic theory, and deaf language acquisition". *Sign Language Studies* 49: 291–342.

Gershkoff-Stowe, L. and Goldin-Meadow, S. 2002. "Is there a natural order for expressing semantic relations". To appear in *Cognitive Psychology*.

Givón, T. 1998. "On the co-evolution of language, mind and brain". *Evolution of Communication* 2: 45–116.

Goldin-Meadow, S. 1997. "The resilience of language in humans". In *Social Influences on Vocal Development*, C.T. Snowdon and M. Hausberger (eds), 293–311. Cambridge: Cambridge University Press.

Goldin-Meadow, S. 1982. "The resilience of recursion: A study of a communication system developed without a conventional language model". In *Language Acquisition: The State of the Art*, E. Wanner and L.R. Gleitman (eds), 51–77. Cambridge: Cambridge University Press.

Goldin-Meadow, S. 1987. "Underlying redundancy and its reduction in a language developed without a language model: The importance of conventional linguistic input". In *Studies in the acquisition of anaphora: Applying the constraints*, Volume II, B. Lust (ed), 105–33. Boston, Mass.: D. Reidel.

Goldin-Meadow, S. 2002a. "Lexical development with out a language model". To appear in *Weaving a Lexicon*, G. Hall and S. Waxman (eds). Cambridge, Mass.: MIT Press.

Goldin-Meadow, S. 2002b. "Thought before language: Do we think ergative". In *Language in Mind: Advances in the Study of Language and Thought*, D. Gentner and S. Goldin-Meadow (eds). Cambridge, Mass.: MIT Press (in press).

Goldin-Meadow, S., Butcher, C., Mylander, C., and Dodge, M. 1994. "Nouns and verbs in a self-styled gesture system: What's in a name?" *Cognitive Psychology* 27: 259–319.

Goldin-Meadow, S., McNeill, D., and Singleton, J. 1996. "Silence is liberating: Removing the handcuffs on grammatical expression in the manual modality". *Psychological Review* 103: 34–55.

Goldin-Meadow, S., and Mylander, C. 1983. "Gestural communication in deaf children: The non-effects of parental input on language development". *Science* 221: 372–4.

Goldin-Meadow, S. and Mylander, C. 1984. "Gestural communication in deaf children: The effects and non-effects of parental input on early language development". *Monographs of the Society for Research in Child Development* 49: 1–121.

Goldin-Meadow, S. and Mylander, C. 1990. "The role of a language model in the development of a morphological system". *Journal of Child Language* 17: 527–63.

Goldin-Meadow, S. and Mylander, C. 1998. "Spontaneous sign systems created by deaf children in two cultures". *Nature* 91: 279–81.

Goldin-Meadow, S., Mylander, C., and Butcher, C. 1995. "The resilience of combinatorial structure at the word level: Morphology in self-styled gesture systems". *Cognition* 56: 195–262.

Goldin-Meadow, S., Yalabik, E., and Gershkoff-Stowe, L. 2000. "The resilience of ergative structure in language created by children and by adults". *Proceedings of the Boston University Conference on Language Development* 24: 343–53.

Goodhart, W. 1984. *Morphological Complexity, ASL and the Acquisition of Sign Language in Deaf Children.* Ph.D. dissertation, Boston University.

Hockett, C.F. 1977. *The View from Language: Selected Essays 1948–74.* Athens, Ga: University of Georgia Press.

Hoffmeister, R. 1978. *The Development of Demonstrative Pronouns, Locatives and Personal Pronouns in the Acquisition of American Sign Language by Deaf Children of Deaf Parents.* Ph.D. dissertation, University of Minnesota.

Hoffmeister, R., and Wilbur, R. 1980. "Developmental: The acquisition of sign language". In *Recent Perspectives on American Sign Language*, H. Lane and F. Grosjean (eds). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hopper, P.J. and Thompson, S.A. 1984. "The iconicity of the universal categories 'noun' and 'verbs'". In *Iconicity in Syntax*, J. Haiman (ed), 151–83. Amsterdam: John Benjamins.

Hopper, P.J. and Thompson, S.A. 1988. "The discourse basis for lexical categories in universal grammar". *Language* 60: 703–52.

Huttenlocher, J. and Smiley, P. 1987. "Early word meanings: The case of object names". *Cognitive Psychology* 19: 63–89.

Huttenlocher, J. and Smiley, P. 1989. "An emerging lexicon: Acquiring words for events". Ms., University of Chicago.

Karmiloff-Smith, A. 1992. *Beyond Modularity: A Developmental Perspective on Cognitive Science.* Cambridge, Mass.: MIT Press.

Kegl, J., Senghas, A., and Coppola, M. 1999. "Creation through contact: Sign language emergence and sign language change in Nicaragua". In *Language Creation and Language Change: Creolization, Diachrony, and Development*, M. DeGraff (ed), 179–237. Cambridge, Mass.: MIT Press.

Lenneberg, E.H. 1964. "Capacity for language acquisition". In *The Structure of Language: Readings in the Philosophy of Language*, J.A. Fodor and J.J. Katz (eds), 579–603. Englewood Cliffs, NJ: Prentice-Hall.

Livingston, S. 1983. "Levels of development in the language of deaf children". *Sign Language Studies* 40: 193–286.

Macnamara, J. 1982. *Names for Things.* Cambridge, Mass.: MIT Press.

MacWhinney, B. 1978. “The acquisition of morphophonology”. *Monographs of the Society for Research in Child Development* 43: 1–22.

Mayberry, R. I. 1992. “The cognitive development of deaf children: Recent insights”. In *Child Neuropsychology*, Volume VII: *Handbook of Neuropsychology*, S. Segalowitz and I. Rapin (eds), 51–68. Amsterdam: Elsevier.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought.* Chicago: University of Chicago Press.

Meadow, K. 1968. “Early manual communication in relation to the deaf child’s intellectual, social, and communicative functioning”. *American Annals of the Deaf* 113: 29–41.

Moores, D. F. 1974. “Nonvocal systems of verbal behavior”. In *Language Perspectives: Acquisition, Retardation, and Intervention*, R. L. Schiefelbusch and L. L. Lloyd (eds). Baltimore: University Park Press.

Morford, J. P., and Goldin-Meadow, S. 1997. “From here to there and now to then: The development of displaced reference in homesign and English”. *Child Development* 68: 420–35.

Newport, E. L. 1984. “Constraints on learning: Studies in the acquisition of American Sign Language”. *Papers and Reports on Child Language Development* 23: 1–22.

Newport, E. L. and Meier, R. P. 1985. “The acquisition of American Sign Language”. In *The Cross-Linguistic Study of Language Acquisition*, Volume I: *The data.* D. I. Slobin (ed). Hillsdale, NJ: Lawrence Erlbaum Associates.

Padden, C. 1983. *Interaction of Morphology and Syntax in American Sign Language.* Ph.D. dissertation, University of California at San Diego.

Petitto, L. A. 1992. “Modularity and constraints in early lexical acquisition: Evidence from children’s early language and gesture”. In *Minnesota Symposium on Child Psychology*, Volume XXV, M. Gunnar (ed), 25–58. Hillsdale, NJ: Lawrence Erlbaum Associates.

Phillips, S. B., Goldin-Meadow, S., and Miller, P. J. 2001. “Enacting stories, seeing worlds: Similarities and differences in the cross-cultural narrative development of linguistically isolated deaf children”. *Human Development* 44: 311–36.

Pye, C. 1992. “The acquisition of K’iche’ Maya”. In *The Cross-Linguistic Study of Language Acquisition*, Volume III, D. I. Slobin (ed). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sapir, E. 1921. *Language: An Introduction to the Study of Speech.* New York: Harcourt Brace Jovanovich.

Schulman, B. W., Mylander, C., and Goldin-Meadow, S. 2001. “Ergative structure at sentence and discourse levels in a self-generated communication system”. *Proceedings of Boston University Conference on Language Development* 25: 815–24.

Senghas, A. 1995. “The development of Nicaraguan Sign Language via the language acquisition process”. *Proceedings of Boston University Child Language Development* 19: 543–52.

Senghas, A. 2000. “The development of early spatial morphology in Nicaraguan Sign Language”. *Proceedings of Boston University Child Language Development* 24: 696–707.

Senghas, A., and Coppola, M. 2001. “Children creating language: How Nicaraguan Sign Language acquired a spatial grammar”. *Psychological Science* 12: 323–8.

Senghas, A., Coppola, M., Newport, E. L., and Supalla, T. 1997. “Argument structure in Nicaraguan Sign Language: The Emergence of grammatical devices”. *Proceedings of Boston University Child Language Development* 21: 550–61.

Silverstein, M. 1976. "Hierarchy of features and ergativity". In *Grammatical Categories in Australian Languages*, R.M.W. Dixon (ed), 112–71. Canberra: Australian Institute of Aboriginal Studies.

Singleton, J.L., Morford, J.P., and Goldin-Meadow, S. 1993. "Once is not enough: Standards of Well-Formedness in Manual Communication Created over Three Different Timespans". *Language* 69:683–715.

Singleton, J.L. and Newport, E.L. 2002. "When learners surpass their models: The acquisition of American Sign Language from impoverished input". To appear in *Cognitive Psychology*.

Supalla, S. 1991. "Manually Coded English: The modality question in signed language development". In *Theoretical Issues in Sign Language Research*, Volume II: *Acquisition*. P. Siple and S. Fischer (eds). Chicago: University of Chicago Press.

Supalla, T., Newport, E.L., Singleton, J.L., Supalla, S., Metlay, D., and Coulter, G. In press. *Test Battery for American Sign Language Morphology and Syntax*. Burtonsville, Md.: Linstok Press.

Suty, K.A., and Friel-Patti, S. 1982. "Looking beyond Signed English to describe the language of two deaf children". *Sign Language Studies* 35:153–68.

Tervoort, B.T. 1961. "Esoteric symbolism in the communication behavior of young deaf children". *American Annals of the Deaf* 106:436–80.

Thompson, S.A. 1988. "A discourse approach to the cross-linguistic category 'adjective'". In *Explaining language universals*, J.A. Hawkins (ed), 167–85. Cambridge, Mass.: Blackwell.

Chapter 15

Language evolution, acquisition and diachrony: Probing the parallels

Dan I. Slobin
University of California, Berkeley

ontogeny. [f. onto- + Gr. -cemeia birth, production.]
1. The origin and development of the individual being; = ontogenesis.
1872 *Microsc. Jrnl.* July 185 'The ontogeny of every organism repeats in brief . . . its phylogeny', i.e. the individual development of every organism . . . repeats approximately the development of its race. (*Oxford English Dictionary*)

Manifestly, the modern learner is in quite a different position to primitive man, and has quite a different task set him: the task of the child is to learn an existing language, but not in the least to frame anything anew. (Jespersen 1921/1964: 417)

In each of the three developmental processes pertaining to human language evolution, acquisition, diachrony — the very same sequence seems to have been involved. (Givón 1998: 102)

Givón has faced us with a challenge: Can we find common processes to account for seeming parallels between the development of language in the individual and changes in languages over time; and can we propose that the same processes might account for the origins of language in genus *Homo* and species *sapiens*? Some 25 years ago I explored some of these parallels in our own species: "Developmental psycholinguistics, in dealing with diachronic processes in the individual, shares much common ground with historical linguistics, with studies of languages in contact, and with the investigation of the evolution of pidgin and creole languages" (Slobin 1977: 185). However, my goal then was to use these various sorts of language change as a window into the dynamic equilibrium arising between competing pressures for semantic transparency, ease of processing, and expressivity. In that framework, Language is a constant, and change serves to illuminate the constancy. The present quest is for the origins of Language itself, and here I am less sanguine about the contributions of child language study. I will examine four longstanding proposals about the role of the child in helping us to understand language evolution and diachrony:

1. The course of language development in the child has parallels with the development of language in our genus and species.
2. The sequence of development of linguistic forms in the child mirrors the diachronic development of those forms in the language that the child is acquiring.
3. Diachronic language change is, in part, the product of the learner.
4. The emergence of new languages — creoles and Nicaraguan Sign Language — is due to child learners.

1. Does ontogeny recapitulate phylogeny?

The *Homo sapiens* child is exposed to an already evolved human language, and has a brain that evolved to make use of such a language, as Jespersen (1921/1964) already pointed out early in the last century. Nevertheless, might it be that the earliest periods of child language reveal the workings of a cognitive and linguistic core that we might share with our hominid, and even pre-hominid ancestors? Bickerton has been explicit about this parallel, on the basis of two sorts of claims: (1) a "proto-language" that is shared by trained apes and toddlers, and (2) human postnatal brain growth:

> We may conclude that there are no substantive formal differences between the utterances of trained apes and the utterances of children under two. The evidence of children's speech could thus be treated as consistent with the hypothesis that the ontogenetic development of language partially replicates its phylogenetic development. The speech of under-twos would then resemble a stage in the development of the hominid line between remote, speechless ancestors and ancestors with languages much like those of today.
>
> Haeckel's claim that ontogeny repeats phylogeny has had a checkered career in the history of biology, and certainly cannot stand as a general law of development. However, it may have application in limited domains. In particular, no one should be surprised if it applies to evolutionary developments that are quite recent and that occur in a species whose brain growth is only 70 percent complete at birth and is not completed until two or more years afterwards. (Bickerton 1990:115)

Givón has long taken a similar position, presented as "Arguments for a neo-recapitulationist view" (1979:273 ff.) and reasserted in 1998, arguing for "proto-grammar" in children, pidgins, and phylogeny (pp. 93 ff.) — albeit in a richly complex "paradigm case of co-evolution, a multi-causal interaction of cultural, behavioral, communicative and neuro-cognitive factors" (p. 105). There are, however, problems in making analogies from child language to simpler ancestral languages. I will point out some of the most salient.

1.1 Trained apes, human toddlers, and proto-language

Cross-species comparisons are difficult, but the temptations to see bits of ourselves in our cousins — or to deny such similarities — are strong and enduring. Savage-Rumbaugh's reports of the accomplishments of bonobos (2001; Savage-Rumbaugh *et al.* 1998) make it clear that many of the prerequisites for human language were already present before the emergence of the hominid line. Bonobos can comprehend spoken English sentences — without instruction. The capacities for segmentation, lexical mapping, and some levels of syntax are thus ancient. Savage-Rumbaugh (2001) even presents evidence for English-based vocal production and writing in bonobos. All of this raises fascinating questions about the evolution and functions of these capacities; but such questions lie outside of the search for parallels in human ontogeny. Certainly, as Savage-Rumbaugh (2001) points out: "These findings render mute old questions regarding the innate limits of the ape brain." They also make it clear that additional factors — both cognitive and social — must have been necessary for the emergence of language.

We cannot predict what new surprises will come from bonobos, but for the moment at least, they have not been given the opportunity to acquire a rich morphological language such as Turkish (agglutinative) or Inuktitut (polysynthetic). Children under 2 who are exposed to such languages do not exhibit the sort of "pre-grammatical" speech described by Bickerton, Givón, and others, such as absence of grammatical morphology and reliance on topic-comment word order. Turkish toddlers show productive use of case inflections on nouns as early as 15 months of age — that is, productive morphology at the one-word stage. They quickly come to use multiple suffixes on nouns, and by the age of 24 months or younger demonstrate full mastery of the nominal inflectional system and much of the verbal paradigm (Aksu-Koç and Slobin 1985; Küntay and Slobin 1999). They also appropriately use pragmatic word-order variation as young as 17 months (Ekmekçi 1986). Aksu-Koç and Slobin provide the following summary of the child studied by Ekmekçi:

> Early control of the functions of word order is reflected in a number of contrastive uses, including the following: (1) Preposed adjectives are used in attributive expressions (e.g. *soğuk su* 'cold water', said at 1;7 when asking for cold water), whereas postposed adjectives are used in predicative expressions (e.g., *çorba sıcak* 'soup hot', said at 2;0 as a complaint). (2) Indefinite or nonreferential direct objects always directly precede the verb (e.g. *kalem getir* 'bring (a) pencil'), whereas definite direct objects (marked by the accusative inflection) can also follow the verb (e.g. both *kalem-i getir* 'pencil-ACC bring' and *getir kalem-i* 'bring pencil-ACC'='bring the pencil') [age 1;10]. (Aksu-Koç and Slobin 1985:856)

Inuktitut children also use morphology productively at the one-word stage, and make appropriate use of nominal ellipsis from age 2 (Allen 1996; Fortescue and Lennert Olsen 1992). Early telegraphic speech (the prototype of "pre-grammar") appears to be characteristic only of certain types of languages.

On the other hand, work by Tomasello (1992, 1999) and Lieven (Lieven *et al.* 1997) on detailed corpora of very early child speech in a number of languages, makes it clear that much of early language is item-based, rather than reflecting productive combinations of the telegraphic or pre-grammatical type. More detailed study of agglutinative and polysynthetic languages may well reveal similar inventories of item-based schemas, though I expect that the evidence of early morphological productivity will stand. That is, early learners are good at extracting salient grammatical devices in the exposure language, as demonstrated, for example, in my work on Operating Principles (Slobin 1973, 1985).

The influence of the exposure language is especially evident in the case of bilingual children. A number of investigators report that such children — as soon as they begin to produce two-word and multiword combinations — differentiate the word-order patterns of their two languages (e.g., Meisel 1989 for French–German bilinguals, De Houwer 1990 for Dutch–English; Deuchar 1992 for Spanish–English bilinguals). These children do not show a standard "pre-grammar" or "proto-language" in which the two languages are differentiated only by choice of lexical items; rather, they are differentially shaped by each of the exposure languages from very early on.

In any event, with further research on both chimpanzees and human children, characterizations of "proto-language" or "proto-grammar" will be in need of revision. And, as Jespersen emphasized (see epigraph), all such reduced linguistic varieties are derived from a developed exposure language. The child or trained chimpanzee is sampling from an existing language, and not creating without input. The structures of early language production are not independent of the structures of the exposure language. (Proposals about creation without or beyond input are taken up with regard to the fourth question, and are discussed in detail in the papers in this volume by Susan Goldin-Meadow and Jill Morford.) While the ways in which children sample from existing languages tells us a great deal about the workings of the human mind, it is not evident that any generalizations can be drawn about pre-human minds from such evidence.

1.2 Heterochrony

Another obstacle to comparing chimps and human babies — despite the huge genetic overlap between the species — is the difference in *timing* of onset and offset of abilities. A quarter-century ago Steven Jay Gould, in *Ontogeny and Phylogeny*, argued for "the evolutionary importance of *heterochrony* — changes in the relative

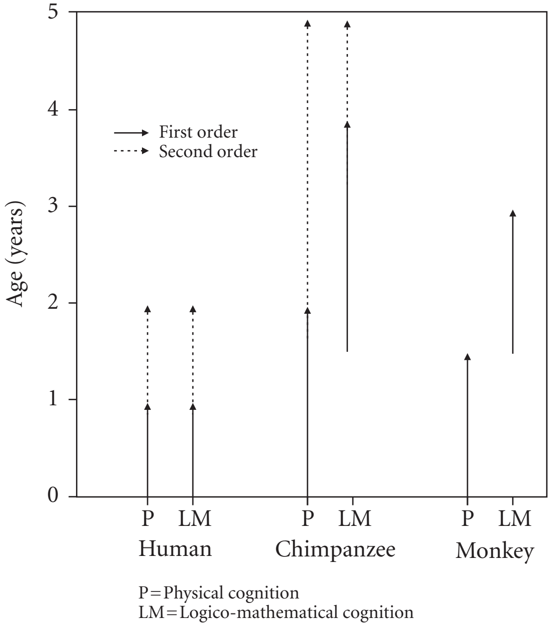


Figure 1. Comparative cognitive development: vectorial trajectories of developmental onset age, velocity, sequence, and organization (but not extent of offset age) (Langer 2001:31)

time of appearance and rate of development for characters already present in ancestors." In his conception, human development is retarded in relation to other primates. However, more recently, Jonas Langer (2000, 2001) has examined heterochrony with regard to several dimensions of cognitive development that are critical for our topic. He applied comparable tests of physical cognition (causality) and logicomathematical cognition (classification) to human infants, two species of great apes (common chimpanzees, bonobos), and two species of monkeys (cebus, macaques). There are two important findings for our purposes: (1) Human cognitive development is accelerated in comparison to the other species (that is, we are not simply neotenized primates). (2) The two sorts of cognition develop in parallel in humans, but asynchronously in apes and monkeys. These heterochronic relations are evident in Figure 1.

The consequence is that physical and logicomathetical cognition can interact from the start in human babies, whereas logicomathematical capacities are not available to apes and monkeys during the early phases of establishing physical cognition. Note also that second-order cognition appears early and synchronously in both domains for humans, allowing for immediate interaction between two types of cognition at a higher level. By contrast, second-order cognition in chimpanzees emerges for physical cognition when the animals have just begun to work out first-order cognition for classification (and second-order cognition has not been observed in monkeys). The opportunities for ontogenetic interaction between cognitive capacities thus varies significantly across species, due to heterochronic effects. Langer's work provides a strong critique of recapitulationism:

> Such phylogenetic displacements in the ontogenetic covariations between the onset, velocity and offset of cognitions in relation to each other disrupt potential repetitions (i.e. recapitulation) of phylogeny in ontogeny. Thus, human, and for that matter chimpanzee, cognitive ontogeny does not simply recapitulate its phylogeny. Instead, heterochronic evolution reorganizes primate species cognitive development. Significant consequences for their respective potential growth follows.
>
> Heterochronic evolutionary reorganization of asynchronic into descendant progressively synchronic development opens up cascading possibilities for cognitions to influence each other and to be influenced by similar environmental influences. (Langer 2000: 374)

Within our species, temporal covariation of cognitive and linguistic abilities shapes the emergence of language. Developing constraints of memory and attention, along with available cognitive structures of all sorts, are responsible for the nature and course of language development. These constraints and their timing vary from species to species, and we cannot know how such factors might have played themselves out in extinct ancestors. Furthermore, as argued below, I suggest that modifications of existing languages generally do not come from very young learners, but from more mature participants in social and linguistic interaction. That is, in humans, much of linguistic innovation is due to individuals who are advanced in cognitive and social development. Comparisons of human toddlers with apes and monkeys are therefore of very limited applicability to the task at hand. (And the long line of more relevant species is, alas, extinct.)

2. Does diachrony recapitulate ontogeny?

There are striking parallels between patterns of language development in individual children and repeated diachronic changes in languages. One is tempted to propose that similar cognitive processes underlie both sorts of development. Grammatical markers originate as lexical items that are gradually specialized in function and reduced in form along a cline of grammaticization. (Consider, for example, the development of the English verb *go* to the future marker *gonna*.) Hopper and Traugott (1993: 7) provide a diagrammatic summary of the cline:

content item > grammatical word > clitic > inflectional affix

The meanings of the content words that begin such diachronic changes tend to be general and basic — for example, verbs such as ‘go’, ‘do’, ‘have’, ‘want’, ‘finish’, ‘take’. Bybee *et al.* (1994: 10) note that such lexical items “encode major orientation points in human experience. They are the most generalized verbs of existence, possession, physical location, or attitude and movement in space.” Eve Clark (1978) has observed that verbs of this sort (“light verbs”) are the most frequent verbs in early child language across a range of languages. Children acquiring English, Finnish, French, Japanese, and Korean make early and frequent use of verbs meaning ‘go’, ‘put’, ‘get’, ‘do’, and ‘make’. The striking overlaps between these two sets of verbs — those that are the sources of grammatical morphemes and those that predominate in early child language — suggest recapitulation. In this case, it seems that diachrony recapitulates ontogeny; that is, languages, like children, start with the basics. I’ve examined this proposal in two domains — case-marking and tense/aspect — and have concluded that the parallels are illusory .

2.1 The origins and extensions of accusative markers

In the histories of Mandarin Chinese, Chickasaw, a number of West African languages of the Benue-Kwa group, and other languages, a verb that means ‘take’ or ‘hold’ develops into some sort of accusative case-marker. In child language development, across a range of languages, the earliest uses of accusative markers are found when the child is talking about prototypical direct manipulation of objects — what I have called “The Manipulative Activity Scene.” The semantics of ‘take’ and ‘hold’ thus correspond to a salient perspective in ontogeny. Could it be that grammatical markers begin as light verbs because languages, like children, build up from basic semantic notions? To answer this question, we have to look at the communicative contexts in which grammaticization and acquisition take place. Lexical items begin on the path towards grammatical elements in the process of conversational inferences carried out by mature speakers, already equipped with complex grammar and lexical semantics. For example, consider the grammaticization of the Mandarin object marker *bǎ*, which starts off in the fifth century BC as a verb meaning ‘take hold of’. In the time of the Tang Dynasty (seventh–ninth centuries AD) it occurs in serial-verb constructions that allow for reanalysis of its function. For example the following sentence can be interpreted in two different ways:

(7) Zuì bǎ zhū-gēn-zǐ xì kàn.
drunk *bǎ* dogwood careful look

The normal serial-verb interpretation would be: ‘While drunk, (I) *took* the dogwood and carefully looked at it'. But a verb meaning 'take' could also be heard as simply reinforcing the act of examining something that is being held, without adding another meaning element: 'While drunk, (I) carefully looked at the dogwood'. That is, the hearer is "invited" to consider a single act, *looking carefully*, rather than two acts, *taking and then looking carefully*. This sort of conversational inference is proposed as the mechanism of grammaticization (e.g., in Modern Mandarin, *bǎ* is no longer a full verb, and it serves as an object marker in the construction SUBJECT *bǎ* DIRECT.OBJECT VERB). For example:

(8) Nǐ bǎ jiǔ màn-màn-de hē.
you *bǎ* wine slowly drink
'You drink the wine slowly!'

Can we use 2-year-old attention to the Manipulative Activity Scene to help us understand this sort of grammaticization? I think not. For a 2-year-old, manipulation is a salient activity, and so a grammatical marker that regularly occurs in reference to manipulation will also become salient. Markers such as accusative inflections are noticed in situations that are referred to by verbs such as 'take' and 'put'. Therefore, for some children, the first accusatives are limited to the direct objects of such verbs (Slobin 1985). Note that the mechanism proposed here is based on 2-year-old interactions with the world. But adult speakers of Tang Dynasty Mandarin did not set out to find a grammatical marker for the basic notion of manipulation, since they already had a fully developed language. As adults, they were able to focus on all sorts of activities and relationships that are not salient to 2-year-olds. When they used a verb meaning 'take' — in serial-verb constructions, in particular interactive contexts — that verb could come to be interpreted as a marker of abstract patienthood. However, this is not because the basic notion of patient is associated with manipulation in the adult mind. As Bybee has pointed out, the verbs that become grammaticized are *general* in their meaning. A verb meaning 'grab', for example, does not become a patient marker. This is due to conversational pragmatics: If a speaker uses a verb like 'grab', the manner of action is in focus, whereas a neutral verb like 'take' conveys no such added meaning. Therefore verbs of general meaning can provide starting points for grammaticization. However, in child language, clauses with highly dynamic verbs like 'grab' are good candidates for accusative marking on the associated patient noun. Clearly, the dynamics of early grammar formation and the dynamics of adult grammaticization are quite different. Children use salient perspectives to discover the meanings of existing grammatical forms, beginning with limited meanings. Adults abstract away from salient perspectives to create general grammatical markers.

2.2 Origins of the present perfect

In many languages, a present perfect tense/aspect form develops as a reanalysis of an attributive construction, in which a resultant state is conceived of as a possession, as in Latin *habeo virgulum fissam* 'I possess a twig that is split' → 'I have split the twig'. In the development of English, and a number of other languages, the perfect emerges from descriptions of *resultant states*: If I have a twig that is split, you might infer that I have split the twig. I have proposed a similar attention to resultant states in early child language across languages:

> A particular salient Perspective focuses on the RESULTS of events In all languages for which there are relevant data, wherever there is an acoustically salient past-tense or perfect marking on the verb, its first use by the child seems to be to comment on an immediately completed event that results in a visible change of state of some object. (Slobin 1985: 1181)

Again, we seem to have a parallel between diachrony and ontogeny. In English, the first uses of the present perfect occur with telic verbs in contexts where the completion of one action provides the grounds for a subsequent action. For example, a British child of 2;2 is negotiating with his mother about putting his bricks away before getting to eat some pie. He uses a present perfect, saying: "When I'*ve picked* the bricks *up*?" That is, he is focusing on a future result and its potential consequences. A focus on resultant state is apparently the historical point-of-entry for the English present perfect as well. Again, we have evidence of uses of a form that allow for two construals, in this case, the established adjectival construal and an emerging perfect construal. Consider the following Old English example:

(9) Ic hæbbe gebunden þone feond þe he drehte.
I have bound that enemy that them afflicted
Adjectival interpretation: 'I have that enemy bound that afflicted them.'
Perfect interpretation: 'I have bound that enemy that afflicted them.'

The established meaning was adjectival: 'I have that enemy, that afflicted them, in a bound state'. But conversational inference would have allowed for an alternative interpretation: 'I have bound that enemy that afflicted them [and now what should we do with the enemy?]'. That is, attention can be drawn to the state of the bound enemy (adjectival) or to the action of the first-person subject (perfect). In both the history of English and the development of the present perfect in contemporary English-speaking children, this is the first meaning of the perfect: a resultant state, brought about by an agent, with ensuing consequences. And in both developmental paths, other meanings of the perfect emerge later, such as the Perfect of Experience (e.g., *I have visited London several times*), and the Continuative Perfect (e.g., *He has sung in the choir for years*). However, for the contemporary child, the full array of uses of the present perfect already exists in the exposure language. The child begins with the resultant state meanings because of the cognitive salience of results. The other uses of the present perfect are less accessible to the cognitive framework of a preschooler, and are acquired later. By contrast, the speakers of Old English did not set off to find a means of marking resultant states that have current relevance; rather, an available construction (possessed attributive participle) was available for reinterpretation in conversational contexts. Once it became grammaticized as a present perfect, the other uses could emerge. In presenting this case study several years ago, I concluded that the apparent parallel between ontogeny and diachrony could not be due to recapitulation:

> It is not surprising that the core meanings of grammatical forms are such 'concrete' notions as visible resultant state and inference to immediately preceding process. Children start with such notions because they are cognitively most simple, natural, and accessible. Older speakers, in the process of grammaticization, play out the metaphoric and metonymic extensions of basic meanings simply because those are the only available materials for such extensions, and not because they are 'recapitulating' developmental processes from early childhood. (Slobin 1994: 128)
>
> In brief, although the diachronic and ontogenetic developments appear to be parallel, they seem to result from different processes. New meanings of grammatical forms arise in adult language use on the basis of pragmatic inferences drawn from existing referential and propositional meanings. Preschool-age children are not yet able to draw most of such inferences, and are limited to core semantic concepts and pragmatic functions. With increasing pragmatic and cognitive competence, they are able to comprehend the extended pragmatic meanings intended by older speakers, and come to express those functions in their speech. I propose, then, that children come to discover pragmatic extensions of grammatical forms, but they do not innovate them; rather, these extensions are innovated diachronically by older speakers, and children acquire them through a prolonged developmental process of conversational inferencing. (Slobin 1994: 129–30)

3. Are child learners the agents of language change?

The preceding discussion also relates to another longstanding proposal for the role of children — namely, that languages change because of the activities of children, either as innovators or as imperfect learners. This proposal has a long history in 19th and 20th century linguistics. Kiparsky, for example, has proposed that simplification in language acquisition leads to language change:

> Normally oversimplified intermediate grammars which the child constructs on its way to adult language eventually give way to the full complexity of the adult system. The linguistic change of simplification takes place on those relatively rare occasions when some feature of these intermediate grammars survives into adulthood and becomes adopted by the speech community as a new linguistic norm. (Kiparsky 1968: 194–5)

The child learner is seen as the source of various sorts of language change, both in generative accounts (e.g. Lightfoot 1988) and functional approaches (e.g. Gvozdanović 1997).

Most changes of this sort, however, are a matter of "cleaning up" an existing grammar, rather than introducing new forms or constructions. And in any event, for a linguistic change to have a lasting effect, it has to be maintained into adulthood; that is, the childish revisions must come to sound normal and acceptable. Sociolinguistic studies, however, show that lasting changes are more likely to be due to usage in adolescent peer groups, rather than in early childhood (e.g. Romaine 1984).

Joan Bybee and I carried out a study of changes in the English past tense, with the title: "Why small children cannot change language on their own" (Bybee and Slobin 1982). We looked for innovations in past-tense verbs forms, such as *builded, hitted,* and *brang*. The data came from three age groups: (1) spontaneous speech records of preschoolers aged 1;6–5, (2) elicited past-tense forms from school-age children aged 8;6–10, and (3) past-tense forms produced under time pressure by adults. All three age groups produced innovative forms (errors, overregularizations), however only the forms produced by school-age and adult speakers mirrored ongoing changes in the English verb system. That is, many of the preschool errors were transient, while many of the errors produced by older speakers showed some chance of becoming part of the language. For example, only older speakers tended to regularize verbs that end in a final dental consonant, such as *hit–hitted*, reflecting an ongoing tendency in English to regularize verbs of this class. For example, verbs such as *started, lifted, fasted, roasted, sweated* did not use to have these overt *-ed* past tenses; the earlier past tense forms were *start, lift, fast, roast, sweat*. The study suggested that, at least in this part of the grammar, early learners are not the innovators:

> The conclusion that must be drawn from the facts is that there is nothing particularly special about the relation between small children's innovative forms and morpho-phonemic change. The innovations of older children and adults may also serve as predictors of change. In fact in some cases where adult innovations differ from early child innovations, such as with the *hit*-class, the adults and older children, who are in better command of the entire system, innovate in ways that manifest more precisely the on-going changes in the system. Thus it appears that

> both socially and linguistically the older children and adults are in control of morpho-phonemic changes. (Bybee and Slobin 1982: 36–7)

4. Are new languages created by children?

4.1 From pidgin to creole

The classic definition of a creole is a pidgin with native speakers. That is, it has long been claimed that children can take an imperfect input language and "nativize" it. I will not attempt to summarize the large and contentious literature on this topic (see, for example, Bickerton 1981; DeGraff 1999; Foley 1988; McWhorter 1997; Muysken 1988; Romaine 1988; Thomason and Kaufman 1988). Arguments in the pages of the *Journal of Pidgin and Creole Languages*, going on since its inception in 1986, provide evidence and counter-evidence for a variety of theories of creole genesis. As I read the literature, there is evidence for considerable influence of substratum languages on emerging creoles, particularly influences of various African languages on Caribbean creoles. Furthermore, demographic data strongly suggest that languages that are considerably more complex than pidgins can arise in interaction between adults, before there are native speakers. Bickerton's proposal that creole genesis reveals an innate "bioprogram" for language seems far less plausible than when it was introduced twenty years ago. (Personally, I am not convinced by any of the evidence or arguments for the bioprogram.)

Most of the world's creole languages arose in the past, under linguistic and social circumstances that will always lack full documentation. But we have at least one contemporary example: the emergence of Tok Pisin as a developed language in New Guinea. Tok Pisin arose out of Melanesian Pidgin, and in the course of some 150 years of use developed a number of grammatical features before it became anyone's first language (Keesing 1991).[1] Much of this development can be attributed to the fact that Tok Pisin was called upon to serve as a standard language of public communication, business, and education, as well as serving as the official language of government proceedings in Papua New Guinea after 1964 (Romaine 1988: 33). Gillian Sankoff was on hand to study the first native speakers of Tok Pisin — a process that she and Suzanne Laberge have aptly referred to as "the acquisition of native speakers by a language." There are two important findings for the purposes of the present argument: (1) A pidgin language can evolve into something like a creole without requiring the hypothesized special intervention of child learners. (2) The first generation of native speakers "smoothes out" the language, rather than innovating new forms. Let us briefly consider the second finding.

The children studied by Sankoff in the 1970s were learning Tok Pisin in families and social situations in which it served as a second language, spoken with some fluency, but also with some grammatical fluidity. The child learners apparently did what children are good at: making a system regular and automatic (what John Haiman [1994] has referred to as "ritualization"). This is evident on the level of speech production, as in the following example (Sankoff and Laberge 1973: 35–6):

> The children speak with much greater speed and fluency, involving a number of morphophonemic reductions as well as reduction in the number of syllables characteristically receiving primary stress. Whereas an adult will say, for the sentence "I am going home,"
>
> (1) Mi gó long háus;
>
> a child will often say
>
> (2) Mì gò l:áus;
>
> three syllables rather than four, with one primary stress rather than two.

Grammatical morphology also changes with native speakers — but, again, they are not the innovators. For example, there was a well-established future marker, *baimbai* (from English *by and by*), which began as an optional adverbial to establish the time frame of a stretch of discourse. But long before there were native speakers, the form was reduced to *bai* and moved to preverbal position within the clause, where it tended to be used redundantly in a series of future predications. What the children did was to make the future marker obligatory, while also reducing it in substance and stress. That is, the child learners played a significant role in accelerating an ongoing process of grammaticization, in which a preverbal clitic, *bə*, moves along a well-established path from a particle to an inflectional prefix. Thomason and Kaufman provide an apt reformulation of the role of the child:

> When Bickerton poses the question of how a child can "produce a rule for which he has no evidence" (1981: 6), he is, in our view, asking the wrong question. We prefer to ask how the child can create grammatical rules on the basis of input data which is much more variable than the input data received by a child in a monolingual environment. (Thomason and Kaufman 1988: 164)

Given the available evidence, I conclude that learning processes of this sort are normal, and do not reveal special capacities of the language-learning child beyond what is already known about the acquisition of "full-fledged" languages. A creole language develops over time, in contexts of expanding communicative use of a limited pidgin language. Child learners help to push the process forward, arriving at a grammar that is more regular and automated — but they do not appear to be the innovators.

4.2 From homesign to Nicaraguan Sign Language

There is little that can be added to Jill Morford's paper in this volume, which provides a thoughtful review and discussion of the remarkable emergence of a new sign language. Her conclusions are much like my evaluation of Tok Pisin: Linguistic structure emerges when people are put together and begin to communicate about a range of topics, using limited resources. Documentation of the emergence of grammatical forms in Nicaraguan Sign Language can be found in Ann Senghas' dissertation . Her data show that all of the grammatical innovations that she studied were already present in the older signers ("first cohort"). These include: handshape classifiers, position/location inflection, person inflection, and number inflection. That is (contrary to claims in the literature), linguistic structure was not the invention of the "second cohort." Rather, what seems to have happened was that younger signers — that is, those who entered a community that already had a developing communication system — used the existing grammatical elements more frequently and fluently. This account of the Nicaraguan situation matches Thomason and Kaufman's response to Bickerton with regard to the emergence of creole languages (with the exception, of course, that the starting point in Nicaragua was a collection of individual homesign systems, rather than two or more existing languages):

> [A]n entirely new language — without genetic affiliation — is created by the first members of the new multilingual community, and further developed and stabilized by later members, both children born into the community and (in many or most cases) newcomers brought in from outside. (Thomason and Kaufman 1988: 166)

Senghas and Coppola come to a similar conclusion with regard to Nicaraguan Sign Language:

> [O]ver their first several years together, the first cohort, as children, systematized [the initial] resources in certain ways, converting raw gestures and homseigns into a partially systematized system. This early work evidently provided adequate raw materials for the second-cohort children to continue to build the grammar. (Senghas and Coppola 2001: 328)

In short, as regular forms begin to develop in a group, younger learners *automate* the language. Morford's discussion of the Nicaraguan situation, as well as her work with late learners of ASL, shows that efficiency in online processing is a critical factor in language mastery. And it is on this dimension that early learners have an advantage, rather than having a special, age-linked capacity "to create language." (Senghas and Coppola [2001] report that children who acquire Nicaraguan Sign Language before the age of 10 sign at a faster rate and are more skilled in comprehending grammatical forms.)

It is clear from the chapters of Morford (this volume) and Goldin-Meadow (this volume) that individual deaf children can innovate linguistic forms, but that it takes an interacting community to push those innovations towards automated, efficient linguistic systems. Therefore, as Morford points out, "the emergence of Nicaraguan Sign Language is better described as a process of *grammaticization* than of innovation" (this volume, p. 337). We would do well to pay close attention to the stages that Morford proposes:

> Thus, the implication of this work is that there may be three distinctive stages in the emergence of language: (1) the emergence of the lexicon, (2) the emergence of system-internal grammatical properties, and (3) the emergence of properties that are dependent upon the development of expectations of co-occurrence patterns in the communicative productions of conspecifics. (this volume, p. 338)

This schematized formulation gives us a promising way of thinking about the emergence of language in evolutionary time, without making false analogies from the capacities and activities of already-evolved *Homo sapiens* children. It also requires us to pay as much attention to the emergence of structure in communicative practice as to the cognitive capacities of the individual — whether innate or developing, whether language-specific or general (Slobin 1997).

5. An interim conclusion

I have briefly examined four longstanding proposals about possible contributions of child language study to questions of linguistic diachrony and evolution, with negative conclusions. My field of developmental psycholinguistics provides insights into the capacities for language, thought, and communication in our species. Children's early formulations of grammar and semantics provide a window into basic operating principles and organizing factors of the human mind. Therefore, ontogenetic theory and data are useful in pinpointing some of the basic concepts and processes that are needed in order to face Givón's challenge. In addition, comparisons with other surviving primate species — their capacities and developmental patterns — give clues about the road that had to be traversed by our ancestors. All of this growing information provides material for a range of speculative scenarios. At best, close attention to biology, development, and linguistic behavior can heighten the plausibility of those scenarios. But we can never have sufficient evidence to scientifically evaluate such narratives, however appealing and plausible some of them may be. I hope that my negative conclusions can at least serve to reasonably constrain our irresistible speculations about who we are and how we got here.

Notes

1. John McWhorter (1995) has made a similar suggestion about Atlantic English-based creoles, suggesting that they derive from an established and elaborated West African pidgin used early in the 17th century. He concludes that this ancestor language "was by no means a rudimentary pidgin, but was, on the contrary, already relatively elaborated by the time of its exportation to the New World. Hence, this contact language exhibited a structural expansion analogous to that of Tok Pisin before creolization, as opposed to the rudimentary structure documented in pidgins of limited social function" (p. 325).

References

Aksu-Koç, A.A., and Slobin, D.I. 1985. "The acquisition of Turkish". In *The Crosslinguistic Study of Language Acquisition*, Volume I: *The Data*, D.I. Slobin (ed), 839–78. Hillsdale, NJ: Lawrence Erlbaum Associates.

Allen, S.E.M. 1996. *Aspects of Argument Structure Acquisition in Inuktitut*. Amsterdam: John Benjamins.

Andersen, R.W. 1983. *Pidginization and Creolization as Language Acquisition*. Rowley, Mass.: Newbury House.

Bickerton, D. 1981. *Roots of Language*. Ann Arbor, Mich.: Karoma.

Bickerton, D. 1990. *Language and Species*. Chicago: University of Chicago Press.

Bybee, J.L. 1985. *Morphology: A Study of the Relation Between Meaning and Form*. Amsterdam: John Benjamins.

Bybee, J.L., and Slobin, D.I. 1982. "Why small children cannot change language on their own: Suggestions from the English past tense". In *Papers from the 5th International Conference on Historical Linguistics*, A. Ahlqvist (ed), 29–37. Amsterdam: John Benjamins.

Bybee, J.L., Perkins, R., and Pagliuca, W. 1994. *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Clark, E.V. 1978. "Discovering what words can do". In *Papers from the Parasession on the Lexicon*, D. Farkas, W.M. Jacobsen, and K.W. Todrys (eds), 34–57. Chicago: Chicago Linguistics Society.

de Houwer, A. 1995. "Bilingual language acquisition". In *The Handbook of Child Language*, P. Fletcher and B. MacWhinney (eds), 219–50. Oxford: Blackwell.

DeGraff, M. 1999. *Language Creation and Language Change: Creolization, Diachrony, and Development*. Cambridge, Mass.: MIT Press.

Deuchar, M. 1992. "Can government and binding theory account for language acquisition?" In *Lenguajes naturales y lenguajes formales VIII*, C. Vide (ed), 273–9. Barcelona: Universitat de Barcelona.

Ekmekçi, Ö. 1986. "The significance of word order in the acquisition of Turkish". In *Studies in Turkish Linguistics*, D.I. Slobin and K. Zimmer (eds), 265–72. Amsterdam: John Benjamins.

Givón, T. 1979. *On Understanding Grammar*. New York: Academic Press.

Givón, T. 1998. "On the co-evolution of language, mind and brain". *Evolution of Communication* 2: 45–116.

Gould, S. J. 1977. *Ontogeny and Phylogeny*. Cambridge, Mass.: Harvard University Press.

Gvozdanović, J. (ed). 1997. *Language Change and Functional Explanations*. Berlin and New York: Mouton de Gruyter.

Haiman, J. 1994. "Ritualization and the development of language". In *Perspectives on Grammaticalization*, W. Pagliuca (ed), 3–28. Amsterdam: John Benjamins.

Hopper, P. J., and Traugott, E. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Jespersen, O. 1921/1964. *Language: Its nature, Development and Origin*. New York: Norton.

Keesing, R. M. 1991. "Substrates, calquing and grammaticalization in Melanesian Pidgin". In *Approaches to Grammaticalization*, Volume I: *Focus on Theoretical and Methodological issues*, E. C. Traugott and B. Heine (eds), 315–42. Amsterdam: John Benjamins.

Kiparsky, P. 1968. "Linguistic universals and linguistic change". In *Universals in Linguistic Theory*, E. Bach and R. T. Harms (eds), 171–202. New York: Holt, Rinehart & Winston.

Küntay, A., and Slobin, D. I. 1999. "The acquisition of Turkish as a native language: A research review. *Turkic Languages* 3: 151–88.

Langer, J. 2000. "The descent of cognitive development". *Developmental Science* 3: 361–88.

Langer, J. 2001. "The mosaic evolution of cognitive and linguistic ontogeny". In *Language Acquisition and Conceptual Development*, M. Bowerman and S. C. Levinson (eds), 19–44. Cambridge: Cambridge University Press.

Lieven, E., Pine, J. M., and Baldwin, G. 1997. "Lexically-based learning and the development of grammar in early multi-word speech". *Journal of Child Language* 24: 187–219.

Lightfoot, D. 1988. "Syntactic change". In *Linguistics: The Cambridge Survey*, Volume I. *Linguistic Theory: Foundations*, F. J. Newmeyer (ed), 303–23. Cambridge: Cambridge University Press

Lord, C. 1982. "The development of object markers in serial verb languages". In *Syntax and Semantics, Volume XV: Studies in Transitivity*, P. J. Hopper and S. A. Thompson (eds), 277–99. New York: Academic Press.

McWhorter, J. 1995. "Sisters under the skin: A case for the genetic relationship between the Atlantic English-based creoles". *Journal of Pidgin and Creole Languages* 10: 289–333.

McWhorter, J. 1997. *Towards a New Model of Creole Genesis*. New York: Peter Lang

Meisel, J. M. 1989. "Early differentiation of languages in bilingual children". In *Bilingualism Across the Lifespan: Aspects of Acquisition, Maturity and Loss*, K. Hyltenstam and L. Obler (eds), 13–40. Cambridge: Cambridge University Press.

Morford, J. P., and Mayberry, R. I. 2000. "A reexamination of 'early exposure' and its implications for language acquisition by eye". In *Language Acquisition by Eye*, J. P. Morford, C. D. Chamberlain, and R. I. Mayberry (eds), 111–27. Mahwah, NJ: Lawrence Erlbaum Associates.

Muysken, P. 1988. "Are creoles a special type of language?" In *Linguistics: The Cambridge Survey*, Volume II: *Linguistic Theory: Extensions and Implications*, F. Newmeyer (ed), 285–301. Cambridge: Cambridge University Press.

Romaine, S. 1984. *The Language of Children and Adolescents: The Acquisition of Communicative Competence*. Oxford: Blackwell.

Sankoff, G. 1979. "The genesis of a language". In *The Genesis of Language*, K.C. Hill (ed), 23–47. Ann Arbor, MI: Karoma.

Sankoff, G., and Laberge, S. 1973. "On the acquisition of native speakers by a language". *Kivung* 6: 32–47.

Savage-Rumbaugh, S., and Taglialatela, J. 2001. Language, apes, and understanding speech. Paper presented at conference "The Rise of Language," Eugene, OR.

Savage-Rumbaugh, S., Shankar, S.G., and Tylor, T.J. 1998. *Apes, Language, and the Human Mind.* Oxford: Oxford University Press.

Senghas, A. 1995. *Children's Contribution to the Birth of Nicaraguan Sign Language.* Doctoral dissertation, MIT.

Senghas, A., and Coppola, M. 2001. "Children creating language: How Nicaraguan Sign Language acquired a special grammar". *Psychological Science* 12: 323–8.

Slobin, D.I. 1973. "Cognitive prerequisites for the development of grammar". In *Studies of Child Language Development*, C.A. Ferguson and D.I. Slobin (eds), 175–208. New York: Holt, Rinehart & Winston.

Slobin, D.I. 1977. "Language change in childhood and in history". In *Language Learning and Thought*, J. Macnamara (ed), 185–214. New York: Academic Press.

Slobin, D.I. 1985. "Cross-linguistic evidence for the Language-Making Capacity". In *The Crosslinguistic Study of Language Acquisition*, Volume II: *Theoretical issues*, D.I. Slobin (ed), 1157–256. Hillsdale, NJ: Lawrence Erlbaum Associates.

Slobin, D.I. 1994. "Talking perfectly: Discourse origins of the present perfect". In *Perspectives on Grammaticalization*, W. Pagliuca (ed), 119–33. Amsterdam: John Benjamins.

Slobin, D.I. 1997. "The origins of grammaticizable notions: Beyond the individual mind". In *The Crosslinguistic Study of Language Acquisition*, Volume V: *Expanding the contexts*, D.I. Slobin (ed), 265–323. Mahwah, NJ: Lawrence Erlbaum Associates.

Tomasello, M. 1992. *First Verbs: A Case Study in Early Grammatical Development.* Cambridge: Cambridge University Press.

Tomasello, M. 1999. *The Cultural Origins of Human Cognition.* Cambridge, Mass.: Harvard University Press.

Traugott, E. 1992. *Old English Syntax.* Cambridge: Cambridge University Press.

Whitney, W.D. (1875). *The Life and Growth of Language: An Outline of Linguistic Science.* New York: Appleton.

In the series TYPOLOGICAL STUDIES IN LANGUAGE (TSL) the following titles have been published thus far:

1. HOPPER, Paul J. (ed.): *Tense-Aspect: Between semantics & pragmatics.* 1982.
2. HAIMAN, John & Pamela MUNRO (eds): *Switch Reference and Universal Grammar. Proceedings of a symposium on switch reference and universal grammar, Winnipeg, May 1981.* 1983.
3. GIVÓN, T.: *Topic Continuity in Discourse. A quantitative cross-language study.* 1983.
4. CHISHOLM, William, Louis T. MILIC & John A.C. GREPPIN (eds): *Interrogativity: A colloquium on the grammar, typology and pragmatics of questions in seven diverse languages, Cleveland, Ohio, October 5th 1981-May 3rd 1982.* 1984.
5. RUTHERFORD, William E. (ed.): *Language Universals and Second Language Acquisition.* 1984 (2nd ed. 1987).
6. HAIMAN, John (Ed.): *Iconicity in Syntax. Proceedings of a symposium on iconicity in syntax, Stanford, June 24-26, 1983.* 1985.
7. CRAIG, Colette (ed.): *Noun Classes and Categorization. Proceedings of a symposium on categorization and noun classification, Eugene, Oregon, October 1983.* 1986.
8. SLOBIN, Dan I. & Karl ZIMMER (eds): *Studies in Turkish Linguistics.* 1986.
9. BYBEE, Joan L.: *Morphology. A Study of the Relation between Meaning and Form.* 1985.
10. RANSOM, Evelyn: *Complementation: its Meaning and Forms.* 1986.
11. TOMLIN, Russel S.: *Coherence and Grounding in Discourse. Outcome of a Symposium, Eugene, Oregon, June 1984.* 1987.
12. NEDJALKOV, Vladimir (ed.): *Typology of Resultative Constructions. Translated from the original Russian edition (1983). English translation edited by Bernard Comrie.* 1988.
14. HINDS, John, Shoichi IWASAKI & Senko K. MAYNARD (eds): *Perspectives on Topicalization. The case of Japanese WA.* 1987.
15. AUSTIN, Peter (ed.): *Complex Sentence Constructions in Australian Languages.* 1988.
16. SHIBATANI, Masayoshi (ed.): *Passive and Voice.* 1988.
17. HAMMOND, Michael, Edith A. MORAVCSIK and Jessica WIRTH (eds): *Studies in Syntactic Typology.* 1988.
18. HAIMAN, John & Sandra A. THOMPSON (eds): *Clause Combining in Grammar and Discourse.* 1988.
19. TRAUGOTT, Elizabeth C. and Bernd HEINE (eds): *Approaches to Grammaticalization, 2 volumes (set)* 1991
20. CROFT, William, Suzanne KEMMER and Keith DENNING (eds): *Studies in Typology and Diachrony. Papers presented to Joseph H. Greenberg on his 75th birthday.* 1990.
21. DOWNING, Pamela, Susan D. LIMA and Michael NOONAN (eds): *The Linguistics of Literacy.* 1992.
22. PAYNE, Doris (ed.): *Pragmatics of Word Order Flexibility.* 1992.
23. KEMMER, Suzanne: *The Middle Voice.* 1993.
24. PERKINS, Revere D.: *Deixis, Grammar, and Culture.* 1992.
25. SVOROU, Soteria: *The Grammar of Space.* 1994.
26. LORD, Carol: *Historical Change in Serial Verb Constructions.* 1993.
27. FOX, Barbara and Paul J. Hopper (eds): *Voice: Form and Function.* 1994.
28. GIVÓN, T. (ed.) : *Voice and Inversion.* 1994.
29. KAHREL, Peter and René van den BERG (eds): *Typological Studies in Negation.* 1994.

30. DOWNING, Pamela and Michael NOONAN: *Word Order in Discourse.* 1995.
31. GERNSBACHER, M. A. and T. GIVÓN (eds): *Coherence in Spontaneous Text.* 1995.
32. BYBEE, Joan and Suzanne FLEISCHMAN (eds): *Modality in Grammar and Discourse.* 1995.
33. FOX, Barbara (ed.): *Studies in Anaphora.* 1996.
34. GIVÓN, T. (ed.): *Conversation. Cognitive, communicative and social perspectives.* 1997.
35. GIVÓN, T. (ed.): *Grammatical Relations. A functionalist perspective.* 1997.
36. NEWMAN, John (ed.): *The Linguistics of Giving.* 1998.
37. RAMAT, Anna Giacalone and Paul J. HOPPER (eds): *The Limits of Grammaticalization.* 1998.
38. SIEWIERSKA, Anna and Jae Jung SONG (eds): *Case, Typology and Grammar. In honor of Barry J. Blake.* 1998.
39. PAYNE, Doris L. and Immanuel BARSHI (eds.): *External Possession.* 1999.
40. FRAJZYNGIER, Zygmunt and Traci S. CURL (eds.): *Reflexives. Forms and functions.* 2000.
41. FRAJZYNGIER, Zygmunt and Traci S. CURL (eds): *Reciprocals. Forms and functions.* 2000.
42. DIESSEL, Holger: *Demonstratives. Form, function and grammaticalization.* 1999.
43. GILDEA, Spike (ed.): *Reconstructing Grammar. Comparative Linguistics and Grammaticalization.* 2000.
44. VOELTZ, F.K. Erhard and Christa KILIAN-HATZ (eds.): *Ideophones.* 2001.
45. BYBEE, Joan and Paul HOPPER (eds.): *Frequency and the Emergence of Linguistic Structure.* 2001.
46. AIKHENVALD, Alexandra Y., R.M.W. DIXON and Masayuki ONISHI (eds.): *Non-canonical Marking of Subjects and Objects.* 2001.
47. BARON, Irene, Michael HERSLUND and Finn SORENSEN (eds.): *Dimensions of Possession.* 2001.
48. SHIBATANI, Masayoshi (ed.): *The Grammar of Causation and Interpersonal Manipulation.* 2002.
49. WISCHER, Ilse and Gabriele DIEWALD (eds.): *New Reflections on Grammaticalization.* 2002.
50. FEIGENBAUM, Susanne and Dennis KURZON (eds.): *Prepositions in their Syntactic, Semantic and Pragmatic Context.* 2002.
51. NEWMAN, John (ed.): *The Linguistics of Sitting, Standing and Lying.* 2002.
52. GÜLDEMANN, Tom and Manfred von RONCADOR (eds.): *Reported Discourse. A meeting ground for different linguistic domains.* 2002.
53. GIVÓN, T. and Bertram MALLE (eds.): *The Evolution of Language from Pre-language.* 2002.
54. AIKHENVALD, Alexandra Y. and R.M.W. DIXON (eds.): *Studies in Evidentiality.* 2003.
55. FRAJZYNGIER, Zygmunt and Erin SHAY: *Explaining Language Structure through Systems Interaction.* 2003.
56. SHAY, Erin and Uwe SEIBERT (eds.): *Motion, Direction and Location in Languages. In honor of Zygmunt Frajzyngier.* 2003.