eScholarship
International Journal of Comparative Psychology

Title
Animal Communication and Human Language: An overview

Permalink
https://escholarship.org/uc/item/3b7977qr

Journal
International Journal of Comparative Psychology, 29(1)

ISSN
0889-3675

Author
Barón Birchenall, Leonardo

Publication Date
2016

DOI
10.46867/ijcp.2016.29.00.07

Copyright Information

Peer reviewed

Animal Communication and Human Language: An overview

Leonardo Barón Birchenall

Aix Marseille Université, CNRS, LPL UMR 7309, 13100, Aix-en-Provence, France

"...il ne s'est toutefois jamais trouvé aucune bête si parfaite, qu'elle ait usé de quelque signe, pour faire entendre à d'autres animaux quelque chose qui n'eût point de rapport à ses passions; et il n'y a point d'homme si imparfait, qu'il n'en use."[1]

René Descartes, Lettre au Marquis de Newcastle

Comparative research has proven to be a fruitful field of study on the ontogenetic and phylogenetic evolution of language, and on the cognitive capacities unique to humans or shared with other animals. The degree of continuity between components of human language and non-human animal communication systems, as well as the existence of a core factor of language, are polemic subjects at present. In this article, we offer an overview of the research on animal communication, comparing the resulting data with the current knowledge on human language development. We try to summarize what is currently known about "language abilities" in multiple animals, and compare those facts to what is known about human language. The aim of the article is to provide an introduction to this particular topic, presenting the different sides of the arguments when possible. A special reference is made to the question of syntactic recursion as the main component of language, allegedly absent among non-human animals. We conclude that the current state of knowledge supports the existence of a certain degree of continuity between different aspects of animal communication and human language, including the syntactic domain.

I. Comparative Linguistics Approach

In a general sense, the term animal communication refers to the transfer of information by an animal that provokes a change in the behavior of the receiver of the information (Preece & Beekman, 2014). Various groups of investigators that believe in the existence of shared characteristics between human language and non-human animal communication have been doing extensive research over the past 25 years, on humans, apes, rodents, and birds, among others. Mainly, the purpose of these studies has been to determine which components of the language faculty are specifically human, which ones could be associated with domain-general mechanisms or could have originally served non-linguistic functions, and which language related capabilities have been inherited by humans from different precursor systems that were already present in ancestral species (Gervain & Mehler, 2010; Ohms, Gill, Heijningen, Beckers, & ten Cate, 2010; Sinnott & Gilmore, 2004). The academic community, however, has not attained a consensus on the validity of the comparative approach on language studies (see Trout, 2001, 2003).

Controversy aside, comparative research has proven to be a fruitful field of study on the ontogenetic and phylogenetic evolution of language, and on the cognitive capacities unique to humans or shared with other animals. This research has revealed a rich set of cognitive capacities on non-human animals; yet, continuity between human language and any other form of animal communication has been refused by many academics. In recent times, one especially persistent argument has been that animals are unable to process a particular syntactic operation: the center-embedded recursion. This specific kind of syntactic processing has been claimed to be the main factor of human language. In the following, we discuss some of the most relevant

---

[1] *"...there has never been known an animal so perfect as to use a sign to make other animals understand something which expressed no passion; and there is no human being so imperfect as not to do so".* René Descartes, Letter to the Marquess of Newcastle

Please send correspondence to Leonardo Barón Birchenall, Aix Marseille Université, CNRS, LPL UMR 7309, 13100, Aix-en-Provence, France. (Email: laescaladesol@gmail.com)

studies on communication systems in different species, and we compare their results with the current data on human language research (particularly on infants' linguistic skills). The aim of this article is to provide a panorama of the relation between human language and non-human animal communication, highlighting the subjects of the uniqueness of language, and the components of the linguistic faculty shared between humans and other animals. The text is broadly organized by species' capacities, but comparisons between species are presented in several sections and an entire section is devoted to the studies on animals' speech perception abilities.

In order to unify the terms presented in the paper, we will use the following classification of animal vocalizations, proposed in Arriaga and Jarvis (2013): *Notes* are the basic acoustic unit of the vocalization. They are formed by a single continuous sound with gradual variations in fundamental frequency. *Calls* and *syllables* are combinations of one or more notes separated by periods of silence. Calls are usually produced in isolation or in short bursts, and may have semantic content on their own, whereas syllables are commonly included in a longer series of rapidly produced vocalizations, and can be void of specific meaning (so syllables not necessarily serve a communicative function if produced in isolation). *Motifs* are series of syllables arranged in a somewhat fixed order, which makes it possible to generate a wide variety of communication units from a limited repertoire of syllables. Finally, *songs* are sets of vocalizations delivered periodically, sometimes with a rhythm. Songs may be produced spontaneously or in response to an external stimulus, and typically contain multiple types of vocalizations.

II. Non-Human Primates (NHP)

A. Broca's Area and Mirror Neurons

Broca's area is a critical brain region for the production of human speech, particularly in motor aspects such as articulation and fluency. Although it has been argued that the NHP brain does not have this area (Suddendorf, Borrello, Allen, & Radick, 2012), many NHP, including great apes, do have a *homologue* of Brodmann's area 44 (which in humans is part of Broca's area): the ventral premotor cortex, known as F5 (Cantalupo & Hopkins, 2001; Pepperberg, 2010; note that *homologous traits* in different organisms are considered as likely derived from a common ancestor, without necessarily serving the same function; by contrast, functionally similar features resulting from independent evolution are termed *analogous traits*). In chimpanzees (*Pan*), bonobos (*Pan paniscus*), and gorillas (*Gorilla*), F5 area is larger in the left hemisphere of the brain than in the right one; a morphological asymmetric pattern similar to the one found in the cortical areas of the human brain counterparts. In addition, the NHP brain exhibits a greater activation of the left hemisphere (with respect to the right one) in response to vocalizations of conspecifics (Ghazanfar & Hauser, 1999).

Area F5 in the NHP brain is the main site containing *mirror neurons*. These neurons discharge both when the monkey performs an action and when it observes another individual performing it, allowing the recognition of others' motor actions by matching them with an internal motor copy. A similar mechanism is present in the human brain, involving Broca's area (among other areas); and although the main function of the mechanism is to recognize other's actions, it may have evolved in humans for intentional communication purposes (Rizzolatti, 1998). The existence of an equivalent of the mirror neuron system in birds and parrots brains has also been proposed (Pepperberg, 2010).

On the other hand, the dorsal pathway in the superior temporal cortex, which connects Brodmann's area 44 and Wernicke's area, is weaker in NHP than in humans (Berwick, Friederici, Chomsky, & Bolhuis,

2013). Furthermore, it has been proposed that although in both human and NHP brain the left Planum Temporale (within Wernicke's area) participates in the processing of non-linguistic sounds, only in humans it also has a specific sensitivity to sound patterns found in natural sign or spoken languages (Petitto, 2005). If this turns out to be the case, NHP would share with humans the ability to process non-linguistic sounds, but not to process linguistic patterns themselves. This important difference could be due to small, but meaningful, organizational differences between the NHP and the human brain.

B. Human Language Learning

Several unsuccessful attempts were made to teach a spoken language to chimpanzees during the first half of the twentieth century. The situation changed drastically in the 60s with the work of Allen and Beatrix Gardner, who toke a different approach: teaching sign language to a female chimp named Washoe (Gardner & Gardner, 1969). According to these authors, after training for near 50 months, Washoe learned how to communicate with signs more than 100 different words, taught spontaneously some of those words to other chimpanzees, and even accomplished to combine words in novel forms. However, although she was able to express simple desires and needs through signs, she never became a fluent American Sign Language user, and her signs were usually repeated, confusing, and without a combinatorial structure. As reported by Savage-Rumbaugh, Shanker and Taylor (1996), Washoe was also able to produce different signs to refer to different objects, and generalize those signs to new objects that were alike in form or belonged to a similar conceptual class. Yet, Washoe only used signs when someone else communicated to her first, and she typically employed sign combinations that were already used by her caregivers.

The results of experiments involving Washoe were controversial, particularly with regard to the possible "mental capabilities" underlying the observed linguistic behaviors. General public, and linguists in particular, become quite skeptical when it was reported that the Gardner couple had not recorded the signs made by Washoe in order, because they were focused on "what she said" rather than "how she said it." While Washoe was learning sign language, Sarah and Lana (two female chimpanzees) were trained to communicate with lexigrams (i.e., arbitrary symbols representing words), which first consisted in plastic magnetic tokens with drawings on them, and later in visual patterns of colors over an electronic panel connected to a computer. Sarah and Lana learned somewhat between 100 and 600 lexigrams, including abstract terms such as logical, mathematical, and grammatical operators (e.g., negation, conditioned relations, and equality; Igoa, in press).

Some years later, a chimpanzee named Nim was raised in a human social environment and was exposed to a sign language outside his home. Despite some erroneous premature claims, Nim's maximum attained level of language development consisted in repeating utterances made by others, employing some "wild card signs" that were appropriate to the context (e.g., "hurry", "me", "Nim"), and communicating almost always in order to obtain a reward, rather than to convey information (Savage-Rumbaugh, Rumbaugh, & Fields, 2009, p. 27). By the same time, two other male chimpanzees (Sherman and Austin) were claimed to be capable of producing and understanding abstract symbols, and using them to represent objects and communicate between each other. These chimps were successfully trained to build up "sentences" by pressing buttons in a keyboard with symbols representing nouns or verbs. However, during their communicative exchanges, they both requested things but neither gave them (Savage-Rumbaugh et al., 2009). Sherman and Austin never understood sentences longer than two words, and their communications were almost always related to emotions and physical needs. Nevertheless, based on observations of the two chimps apparently "monitoring" the effectiveness of their communications, their caregivers assumed the occurrence of some kind of conversations between them, integrated with the rest of the interaction behaviors.

In contrast to the chimpanzees mentioned above, all of which received explicit language training, the bonobos Kanzi and Panbanisha (male and female, respectively) did not. These bonobos were kept from their birth in an English-speaking environment, in which speech and lexigrams were used to communicate with them. When Kanzi or Panbanisha wished to communicate, they pressed a lexigram in a board and thus a computer emitted the correspondent word (Savage-Rumbaugh, McDonald, Sevcik, Hopkins, & Rubert, 1986; a paper version of the lexigrams could also be used when being outside). According to Savage-Rumbaugh et al. (1986), these conditions were sufficient for the acquisition and understanding of the spoken language, in a level similar to that of a three-year-old child. That includes understanding of future and past tense sentences, understanding of embedded constructions, learning and differentiation of written symbols, association of symbols with spoken words, and spontaneous communication of desires and thoughts through lexigrams. In the case of Kanzi and Panbanisha, however, comprehension greatly exceeded production, because even though these bonobos could understand utterances from four to eight words, sometimes the first time they heard it, their production was usually limited to a single word (lexigram) accompanied by gestures or vocalizations. In any case, the findings of the Savage-Rumbaugh team should be taken with care, given that several of them are based on anecdotal evidence rather than on experimental data.

Additionally, whereas children's usage of language appears to follow some kind of natural categories (i.e., events, actions, and objects; see Keil, 1983), overall NHP do not display this kind of sensitivity to differences among natural kinds, even after years of training and communication with humans. Chimpanzees, for example, use the sign for "apple" to refer to the action of eating apples, to the place where apples are stored, and to events and objects related to apples. In the words of Petitto: "chimps do not really have 'names for things' at all. They have only a hodge-podge of loose associations with no ... internal constraints or categories and rules that govern them" (Petitto, 2005, p. 86-87). Likewise, although NHP demonstrate capabilities to associate words (signs) with concrete meanings and events in their environment, they seem to have a very poor ability to associate words with abstract meanings.

C. Natural Communication and Syntax

In their natural environment, NHP exhibit a complex way of communicating that resembles human language in different aspects, including: (a) use of temporal features for identifying different types of calls; (b) use of calls to refer to objects and events in the environment; (c) use of calls to signal events concerning food, predators, and social relations; and (d) moderate flexibility in production and understanding of vocalizations (though very restricted in NHP; Ghazanfar & Hauser, 1999). The wild silvery gibbon (*Hylobates hoolock moloch*), for example, produces songs consisting in sequences of notes combined in complex structures (with variety and repertoire highly reduced with respect to songbirds). These songs serve functions such as positioning among groups, territorial and food defense, mate attraction, and strengthening of the pair bond. Songs may be interpreted individually or in duets, and have different structures depending on the sex of the vocalizer (Geissmann & Nijman, 2006).

Another example is the grivet monkey (*Chlorocebus aethiops*), which uses a limited repertory of differentiated vocal signals to alert others to different types of predators, without apparently noticing if the recipient of the message is or is not aware of the presence of the predator (Manser, 2013). In this regard, consider also the Vervet monkeys (*Chlorocebus pygerythrus*), which continue to produce alarm calls even when the entire group has already seen the predator and run to safety (Fitch, 2005). Together, these data indicate that monkeys do not intend to transmit information to others when communicating, and do not arrange their messages in accordance with the recipient's state of knowledge.

On the other hand, unlike other NHP, Chimpanzees have a multimodal form of communication that includes hand gestures, glances, facial expressions, postures, and vocalizations (the last-mentioned constituting a low developed form of communication compared to their gestural communication). These communicative elements, used in a coordinated fashion, serve different functions, including group cohesion, conflict resolution, and intersubjective contact (Mitani, Hasegawa, Gros-Louis, Marler, & Byrne, 1992). According to Mitani et al. (1992), although chimpanzees can use vocalizations to establish contact or maintain distance between individuals, or to constitute cooperative alliances between distant males, their content tends to be of an emotional character, involving a mostly genetically determined transmission of information.

Otherwise, it is a polemic matter within language studies if the combination of primate vocalizations follows some kind of structural rules. According to Bickerton (1990), none of the NHP' communication systems have something comparable with the grammatical elements found in human language. Therefore, communicative units used by NHP could not be altered to include different meanings or nuances, nor be decomposed into their constituent parts. Conversely, Ouattara, Lemasson and Zuberbühler (2009), propose that although there is no syntax per se in non-human animal communication, this does not imply that this type of communication does not follow at all certain combinatorial principles. For instance, Adult male Campbell's monkeys (*Cercopithecus campbelli*) produce six different types of loud calls, which can be combined in nine different sequences. Resulting sequences are then associated with specific external events, such as degree of predation threat, spatial relations within the group, specific predator classes, and falling trees (Ouattara et al., 2009). Campbells would also adhere to several combinatorial principles when linking together call sequences, including: (a) production of sequences composed of calls that already carried narrow meanings (sequence and call meanings being identical); (b) production of meaningful sequences composed of calls with unspecific meanings; (c) combination of two meaningful sequences into a more complex one with a different meaning; (d) addition of meaningless calls to an already meaningful sequence, thus changing its meaning; and (e) addition of meaningful calls to an already meaningful sequence, thus refining its meaning. Note that although Ouattara et al. (2009) do not imply that the combinatorial principles present in Campbells' vocalizations are equivalent to the syntax of human language, they do argue that those combinatorial principles represent a trace of the existence of a *proto-syntax* in the animal kingdom (yet they concede it is unlikely that monkeys intentionally produce these combinations).

Putty-nosed free-ranging monkeys (*Cercopithecus nictitans*), in turn, combine two types of calls in different sequences, which are associated with specific external events, such as predation threat and imminent movements of the group (Arnold & Zuberbühler, 2006, 2008). Apparently, the different sequences of calls of these monkeys can communicate at least three types of information: the quality of the event witnessed by the male, the caller identity, and whether the caller intends, or not, to advance. The way in which calls are arranged, rather than some acoustic feature of the individual call, would convey the meaning of these monkeys' communications. Further evidence of combination of calls into complex sequences, and association of these sequences with external events in a significant way for the recipient, has also been found in black-fronted titi monkeys (*Callicebus nigrifrons*; Cäsar, Zuberbühler, Young, & Byrne, 2013), adult Diana monkeys (*Cercopithecus diana*; Candiotti, Zuberbühler, & Lemasson, 2012), and White-handed gibbons (*Hylobates lar*; Clarke, Reichard, & Zuberbühler, 2006). In this regard, Tomasello (2005) argues that no species of ape has specific alarm calls or any other vocalizations that appear to be referential (such as those being discussed); hence, monkey's referential calls could not be the direct precursor of human language. In contrast, Hurford (2012) suggest that human voluntary vocalized words evolved from involuntary ape cries.

With respect to great apes, according to her caregivers, Lana was able to produce utterances combining several lexigrams. She "was acquiring not only words, but complete 'stock sentences' with embedded grammatical rules, such as 'Please machine give piece-of bread'" (Savage-Rumbaugh et al., 2009, p. 26). Additionally, Savage-Rumbaugh argues that although Lana required a lot of explicit training for understanding

complex utterances, she managed to produce some novel structures with a simple grammar basis. Conversely, as believed by Petitto (2005), the structural organization of NHP' communications and the syntax of human language are different in a fundamental matter: Although NHP can link together one or two signs in ways that seem to follow structural patterns, they are unable to build sequences of patterns of three, four, or more signs, as a four-year-old child exposed to a sign language is perfectly capable to do. (We revisit the issue of this section later, when discussing birds' syntactic capacities.)

D. Pointing Gesture and Displacement

The ability to refer to entities distant in time or space, or both, is called *displacement*, and is one of the key features of human language (Hockett, 1969). This ability arises in infants between 11 and 12 months, in the form of the *pointing gesture*, and it is used henceforth for the rest of the life (Elgier & Bentosela, 2009). Depending on the role played in the interaction processes, the pointing gesture can be *imperative pointing* or *declarative pointing*. The imperative pointing is a way to make someone do something, by using him as if he were some kind of "human tool", whereas the declarative pointing is a way to direct someone's attention toward some object or event.

The act of signaling, however, is not exclusive to humans: most primate species do it. The pointing gesture is also very common in apes in captivity, yet rare in their natural habitats (if not absent). During captivity, apes perform spontaneously imperative pointing, usually to indicate their desire to eat or to reach an object. Nonetheless, the structuration of the sign tends to be different from that of humans, because apes extend the entire hand rather than just the index finger (Leavens, Hopkins, & Bard, 1996; Liszkowski, Schäfer, Carpenter, & Tomasello, 2009).

Furthermore, as concluded by Liszkowski et al. (2009), overall animal communication does not appear to involve references to absent entities. Should this be the case, men and NHP would not share the displacement ability, which in turn suggest that this particular skill appeared in human evolution only after the divergence with great apes, about six million years ago, prior to the emergence of language (see Ghazanfar & Hauser, 1999). Conversely, Lyn et al. (2014) have found that chimpanzees and bonobos are indeed capable of referring to absent and displaced entities. The difference between these two arguments, according to Lyn et al. (2014), is due to methodological flaws in Liszkowski et al.'s (2009) work, aggravated by errors of interpretation (we will return to the topic of displacement in the section about honeybee dances).

E. Intentionality

NHP' vocal behavior is often seen as irrelevant when explaining the evolution of human language, mainly because NHP have a very restricted vocal control, and their calls apparently lack intentionality (Ouattara et al., 2009). Moreover, although some species of NHP can assume the presence of an unseen predator from their conspecifics vocalizations, the vocalizer must always see or hear the predator in order to call (Liszkowski et al., 2009). This being the case, primate calls would be a direct reaction to an event or a physical state, without any intention to inform others about a referent perceptually absent for them. Contrariwise, several studies indicate that chimpanzees do possess the ability to recognize the intentional states of others, at least at some extent, and that this ability improves in competitive contexts (Tomasello, Call, & Hare, 2003, and references therein). Chimpanzees would thus gather some information about what others see and not see, or saw or did not see in the immediate past. They also would consider the goals of others'

behaviors when planning their own behavior during interactions. Note that the fact that a competitive context favors this kind of "mind reading" skill suggests that chimps can control this ability in a certain degree.

Furthermore, Povinelli, Bierschwale and Cech (1999) found that chimpanzees and 3-year-old human children can both use the gaze direction and the whole head orientation of a person to find hidden food, but cannot rely for this purpose only in the gaze direction. In these experiments, chimps were more effective than children finding the hidden food when the person's gaze was directed approximately toward the place in which the food was but away from its actual container. According to Lyn et al. (2014), Povinelli and colleagues (1999) erroneously interpreted these findings as evidence for apes' inferiority with respect to cognitive representations of visual perspective.

Nevertheless, different interpretations of what intentionality means have been proposed. Tomasello, for example, suggests the existence among humans of a *shared* (or *we*) *intentionality* (Tomasello, Carpenter, Call, Behne, & Moll, 2005). This type of intentionality implies collaborative interactions in which participants have a shared goal, coordinated behaviors, and joint attentional processes. Interactions of this type "require not only an understanding of the goals, intentions, and perceptions of other persons, but also, in addition, a motivation to share these things in interaction with others – and perhaps special forms of dialogic cognitive representation for doing so" (Tomasello et al., 2005, p. 2). Fitch (2010), in turn, suggests the existence of *first-* and *second-order* intentionality. First- (or zero-) order intentionality implies a link between a mental representation and something that exists in the world, without involving a specific intent to inform another individual of something, modifying thus his internal representations. Second-order intentionality also requires the goal of changing another individual's state of knowledge when communicating.

About non-human animal's cognitive skills and possession of concepts, Fitch comments:

Many capabilities that were long thought to be unique to humans have now been demonstrated convincingly in animals. These include cross-modal association, episodic memory, anticipatory cognition, gaze following, basic theory of mind, tool use, and tool construction. With these data, we can answer the old question of whether animals think, and have concepts, affirmatively. If by "concepts" we simply mean "mental representations, not necessarily conscious," few scientists today question the notion that animals have concepts, at some level, and contemporary cognitive ethologists and comparative psychologists are providing an ever more impressive catalog of the types of concepts that non-linguistic creatures possess and manipulate. (Fitch, 2010, p. 171-172)

If Fitch is right, then we can conclude that apes' communicative behaviors are intentional, in the sense that they are goal-oriented and governed by a purpose. They belong to the zero-order intentionality because they implicate a relation between (basic, non-conscious) internal representations and things in the world. However, given that reports of apes' declarative communications, as far as we know, are absent in the scientific literature, the mainly intention of these behaviors appears to be imperative. Furthermore, NHP do not seem to envisage the modification of others' internal representations when communicating with them, neither seem to be able to share their mental lives with others (therefore, no second-order or shared intentionality). One possible explanation of this phenomenon is provided by Vygotsky's (1934/1986) hypothesis, according to which, only human beings can talk about what they think and solve problems with the help of their thoughts, due to a connection between thought and language that occurs around the second year of life. In this sense, non-human animals may have basic thoughts and (sometimes elaborate) problem-solving skills, but they seem unable to functionally link both of them (i.e., establish a connection between concepts and phonetic or motor representations that support the execution of words and signs).

III. Speech-Related Perceptual Capacities

Perceptual capabilities related to speech are one of the most studied topics in comparative linguistics. Most research has been done on NHP, but studies have also been conducted in rats, birds, and other species. In this section, we present and compare briefly some key results related to prosodic and phonologic discrimination, and detection of distributional regularities within the speech signal.

Cotton-top tamarin monkeys (*Saguinus oedipus*) can discriminate between two languages belonging to different rhythmic categories (in this context, *to discriminate* means to react different to). This behavioral response is equivalent to the response toward the same stimuli of two- to five-day-old newborns (Ramus, Hauser, Miller, Morris, & Mehler, 2000) and Long-Evans rats (*Rattus norvegicus*; Toro, Trobalon, & Sebastián-Gallés, 2003). Nonetheless, tamarins are unable to differentiate between two languages belonging to the same rhythmic category (Tincoff et al., 2005), neither five-day-old newborns (Nazzi, Bertoncini, & Mehler, 1998), nor two-month-old infants (Mehler et al., 1988). It is only around their fourth birthday that infants achieve the ability to perform such discrimination, as long as one of the languages is their mother tongue (Bosch & Sebastián-Gallés, 1997).

Tamarins and one-year-olds can both detect simple grammatical patterns (similar to those found in natural languages) and *predictive dependencies* (PDs) between elements within the speech signal (PDs are relations between types of words within a sentence. Some words, such as an article, require the presence of another word, such as a noun. Thus, the presence of one of these words within a sentence predicts the presence of the other). Infants can detect those dependencies between single items and between categories, whereas tamarins can only do it between categories (Saffran et al., 2008). For one-year-olds, however, it is difficult to learn grammatical patterns different from those found in natural languages; which suggests the influence of general learning constraints on the acquisition of linguistic structures.

Tamarins are also able to identify elements within the speech signal by detecting *transitional probabilities* (TPs) between them (Hauser, Newport, & Aslin, 2001; TPs are properties of distribution, such as the probability that a given syllable follows another one at the end of a word or a sentence). Eight-month-old infants exhibit an equivalent capacity, which might be a key factor for segmenting continuous speech into words during language acquisition (Saffran, Aslin, & Newport, 1996). Long-Evans, in turn, can track distributional regularities between elements in the speech signal detecting the frequency of co-occurrence of the syllables, but are unable to do it detecting TPs, or when the distributional regularities are implemented between non-consecutive elements of the signal (Toro & Trobalon, 2005). On the other hand, human adults and tamarins can segment the speech signal detecting TPs between non-consecutive elements. This detection, however, is relatively easy for humans when TPs are implemented between consonants or vowels, but becomes difficult when implemented between syllables. Conversely, for tamarins, detection of TPs is facilitated when they are implemented between syllables and vowels, but becomes harder when implemented between consonants (Newport & Aslin, 2004; Newport, Hauser, Spaepen, & Aslin, 2004; Peña, Bonatti, Nespor, & Mehler, 2002).

Approximately at seven months, infants can detect and generalize simple algebraic rules within the speech signal (Marcus, Vijayan, Bandi Rao, & Vishton, 1999). Long-Evans are also capable of detecting this kind of rules when they are implemented on sound sequences (Murphy, Mondragon, & Murphy, 2008), but not when implemented on word sequences (Toro & Trobalon, 2005). Eleven-month infants, however, can detect and generalize simple algebraic rules when they are implemented on vowels, but not when they are implemented on consonants (Pons & Toro, 2010). As for the NHP, experimental data indicate that tamarins can learn rules formally similar to the rules governing affixation processes in natural languages (Endress, Cahill, Block, Watumull, & Hauser, 2009).

With respect to phonological differentiation capabilities, Japanese macaque (*Macaca fuscata*) can discriminate between two stimuli within a computer-generated continua ranging from voiced to voiceless syllables (Kuhl & Padden, 1982). As human adults and infants, macaques perform better on the discrimination tasks when the stimuli are on opposite sides of a phoneme boundary than when both stimuli are in the side corresponding to one of the phonemes. This phenomenon, known as *phoneme boundary effect*, is apparently not exclusive to humans. Another speech perception related phenomenon not exclusive to humans is the *perceptual magnet effect*; according to which, a prototype token of a speech category allows significantly greater generalization toward similar sounds than a non-prototype one. This effect has been reported in human adults and six-month-old infants, but has not been found in Rhesus macaques (*Macaca mulatta*; Kuhl, 1991). These data indicate that the internal organization of the phonetic categories around prototypical tokens is a specific aspect of human speech, which appears early during the ontogenetic development. Nevertheless, main characteristics of the effect have been found in an avian species, suggesting that this particular perceptual phenomenon is unique to humans, among primates (Fitch, Hauser, & Chomsky, 2005).

Japanese macaques can also discriminate between tokens of natural consonant-vowel syllables produced in different vocal contexts and uttered by different male and female speakers (Sinnott & Gilmore, 2004). This discrimination process, however, exhibits a contextual effect with respect to the same stimuli, which is inexistent in humans. Macaques perform better when discriminating back vowels /a/ and /u/ with respect to front vowels /i/ and /e/, whereas humans' performance is equivalent in both cases. Likewise, after a long training, Zebra finches (*Taeniopygia guttata*) can categorize monosyllabic words that differ in a single vocal, and then apply this categorization to the same words spoken by new talkers, regardless of their gender (Ohms et al., 2010; in this context, *to categorize* means to react in a different fashion toward a group of stimuli with regard to another one). Ohms et al. (2010) think that Zebra finches could be using some perceptual strategies equivalent to those used by humans when categorizing the same type of words (see Ohms, Escudero, Lammers, & ten Cate, 2012). In that way, birds and humans would achieve to normalize the speech signal despite variations within the acoustic source and between different sources.

Lister hooded rats (*Rattus norvegicus*) are also able to discriminate between syllables in a human-like way. Like humans, these rodents rely on the syllables *rise time* (i.e., the time between the beginning of the syllable and its maximum amplitude) to discriminate between them (Reed, Howell, Sackin, Pizzimenti, & Rosen, 2003). Similar discrimination capabilities have been found in chinchillas (Kuhl & Miller, 1975), budgerigars (Dent, Brittan-Powell, Dooling, & Pierce, 1997), and quail (Trout, 2003). In Pepperberg's (2010) opinion, these perceptual abilities are likely to be a general vertebrate trait.

IV. Vocal Learning in Mice and Bats

House mice (*Mus musculus*) and brown rats (*Rattus norvegicus*) communicate themselves through *ultrasonic vocalizations* (USVs), produced in frequencies ranging from 30 to 110 kHz (Arriaga & Jarvis, 2013). In the case of mice, adult male USVs seem to reflect internal states and facilitate social communication during non-aggressive encounters. These vocalizations contain syllable-like sounds produced in regular temporal patterns, which differ between individuals. Making changes to a preexistent template, mice can modify the spectral-temporal structure of their vocalizations depending on external circumstances (Arriaga, Zhou, & Jarvis, 2012). This basic flexibility during communication represents a limited form of vocal learning, with respect to that of humans and songbirds. In humans, however, vocal learning takes place in childhood and adulthood, whereas in birds–and apparently in mice as well–vocal learning is restricted to a very specific period of time.

Note that the existence of vocal modification skills in mice is an outstanding fact, given that mammals in general, including NHP and excluding men and cetacean, have been traditionally considered poor vocal learners and imitators. NHP, particularly, change their vocalization features mostly modifying innate calls by altering the position of mouth and lips rather than controlling movements of the larynx (Colbert-White, Corballis, & Fragaszy, 2014). Conversely, animals that have proven vocal learning capabilities include hummingbirds, songbirds, parakeets, cetaceans, bats, elephants, and some pinnipeds (such as the elephant seal). These animals can imitate sounds from other species, but most of the time they build their vocal repertoire imitating sounds of their own species.

In most cases, birds' and cetaceans' vocal mimicry is made by males, which tend to use it for mating purposes. Nonetheless, vocal imitation also occurs in females, and for different purposes, such as group cohesion and bonding. In some cetaceans, such as the humpback whale (*Megaptera novaeangliae*), songs are produced by males in mating contexts, but in some others, such as dolphins (*Delphinidae*), songs are produced by both males and females (Fitch, 2005). In the case of humans, the amount of vocalizations tends to be similar in men and women, both in infants (Reimchen, 2013), and in adults (Mehl, Vazire, Ramirez-Esparza, Slatcher, & Pennebaker, 2007). Then, of course, not all animals communicate through vocalizations (i.e., emissions produced by a vocal organ: the syrinx, in birds, and the larynx, in frogs and most mammals). Some insects, for instance, are able to produce sounds with their legs and their wings to create courtship songs.

There is yet another mammal that seems to have considerable vocal imitation abilities, with which he can modify song structures depending on the surrounding context: the bat (Bohn, Smarsh, & Smotherman, 2013). Like birds, the majority of singing bats are males that tend to sing during mating season to court females and to defend their territory. Male bats, however, also sing at other times of the year, and in colonies in which no females inhabit (songs would then also serve purposes different from mating). Bohn et al. believe that bat songs are probably not innate, therefore requiring vocal learning, as in the case of bird songs and human speech.

In particular, Brazilian free-tailed bat (*Tadarida brasiliensis*) sings in a similar way to that of birds, producing hierarchical sound structures that vary in order and number of phrases between executions, following some specific rules of organization (Bohn et al., 2013). These bats' songs have a hierarchical structure with three types of phrases: chirps, trills, and buzzes; which in turn are composed of four types of syllables: chirp A, chirp B, trill, and buzz. A few types of elements (in this case, phrases) can be combined to form a set of potentially enormous unique units. Free-tailed bats (*Molossidae*), in turn, can quickly change the structure of their songs in response to contextual circumstances. In less than 100 milliseconds, these bats can go from a courtship singing into a territorial singing if another male echolocates nearby. Bat songs might even have simple rules for constructing melodies, such as *only chirp syllables can be included in chirp phrases*, *and buzz syllables in buzz phrases* (Morell, 2014).

Additionally, the development of vocal imitation capacities in bats follows some specific patterns. For instance, when raised in a colony, Egyptian fruit bat (*Rousettus aegyptiacus*) pups acquire the adult vocalizations repertoire, but when raised in acoustical isolation they do not develop normal adult vocalizations (Prat, Taub, & Yovel, 2015). If pups raised in isolation are exposed to abnormal (manipulated) bat vocalizations recordings, they develop a repertoire similar to such vocalizations. Note that in this study, pups raised in isolation were kept with their mothers during 80 days before the period of near five months of complete separation; but, according to the authors, in the absence of other adults, the mother does not vocally interact with her pup, not even in response to the pups' isolation calls. Overall, these results indicate that the acquisition of the bat vocal repertoire implies learning, and that exposure to adult vocalizations is crucial during the process.

During isolation, however, bat pups in the Prat et al. (2015) study started to incorporate adult-like segments into their calls. At the end of the experiment, the isolated group was mixed with the control group (bats already performing adult vocalizations). Then, the formerly isolated pups vocalizations became normalized (adult-like) within a month. This developmental pattern indicates that there is not a critical period for vocal learning such as the one described for songbirds (i.e., usually between ten days and four months after birth; Doupe & Kuhl, 1999).

Bat songs might even have semantic content. Male European Pipistrellus nathusii (*Pipistrellus nathusii*), for instance, conveys different types of information through different phrases within a song, including information about the bats' population, where to land, and an individual *vocal signature* (Morell, 2014; compare with dolphins' signature whistles below). Furthermore, bat songs seem to follow a specific developmental pattern, as has been found in male and female sac-winged bat (*Saccopteryx bilineata*) pups. These pups learn songs by babbling and imitating early in life, as the offspring of many species of birds (Morell, 2014), and human infants acquiring sign or spoken languages (Petitto & Marentette, 1991).

V. Dolphins: Conversation or Communication

Dolphins use whistles as contact calls to facilitate encounters between separate individuals, and it is possible that they also use sounds and body postures to indicate their intention to play, and clicks to signal their emotional states and their intention to communicate (Kuczaj, 2013). As birdsongs, dolphin's whistles follow a specific developmental pattern. For example, neonate bottlenose dolphins (*Tursiops truncatus*) practice whistling by overproducing different whistle segments before they begin to produce stereotyped adult-like whistles. The newborn calf's whistle often follows the mother's whistle, as if the calf were imitating a model (Jones, 2014, and references therein). According to Jones (2014), the bottlenose's early vocalizations fit the criteria of infant babbling; but, unlike human newborns, such vocalizations do not seem to elicit specific reactions from the mother (such as the elicitation of an infant-caregiver interaction).

Once whistling has been learned, bottlenose dolphins become able to address individual conspecifics with learned signals, by matching another dolphin's individualized *signature whistle* (i.e., particular frequency modulation patterns of the whistles, independent of general voice features; King, Harley, & Janik, 2014). The purpose of this behavior is seen mainly as an affiliative signal, but it could be also used to manage aggression. As in humans, the capacity to address a particular conspecific allows an individual to solicit the attention of another and to direct information toward an intended recipient. According to King et al. (2014), some male songbirds also can selectively address another bird, but they do it specifically as a sign of aggression. For doing this, the male songbird performs the song from his own repertoire that most closely resembles that of the other male he wishes to address (despite the latter being a neighbor or a stranger).

Nonetheless, given that conversation involves communication, but not vice versa, it is far to be established if dolphin communication involves some kind of conversation. Communication implies transmission of information between individuals, despite their intention to communicate. As long as the information is transmitted, communication has taken place. Conversations between humans, conversely, consist in joint activities that require cooperation among interlocutors, so that both (all) of them sufficiently understand the meaning of the dialog as a whole (Garrod & Pickering, 2004). Moreover, the capacity to send and receive messages during a conversation (termed *interchangeability*) is considered one of the hallmarks of language (Hockett, 1969).

Interchangeability is heavily influenced by another universal feature of human language: *turn-taking* during interactions (i.e., exchanges of speech with intervals of silence and minimal overlapping). Although it is believed that three-month-old infants already take part in vocal exchanges that are similar to adult conversations, it is not until around four months that turn-taking occurs in these exchanges (Reimchen, 2013). Among NHP, marmoset monkeys (*Callithrix jacchus*) have been found to take turns in their vocal exchanges, waiting for the first vocalizer to finish the call before responding (Takahashi, Narayanan, & Ghazanfar, 2013). One of the monkeys involved in the interaction makes the first call, and the second one waits for the end of the call and responds after a specific interval of about three to five seconds. If one of the monkeys accelerates or slows down the timing of the vocalization, the other does it as well. According to Takahashi et al. (2013), this vocal turn-taking behavior appears to be based on dynamics of coupled oscillators, similar to the dynamics of turn-taking in human conversations (see Cummins, 2009).

Returning to dolphins, Kuczaj (2013) points out that although it has been largely speculated that communication between them and humans is possible, no scientific evidence exists indicating the possibility of a significant conversations between dolphins (or between dolphins and humans) involving mutual exchange of information. In addition, it is unknown yet what the dolphin calls mean for both the sender and the recipient, or if dolphins can perceive the emotional state of another dolphin, or a human (which is a key issue during a conversation). Furthermore, several technical complications occur in this type of research, such as the need for special equipment to capture sounds (emitted by dolphins) that humans do not perceive, the difficulty to find which of the dolphins involved in an interaction produced a given sound, and the indeterminacy of the communication units and their referents.

VI. Dogs and Fast Mapping

Infants rapidly form a crude assumption about the meaning of a new word after being exposed to it only a few times. This phenomenon is known as *fast mapping*, and unlike syntax acquisition, it has been observed in both adults and infants (therefore it is not completely limited to a critical period of time during childhood; Markson & Bloom, 1997). Fast mapping facilitates the occurrence of the *lexical explosion* in infants between 18 to 24 months and five years. During the lexical explosion, a significant increase in the child's comprehensive and expressive vocabulary takes place, with a new word being learned almost every waking hour (Keil, 1983). Fast mapping has been also documented in infants as young as 13 months, as well as in children with Down syndrome, specific language impairment, and memory and attentional limitations (Gershkoff-Stowe & Hahn, 2007).

The process of fast mapping develops itself in two phases: at first, an initial connection is rapidly established between a word and a referent, allowing infants to acquire a partial knowledge of the meaning of such word. In the second phase, or extended mapping, the meaning of the word increasingly resembles that of adults, as the infants' experience with the referred object augments (Carey, 2010). In Carey's opinion, extended mapping consists in a process of testing hypotheses from a large space of syntactic and semantic primitives, generating new features that are subsequently associated with words.

Fast mapping, however, is apparently not exclusive to humans. According to Kaminski, Call and Fischer (2004), this process is likely to be mediated by memory and learning general mechanisms shared by human and other animals, rather than by a language-exclusive acquisition device. To test this hypothesis, a border collie's ability to associate words with their referents was measured, by analyzing his retrieval behavior of known and unknown objects. Kaminski et al. claim that the dog learned over 200 words corresponding to different objects. Some words would have been learned by direct teaching (i.e., showing the object and letting the dog play with it while repeating the corresponding word), others through some kind of learning-by-

exclusion mechanism (i.e., the dog was asked to bring an object whose name he did not know, from a place in which that object was among others whose name the dog already knew).

Given that Kaminski et al.'s (2004) study received critics due to methodological flaws, a new series of experiments was carried out by Pilley and Reid (2011), with another border collie. According to these authors, after intensive training the dog learned the proper nouns corresponding to 1022 objects, and retained the knowledge for a period of three years. The dog would have even understood the meanings of three different orders and three proper nouns, which were randomly matched, and also three common nouns representing categories (in this context, *to understand* means to exhibit an adequate behavior of recovery in response to a given order). One word to diverse objects mapping, and vice versa, were also observed in the dog's behavior, as well as the learning by exclusion claimed by Kaminski et al. Nonetheless, unlike Kaminski's study, Pilley and Reid (2011) found that words learned by exclusion were forgotten after 24 hours.

In any case, one should be careful when drawing conclusions regarding fast mapping abilities in non-human animals. Concerning dogs, It may be objected that direct teaching and learning by exclusion after intensive training are not examples of fast mapping, given that children perform this process in an apparently automatic and undirected manner (at least during the first phase). Consider also the so-called *circus problem*, which consists in stating the existence of shared biological mechanisms or functional equivalences between two species based only on finding similar behavioral responses to similar stimuli (Trout, 2001). After all, as Trout says, maybe with a little too much irony, "something exciting always happens when the circus comes to town" (Trout, 2001, p. 531).

VII. Displacement in Honeybee Dances

The dance of a female worker honeybee (genus *Apis*) can indicate to her sisters the finding of a patch of flowers that provides pollen and nectar, the finding of water, the finding of waxy materials (in cases where the hive needs repair), or even the finding of a new place for living (when a part, or the entire colony, needs to be relocated; Crist, 2004; Fitch, 2010). "On return to the nest some successful foragers perform a linear 'waggle' in which the dancer shakes her abdomen from side to side vigorously [*waggle run*], then turns to the left or right and circles back to repeat the waggle [*return phase*], in the process tracing a figure of eight pattern" (Preece & Beekman, 2014, p. 20).

Crist (2004) noted that honeybee dances are subject to different rules, including the following: (a) a conventional template must be followed in order to indicate distance, direction, and attractiveness of a source; (b) dances must indicate a relevant source for the most urgent needing in the hive. If nothing special is required, the dance indicates patches of flowers; (c) when sources of equivalent quality are found, the dance indicates the nearest of them; (d) if the resources can be found through smell, it will not be a dance; (e) non-abundant resources are indicated only when there is an urgent need in the hive; (f) the dance takes place in a specific location near the hive entrance, called the "dance floor"; and (g) dances are never performed without an audience.

The honeybee dance may be understood as a symbolic system of referential communication, as it provides information about a state of things distant in time and space (see *displacement* above). The message conveys a *functional referential signal*, because it provides information about external events to a receiver, but is not necessarily the signaler's intention to do so (Fitch, 2010). The transmitted message is composed of discrete meaningful units (waggle runs). Unlike the typical arbitrariness found in human speech, such units seem to be *deictic* and *iconic* (Igoa, in press): deictic, because the direction of the waggle depends on the angle of the resource relative to the sun, when it is visible, and on gravity, when dancing in the dark; iconic, because

there is a natural relation between the waggle run duration and the distance toward the entity signaled (the more distant the source, the more longer the waggling). The vigor of the tail-wagging and the speed of the turns are also related to the quality of the source found. Consequently, finding a good resource generates a long and vigorous dance, attracting thus more fellow bees to the search (Preece & Beekman, 2014).

An alternative view on the matter is that the behavior of the dancing bee consists in a canonical transformation of its own path traveled in search of food, rather than an attempt to draw other bees' attention toward a perceptually absent element (Liszkowski et al., 2009). In this sense, honeybee dances would not fulfill the referential function that is often attributed to them. Conversely, chemical signals present in the smell of the flowers would permeate the bee, allowing other bees to arrive at the right place. Moreover, when the site signaled is near and abundant, spatial and scent memory could be used to find the desired place, without needing the information conveyed by the dance (Preece & Beekman, 2014).

VIII. Birds

A. Nature versus Nurture

The largest and most studied vocal imitators in the animal kingdom are the 4,000+ species of songbirds (*Passerines*), followed by parrots (*Psittacidae*), and some hummingbirds (*Trochilidae*). In songbirds, as in the case of bats, singing is mostly made by males, and functions as a reproductive exhibition to repel rivals and attract females. Yet, as is the case of some NHP as well, several species of songbirds females also sing alone or in complex duets with males (Fitch, 2005). With respect to song learning, different findings indicate that the newborn songbird brain is not a tabula rasa, including: (a) although birds bred in acoustic isolation produce far less complex songs than normal, these songs still contain species-specific structures; and (b) without having been exposed to songs, young birds exhibit greater changes in heart rate and more begging calls in response to conspecific songs with regard to songs of other species (Brainard & Doupe, 2002). A similar phenomenon has been observed in human newborns, who can discriminate between the voice of their own mother and other women's voices, and between their mother tongue and a different language (Gervain & Mehler, 2010; other newborn animals, such as ducklings and lambs, can also distinguish their mother's voice from another female's voice; Fitch, 2010). Nonetheless, unlike human learning in utero, songbirds' experiences before and immediately after hatching, concerning other birds singing, do not seem to have serious effects on subsequent song learning.

Diverse contextual aspects influence singing development in birds. For instance, songbirds early removed from their environment and exposed later to unrelated conspecific adult songs (natural or recorded), eventually produce normal songs matching those that they have heard. Furthermore, young birds only retain (or *crystallize*) a portion of the songs learned. Some birds will end up singing selected tunes of their surrounding tutors, others will end up singing tunes that are more efficient to attract surrounding females. Chosen songs are repeated and memorized–apparently, as is the case with humans, partially during sleep–and eventually become the only tunes that the birds will be singing from then on (Brainard & Doupe, 2002; Doupe & Kuhl, 1999). In short, in songbirds' vocal learning, experience seems to interact with innate predispositions. Nevertheless, there is an ongoing controversy between instructive models, which argue that bird songs are learned, and selective models, which claim that, through sensorial experience, birds select the sounds of their songs from among an extensive set of preexistent possibilities.

In any case, birds' ability to learn songs is restricted to a critical period of development, which varies among different species, usually between ten days and four months after birth. During this critical period, birds

must hear songs of an adult bird to form a template, which serves later for practicing without the presence of the tutor. Given that listening to the own vocalizations is also necessary for an appropriate development of singing, if a bird becomes deaf after acquiring the template, but before practicing, it will end up producing strange-shaped tunes (Doupe & Kuhl, 1999). The critical acquisition period is divided in two phases: *sensory learning* (listening) and *sensory-motor learning* (listening-repeating). During sensory-motor learning, a *subsong* is produced by the young bird. These subsongs are generic among individuals, yet comprising variations, in a similar way of that of infants' babbling and mice USVs. Subsongs eventually become *plastic songs*, which vary greatly between one implementation and another, while gradually incorporating some recognizable elements of the tutor song. Plastic song remains until the bird crystallizes it in a stable mature song, and then new songs cannot be learned. In some bird species, however, phases of acquisition may overlap, and in the case of the common canary (*Serinus canaria*), the sensory learning period continues through the entire adulthood.

B. Brain Asymmetry and Specialized Neurons

In general terms, humans, songbirds, parrots, and apes exhibit a cerebral hemispheric asymmetry in areas related to communication. In songbirds, lateralization of the cerebral areas controlling the singing behavior usually favors the left hemisphere. This has been attributed to a peripheral muscle inhibition of the syrinx, whereas the higher centers of the brain responsible for singing maintain a bilateral activation (Doupe & Kuhl, 1999). Such higher centers of control consist in a complex hierarchy of specialized areas in the forebrain, where the motor and auditory centers closely interact to regulate the lower vocal motor areas (also found in non-vocal learning birds). Nonetheless, in some birds, such as the Zebra finch, the cerebral asymmetry favors the right hemisphere (Colbert-White et al., 2014).

Within the songbird brain, specific groups of neurons are dedicated to control singing behaviors. These neurons exhibit a stronger activation to the sound of the bird own songs, and sometimes to the tutor's song, with respect to other complex auditory stimuli, such as songs of conspecifics, or the own song presented backwards or disarranged (Brainard & Doupe, 2002). In addition to responding to complex sensory acoustic signals, the song-specific neurons in the birds' brain are also activated during motor production. This indicates that such neurons support, at least in part, the association between sensory and motor representations related to the singing.

Nonetheless, although the predominant left hemisphere lateralization for communication in birds' brain seems to imply a similarity with the case of human speech, lateralization of language in the human brain is currently a controversial issue. At this respect, it has been proposed that syntactic and semantic speech processing takes place in the left hemisphere, whereas prosodic related processes (rhythm and intonation) are treated in the right hemisphere (Friederici & Alter, 2004). At all events, it appears to be false that human speech is entirety controlled by the left hemisphere of the brain.

C. Birdsongs and Human Language

In addition to those already mentioned, several similarities between bird singing and human language have been found, including the following (see Aitchison, 2000 for a concise review). Many birds produce two kinds of sounds: *calls*, which are mostly innate and serve functions such as alerting and congregating, and *songs*, which usually imply early learning. Humans have a few innate calls as well: the babies' cries. Among those, at least two are believed to be universal: the pain cry and the hunger cry–compare also with the bat

pups' innate isolation calls, which appear at the first day after birth and gradually develop into adult-like calls (Prat et al., 2015). In bird songs, notes alone are meaningless, as well as the sound segments that comprise words within speech: they both acquire their meaning as a part of a sequence (see *duality of patterning* in Hockett, 1969). In both bird songs and human speech, apparently in humpback whales as well, segments of sounds are embedded into general patterns of rhythm and intonation, which vary between individuals of the same species, generating dialects (Fitch, 2010). Further similarities between bird singing and human speech include: (a) development of complex vocalizations from early stages of life; (b) strong dependence on hearing and imitating adults of the species, and hearing oneself while practicing, for the correct development of vocalizations; (c) dependence on innate predispositions related to perception and learning; and (d) involvement of the FOXP2 gene in the vocalization processes (Berwick et al., 2013).

Note that FOXP2 is strongly related to human speech, and is one of the few genes that differ between chimpanzees and humans (Lieberman, 2013). According to Lieberman, a protein known as SRPX2, modulated by FOXP2, promotes vocalizations in mammals, controlling the formation of synapses in the basal ganglia. Introducing a form of FOXP2 in mice offspring increases their brain plasticity and connections between neurons within the basal ganglia. Reducing the expression of SRPX2 in the brain of the mice offspring generates a reduction of vocalizations. On the other hand, given that in birds FOXP2 is also involved in vocal learning, and considering that there is not vocal learning in primates, it has been proposed that the evolution of this particular type of learning capacity occurred independently in birds and in humans (Fitch et al., 2005). At this regard, Fitch (2010) points out: "although vocal imitation has convergently evolved in humans and birds, the same gene is playing a closely analogous role, in the same brain regions" (p. 57).

Differences between bird singing and human speech include: (a) for the most part, only male birds sing; females do not, unless they are injected with the male hormone testosterone (Aitchison, 2000); (b) bird vocalizations have a greater spatial scope compared to human speech; (c) the purpose of bird singing is very restricted: songs are usually intended to attract females or to repel intruders; (d) birds cannot transmit complex or referential meanings (Igoa, in press); and (e) whereas human language allows meanings, abstractions, and flexible associations, bird songs do not (Doupe & Kuhl, 1999).

D. Parrots

Parrots quickly learn to imitate speech, and with proper training, they can associate many words with their referents, as dolphins and dogs do (Fitch, 2005). Then again, parrots are a special case among non-human animals, due to their ability to use an acoustic communicative system that humans can understand. For doing this, parrots spontaneously combine syllables and build new words with them. Such words can refer to objects, properties of objects, and even to numbers, in a context-appropriate fashion (Pepperberg, 1998). Furthermore, as in human speech, Grey parrots' (*Psittacus erithacus*) vocalization patterns exhibit *contextual substitutability* (i.e., the possibility of two words to fit equally well in a sentence)–at least in the case of the one parrot studied by Kaufman, Colbert-White and Burgess (2013). According to these authors, in addition to learning set phrases, parrots can learn portions of vocabulary as individual words, and then integrate those words into known phrases (if the words are synonyms, Greys would be able to use them in an interchangeably way). Kaufman et al. suggests that these findings indicate the existence of higher-order cognitive skills, both in infants and in Grey parrots.

Likewise, similar to infants' babbling, Grey parrots practice their vocalizations using *sound play* (i.e., repetitions and variations of vocal patterns). This allows them to derive new speech sequences spontaneously from already existing ones (Pepperberg, 2010). During sound play, parrots perform acoustic transients that seem to follow some kind of progression rules: "We find progressions like 'grey,' 'grain,' 'chain,' and 'cane,'

but not 'achn'" (Pepperberg, 1998, p. 526). Furthermore, like humans and bonobos, parrots learn referential speech labels better with the help of live tutors than with videotapes (even if they are reinforced to see the videos, or if their attention is directed by human tutors; Kaufman et al., 2013). In this regard, as young children do, Grey parrots develop communication skills most effectively when teaching is *referential* (presenting the objects of which one is speaking), *functional* (providing the opportunity to use the words learned to ask for objects during training), and within a social context (Pepperberg, 2010; note that several conclusions attained by Pepperberg in the aforementioned studies are drawn from the behavior of one single parrot).

In addition, as claimed by Colbert-White et al. (2014), parrots share with humans, NHP, and songbirds, features of basic sociability, such as frequent interactions with conspecifics, individual recognition of group members, and extensive parental care of young. Yet, unlike songbirds, parrots also exhibit a discrete behavioral repertoire for affiliative nonsexual social interaction with conspecifics, social correlates of intelligence, and hierarchical relationships among group members. This being the case, it would be advisable to study the communication in parrots as a comparative standard regarding human speech, in an amount equivalent to the actual amount of study on bird singing and NHP' communication. At all events, Colbert-White and colleagues (2014) acknowledge that it is far to be clear if parrots use their vocalizations to convey information, rather than just for socialization purposes.

IX. Birdsong Syntax and the Core Component of Language

As in the case of NHP, several studies on birds have found organizational capabilities related to the structural units of communication. It has been claimed, for example, that Bengalese finch (*Lonchura striata domestica*) can use information resulting from the sequential processing of syllables to discriminate between conspecifics songs (Abe & Watanabe, 2011). As maintained by these authors, Bengalese finches are also able to extract rules from an artificial grammar composed of synthesized syllable strings, and use such rules to classify novel acoustic information. Abe and Watanabe conclude that songbirds spontaneously acquire, after birth, the ability to discriminate syntactic hierarchical structures by means of interactions with conspecifics. Such ability would thus be a feature of language shared between humans and birds. However, Abe and Watanabe's study has been accused of having a flawed experimental design, and their conclusions have been questioned (see Berwick et al., 2013).

Controversy aside, some contemporary language theories reduce syntactic capacities to the recursive capacity. It is assumed then that recursion is a condition unique to the human cognitive system, and that its absence among animal communication systems precludes the existence of any human-like system of communication in the animal kingdom (Berwick et al., 2013; Chomsky, 2011; Hauser, Chomsky & Fitch, 2002). Supporting this view, several authors find that at present no convincing evidence indicates that, besides humans, any animal possesses the ability to process recursive communicative structures (Bickerton, 1990; Fitch, 2005; Uriagereka, Reggia & Wilkinson, 2013). In addition, it is claimed that among non-human animals the combination of vocalizations, or signs, is mostly restricted to a few elements in a poorly organized and random repetitive fashion.

The existence of a computational mental device, unique to humans, is assumed: the *faculty of language in the narrow sense* (FLN), which would be part of a larger linguistic structure named *faculty of language in the broad sense* (FLB). For a mechanism to be part of FLN, it must operate exclusively in the linguistic domain and not have "clear" homologues or analogs in other species (Fitch et al., 2005, p. 191). The main function of FLN would be the processing of *center-embedded recursion* (CER) during language production and comprehension (we will return to that point in a moment). FLB would include all mechanisms involved in language use and acquisition that are not unique to humans and that have other functions apart from language

(e.g., theory of mind and memory). Moreover, it has been proposed that FLN appeared due to a "slight rewiring of the brain" in a particular individual in East Africa, about 75,000 years ago. From there, it would have spread over the world, without any major changes, constituting the universal core element shared by the approximate 6,000 existing spoken languages and the 130, or so, sign languages (Berwick et al., 2013; Chomsky, 2011; Hauser et al., 2002). Note that defining language as FLN, leaving aside all the other aspects *related* to language itself (FLB), and then claiming that FLN is inexistent among animal communications, implies a categorical refusal of any continuity between non-human animal communication and human language.

Many authors object this account of the composition and development of human language, and currently diverse theories propose different explanations of the subject. Pinker and Jackendoff (2005), for instance, doubt that recursion is the unique specifically human language component, because, they point out, there are other linguistic aspects that are not recursive and might also be exclusively human (e.g., particular properties of words, aspects of phonology and morphology, case, and agreement). Lipkind et al. (2013) suggests the existence of a generative process, common for different species, which governs vocal combinations development through a gradual process of transitions. During these stepwise transitions, syllables could be rearranged, and new syllables could be inserted into a string, slowly generating new strings. Such acquisition of vocal transitions would take place during early stages of development in different species, including birds and humans. Tomasello (2005), in turn, proposes that linguistic constructions, including words, syntactic simple and complex structures, and discursive expressions, are learned. For learning them, young children would rely on a varied repertoire of domain-general cognitive skills (e.g., analogy, generalization, pragmatic inferences, and statistically based distributional analyses) that would be shared at least with the rest of primates. In addition, intention-reading skills would facilitate language learning. Tomasello (2005) believes that these intention-reading capacities are a uniquely human biological adaptation, but not a language-specific adaptation, given that they would also support other cultural practices. From this perspective, a set of historical and ontogenetic processes allow the emergence of the syntactic characteristics of a particular language, as meaningful concatenations of symbols used to communicate between individuals. Later, those concatenations would consolidate into grammatical constructions, in a process referred to as *grammaticalization*.

In this regard, Goldberg (2006) broadly defines linguistic constructions as conventionalized pairings of form and functional meanings (including grammatical patterns, words, utterances, and expressions). From a functionalist point of view, Goldberg proposes an explanation of how those constructions can be learned from simple linguistic input and then be generalized until attaining more complex forms. Furthermore, Hurford (2012) argues that there is not a firm distinction between lexical items and grammatical constructions (the syntax-lexicon continuum hypothesis). According to this author, the grammatical component of language evolved in the human species in a similar way to that of the human ontogenetic development: putting words together in simple ways, and then moving on to increasingly more complex ways. Also from a functionalist point of view, Dor (2015) suggests that human language is a socially constructed, imagination-instructing communication technology: the product of a collective process of invention and development that resides between speakers at a level of organization and complexity that cannot be reduced to the individual mind. Additional hypotheses on the evolution of syntax and other aspects of language are reviewed in Fitch (2010), Hurford (2012), and Aitchison (2000).

Returning to the syntactic organization of non-human animal communication, in the words of Fitch et al. (2005), "the demonstration of recursion in birds would mean that it is not uniquely human, just as surely as the same finding in chimpanzees" (p. 197). Nonetheless, it is a specific kind of recursion that has been claimed to be the uniquely human language trait: the *center-embedded recursion* (CER). In plain words, recursion is a process of repetition (or iteration) of equal or similar elements, completely or in part, within a sequence. In this process, a procedure is re-applied to its own output, such that the output of one application becomes the input

for the next one (for the difference between *iteration* and *recursion*, see Hurford, 2012). Recursion can be divided in *tail recursion* and CER. Tail recursion is a repetitive process in which constituents of a sequence are attached to an end of the sequence, such as in the following example (taken from Corballis, 2007): This is the house that Jack built. → This is the malt that lay in the house that Jack built. → This is the rat that ate the malt that lay in the house that Jack built. According to Corballis, this kind of iterated sequences is a common feature of bird and NHP calls; therefore, tail recursion would not be the main unique factor underlying human language. CER, on the other hand, is an iterative process in which constituents of a sequence are attached within the sequence (e.g., The malt that the rat ate lay in the house that Jack built. → The malt that the rat that ate the cat killed lay in the house that Jack built). CER implies an additional processing burden with respect to tail recursion, because the procedure must resume from where it left off. This requires some kind of memory device, such as a marker indicating where to resume the procedure after embedding a new constituent.

CER is considered the key component of a *context-free grammar*, sometimes referred as *phrase structure grammar*, which is argued to be exclusive to humans. Such grammars are completely independent of the context of use, and depend only on "rules [that] are all restricted to be in the form X → w, where X is a single phrase name (such as VP or NP), and w is some string of phrase names or words" (Berwick et al., 2013, p. 89). Animal communication systems, by contrast, are normally explained in terms of a *finite-state grammar*; in which, unlike context-free grammars, all constituents are fully specified by the transitional probabilities between a certain number of states (such as words or calls; Gentner, Fenn, Margoliash, & Nusbaum, 2006). In addition to link items together, like a finite-state grammar, context-free grammars "can embed strings within other strings, thus creating complex hierarchical structures ('phrase structures'), and long-distance dependencies" (Fitch & Hauser, 2004, p. 378).

In spite of the above, the ability to process CER has been tested in European starlings (*Sturnus vulgaris*) by Gentner et al. (2006). According to these authors, European starlings can recognize syntactically formed acoustic strings, including strings that use a recursive center-embedding rule. Moreover, starlings would be able "to classify new patterns defined by the grammar and reliably exclude agrammatical patterns" (Gentner et al., 2006, p. 1204). This being the case, CER would not be the factor of language exclusive to humans. Nonetheless, several authors disagree with the results presented by Gentner et al. (2006), arguing that the characteristics of the study are not appropriate to test the processing capacity of a CER structure (Ojima & Okanoya, 2014; Uriagereka et al., 2013). The critics recommend an increasing of the comparative research in the particular subject of structural combinatorial capacities (see Ojima & Okanoya for an overview on this topic).

As for NHP, in a Fitch and Hauser's (2004) study, tamarin monkeys failed to learn an artificial context-free grammar that generated center-embedded utterances. Human adults, otherwise, presented no major problems when learning such grammar. Perruchet and Rey (2005), however, disagree with the results of this study, and report that human adults are indeed able to learn the language used by Fitch and Hauser, without needing to process CER. When Perruchet and Rey (2005) changed the procedure, making the CER processing obligatory, participants did not present evidence of learning (these results are attributed by Hurford [2012] to the lack of meaning of the stimuli). Additionally, a study on guinea baboons (*Papio papio*) has shed some more light on the subject. With an operant conditioning technique, baboons were trained to associate in order different pairs of visual shapes presented on a touch screen (Rey, Perruchet, & Fagot, 2012). The experiment involved several series of increasing difficulty tasks and the presentation of visual distractors. The capacity of the baboons to process a one-level-embedded pair of shapes was then evaluated. Rey and colleagues conclude that baboons exhibit a spontaneous preference "for producing responses consistent with a CE [CER] structure, without such a structure being specifically reinforced" (Rey et al., 2012, p. 182). This preference is explained as an incidental product of associative learning and working memory constraints. Note

that in this study the measurement was made on visual stimuli, suggesting the possibility of an intermodal mechanism responsible for the CER processing.

It must also be noted at this point that CER is essentially absent from naturally occurring child speech (Diessel & Tomasello, 2005), and that it does not seem to appear during first language acquisition until the seventh year of life (Karlsson, 2007). Moreover, although CER in written language is found, usually with a maximum of three embedded clauses, in adult speech production is essentially absent (Karlsson, 2007). CER is difficult to learn even upon request (Ojima & Okanoya, 2014), and it is not manageable in perception whenever the depth of embedding exceeds one or two levels (Perruchet & Rey, 2005). Note also that the existence of at least two languages with apparently no recursive sentential patterns has been claimed: the Amazonian language Pirahã and the Indonesian language Riau (Everett, 2012a, b; for an analysis of Riau see Hurford, 2012). According to Everett, Pirahã has no numbers of any kind, neither a concept of counting, terms for quantification, colors, or embedding structures. Everett considers that every language is a learned tool shaped by cultural and communicative demands: an "instrument created by hominids to satisfy their social need for meaning and community" (Everett, 2012a, p. xi). Consequently, he claims that recursion is not a necessary condition for human syntax, but a general cognitive ability that may or may not be present in a given language.

In this regard, Evans and Levinson (2009) comment: "In this context where recursion has been suggested to be the criterial feature of the human language capacity, it is important for cognitive scientists to know that many languages show distinct limits on recursion ... or even lack it altogether" (p. 442). The authors claim that significant recurrent patterns in linguistic organization are not universals, but stable engineering solutions for functional pressures within specific cultural-historical contexts. In such contexts, population's shared cognitive skills (e.g., memory, action control, sensory integration) would have served to solve environmental and communicative challenges through language uses, which in turn would have crystallized over time. In addition, due to diverse historical and cultural differences, a spectrum of cross-linguistically variables would have emerged, causing that different languages developed different types of grammatical constructions. Evans and Levinson (2009) conclude that the unique and most remarkable human language trait is the diversity present in all linguistic levels, both in form and in content. Furthermore, Tomasello (2005) proposes that there are very few, if any, specific grammatical categories present in all languages. From this point of view, many languages have grammatical categories apparently unique to them, which do not correspond to any of the European categories as these have been defined through history. Tomasello (2005) suggests that language universals are universals of communication, cognition, and human physiology, which have arisen due to humans' similar needs for solve similar kinds of communicative tasks, during their similar social lives. An alternative view on Pirahã's grammatical characteristics is that of Hurford (2012). He proposes that such language is "not as categorically different from (most) other languages, but rather at one end of a continuum of languages ranked according to the depth of embedding that they allow" (p. 394). In the opinion of Hurford, given that Pirahã productively combines words more than once, it does indeed have recursion.

In response to the possible existence of languages without recursion, Fitch et al. (2005, p. 203) claim that such fact would "in no way" alter the explanatory landscape of recursion as the core factor of human language. As reported by Tomasello (2009), Chomsky's comments on non-recursive languages have been that they "do have recursive structures, it is just that one cannot see them on the surface. But even if they do not have such structures, that is fine because the components of universal grammar do not all apply universally" (p. 471). This strategy, concludes Tomasello (2009), "is the most effective because it basically immunizes the Universal Grammar (UG) hypothesis from falsification" (p. 471). So far, the controversy in the field of linguistics over non-recursive languages (and on Everett's work itself) has not ended, and the final word has not been spoken yet (see Nevins, Pesetsky, & Rodrigues, 2009).

X. Conclusions

Despite the difficulties inherent in the comparative study of language, and the controversies around it, research carried out mostly over the last 25 years has gradually helped to clarify the fundamental features of human language, deepening at the same time the knowledge about the communicative abilities and the cognitive processes of non-human animals. At this stage of the research, some things are becoming clear. On the one hand, there seems to be an asymmetry between what non-human animals know and what they can express. NHP, for example, surely know a lot about predators, food, trees, and many other things from their environment, but they seem unable to intentionally communicate this knowledge to others (at least through vocalizations). Overall animal communicative emissions appear to be highly restricted to physical needs, moods, and survival related matters, such as sources of food, presence of predators, and coupling. However, communication for group cohesion and coordination purposes, such as in the case of bats and dolphins, cannot be ruled out completely.

Additionally, non-human animals apparently do not intend to convey information to others when communicating. With the possible exception of great apes (a case currently under discussion), animals appear to perform their communicative emissions automatically in response to specific environmental stimuli (that could include the presence of conspecifics), transmitting only messages related to the survival and welfare of the group. When vocalizations are involved, the sender of the message does not appear to structure it in a way relevant to the knowledge (or lack of knowledge) of the recipient, as in the case of the grivet and Vervet monkeys mentioned earlier. This kind of behavior has been largely attributed to the inability of animals to infer other individuals' internal states based on explicit behaviors (another controversial issue of investigation on NHP, and an interesting field of study in cetaceans).

On the other hand, it has been proposed that animal communication is restricted to convey information about things that can be perceived through the senses. As a result, animals might not transmit information regarding nonexistent or abstract entities (Bickerton, 1990; Ghazanfar & Hauser, 1999; Liszkowski et al., 2009). Nonetheless, it is not the same case with respect to *non-present concrete entities*: things that have been perceived through the senses in the past, but are not present when transmitting the message. Consider honeybees dancing about scenarios distant in time and space, and chimpanzees and bonobos referring in some amount to displaced entities. The things about which they communicate have been perceived prior to the act of communication; hence, some kind of representation of the thing perceived is responsible for the transmitted information, thus the senses are not crucial for the conveyance in that precise moment. Although this of course does not imply that bees and chimps can communicate about abstract or nonexistent entities, the particular capacity to refer to such entities can arguably have evolved in humans from the ability of animals to communicate about non-present concrete entities.

With respect to the repertoire of elements used for communicative purposes (sounds or signs), it is estimated that the "vocabulary" of animals does not get anywhere close to that of humans. That is, according to a conservative estimate, a child usually produces around 500 different words at about three years of age (but comprehends at least twice that amount) (Gershkoff & Hahn, 2007). At the age of six, children produce about 6,000 words (Carey, 2010). In turn, an average adult English native speaker, graduated from university, has a repertoire of about 20,000 word families. In comparison, great apes, parrots, and songbirds, such as the black-capped chickadee (*Poecile atricapillus*), have repertoires of less than 100 different vocal types (songs or calls). As an exception, the repertoire of the common nightingale (*Luscinia megarhynchos*) includes more than 200 different elements (Colbert-White et al., 2014). According to these numbers, non-human animals do not appear to have such a thing as a lexical explosion period, which allows infants to acquire a large amount of vocal or sign communicative elements in an expeditious manner and without needing explicit training. Yet, general

learning mechanisms could support the association of a considerable amount of verbal labels with their referents, as it seems to happen in dogs and parrots, maybe also in dolphins. Consequently, fast mapping would not necessarily be the unique factor behind the lexical explosion in infants, but, nevertheless, a capacity unique to humans.

Concerning the perception of human speech, despite the existence of significant similarities between men and animals, there is also a strong asymmetry in non-human animal cognition between a wide-ranging and complex perceptual capacity and a poor and restricted productive ability. This being the case, it is possible that the phylogenetic development of human language relies on ancient perceptual capacities already present in birds and some mammals. Those capacities would have converged at a certain point of the evolution with skills for structural organization and conceptual abstraction, already existent in humans, and not necessarily present in the species of which perceptual abilities were inherited (which is basically Kuhl's *auditory hypothesis;* reviewed in Heimbauer, 2012).

As for the core component of language, we have seen in this review that, though polemical, there is more empirical evidence of recursion in animal communication than evidence of second-order or shared intentionality among non-human animals. In the case of empirical evidence of animals communicating about nonexistent or abstract entities, as far as we know, it simply does not exist. Accordingly, both abstract thinking and intentionality–understood as a capacity to share internal states and modify others' internal states, on purpose–are likely to be exclusively human skills, related "to some interesting extent" to the uniqueness of human language.

With respect to the inability of non-human animals to communicate about nonexistent or abstract entities, two possibilities come to mind: (1) animals are incapable of creating abstract representations or representations about nonexistent things; or (2) they are unable to "connect" such kind of representations with sounds or signs, so they can share them with others. Given the evidence of honeybees and chimpanzees referring to non-present concrete entities, and the generalized consensus on animals possessing some kind of basic non-conscious representations about the things in the world, the second choice does not seem irrational. In fact, as mentioned earlier, Vygotsky (1934/1986) proposed that humans are the only animal in which thought and language are connected between each other (understanding language as a set of sounds). This would allow humans to communicate their thoughts through words and to use words to think. Should this be the case, the main factor impeding the existence of a human-like system of communication among animals could be the lack of a connection between their signs/vocalizations and their conceptual representations (i.e., the impossibility of establishing a relation between conceptual representations and phonetic or motor representations supporting the execution of words and signs). It should be noted that, although judged improbable, Chomsky and colleagues admit that FLN could prove to be an empty category, implying only a connection between the *conceptual-intentional* and the *sensory-motor systems* (which manage concepts, phonological, and motor representations; Fitch et al., 2005). For us, this scenario helps to understand the asymmetry between possessed knowledge and transmitted information found in animals, the inexistence of conversations between them (or between animals and humans), the lack of a lexical explosion period (when fast mapping is possible), and the lack of communication about abstract or non-existent entities, in the non-human animal kingdom.

In summary, the existence of a certain degree of continuity between different aspects of human language and animal communication systems, in different domains, including the syntactic one, is likely to be the case. Additionally, CER does not seem to be the core factor of the faculty of language, ubiquitous in all human spoken and sign languages, and inexistent among non-human animals. Naturally, animals speaking or signing just as humans are not to be found, but one must not lose sight of the fact that the human's linguistic ability is constituted by a multiplicity of factors, which are shared to one degree or another with other species.

References

Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neurosciences, 14*(8), 1067-1074.

Aitchison, J. (2000). *The seeds of speech: Language origin and evolution*. Cambridge, UK: Cambridge University Press.

Arnold, K., & Zuberbühler, K. (2006). Semantic combinations in primate calls. *Nature*, *441*, 303-303.

Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology*, *18*(5), R202-R203.

Arriaga, G., & Jarvis, E. (2013). Mouse vocal communication system: Are ultrasounds learned or innate? *Brain and Language*, *124*, 96-116.

Arriaga, G., Zhou, E., & Jarvis, E. (2012). Of mice, birds, and men: the mouse ultrasonic song system has some features similar to humans and song-learning birds. *PLoS ONE*, *7*(10), 1-15.

Berwick, R., Friederici, A., Chomsky, N., & Bolhuis, J. (2013). Evolution, brain, and the nature of language. *Trends in Cognitive Sciences*, *17*(2), 89-98.

Bickerton, D. (1990). *Language and species*. Chicago, IL: University of Chicago Press.

Bohn, K., Smarsh, G., & Smotherman, M. (2013). Social context evokes rapid changes in bat song syntax. *Animal Behaviour*, *85*, 1485-1491.

Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition, 65*, 33-69.

Brainard, M., & Doupe, A. (2002). What songbirds teach us about learning. *Nature*, *417*(6886), 351-358.

Candiotti, A., Zuberbühler, K., & Lemasson, A. (2012). Context-related call combinations in female Diana monkeys. *Animal Cognition*, *15*(3), 327-339.

Cantalupo, C., & Hopkins, W. (2001). Asymmetric Broca's area in great apes. *Nature*, *414*(6863), 505.

Carey, S. (2010). Beyond fast mapping. *Language Learning and Development*, *6*(3), 184-205.

Cäsar, C., Zuberbühler, K., Young, R., & Byrne, R. (2013). Titi monkey call sequences vary with predator location and type. *Biology Letters*, *9*, 1-5.

Chomsky, N. (2011). Language and other cognitive systems. What is special about language? *Language Learning and Development*, *7*, 263-278.

Clarke, E., Reichard, U., & Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS ONE*, *1*(e73), 1-10.

Colbert-White, E., Corballis, M., & Fragaszy, D. (2014). Where apes and songbirds are left behind: A comparative assessment of the requisites for speech. *Comparative Cognition and Behavior Reviews*, *9*, 1-28.

Corballis, M. (2007). Recursion, language, and starlings. *Cognitive Science*, *31*, 697-704.

Crist, E. (2004). Can an insect speak? The case of the honeybee dance language. *Social Studies of Science, 34*(1), 7-43.

Cummins, F. (2009). Rhythm as an affordance for the entrainment of movement. *Phonetica*, *66*(1-2), 15-28.

Dent, M., Brittan-Powell, E., Dooling, R., & Pierce, A. (1997). Perception of synthetic /ba/–/wa/ speech continuum by budgerigars (*Melopsittacus undulatus*). *Journal of Acoustical Society of America, 102*(3), 1891-1897.

Diessel, H., & Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language*, *81*(4), 882-906.

Dor, D. (2015). *The instruction of imagination: Language as a social communication technology*. New York, NY: Oxford University Press.

Doupe, A., & Kuhl, P. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience, 22*, 567-631.

Elgier, A., & Bentosela, M. (2009). El gesto de señalar: Una llave para la comunicación entre especies [The pointing gesture: A key for communication between species]. *Interdisciplinaria*, *26*(2), 157-182.

Endress, A., Cahill, D., Block, S., Watumull, J., & Hauser, M. (2009). Evidence of an evolutionary precursor to human language affixation in a non-human primate. *Biology Letters, 5,* 749-751.

Evans, N., & Levinson, S. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429-448.

Everett, D. (2012a). *Language: The cultural tool*. New York, NY: Pantheon Books.

Everett, D. (2012b). What does Pirahã grammar have to teach us about human language and the mind? *WIREs Cognitive Sciences*, *3*, 555-563.

Fitch, W. (2005). The evolution of language: A comparative review. *Biology and Philosophy*, *20*, 193-230.

Fitch, W. (2010). *The evolution of language*. New York, NY: Cambridge University Press.

Fitch, W., & Hauser, M. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*, 377-380.

Fitch, W., Hauser, M., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition*, *97*(2), 179-210.

Friederici, A., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language*, *89*(2), 267-276.

Gardner, R., & Gardner, B. (1969). Teaching sign language to a chimpanzee. *Science*, *165*(3894), 664-672.

Garrod, S., & Pickering, M. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8-11.

Geissmann, T., & Nijman, V. (2006). Calling in wild silvery gibbons (*Hylobates moloch)* in Java (Indonesia): Behavior, phylogeny, and conservation. *American Journal of Primatology*, *68*(1), 1-19.

Gentner, T., Fenn, K., Margoliash, D., & Nusbaum, H. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, *440*(7088), 1204-1207.

Gershkoff-Stowe, L., & Hahn, E. (2007). Fast mapping skills in the developing lexicon. *Journal of Speech, Language, and Hearing Research*, *50*, 682-697.

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology, 61,* 191-218.

Ghazanfar, A., & Hauser, M. (1999). The neuroethology of primate vocal communication substrates for the evolution of speech. *Trends in Cognitive Sciences, 3*(10), 377-384.

Goldberg, A. (2006). *Constructions at work: The nature of generalization in language*. New York, NY: Oxford University Press.

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.

Hauser, M., Newport, E., & Aslin, R. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition, 78,* B53-B54.

Heimbauer, A. (2012). Investigating speech perception in evolutionary perspective: Comparisons of chimpanzee (*Pan troglodytes*) and human capabilities (Unpublished doctoral dissertation). Georgia State University, Georgia.

Hockett, C. (1969). The problem of universals in language. In J. Greenberg (Ed.), *Universals of Language* (pp. 1-29). Cambridge, MA: MIT Press.

Hurford, J. (2012). *The origins of grammar: Language in the light of evolution II*. New York, NY: Oxford University Press.

Igoa, J. (in press). Lenguaje humano y comunicación animal [Human language and animal communication]. In R. Pellón (Ed.), *Cognición Comparada* [Comparative Cognition]. Madrid: UNED.

Jones, B. (2014). *The ontogeny of whistle production in infant Atlantic Bottlenose Dolphins (Tursiops truncatus) during the first thirty days of life* (Unpublished master's thesis). The University of Southern Mississippi, Hattiesburg, Mississippi.

Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog: Evidence for "fast mapping". *Science*, *304*(5677), 1682-1683.

Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. *Journal of Linguistics*, *43*(2), 365-392.

Kaufman, A., Colbert-White, E., & Burgess, C. (2013). Higher-order semantic structures in an African Grey parrot's vocalizations: Evidence from the hyperspace analog to language (HAL) model. *Animal Cognition*, *16*(5), 789-801.

Keil, F. (1983). Semantic inferences and the acquisition of word meaning. In T. Seiler & W. Wannenmacher (Eds.)*, Concept Development and the Development of Word Meaning* (pp. 103-124). Berlin: Spriner-Verlag.

King, S., Harley, H., & Janik, V. (2014). The role of signature whistle matching in bottlenose dolphins, *Tursiops truncatus*. *Animal Behaviour*, *96*, 79-86.

Kuczaj, S. (2013). Are conversations between dolphins and humans possible? *International Journal of Comparative Psychology*, *26*, 114-123.

Kuhl, P. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, *50*(2), 93-107.

Kuhl, P., & Miller, J. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190*(4209), 69-72.

Kuhl, P., & Padden, D. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception and Psychophysics*, *32*(6), 542-550.

Leavens, D., Hopkins, W., & Bard, K. (1996). Indexical and referential pointing in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *110*(4), 346-353.

Lieberman, P. (2013). Synapses, language, and being human. *Science*, *342*, 944-945.

Lipkind, D., Marcus, G., Bemis, D., Sasahara, K., Jacoby, N., Takahasi, M., Suzuki, K., Feher, O., Ravbar, P., Okanoya, K., & Tchernichovski, O. (2013). Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature*, *498*(7452), 104-108.

Liszkowski, U., Schäfer, M., Carpenter, M., & Tomasello, M. (2009). Prelinguistic infants, but not chimpanzees, communicate about absent entities. *Psychological Science*, *20*(5), 654-660.

Lyn, H., Russell, J., Leavens, D., Bard, K., Boysen, S., Schaeffer, J., & Hopkins, W. (2014). Apes communicate about absent and displaced objects: Methodology matters. *Animal Cognition*, *17*(1), 85-94.

Manser, M. (2013). Semantic communication in vervet monkeys and other animals. *Animal Behaviour*, *86*(3), 491-496.

Marcus, G., Vijayan, S., Bandi Rao, S., & Vishton, P. (1999). Rule learning by seven-month-old infants. *Science*, *283*, 77-80.

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, *385*(6619), 813-815.

Mehl, M., Vazire, S., Ramírez-Esparza, N., Slatcher, R., & Pennebaker, J. (2007). Are women really more talkative than men? *Science*, *317*(5834), 82-82.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*, 143-178.

Mitani, J., Hasegawa, T., Gros-Louis, J., Marler, P., & Byrne, R. (1992). Dialects in wild chimpanzees? *American Journal of Primatology*, *27*(4), 233-243.

Morell, V. (2014). When the bat sings. *Science*, *344*(6190), 1334-1337.

Murphy, R., Mondragón, E., & Murphy, V. (2008). Rule learning by rats. *Science*, *319*, 1849-1851.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 756-766.

Nevins, A., Pesetsky, D., & Rodrigues, C. (2009). Evidence and argumentation: A reply to Everett (2009). *Language*, *85*(3), 671-681.

Newport, E., & Aslin, R. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*, 127-162.

Newport, E., Hauser, M., Spaepen, G., & Aslin, R. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, *49*, 85-117.

Ohms, V., Escudero, P., Lammers, K., & ten Cate, C. (2012). Zebra finches and Dutch adults exhibit the same cue weighting bias in vowel perception. *Animal Cognition*, *15*(2), 155-161.

Ohms, V., Gill, A., Heijningen, C., Beckers, G., & ten Cate, C. (2010). Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proceedings of the Royal Society B*, *277*, 1003-1009.

Ojima, S., & Okanoya, K. (2014). The non-hierarchical nature of the Chomsky hierarchy-driven artificial-grammar learning. *Biolinguistics*, *8*, 163-180.

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, *106*(51), 22026-22031.

Peña, M., Bonatti, L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*, 604-607.

Pepperberg, I. (1998). Out of the mouths of babes . . . And beaks of birds? A broader interpretation of the frame/content theory for the evolution of speech production. *Behavioral and Brain Sciences, 21*(4), 526-527.

Pepperberg, I. (2010). Vocal learning in Grey parrots: A brief review of perception, production, and cross-species comparisons. *Brain and Language*, *115*, 81-91.

Perruchet, P., & Rey, A. (2005). Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin & Review*, *12*(2), 307-313.

Petitto, L. (2005). How the brain begets language. In J. McGilvray (Ed.), *The Cambridge companion to Chomsky* (pp. 84-101). Cambridge, UK: Cambridge University Press.

Petitto, L., & Marentette, P. (1991). Babbling in the manual mode: Evidence for the ontogeny of language. *Science*, *251*(5000), 1493-1496.

Pilley, J., & Reid, A. (2011). Border collie comprehends object names as verbal referents. *Behavioral Processes*, *86*(2), 184-195.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, *95*, 201-236.

Pons, F., & Toro, J. (2010). Structural generalizations over consonants and vowels in 11-month old infants. *Cognition*, *116*, 361-367.

Povinelli, D., Bierschwale, D., & Cech, C. (1999). Comprehension of seeing as a referential act in young children, but not juvenile chimpanzees. *British Journal of Developmental Psychology*, *17*, 37-60.

Prat, Y., Taub, M., & Yovel, Y. (2015). Vocal learning in a social mammal: Demonstrated by isolation and playback experiments in bats. *Science Advances*, *1*, e1500019.

Preece, K., & Beekman, M. (2014). Honeybee waggle dance error: Adaption or constraint? Unravelling the complex dance language of honeybees. *Animal Behaviour*, *94*, 19-26.

Ramus, F., Hauser, M., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288*, 349-351.

Reed, P., Howell, P., Sackin, S., Pizzimenti, L., & Rosen, S. (2003). Speech perception in rats: use of duration and rise time cues in labeling of affricate/fricative sounds. *Journal of the Experimental Analysis of Behavior, 80,* 205-215.

Reimchen, M. (2013). *Infant vocal responses to questions and declaratives in maternal speech* (Unpublished master dissertation). University of Manitoba, Winnipeg.

Rey, A., Perruchet, P., & Fagot, J. (2012). Centre-embedded structures are a by-product of associative learning and working memory constraints: Evidence from baboons (*Papio Papio*). *Cognition*, *123*, 180-184.

Rizzolatti, G. (1998). What happened to Homo habilis? (Language and mirror neurons). *Behavioral and Brain Sciences*, *21*(4), 527-528.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., & Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition*, *107*, 479-500.

Savage-Rumbaugh, S., McDonald, K., Sevcik, R., Hopkins, W., & Rubert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *Journal of Experimental Psychology: General*, *115*(3), 211-235.

Savage-Rumbaugh, S., Rumbaugh, D., & Fields, W. (2009). Empirical Kanzi: The ape language controversy revisited. *Skeptic*, *15*(1), 25-33.

Savage-Rumbaugh, S., Shanker, S., & Taylor, T. (1996). Apes with language. *Critical Quarterly*, *38*(3), 45-57.

Sinnott, J., & Gilmore, C. (2004). Perception of place-of-articulation information in natural speech by monkeys versus humans. *Perception & Psychophysics*, *66*(8), 1341-1350.

Suddendorf, T., Borrello, M., Allen, C., & Radick, G. (2012). If I could talk to the animals. Book symposium on Gregory Radick: The simian tongue: The long debate about animal language. *Metascience, 21,* 253-267.

Takahashi, D., Narayanan, D., & Ghazanfar, A. (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. *Current Biology*, *23*, 2162-2168.

Tincoff, R., Hauser, M., Tsao, F., Spaepen, G., Ramus, F., & Mehler, J. (2005). The role of speech rhythm in language discrimination: Further tests with a non-human primate. *Developmental Science, 8*(1), 26-35.

Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2009). Universal grammar is dead. *Behavioral and Brain Sciences*, *32*(5), 470-471.

Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees understand psychological states – the question is which ones and to what extent. *Trends in Cognitive Sciences*, *7*(4), 153-156.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 1-17.

Toro, J., & Trobalon, J. (2005). Statistical computations over a speech stream in a rodent. *Perception and Psychophysics*, *67*, 867-875.

Toro, J., Trobalon, J., & Sebastián-Gallés, N. (2003). The use of prosodic cues in language discrimination tasks by rats. *Animal Cognition, 6*, 131-136.

Trout, J. (2001). The biological basis of speech: What to infer from talking to the animals. *Psychological Review, 108*(3), 523-549.

Trout, J. (2003). Biological specializations for speech: What can the animals tell us? *Current Directions in Psychological Science, 12*(5), 155-159.

Uriagereka, J., Reggia, J., & Wilkinson, G. (2013). A framework for the comparative study of language. *Evolutionary Psychology*, *11*(3), 470-492.

Vygotsky, L. (1934/1986). *Thought and language* (A. Kozulin, Trans.). Cambridge, MA: MIT Press.