



Call Combinations in Great Apes and the Evolution of Syntax

Maël Leroux^{1,2*} and Simon W. Townsend^{1,2,3}

¹Department of Comparative Language Science, University of Zurich, Switzerland

²Center for the Interdisciplinary Study of Language Evolution (ISLE), University of Zurich, Switzerland

³Department of Psychology, University of Warwick, UK

*Corresponding author (Email: mael.leroux@uzh.ch)

Citation – Leroux, M., & Townsend, S. W. (2020). Call combinations in great apes and the evolution of syntax. *Animal Behavior and Cognition*, 7(2), 131-139. doi: <https://doi.org/10.26451/abc.07.02.07.2020>

Abstract – Following the observation that vervet monkeys are capable of labelling different predator types with their vocalizations, comparative research in language evolution gained increasing interest. Over the last four decades, an impressive body of data has since accumulated demonstrating that many features of language can be found in the communication systems of nonhuman primates. One stumbling block to the phylogenetic reconstruction of language, however, has been language’s syntactic layer. We specifically highlight that, whilst current studies provide promising evidence for syntactic-like structures in the communication systems of monkeys, reconstructing the evolutionary origins of syntax hinges on comparable data from our closest-living relatives, the great apes. We critically assess existing data on potential candidates for combinatorial structures in the great ape clade and conclude that further experimental investigation is crucial to validating preliminary observational findings.

Keywords – Combinatoricity, Syntax, Language evolution, Nonhuman primates, Communication

Four decades ago, Seyfarth et al. (1980a, b) published their seminal research demonstrating vervet monkeys (*Chlorocebus aethiops*) can use vocalizations to refer to external threats in the environment, akin to the referential nature of human language. Not only did this work provide striking insights into the complexities underlying primate vocal communication and “how monkeys see the world” (Cheney & Seyfarth, 1990), but it also played a central role in catalyzing a newly emerging research discipline – namely the comparative study of language evolution.

Since this groundbreaking and influential work, a significant body of data has materialized probing the extent to which core components of language can be found in the communicative and cognitive systems of nonhuman primates. For example, capuchin monkeys (*Cebus apella*; Di Bitetti, 2003), titi monkeys (*Callicebus nigrifrons*; Cäsar & Zuberbühler, 2012), Campbell’s monkeys (*Cercopithecus campbelli*; Ouattara, Zuberbühler et al., 2009), bonobos (*Pan paniscus*; Clay & Zuberbühler, 2009) and chimpanzees (*Pan troglodytes*; Slocombe & Zuberbühler, 2005) amongst others, have been shown to produce referential signals (for a review: Townsend & Manser, 2012), whilst intentional communication has been highlighted in the vocal systems of rhesus macaques (*Macaca mulatta*; Santos et al., 2006) and chimpanzees (Crockford et al., 2015; Schel, Townsend et al., 2013, for a review: Zuberbühler, 2018b). Turn-taking has been reported in Japanese macaques (*Macaca fuscata*; Katsu et al., 2019), spider monkeys (*Ateles geoffroyi*; Briseño-Jaramillo et al., 2018) and gorillas (*Gorilla gorilla*; Lemasson et al., 2018, for a review, Pika et al., 2018) and finally, vocal flexibility has been

reported in rhesus and Japanese macaques (Masataka & Fujita, 1989), orangutans (*Pongo pygmaeus*; Lameira & Shumaker, 2019) and chimpanzees (Watson et al., 2015a, b; but see also Fischer et al., 2015, for a review: Lameira, 2017).

Together, this work has helped begin to reconstruct the evolutionary origins of language: rather than being a recent trait that evolved de novo in humans (Lieberman, 1991), language seems to be evolutionarily stratified, comprised of many components that have likely emerged gradually, building on pre-existing capabilities in our primate ancestors (Fröhlich et al., 2019; Levinson & Holler, 2014). However, one component whose evolutionary path is less clear is language's syntactic structure.

Syntax: Uniquely Human?

Syntax is defined as the ability to systematically combine meaning-bearing units together into larger meaningful structures or phrases (Hurford, 2011; Suzuki & Zuberbühler, 2019). Syntactic structures can take several forms; configurations in which the meaning of the whole can be related to the meaning of its parts (suggested to be termed “compositional syntax”, e.g., *red car* or *biologists and linguists*, Hurford, 2011; Townsend et al., 2018), but also configurations where the meaning of the whole is not (at least unambiguously) related to the meaning of its parts (suggested by Hurford, 2011, to be termed “combinatorial syntax”) such as compounds (e.g., *treehouse*) or idiomatic expressions (e.g., *kick the bucket*, Hurford, 2011; Townsend et al., 2018). Without syntax, humans would be hugely constrained in their communicative potential and therefore, unpacking its origins is central to a holistic understanding of language and its evolution. For many years, language's syntactic layer was considered to be absent from animal communication (Hurford, 2011; Marler, 1977). This lack of comparative data has been previously used to support the conclusion that syntax is perhaps the core feature distinguishing language from nonhuman communication systems (Bolhuis et al., 2014; Fitch, 2018). However, emerging data in monkeys has recently cast doubt on this assumption, with a growing number of species demonstrating the capacity to combine meaningful calls together into larger meaningful structures.

Campbell's monkeys, for instance, have two different alarm calls: the “*krak*”, given when detecting a high-urgency threat from the ground (e.g., leopard) and the “*wak*”, when detecting a danger from the sky (e.g., eagle, Ouattara et al., 2009a). Detailed behavioral observations and predator presentation experiments showed that, in less-specific aerial or terrestrial disturbance contexts, both alarm calls are affixed with an acoustically-invariable “-oo” call (Ouattara et al., 2009b). Moreover, playback experiments demonstrated a reduced antipredator response in subjects when exposed to the affixed alarm calls, confirming that the addition of the “-oo” affix serves to modify the semantic content of alarms, changing them from specific (“*krak*” and “*wak*”) to more general (“*krak-oo*” and “*wak-oo*”; Coye et al., 2015). Given the same affix is used in different sequences and has a predictable effect on call meaning, this example has been forwarded as evidence for a rudimentary form of compositional syntax (Collier et al., 2014).

Diana monkeys (*Cercopithecus diana*) have also been shown to concatenate calls into compositional-like structures. Females produce three distinct social calls (e.g., L, R and A), which they combine into larger sequences (i.e., L-A or R-A). Both L and R calls co-occur with external social events (positive or negative situations respectively) whilst A calls seem to be individually-specific calls (Candiotti et al., 2012a, b), potentially cuing information regarding the identity of the signaler. Playback experiments of L-A versus R-A combinations confirmed that receivers are capable of extracting contextual information from the combination (Coye et al., 2016). Furthermore, through artificially switching the A-unit from a group member with one from a neighboring group, the authors verified that receivers are also capable of extracting identity-related information from the combination (Coye et al., 2016). These results suggest the meaning of the combination is derived directly from its parts and therefore, superficially resembles compositional syntax in humans.

Whilst syntactic structures with compositional meaning are undoubtedly core to language, they are not the only forms of meaningful combinations present in human language. Indeed, prefabricated,

fixed combinations such as idiomatic expressions or formulaic structures represent up to half of the phrases used in language (Townsend et al., 2018; Van Lancker-Sidtis & Rallon, 2004).

Comparative work in putty-nosed monkeys (*Cercopithecus nictitans*) suggests similar abilities also exist in the communication systems of our primate relatives. Putty-nosed monkeys produce two acoustically distinct calls, the “pyow” and the “hack.” Series of “pyows” are given to a range of disturbances on the ground, while series of “hacks” are given more specifically to eagles (Arnold & Zuberbühler, 2006b). Observational data suggest “pyows” and “hacks” are also combined together into a larger series, primarily in travel contexts and playback experiments have confirmed the combination is meaningful to receivers and functions to initiate group travel (Arnold & Zuberbühler, 2006a). In light of the discontinuity between the meaning of the sequence (travel) and its relative parts (predators, disturbances) a compositional analysis is insufficient. Rather, this structure has been argued to better resemble more combinatorial or formulaic syntactic structures in language (Arnold & Zuberbühler, 2012).

Lastly and most recently, evidence for meaningful call combinations has been reported in titi monkeys. Like Campbell’s and putty-nosed monkeys, titi monkeys produce acoustically distinct alarm calls for different classes of predators: “A” calls for aerial threats and “B” calls for ground dangers (Cäsar & Zuberbühler, 2012). Under certain circumstances, they also combine both calls into larger sequences that encode both the type but also the location of the predator (Cäsar et al., 2013). Detailed statistical analyses of the sequence structure and playback experiments have further demonstrated that, for this call combination, information appears to be encoded probabilistically, rather than compositionally or combinatorially, with the meaning being related to the proportion of certain call pairs or “bigrams” (Berthet et al., 2019) and could therefore, represent a potentially unique combinatorial mechanism in animals.

Given the growing number of examples of syntactic-like structuring in both our Old- (Catarrhini) and New-world (Platyrrhini) monkey cousins, one interpretation is that the cognitive building blocks underlying the capacity to sequence vocalizations can be traced back potentially as far as the last common ancestor between monkeys and humans (approx. 45 million years ago; Pozzi et al., 2014, see Figure 1). However, an alternative scenario worth considering is that, given the phylogenetic distance between monkeys and humans, any similarities at the level of combinatoriality merely represents an instance of convergent evolution. Indeed, similar combinatorial phenomena have been reported in more distantly related species to humans including birds (Engesser et al., 2016) and meerkats (Collier et al., 2017) suggesting convergence is a plausible scenario.

In order to disentangle these two potential hypotheses and, in-turn, better understand the evolutionary progression of syntactic structures in the human lineage, comparative data demonstrating similar rudimentary forms of syntax in our closest living relatives, the great apes, is central. Yet, to our knowledge, there exists a paucity of evidence for such capacities in this clade. In this next section, we review the current state-of-the art here and discuss future research that is key in furthering understanding of the evolution of syntax.

Call Combinations in Great Apes

Initial evidence for call combinations in the vocal system of great apes came from observations in wild chimpanzees (Crockford & Boesch, 2005). Previous work has shown that chimpanzees possess a vocal repertoire comprising around 15 individual call types (Goodall, 1986, see also Slocombe & Zuberbühler, 2010) and behavioral observations conducted over a period of 15 months in the Taï Forest, Côte d’Ivoire, further reported that chimpanzees concatenated these 15 calls into at least 88 different types of combinations, representing 28% of their total vocal production (Crockford & Boesch, 2005). What functions this impressive variety of combinations potentially serve are, however, unclear. Crockford and Boesch (2005) propose a number of possibilities from modification and intensification to conjunction, communication to different audiences and signaling of identity (Crockford & Boesch, 2005). In order to differentiate between these alternatives, and to establish what form of syntactic structure these

promising combinations potentially represent (i.e., compositional vs. idiomatic), playback experiments are critical, though to date, such data are missing.

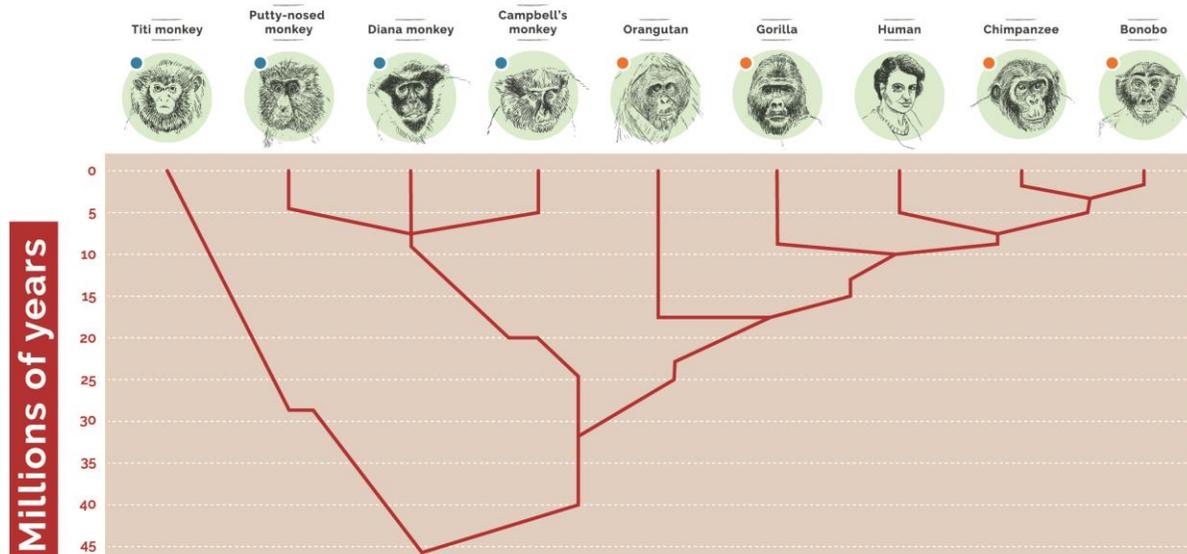


Figure 1. Phylogenetic tree illustrating those primate species for which syntactic-like combinatorial structures have been investigated. Blue circles indicate experimental validation for a syntactic-like structure, orange circles indicate incomplete or missing experimental validation for a syntactic-like structure. Estimated dates of divergence between the nine primate species are extracted from Pozzi et al. (2014).

One particular combinatorial structure that has received research attention in chimpanzees is the long-distance pant-hoot call. Pant-hoots are composed of four distinct “phases:” the introduction, the build-up, the climax and the let-down (Slocombe & Zuberbühler, 2010). Fedurek et al. (2016) have investigated the information content encoded across the distinct phases of this structure. Specifically, machine-learning-based analyses suggest that the introduction encodes mainly identity-related information while the age of the signaler is encoded in both the introduction and the build-up. The climax once more encodes identity, in addition to the signaler’s social status. Finally, the let-down phase varies reliably with the context of production – specifically, pant-hoots produced in a feeding context were less likely to contain a let-down phase compared to pant-hoots produced in a travelling context (Fedurek et al., 2016). Because existing data suggest that primates are limited to the concatenation of two call-types (Miyagawa & Clarke, 2019, though see Ouattara et al., 2009a), such a multi-component call would therefore represent a striking, complex combinatorial structure. However, two additional steps are required if we are to feasibly compare this structure with monkey call combinations and indeed human syntax. First, it is imperative to establish whether the phases of the pant-hoot are constrained to this sequence or if they also occur as independent call types (Suzuki & Zuberbühler, 2019), though existing data suggest that some phases of the pant-hoot (e.g., hoos and screams) might occur in isolation and thus be potentially meaningful (Crockford et al., 2018; Slocombe et al., 2009). Secondly, and building on this, playback experiments are needed to confirm whether information encoded at the sequence level is salient to receivers and how the individual comprising calls contribute to this.

Similarly to pant-hoots in chimpanzees, Spillmann et al. (2010) reported that orangutans also produce multi-component long distance calls. Long calls in orangutans are composed of three main calls: “grumbles,” followed by “pulses” and terminating with “bubbles” (Spillmann et al., 2010). The authors additionally distinguish between five different subtypes of pulses: roar, huius, intermediary, sigh, and bubble, suggesting a considerable degree of acoustic complexity underlies the long call structure. In line with pant-hoots, acoustic analyses of the comprising calls also indicated the long calls have the potential to cue information on both the identity of the caller and the accompanying context of production

(specifically whether spontaneous or elicited, Spillmann et al., 2010). When produced spontaneously, long calls were slower, contained pulses of longer duration and had a higher number of pulses and bubbles within a sequence compared to the long calls emitted in elicited contexts (e.g., male displaying or hearing another male displaying; Spillmann et al., 2010). In addition, after hearing a long call, females were observed to move further away if the long call was produced spontaneously compared to if it was an elicited variant (Spillmann et al., 2010). These behavioral data provide preliminary evidence that receivers are capable of distinguishing between the different long call types suggesting these structures are indeed meaningful. However, as for the pant-hoot in chimpanzees, it remains unclear to what extent the components of the long call, other than grumbles, occur in isolation (Hardus et al., 2009; Kershenbaum et al., 2014) and hence, might be meaning-bearing, and exactly how the individual call units contribute to the overall meaning of the sequence. Further observational and experimental work is therefore central to shedding light on these outstanding issues and elucidating whether the vocal system of orangutans is also characterized by syntactic-like call combinations.

A potentially more revealing example of call combinations in great-apes comes from the contact calls of mountain and lowland gorillas (Hedwig et al., 2014, 2015). Gorillas produce five different close-distance contact calls (i.e., A1, T1, T2, T3, T4) both alone and in flexible, non-random combinations (e.g., A1-T4, T2-T4). Individual grunts (i.e., A1 and T2, atonal, and tonal grunts respectively) are produced during resting when in the presence of conspecifics, while grumbles (i.e., T4) are produced when foraging alone. Given that both grunts and the combinations occur in resting contexts, there seems to be some contextual overlap between the singly produced calls and their combinations, and hence, the meaning of the combination might be compositionally related to the comprising calls. However, there also exists contextual discontinuity as grumbles in isolation occur in qualitatively different contexts to the combinations (i.e., foraging and resting respectively). In terms of function, observational data indicates both call combinations are produced when vocally responding to group members, suggesting that such call combinations might play a role in vocal exchanges in gorillas (Hedwig et al., 2015). Although this system seems promising, careful experimental validation is still critical to confirm both the function and whether, like in monkey species, gorillas process these combinations as syntactic-like structures.

Finally, of all the apes, bonobos have received perhaps the most attention with regards to their combinatorial capacities, with work in the wild and captivity indicating that they possess a diverse repertoire of combinations. Initial work by Clay and Zuberbühler (2009) showed that, during feeding contexts, bonobos combine acoustically distinct food-associated calls (e.g., peep, yelp, peep-yelp, grunt, and bark) together into larger sequences that encode food quality: sequences made of barks and peeps are given to highly preferred food while sequences composed mainly of peep-yelps and yelps are given to low preference food. Furthermore, systematic playback experiments confirmed that bonobos are capable of extracting information regarding the food quality from the sequence (Clay & Zuberbühler, 2011), though exactly how the meaning of the sequence is derived from the meaning of its components remains unclear.

More recently, work from a wild population of bonobos suggests they are capable of combining other calls from their repertoire, namely whistles and high-hoots. Detailed behavioral observations focusing on receiver responses suggest that bonobos are more likely to switch parties following the production of a call combination compared to when they produce high-hoots alone (Schamberg et al., 2016). Call combinations might therefore serve a coordination function in bonobo social lives, though here again, experimental validation is needed to confirm this hypothesis and investigate how meaning is ultimately attributed to this combination.

Future Directions

Since the seminal work of Seyfarth et al. (1980a, b), comparative approaches to the evolution of language has flourished as a research field, with data collected over the last 40 years demonstrating striking evidence for language-like attributes in the communicative and cognitive systems of nonhuman primates (Fedurek & Slocombe, 2011; Fröhlich et al., 2019; Seyfarth & Cheney, 2010; Zuberbühler &

Lemasson, 2014). Most recently, much attention has turned to the syntactic capabilities of primates (Zuberbühler, 2018a), given that this attribute has been previously argued to represent the defining feature of language (Bolhuis et al., 2014). Although observational and experimental work in monkeys is suggestive that the capacity to sequence meaning-bearing units together into larger structures is evolutionarily ancient, comparable data in great apes, central to testing this claim through resolving the more recent evolutionary history of syntax, is mostly missing (see Figure 1). Specifically, experimental verification is vital to ascertain: i) if identified combinations have a communicative function and are not just “read-outs” of current behavioral context and ii) the mechanisms by which receivers unpack meaning i.e., are combinations compositional or more formulaic and idiomatic in their meaning? Such experiments are far from straightforward to implement under natural conditions, and often come with a range of logistical and ethical considerations that must be considered and accounted for (see Schel, Machanda et al., 2013, and Slocombe et al., 2009 for details). Nevertheless, preliminary work we have been conducting on call combinations in wild chimpanzees suggests playback experiments are feasible and elicit reliable behavioral responses from receivers (Leroux et al., unpublished data). It is important to note however, that, despite being the gold standard for assessing the meaning of vocalizations, playbacks may not always be possible. In such situations, other more behaviorally-based proxies of call meaning, such as assessing whether receivers respond in ways suggesting that the signaler’s goal is met, might become particularly important (see Hobaiter & Byrne, 2014, for more details). Furthermore, whilst our focus in this review has been to highlight the various combinations of vocalizations in our closest living relatives and the extent to which they can be analyzed as syntactic-like structures, emerging data suggest that great apes may also be capable of combining signals from different modalities. As is the case in the vocal domain, such cross-modal concatenation may help to refine or augment meaning and hence, these communicative structures might also be relevant in unpacking the evolutionary progression of syntax more generally.

Finally, to ensure that comparative data can reliably inform our understanding of syntax’s evolutionary origins, we argue it is vital that researchers in animal communication establish a dialogue with language scientists working on similar questions and data in humans. Such cross-disciplinary discourse can help guide hypotheses and develop robust models for the evolution of syntax in human language (e.g., Bolhuis et al., 2018a, b; Griesser et al., 2018; Townsend et al., 2018). Without this, we risk making imprecise or even unfair comparisons between nonhuman animal communication systems and human language, which has arguably been one of the key factors limiting recent progress in this field.

Acknowledgments

We thank Loïc Pougnault and Stuart Watson for comments on the manuscript, Adèle Trevilly for illustrations and Cassandre Trevilly for the design of the figure. This work was supported by the Swiss National Science Foundation (31003A_153065 and PP00P3_163850) to S.W.T. We declare no conflict of interest.

References

- Arnold, K., & Zuberbühler, K. (2006a). Semantic combinations in primate calls. *Nature*, *441*, 303.
- Arnold, K., & Zuberbühler, K. (2006b). The alarm-calling system of adult male putty-nosed monkeys, *Cercopithecus nictitans martini*. *Animal Behaviour*, *72*, 643–653.
- Arnold, K., & Zuberbühler, K. (2012). Call combinations in monkeys: Compositional or idiomatic expressions? *Brain and Language*, *120*, 303–309.
- Berthet, M., Mesbahi, G., Pajot, A., Cäsar, C., Neumann, C., & Zuberbühler, K. (2019). Titi monkeys combine alarm calls to create probabilistic meaning. *Science Advances*, *5*, eaav3991.
- Bolhuis, J. J., Beckers, G. J. L., Huybregts, M. A. C., Berwick, R. C., & Everaert, M. B. H. (2018a). Meaningful syntactic structure in songbird vocalizations? *PLoS Biology*, *16*, e2005157.
- Bolhuis, J. J., Beckers, G. J. L., Huybregts, M. A. C., Berwick, R. C., & Everaert, M. B. H. (2018b). The slings and arrows of comparative linguistics. *PLoS Biology*, *16*, e3000019.

- Bolhuis, J. J., Tattersall, I., Chomsky, N., & Berwick, R. C. (2014). How could language have evolved? *PLoS Biology*, *12*, e1001934.
- Briseño-Jaramillo, M., Ramos-Fernandez, G., Palacios-Romo, T. M., Sosa-López, J. R., & Lemasson, A. (2018). Age and social affinity effects on contact call interactions in free-ranging spider monkeys. *Behavioral Ecology and Sociobiology*, *72*, 192.
- Candiotti, A., Zuberbühler, K., & Lemasson, A. (2012a). Context-related call combinations in female Diana monkeys. *Animal Cognition*, *15*, 327–339.
- Candiotti, A., Zuberbühler, K., & Lemasson, A. (2012b). Convergence and divergence in Diana monkey vocalizations. *Biology Letters*, *8*, 382–385.
- Cäsar, C., & Zuberbühler, K. (2012). Referential alarm calling behaviour in New World primates. *Current Zoology*, *58*, 680–697.
- Cäsar, C., Zuberbühler, K., Young, R. J., & Byrne, R. W. (2013). Titi monkey call sequences vary with predator location and type. *Biology Letters*, *9*, 20130535.
- Cheney, D. L., & Seyfarth, R. M. (1990). *How monkeys see the world: Inside the mind of another species*. University of Chicago Press.
- Clay, Z., & Zuberbühler, K. (2009). Food-associated calling sequences in bonobos. *Animal Behaviour*, *77*, 1387–1396.
- Clay, Z., & Zuberbühler, K. (2011). Bonobos extract meaning from call sequences. *PLoS ONE*, *6*, e18786.
- Collier, K., Bickel, B., van Schaik, C. P., Manser, M. B., & Townsend, S. W. (2014). Language evolution: Syntax before phonology? *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20140263.
- Collier, K., Townsend, S. W., & Manser, M. B. (2017). Call concatenation in wild meerkats. *Animal Behaviour*, *134*, 257–269.
- Coye, C., Ouattara, K., Zuberbühler, K., & Lemasson, A. (2015). Suffixation influences receivers' behaviour in nonhuman primates. *Proceedings of the Royal Society B: Biological Sciences*, *282*, 20150265.
- Coye, C., Zuberbühler, K., & Lemasson, A. (2016). Morphologically structured vocalizations in female Diana monkeys. *Animal Behaviour*, *115*, 97–105.
- Crockford, C., & Boesch, C. (2005). Call combinations in wild chimpanzees. *Behaviour*, *142*, 397–421.
- Crockford, C., Gruber, T., & Zuberbühler, K. (2018). Chimpanzee quiet hoo variants differ according to context. *Royal Society Open Science*, *5*, 172066.
- Crockford, C., Wittig, R. M., & Zuberbühler, K. (2015). An intentional vocalization draws others' attention: A playback experiment with wild chimpanzees. *Animal Cognition*, *18*, 581–591.
- Di Bitetti, M. S. (2003). Food-associated calls of tufted capuchin monkeys (*Cebus apella nigritus*) are functionally referential signals. *Behaviour*, *140*, 565–592.
- Engesser, S., Ridley, A. R., & Townsend, S. W. (2016). Meaningful call combinations and compositional processing in the southern pied babbler. *Proceedings of the National Academy of Sciences*, *113*, 5976–5981.
- Fedurek, P., & Slocombe, K. E. (2011). Primate vocal communication: A useful tool for understanding human speech and language evolution? Tools for tackling language and speech evolution. *Human Biology*, *83*, 153–173.
- Fedurek, P., Zuberbühler, K., & Dahl, C. D. (2016). Sequential information in a great ape utterance. *Scientific Reports*, *6*, 38226.
- Fischer, J., Wheeler, B. C., & Higham, J. P. (2015). Is there any evidence for vocal learning in chimpanzee food calls? *Current Biology*, *25*, R1028–R1029.
- Fitch, T. W. (2018). Bio-linguistics: Monkeys break through the syntax barrier. *Current Biology*, *28*, R695–R697.
- Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & van Schaik, C. P. (2019). Multimodal communication and language origins: Integrating gestures and vocalizations. *Biological Reviews*, *94*, 1809–1829.
- Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Harvard University Press.
- Griesser, M., Wheatcroft, D., & Suzuki, T. N. (2018). From bird calls to human language: exploring the evolutionary drivers of compositional syntax. *Current Opinion in Behavioral Sciences*, *21*, 6–12.
- Hardus, M. E., Lameira, A. R., Singleton, I., Morrogh-Bernard, H. C., Knott, C. D., Ancrenaz, M., Utami Atmoko, S. S., & Wich, S. A. (2009). A description of the orangutan's vocal and sound repertoire, with a focus on geographic variation. In S. A. Wich, T. Mitra Setia, & C. P. van Schaik (Eds), *Orangutans: Geographic variation in behavioral ecology and conservation* (pp. 49–60). Oxford University Press.
- Hedwig, D., Mundry, R., Robbins, M. M., & Boesch, C. (2015). Contextual correlates of syntactic variation in mountain and western gorilla close-distance vocalizations: Indications for lexical or phonological syntax? *Animal Cognition*, *18*, 423–435.

- Hedwig, D., Robbins, M. M., Mundry, R., Hammerschmidt, K., & Boesch, C. (2014). Acoustic structure and variation in mountain and western gorilla close calls: A syntactic approach. *Behaviour*, *151*, 1091–1120.
- Hobaiter, C., & Byrne, R. W. (2014). The meanings of chimpanzee gestures. *Current Biology*, *24*, 1596–1600.
- Hurford, J. (2011). *The origins of grammar: Language in the light of evolution II*. Oxford University Press.
- Katsu, N., Yamada, K., Okanoya, K., & Nakamichi, M. (2019). Temporal adjustment of short calls according to a partner during vocal turn-taking in Japanese macaques. *Current Zoology*, *65*, 99–105.
- Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., Bohn, K., Cao, Y., Carter, G., Cäsar, C., Coen, M., DeRuiter, S. L., Doyle, L., Edelman, S., Ferrer-i-Cancho, R., Freeberg, T. M., Garland, E. C., Gustison, M., Harley, H. E.,...Zamora-Gutierrez, V. (2014). Acoustic sequences in nonhuman animals: A tutorial review and prospectus. *Biological Reviews*, *91*, 13–52.
- Lameira, A. R. (2017). Bidding evidence for primate vocal learning and the cultural substrates for speech evolution. *Neuroscience & Biobehavioral Reviews*, *83*, 429–439.
- Lameira, A. R., & Shumaker, R. W. (2019). Orangutans show active voicing through a membranophone. *Scientific Reports*, *9*, 5–10.
- Lemasson, A., Pereira, H., & Levréro, F. (2018). Social basis of vocal interactions in western lowland gorillas (*Gorilla g. gorilla*). *Journal of Comparative Psychology*, *132*, 141–151.
- Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 20130302.
- Lieberman, P. (1991). *Uniquely human: The evolution of speech, thought, and selfless behavior*. Harvard University Press.
- Marler, P. (1977). The structure of animal communication sounds. In T. H. Bullock (Ed), *Recognition of complex acoustic signals: Report of the Dahlem workshop*. Abakon Verlagsgesellschaft.
- Masataka, N., & Fujita, K. (1989). Vocal learning of Japanese and rhesus monkeys. *Behaviour*, *109*, 191–199.
- Miyagawa, S., & Clarke, E. (2019). Systems underlying human and old world monkey communication: One, two, or infinite. *Frontiers in Psychology*, *10*, 1–11.
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009a). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, *106*, 22026–22031.
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009b). Campbell's monkeys use affixation to alter call meaning. *PLoS ONE*, *4*, e7808.
- Ouattara, K., Zuberbühler, K., N'goran, E. K., Gombert, J.-E., & Lemasson, A. (2009). The alarm call system of female Campbell's monkeys. *Animal Behaviour*, *78*, 35–44.
- Pika, S., Wilkinson, R., Kendrick, K. H., & Vernes, S. C. (2018). Taking turns: Bridging the gap between human and animal communication. *Proceedings of the Royal Society B: Biological Sciences*, *285*, 20180598.
- Pozzi, L., Hodgson, J. A., Burrell, A. S., Sterner, K. N., Raaum, R. L., & Disotell, T. R. (2014). Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution*, *75*, 165–183.
- Santos, L. R., Nissen, A. G., & Ferrugia, J. A. (2006). Rhesus monkeys, *Macaca mulatta*, know what others can and cannot hear. *Animal Behaviour*, *71*, 1175–1181.
- Schamberg, I., Cheney, D. L., Clay, Z., Hohmann, G., & Seyfarth, R. M. (2016). Call combinations, vocal exchanges and interparty movement in wild bonobos. *Animal Behaviour*, *122*, 109–116.
- Schel, A. M., Machanda, Z., Townsend, S. W., Zuberbühler, K., & Slocombe, K. E. (2013). Chimpanzee food calls are directed at specific individuals. *Animal Behaviour*, *86*, 955–965.
- Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., & Slocombe, K. E. (2013). Chimpanzee alarm call production meets key criteria for intentionality. *PLoS ONE*, *8*, e76674.
- Seyfarth, R. M., & Cheney, D. L. (2010). Primate vocal communication. In M. Platt & G. Asif (Eds), *Primate neuroethology* (pp. 84–97). Oxford University Press.
- Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980a). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, *210*, 801–803.
- Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980b). Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Animal Behaviour*, *28*, 1070–1094.
- Slocombe, K. E., Townsend, S. W., & Zuberbühler, K. (2009). Wild chimpanzees (*Pan troglodytes schweinfurthii*) distinguish between different scream types: Evidence from a playback study. *Animal Cognition*, *12*, 441–449.
- Slocombe, K. E., & Zuberbühler, K. (2005). Functionally referential communication in a chimpanzee. *Current Biology*, *15*, 1779–1784.
- Slocombe, K. E., & Zuberbühler, K. (2010). Vocal communication in chimpanzees. In E. V. Lonsdorf, S. R. Ross, &

- T. Matsuzawa (Eds), *The mind of the chimpanzee* (pp. 192-207). The University of Chicago Press.
- Spillmann, B., Dunkel, L. P., van Noordwijk, M. A., Amda, R. N. A., Lameira, A. R., Wich, S. A., & van Schaik, C. P. (2010). Acoustic properties of long calls given by flanged male orang-utans (*Pongo pygmaeus wurmbii*) reflect both individual identity and context. *Ethology*, *116*, 385–395.
- Townsend, S. W., Engesser, S., Stoll, S., Zuberbühler, K., & Bickel, B. (2018). Compositionality in animals and humans. *PLoS Biology*, *16*, e2006425.
- Townsend, S. W., & Manser, M. B. (2012). Functionally referential communication in mammals: The past, present and the future. *Ethology*, *119*, 1–11.
- Van Lancker-Sidtis, D., & Rallon, G. (2004). Tracking the incidence of formulaic expressions in everyday speech: Methods for classification and verification. *Language and Communication*, *24*, 207–240.
- Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., West, V., & Slocombe, K. E. (2015a). Reply to Fischer et al. *Current Biology*, *25*, R1019–R1031.
- Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., West, V., & Slocombe, K. E. (2015b). Vocal learning in the functionally referential food grunts of chimpanzees. *Current Biology*, *25*, 495–499.
- Zuberbühler, K. (2018a). Combinatorial capacities in primates. *Current Opinion in Behavioral Sciences*, *21*, 161–169.
- Zuberbühler, K. (2018b). Intentional communication in primates. *Revue Tranel*, *68*, 69–75.
- Zuberbühler, K., & Lemasson, A. (2014). Primate communication: Meaning from strings of calls. In F. Lowenthal & L. Lefebvre (Eds), *Language and recursion* (pp. 115-125). Springer Science.