22

Neural Substrates for String-Context Mutual Segmentation: A Path to Human Language

Kazuo Okanoya and Bjorn Merker

22.1. Introduction

Linguistic structures are products of biological prerequisites and historical processes. Here we consider a number of neural, behavioral, and learning mechanisms that serve necessary or facilitating roles in the initiation of historical processes. We hypothesize that if mutual segmentation of strings and contexts is promoted by particular biological adaptations and ecological pressures, this could initiate a subsequent historical process of linguistic elaboration. To enable this mutual segmentation, three biological sub-faculties are indispensable: vocal learning, string segmentation, and contextual segmentation. Vocal learning enabled intentional control of vocal output via the direct connection between face motor cortex and medullary vocal nuclei. String segmentation became possible by bottom-up statistical learning by basal ganglia and top-down rule extraction by the prefrontal cortex. Contextual segmentation was implemented also by bottom-up induction of situational correlations at hippocampal and related limbic structures and top-down segmentation of perceived states by the prefrontal cortex. Taken together, we propose that string-context mutual segmentation got its start through the interaction of the prefrontal-hippocampal and prefrontal-striatal parallel loops.

22.1.1. String-Context Mutual Segmentation

In the previous chapter we presented a hypothesis (first suggested by Darwin) for the possible origin of human language. Briefly, we assume our ancestor was a singing ape. Song display was an honest indicator of the singer's fitness. Through the handicap principle, song displays became complex and lengthy (Zahavi & Zahavi, 1996) and included many vocal tokens that were shared with the group members through imitation. Song was initially used for attracting mates and repelling rivals, but gradually different variations of songs were used in different behavioral context. Since songs were learned and song elements were shared within society, a fixed sub-part of these songs appeared in several situation-specific songs. These sub-strings began to be associated with the sub-commonality in the situations within which songs were sung. As processes were repeated and transmitted through generations, linguistic structures would emerge as historical consequences (Figure 22.1).

We divided the requirements for linguistic structures into biological prerequisites and the historical process and then defined what are required as biological preparations to initiate the historical process of linguistic elaboration. The historical process may simply be stated as the process of string-context mutual segmentation: song strings and behavioral contexts are mutually segmented during social interactions. Over generations, this would lead to the emergence of linguistic structure. The natural history up until the historical process was initiated has been described in detail in the previous chapter. The aim of the present chapter is thus to provide neural substrates for the biological prerequisites required to start the process of mutual segmentation.

22.1.2. Sub-Faculties for Mutual Segmentation

Sub-faculties we consider here are vocal learning, string segmentation, and context segmentation (Figure 22.2). Vocal learning is strictly defined as a postnatal acquisition of a novel motor pattern for a novel vocal performance (Jarvis, 2004). A mere modification of innate vocal patterns, which is often seen in vocal animals, should be referred to as vocal plasticity rather than vocal learning. Vocal learning should not be confused with auditory learning that refers to an acquisition of the association between a specific auditory stimulus and a specific behavior: auditory learning is a form of general associative learning with stimuli being auditory, but vocal learning is a specific imitative learning that requires dense vocal-auditory interactions.

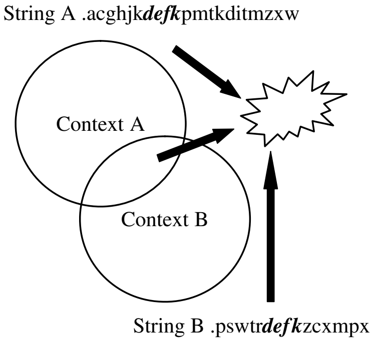

FIGURE 22.1. The string-context mutual segmentation hypothesis.

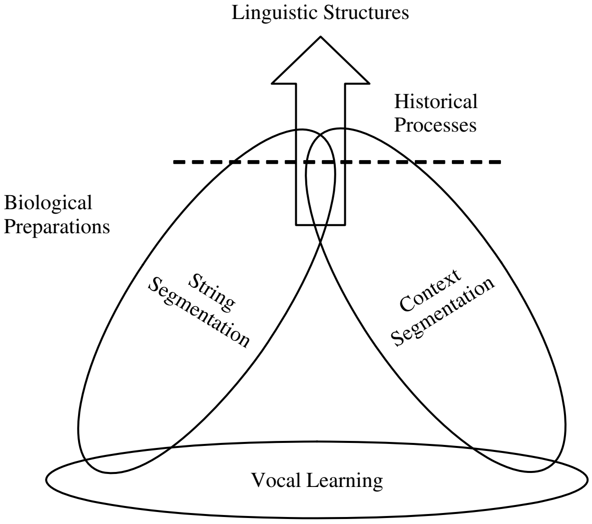

FIGURE 22.2. Biological preparations for mutual segmentation.

String segmentation is an ability to divide a continuous stream of sensory stimuli into smaller functional units. Strings can be mediated by any sensory modalities including visual, auditory, olfactory and tactile, but we specifically deal with auditory segmentation in this chapter, since the primary mode of both song and language is auditory, although language could be expressed in other domains. Auditory segmentation could possibly be based on several cues such as pauses between auditory tokens, sudden discontinuities or transitions in pitch or intensity continua, and probabilistic relations between auditory tokens. In first language acquisition, infants can use any of these cues to segment words from the continuous speech stream (Gomez & Gerken, 2000). This is a first step towards acquiring the syntax.

Context segmentation is an ability to divide a continuous contextual/cognitive space into smaller (functional) units. Contextual space here is loosely defined as a multivariate space whose dimensions might include sensory modalities, internal states including memory, and reinforcement contingencies. Animals explicitly or implicitly can learn and discriminate complex context and learn to select adaptive behavior accordingly. Undoubtedly, this form of learning is widespread among vertebrates. This is the very first step in acquiring semantics.

22.2. Vocal Learning

22.2.1. Species with Vocal Learning

Although auditory learning is widespread among animals possessing an auditory system, vocal learning is scarce. Among primates, we humans are the only species that show vocal learning. There is evidence of vocal plasticity in non-human primates (Hihara et al., 2003), but these exemplify the degree of plasticity that cannot be classified as an acquisition of novel vocal patterns (Janik and Slater, 1997). In mammals, only three phylogenic lines of animals show vocal learning. These are primates (only humans), bats (Boughman, 1998), and cetaceans (Payne, 2000). Coincidently, also in birds, only three classes of animals show vocal learning; passerines (songbirds), psittacines (parrots), and swifts (humming birds) (Jarvis, 2004). It is important to note that these classes of animals are only distantly related when considering evolutionary origins of vocal learning.

This observation provides three possibilities as to the evolution of vocal learning. First, vocal learning independently evolved in these animals. Thus, neural substrates for vocal learning are different among animals with vocal learning. Second, vocal learning independently disappeared from many classes of animals and vocal learning was actually widespread in the past. In this case, vocal learning was present in a common ancestor of birds and mammals and we should consider examining some existing reptiles for vocal learning. Also, in this case, neural substrates utilized for vocal learning should be common among mammals and birds. This is unlikely considering the vast divergence of telencephalic architecture between mammals and birds.

Third, and most likely, vocal learning evolved independently, but out of common plans of brain circuitry under similar selection pressures. A general plan for intentional motor execution and a general plan for sensory-motor interface were probably precursors for vocal learning. In fact, Jarvis (2000) observed similar topography and connectivity as well as the pattern of gene expression in the forebrain nuclei responsible for vocal learning in three lines of avian vocal learners. If vocal learning evolved independently out of common sensory-motor systems, under common selection pressures, we need to consider the neural substrates and pre-adaptations for vocal learning.

22.2.2. Neural Substrates for Vocal Learning

Are there any specific anatomical substrates that correlate with the faculty of vocal learning? One candidate substrate would be the direct cortical-medullary pathway for articulation and breathing. In humans, a part of the motor cortex directly projects to the nucleus retro-ambiguous of the medulla (Kuypers, 1958). This projection is absent in the squirrel monkey and chimpanzee and Jurgens (2002) thus assumes that this projection exists only in humans among primates. Similarly, there is the same direct cortical-medullary pathway for articulation and breathing in the zebra finch, a species of songbirds, but a similar projection in pigeons does not exist and all of pigeon vocalizations are innate (Wild, 1993; Wild et al., 1997). Considering this evidence, we can hypothesize that this projection exists in those species that show vocal learning while it is absent in the species without vocal learning (Okanoya et al., 2004).

While this projection exists only in a limited number of species, there is still a possibility that a very faint projection of this kind is nevertheless present in most species. Deacon (1997) introduced the anecdotal story of an orphaned harbor seal that learned to mimic the speech of the fisherman who raised the seal in his home. This animal had a brain inflammation when young and Deacon suggested that during the process of recovery the cortical medullar projection might have been reinforced in this particular animal. If so, the possibility arises that animals trained to perform spontaneous vocalizations while young might have this pathway reinforced, thus inducing vocal learning in a species that would not ordinarily possess this capacity.

22.2.3. Pre-Adaptations for Vocal Learning

Among vocal learners, birds, bats, and whales seem to have reasonable pre-adaptation for vocal learning. Birds and bats have to control their breathing while they are flying and this requirement was probably the pre-adaptation for intentional vocal control. Similarly, whales need to control breathing while they are submerged and this led to intentional vocal control (Janik & Slater, 2000). The issue of the origin of vocal learning in humans is a major unsolved issue, as noted in our previous chapter (this volume). Here too, intentional control of respiration may have played a significant role, and that in the context of the special circumstances surrounding the immaturity of the human neonate and its need for parental care (Okanoya et al., 2002).

Human babies are conspicuous among primates in that they emit high-intensity, long-lasting cries right after birth. Such cries are obviously maladaptive in wild animals in that it could easily attract predators. We suggest therefore that the infant cry must be a behavior obtained after ancestral humans acquired social and cultural skills to protect themselves from predation pressure. The first cry after the birth has a function to eliminate amniotic fluid, but crying continues long after that. Likewise, nidicolous species of birds emit very loud begging calls and isolation calls that recruit parental behavior including feeding and protection of the hatchlings. Begging calls are so loud that they easily can attract predators. Therefore, parents have to engage in strong parental behavior to avoid predation of hatchlings and themselves (cf. Trivers, 1974). Chicks can thus manipulate parents because they acquire cortical control of breathing to adapt for flying. In this regard, the fact that humans and nidicolous songbirds share helplessness in infancy may also be related to this pre-adaptation (Alexander, 1990).

We analyzed syntactical and phonological developmental changes in the baby cry to show how the pattern becomes complex as the baby grows (Okanoya et al., 2002). At least three stages of cry development could be identified; each may be associated with respective anatomical changes. Baby cries begin as a regular repetition of a stereotyped vocal unit. At this stage, the cry probably is controlled by the midbrain vocal center only. This type of vocalization continues for two post-natal weeks after which cries begin to be more irregular, showing variable patterns of phonology and rhythmic patterns. At this stage, limbic influences are presumably gradually growing stronger. Mothers often can identify what babies want based on the pattern of their cries. During the interaction with mothers, the cry probably became more adaptive by allowing cortical control so that mothers can be more precisely governed by the baby's crying. Infant cries thus may function to train cortical-midbrain connections necessary for more intentional vocal output that eventually results in speech competence.

22.3. String Segmentation

22.3.1. Behavioral Evidence for String Segmentation

It is essential for our sensory systems to segment continuous streams of stimuli into functional units. Especially, the very beginning of language acquisition involves segmentation of words out of a continuous speech stream. Among several cues available for string segmentation, let us examine the statistical cue. Learning of sequence statistics or "statistical learning" has attracted a great deal of attention in recent years (Gomez & Gerken, 2000). When a nonsense auditory stream consisting of 6 tri-syllable "words" were continuously presented, the transitional probability within a "word" was always 1 but that between "words" was 1/6. Without knowing each word and without knowing the structure of the stimulus, we only hear a nonsense stream of syllables but gradually begin to appreciate word boundaries because of statistical learning. When asked the familiarity of the stimulus "word" that was presented in the task versus a "pseudo word" that was never presented, we can usually identify which was more familiar.

As early as at eight months after birth, babies can detect transitional probabilities between on-going speech syllables and use those statistics to find word boundaries, as shown by a habituation paradigm (Saffran et al., 1996). The same "word boundary" detection was possible with non-linguistic stimuli of pure tones (Saffran et al., 1999). Similar procedures were used to examine statistical learning in a New World monkey, the cotton-top Tamarin (Hauser et al., 2001) showing that this ability is not specific to humans. These were taken as evidence that word segmentation learning may not require any special language device but rather is made possible by a domain general ability of statistical learning (Bates & Elman, 1996).

22.3.2. String Segmentation in Birdsong

Songbirds learn courtship songs from adult males (reviewed in Catchpole & Slater, 1995; Zeigler & Marler, 2004). Their songs consist of temporally isolated song elements and their elements are arranged in a fixed order or in accordance with a specific syntactical rule. Most of these song elements are learned from their fathers and conspecific neighbors. When song learning occurs in a multi-tutor environment, learners often splice parts of songs that are sung by different males based on prosodic cues such as silent intervals between song notes, changes in element types, or at the point of interruption (Williams & Staples, 1992). In a species with more complex song syntax, birds also use statistical information in conspecific songs: chunking of song notes often occurs at junctions of lower transition probability (Takahasi & Okanoya, unpublished data). Thus, string segmentation occurs in songbirds during song learning and song production.

When perceiving speech streams, a phrase structure is processed as a perceptual unit in humans. When a sentence like “I love you” is presented into one ear and a click is presented into the other in a temporal position that overlaps with the word “love”, the perceptual position of the click is likely to shift in between “I” and “love.” That is, the perceptual position of an embedded click moves outside of the phrase structure (Fodor & Bever, 1965). We examined whether a similar phenomenon could be observed in songbirds. We trained Bengalese finches, a species of songbirds with complex song syntax, in a click-detection experiment. Birds were trained to peck a key when they heard a click in an operant conditioning task with food reinforcement. When a background of his own song was played in the test box, the reaction time of the subject bird to detect the click was longer than without the background song or with the background song played in reverse: detection of clicks is postponed until a chunk of song notes is processed. Thus, chunks of birdsong, like chunks of linguistic elements, are processed as a cognitive unit (Suge & Okanoya, under revision).

These data suggest that one essential prerequisite for the type of historical structural transformation dealt with in our previous chapter – namely the capacity for segmentation (cf. Hurford, 2000) – is already present in the vocal learning of some singing species. The capacity for segmentation might also lead to the hierarchical nature of songs in some species (Okanoya, 2004).

22.3.3. Neural Substrates for String Segmentation

If statistical learning is possible with domain-general brain mechanisms, language areas in the brain may not be required for such processes. However, brain imaging studies on learning of sequential rules in music have shown the contrary: Broca’s area was active during such tasks (Maess et al., 2001; Patel, 2003). More directly, we have used a similar statistical stimulus and the stimulus with the same token but without statistical rules to compare brain activation by near-infrared optic spectroscopy and found that Broca’s area was active only during statistical stimulation (Abla & Okanoya, in preparation). Thus, statistical learning probably utilizes Broca’s areas but it is also possible without such specialized areas as shown in the new world monkey, tamarins. Perhaps, results of statistical learning is expressed in Broca’s area in humans, but computation of statistical information could be processed by some other systems as well.

Here again, results from birdsongs may be suggestive (Okanoya, 2004). Most neural substrates for song learning, song maintenance, and song performance have been identified in the bird brain (Zeigler & Marler, 2004). Briefly, birdsong is controlled by two major forebrain pathways: anterior pathway and posterior pathway. Auditory information is conveyed up to a part of the forebrain equivalent to the primary auditory cortex, Field L (a part of the primary auditory cortex). From there, processing for more specific information occurs in NCM (nidopallium code-laterale, a part of the secondary auditory cortex) and cHV (caudal Hyperpallium ventrale, a part of the secondary auditory cortex). Programming for song motor patterns occurs in a region of the face motor cortex equivalent, the RA (robustus archipallium, a part of the primary motor cortex), which directly projects to the medullary respiratory and motor centers, or via the midbrain vocal area, the DM (dorsal medial nucleus of the midbrain), which is equivalent to the periaquadactal gray in mammals. The auditory information is fed into NIf (nucleus interfacialis, a part of the association cortex) and then to HVC (hyperpallium ventrale pars caudalis, a part of the association cortex). HVC has direct projection to RA and also to Area X, which is a part of basal ganglia. Area X then projects to a thalamic nucleus DLM and from there, projection comes back to a telencephalic nucleus LMAN (lateral magnocellular nucleus of the nidopallium, a part of the sensory/motor cortex). LMAN then projects to RA, completing the indirect connection between HVC and RA. The HVC-RA-medullar connection is called the posterior motor pathway and the HVC-Area X-DLM-LMAN-RA connection is called the anterior pathway.

Thus, when an analogy between human language and birdsong is considered, we can propose the followings: NIf may be comparable with Wernicke’s area and HVC may be comparable with Broca’s area, because NIf receives higher auditory information and sends to HVC that controls the face-motor cortex equivalent, the RA. Also, LMAN may be comparable with the anterior cingulate cortex, because LMAN receives projection from the basal ganglia and also sends projection to RA, the motor cortex equivalent (Figure 22.3).

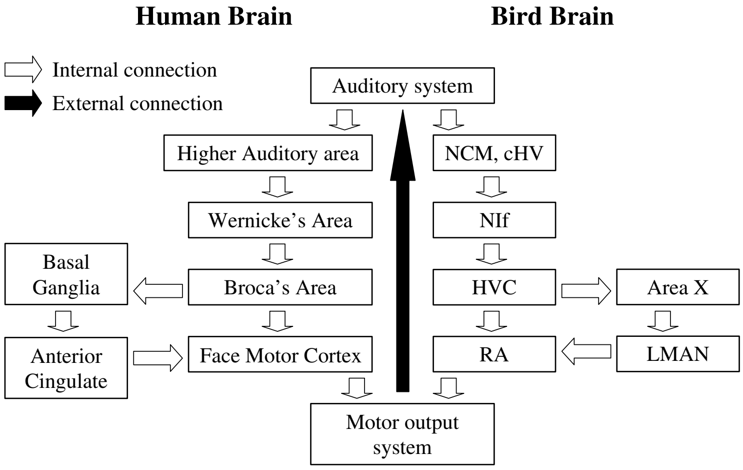

FIGURE 22.3. Analogy for string segmentation by human brain and by bird brain.

In an ERP study with human subjects, we found that string segmentation was associated with a negative brain potential that was strongest at an electrode placed near the anterior cingulate cortex (Abla et al., in preparation). When a source estimation algorithm was applied on the same data recorded from 32 channels, the basal ganglia was the source of the signal. With the data from the near infrared spectroscopy stated earlier, these data suggest that the prefrontal cortex – basal ganglia loop maybe responsible for statistical segmentation of the continuous auditory stream. The same loop may be responsible for controlling complex serial behavior involving string segmentation of a continuous stream, such as piano playing or type writing (Parsons et al., 2005).

The same loop is also known to be used in birdsong learning and maintenance. When LMAN was lesioned in juvenile zebra finches, their songs became stereotyped without further elaboration (Bottjer et al., 1984). On the other hand, when Area X was lesioned in adult Bengalese finches, segmental structure of the song was impaired by causing stuttering (Kobayashi et al., 2001). Furthermore, when pharmacological manipulation was made to suppress neural activities in LMAN, variations in juvenile song disappeared but it recovered as the drug effect waned (Ölveczky et al., 2005). Thus, segmentation of birdsong strings may also be governed by the cortex-basal ganglia pathway.

Both HVC in songbirds and Broca’s area in humans are song- or language-specific organs in the sense that these could not be identified in the brains of related species (i.e., non-songbirds or non-human primates) without vocal learning. In both humans and birds, these structures appeared on the motor side of the sensory-motor junction in the forebrain. These structures may be a specialized form of prefrontal structures that promote statistical learning and rule learning. In this regard, it may be interesting to examine the efficiency of statistical learning in animals with and without these specialized structures. The controversy between the nativist and the empiricist could partially reside in how the modularity and domain specificity of these structures are seen. Are they structures specially evolved for that specific purpose or are they simply a specialization of existing structures? This is an essential question to be solved by comparative neuroethological investigations in birds and humans.

22.4. Context Segmentation

We routinely segment behavioral context based on multiple environmental and internal cues and this would be a rudimentary form of semantics. A somewhat simpler example is spatial navigation. The idea that the hippocampus may function as a spatial map was first presented by O’Keefe and Dostrovsky in 1971 with some experimental evidence. When a rat navigates a novel environment freely, a particular neuron in the hippocampus fires when the rat is exploiting a particular part of the environment. This finding was later expanded to include modeling research. In general, it was shown that with Hebbian learning and lateral inhibition, a network of suitably interconnected neurons begins to fire when a specific class of inputs is present. Such a network can establish attractor dynamics and several attractors can specify specific locations of the environment. Not only that, since the hippocampus receives sensory, emotional, and reinforcement information from different brain areas, this structure is suitable for segmenting a multi-dimensional behavioral context (Gluck & Myers, 1993).

In fact, a recent experiment by O'Keefe's group showed that these cells not only respond to specific locations, but also begin to behave as attractors for specific environmental shapes (Wills et al., 2005). This is in good agreement with our current hypothesis that the hippocampus could function as a bottom-up categorizer for behavioral context in general. In common with many other learning networks, such a categorizing network will show the capacity for generalization (Ghirlanda & Enquist, 2003), supplying the other essential prerequisite for the type of historical structural transformation dealt with in our previous chapter (cf. Hurford, 2000).

While the hippocampus functions as a bottom-up categorizer, top-down effects are probably governed by the prefrontal cortex. While the learning rate of the neocortex is slow, the hippocampus learns rapidly based on concurrent, local, and time-limited information (O'Reilly & Rudy, 2000; O'Reilly & Norman, 2002). Thus, the hippocampus can segment contextual parameters quickly, and the result of segmentation is tested by statistical or rule-based prediction by the prefrontal cortex. This is similar to the case of string segmentation in which basal ganglia give rise to short term statistics of on-going stimuli, the prefrontal cortex is more likely to function at a slower rate with a longer time constant.

22.5. Putting them Together

22.5.1. Mutual Categorization of Vocalizations and Behavioral Contexts: A Rudimentary Step

It is not easy to show the process of string-context mutual segmentation experimentally since it is a historical process. But even in the short run, similar processes might occur. Precedents for a rudimentary form of acquired "naming" achieved through association between behavioral contexts and particular behavioral tokens exists in several animals. When this process is enhanced by the ability to learn new behavioral tokens and combine them freely, that is, for example, by vocal learning, we are almost at the stage to begin the string-context mutual segmentation.

We observed that when trained to use a rake to retrieve distant food, monkeys spontaneously began to vocalize "coo" calls. They did so especially when the preparation of the rake tool by the experimenter was delayed (Hihara et al., 2003).

To further investigate this phenomenon, we systematically manipulated behavioral contexts by presenting the tool or food whenever the monkey made a vocalization *irrespective of the type of call emitted*. In one experimental situation, the experimenter placed a food morsel at a distance when the monkey produced a coo call (Call occasion A). By the second coo call (Call occasion B) the experimenter gave a rake tool to the monkey. The monkey could retrieve the food by the rake. In another, the experimenter gave the rake to the monkey beforehand. A piece of food was placed at a distance when the monkey vocalized a coo call (Call occasion C). Again, no attempt was made to differentiate the calls by type in these situations (A, B, and C). After 5 sessions of trainings, the monkeys eventually used acoustically distinct types of calls when they asked for the tool (Call B) or food (Call A and C). The calls used to ask for the tool was longer and higher pitched than the ones used to ask for the food. Calls had become correlated with context without being differentially reinforced to this end. We note the relevance of this finding to the process of segregation of song-strings by context discussed in our previous chapter.

As a possible explanation for this finding we suggest that the different reward conditions (food or tool) set up different emotional contexts for the monkeys. Different emotional contexts, in turn, affected the production of coo calls differently for the tool or food situations. Since the tool training can be assumed to activate the neocortex very highly, the calls were associated with different behavioral contexts. Thus, the calls became categorized and emotionally differentiated calls gradually became categorical vocalizations. Through this process, we suspect the emotional coo calls changed into categorical labels denoting the behavioral situation. We speculate that this categorization of vocal tokens may be related to highly specified behavioral situations. Such specified behavioral situations would evoke specific emotional content in the limbic system including amygdala and hippocampus. States of excitation in the limbic system may thus be labeled by the emotional vocalizations associated with the situations.

Naturally, our macaque example is limited to a contextual association between *call* types and context, but the principle is no less applicable to the association of contexts with full songstrings produced by a singing animal, as discussed in the previous chapter.

22.6. Conclusion

Taking these speculations together, we propose that longer strings are segmented by the action of the cortex-basal ganglia pathway and each part-string is represented by a mnemonic. In either case, the parallel, complex behavioral contexts are segmented by the prefrontal- hippocampal loop and each part-context may also be represented by a different mnemonic. These two sets of mnemonics can interact together to represent both part-string and part-context. Alternatively, if the process advanced slowly, the two sets of mnemonics may not be necessary and string and context segmentation would proceed with one common mnemonic.

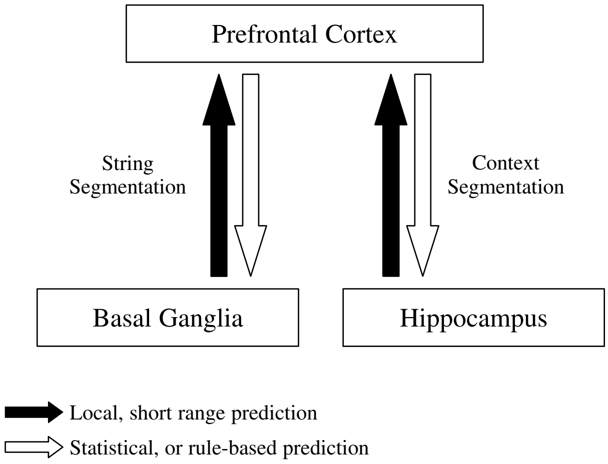

FIGURE 22.4. Neural substrates for string-context mutual segmentation

In any case, the parallel operation of these two systems would then enable mutual segmentation and matching of behavioral context and song strings, a process tantamount to an incipient language system (Merker & Okanoya, 2005).

In this regard, language may be possible without assuming a special "recursion" device suggested by Hauser, Chomsky and Fitch (2002). The recursive function might arise secondarily from the interaction between the prefrontal-basal ganglia loop and the prefrontal-hippocampus loop (Figure 22.4); that is, although the recursive function reflects the formal aspect of language, it may actually be the outcome of syntax-semantics interaction. Once a system of mutual segmentation is in place, the historical process itself may generate the remaining formal linguistic devices and structures. As shown by Kirby (2002) even recursion can emerge from the historical process, given "predicates that can take other predicates as arguments."

References

Alexander, R. D. (1990). How Did Humans Evolve? Reflections on the Uniquely Unique Species. *University of Michigan Museum of Zoology Special Publication, 1,* 1–38.

Bates, E. & Elman, J. (1996) Learning rediscovered. *Science*, 274, 1849–1850.

Bottjer, S. W., Miesner, E. A. & Arnold, A. P. (1984). Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science*, 224, 901–3.

Boughman, J. W. (1998). Vocal learning by greater spear-nosed bats. *Proc. R. Soc. Lond.B*, 265, 227–233.

Catchpole, C. K., & Slater, P. J. B. (1995). *Bird song: Biological themes and variations.* Cambridge, UK: Cambridge University Press.

Deacon, T. W. (1997). *The symbolic species.* Norton & Company, New York.

Fodor, J, & Bever, T (1965). The psychological reality of linguistic segments. *J verbal learn verbal behav* 4, 414–420.

Ghirlanda, S. and Enquist, M. (2003). A century of generalization. *Animal Behaviour*, 66, 15–36.

Gluck, M. A., Myers, C. E. (1993). Hippocampal mediation of stimulus representation: a computational theory. *Hippocampus*, 3, 491–516.

Gomez, R. L. & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, 4, 178–186.

Hauser, M. D., Newport, E. L. & Aslin, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: statistical learning in cotton-top tamarins. *Cognition*, 78, B53–B64.

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

Hihara S, Yamada H, Iriki A, Okanoya K. (2003). Spontaneous vocal differentiation of coo-calls for tools and food in Japanese monkeys. *Neurosci Res.* 45, 383–9.

Hosino, T. & Oknanoya, K. (2000). Lesion of a higher order song nucleus disrupts phrase level complexity in Bengalese finches. *NeuroReport,* 11, 2091–2095.

Hurford, J. R. (2000). The emergence of syntax. Editorial introduction to the section on syntax in C. Knight, M. Studdert-Kennedy and J. Hurford (Eds.), *The Evolutionary Emergence of Language: Social function and the origins of linguistic form*, pp. 219–230. Cambridge: Cambridge University Press.

Janik, V. M. and Slater, P. J. B. (1997). Vocal learning in mammals. *Advances in the Study of Behavior, 26*, 59–99.

Jarvis, E. D. et al. (2000). Behaviourally driven gene expression reveals song nuclei in hummingbird brain. *Nature*, 406, 628–632.

Jarvis ED (2004). Learned birdsong and the neurobiology of human language. In HP Ziegler and P Marler, eds. *Behavioral Neurobiology of Birdsong*, pp. 749–777. *Annals of the New York Academy of Sciences*, 1016.

Jurgens U. (2002). A study of the central control of vocalization using the squirrel monkey. Med. Eng. Phys., 7-8, 473–477.

Kirby, S. (2002). Learning, bottlenecks, and the evolution of recursive syntax. In T. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, pp. 173–204. Cambridge University Press.

Kobayashi, K., Uno, H. & Okanoya, K. (2001). Partial lesions in the anterior forebrain pathway affect song production in adult Bengalese finches. *NeuroReport*, 12, 353–358.

Kuypers, H. G. J. M. (1958). Corticobulbar connection to the pons and lower brain-stem in man. Brain, 81, 364–88.

Maess, B., Koelsch, S., Gunter, T. C. & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. Nature Neuroscience 4, 540–545.

Merker, B. & Okanoya, K. (2005). Contextual semanticization of song strings syntax: a possible path to human language. *Proceedings of Second International Symposium on the Emergence and Evolution of Linguistic Communication*, Universisty of Hertfordshire, April 2005, pp. 72–75.

Okanoya, K. (2002). Sexual Display as a Syntactic Vehicle: The Evolution of Syntax in Birdsong and Human Language through Sexual Selection. In A. Wray (Ed.), *The transition to language*, pp. 44–64. Oxford: Oxford University Press.

Okanoya, K. (2004). Song syntax in Bengalese finches: proximate and ultimate analyses. Advances in the study of behaviour, 34, 297–345.

Okanoya, K., Ichii, H., Ushijima, H. (2002). Baby cry as a pre-adaptation to language. *Proceedings of Fourth International Conference on the Evolution of Language*, Harvard, March 2002, pp. 86.

Okanoya, K., Hihara, S., Tokimoto, N., Tobari, Y. & Iriki, A. (2004). Complex vocal behaviour and cortical-medullar projection. *Proceedings of First International Symposium on the Emergence and Evolution of Linguistic Communication*, Kanazawa, Japan, May 2004, pp. 65–69.

O'Keefe, J. & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34, 171–175.

Ölveczky, B. P., Andalman, A. S. & Fee, M. S. (2005). Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol.* 3, e153.

O'Reilly, R. C. & Rudy, J. W. (2000). Computational principles of learning in the neocortex and hippocampus. *Hippocampus,* 10, 389–397.

O'Reilly, R. C. & Norman, K. (2002). Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends in Cognitive Science*, 505–510.

Parsons, L. M., Sergent, J., Hodges, D. A. & Fox, P. T. (2005). The brain basis of piano performance. *Neuropsychologia*, 43, 199–215.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6, 674– 681.

Pepperberg, I. (1999). The Alex studies: cognitive and communicative abilities of grey parrots. Harvard University Press, Cambridge.

Payne, K. (2000). The progressively changing songs of humpback whales: A window on the creative process in a wild animal. In N.L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music,* pp. 135–150. Cambridge, MA: The MIT Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, 274, 1926–1928.

Saffran, J.R. et al. (1999). Statistical learning of tonal structure by adults and infants. *Cognition*, 70, 27–52

Suge, R. & Okanoya, K. (under revision). Perceptual chunking in self-produced song by Bengalese finches. *Animal Cognition.*

Trivers, R. L. (1974). Parent-offspring conflict. *Americal Zoologist*, 14, 249–264.

Wild, J. M. (1993). Descending projections of the songbird nucleus robustus archistriatalis. *J. Comp. Neurol.* 338, 225–241.

Wild, J. M., Li, D. & Eagleton, C. (1997). Projections of the dorsomedial nucleus of the intercollicular complex (DM) in relation to respiratory – vocal nuclei in the brainstem of pigeon and zebra finch. *J. Comp. Neurol.* 377, 392–413.

Williams, H. & Staples, K. (1992). Syllable chunking in zebra finch (Taeniopygia guttata) song. *Journal of Comparative Psychology*, 106, 278–286.

Wills, T. J., Lever, C., Cacucci, F., Burgess, N. & O'Keefe, J. (2005). Attractor dynamics in the hippocampal representation of the local environment. *Science*, 308, 873–876.

Zahavi, A. and Zahavi, A. (1997). The handicap principle: A missing piece of Darwin's puzzle. Oxford University Press, Oxford.

Zeigler, H. P. & Marler, P. (Eds.) (2004). Behavioral Neurobiology of Birdsong. *Annals of the New York Academy of Sciences, 1016.*