Evolutionary roads to syntax

Klaus Zuberbühler

Evolutionary roads to syntax

Abstract

Syntax is habitually named as what sets human language apart from other communication systems, but how did it evolve? Comparative research on animal behaviour has contributed in important ways, with mainly three sets of data. First, animals have been subjected to artificial grammar tasks, based on the hypothesis that human syntax has evolved through advanced *computational capacity*. In these experiments humans generally outperform animals, but there are questions about validity, as experimental stimuli are (deliberately) kept devoid of semantic content. Second, animal communication has been compared in terms of the *surface structures* with the aim of developing a typology of animal syntax, based on the hypothesis that syntax is an evolutionary solution to the constraints of small signal repertoires. A wide range of combinatorial phenomena has been described, mainly in non-human primates, but there is little support for the hypothesis that syntax has emerged due to repertoire size constraints. A third way of studying the evolution of syntax is to compare how animals perceive and communicate about external events, the mental deep structure of syntax. Human syntax is closely aligned with how we perceive events in terms of agency, action, and patience, each with subsidiary functions. The event perception hypothesis has been least explored in animals and requires a serious research programme.

Theories of syntax

Studying the evolution of language is notoriously difficult. Neither brains nor behaviours fossilise, such that the archaeological record can only offer little insight into much of what makes humans unique. The comparative approach has turned out to be a viable alternative, by which the behaviours and cognition of closely related animal species are compared in order to draw inferences about the past evolutionary history. The assumption is that behaviour, and its underlying cognitive governance, has a heritable, genetic basis that can be traced phylogenetically. If the topic concerns a human-specific trait, such as syntax, non-human primates naturally play a key role in such endeavours, something that is also reflected in this opinion piece.

What is syntax and how did it evolve? The standard dictionary entry for syntax is something like "a set of rules, principles and processes that determine how sentences are formed from words and phrases in a language". For evolutionary studies, however, this definition is unsatisfactory because it presupposes linguistic units, i.e., words, phrases, sentences, which are themselves not available to animals. It is possible, however, to modify the standard syntax definition by using functional placeholders (word = meaningful unit; sentence = utterance that conveys a statement, question, exclamation, or command). Hence, syntax is *the set of rules, principles and processes that determine how statements, questions, exclamations or commands are formed from meaningful units.*

How did syntax evolve? The prevailing view, at least amongst biologically trained scholars, is that the evolution of syntax was a gradual process to the effect that its evolutionary history can be reconstructed by comparative evidence. Three lines of investigation have produced relevant data for evolutionary considerations: syntax as a computational capacity (the computational hypothesis), syntax as way to evade repertoire constraints (the surface structure hypothesis), and syntax as a reflection of event perception (the deep structure hypothesis).

Syntax as computational capacity

A first road to study the evolution syntax derives from theoretical linguistics, which seeks to describe language in terms of formal, artificial grammars with increasing complexity (the 'Chomsky hierarchy') using computer science tools (Chomsky, 1956). The assumption is that computationally simple syntax, such as finite state grammar, requires fewer computational operations and thus fewer cognitive resources than complex syntax, such as phrase structure grammar. In finite state grammar, the meaning of a sentence emerges from taking into account the relations of adjacent words, i.e. decisions are taken serially. In natural languages the finite state grammar cannot explain the entire range of phenomena, mainly because there are also dependencies between non-adjacent words, requiring more complex phrase structure grammar (Chomsky, 1957).

In behaviour experiments, subjects are exposed to stimulus sequences that comply with (or are in breach of) the grammar under investigation. The prediction is that successful processing enables a subject to perceive syntactic violations, measured by increased attention (Fitch & Hauser, 2004). This reasoning is analogous to how linguists examine natural grammars, by asking native speakers to make grammaticality judgements (Chomsky 1957). For example, the sentence "colourless green ideas sleep furiously" is typically judged as grammatical, despite the fact that it is semantically nonsensical. In artificial grammar research, however, 'sentences' are usually represented by sequences of tones, speech sounds or vocalisations that do or do not comply with the respective grammar under study. The main conclusion from this research has been that only humans can deal with complex artificial grammars (Fitch & Hauser, 2004; Wilson, Smith, & Petkov, 2015) due to the limited computational power of animal brains (Friederici, 2004), but see (Gentner, Fenn, Margoliash, & Nusbaum, 2006; van Heijningen, de Visser, Zuidema, & ten Cate, 2009). The evolution of syntax, in this view, is a direct consequence of the evolution of computational power required for syntactic processing.

One issue with artificial grammar research is that stimulus sequences are usually meaningless simple tones. This is a deliberate choice so that the syntactic apparatus can be investigated in its pure state, uncontaminated by semantics. Although the logic is pertinent, there are questions about the 'ecological' validity of this approach. 'Colourful green ideas sleep furiously' may be nonsensical but the sentence is still composed of meaningful units, which may trigger processing in different brain areas than processing of meaningless tone sequences. One debate therefore is whether artificial grammar experiments reveal something relevant for evolutionary theories of syntax or whether they are more informative regarding acoustic pattern recognition (Hochmann, Azadpour, & Mehler, 2008; Zuberbuhler, 2018). Brain imaging studies would provide valuable input towards this question.

Syntax in surface structures

Another influential hypothesis is that syntax evolves as soon as lexicons reach their limits, because of memory or production limits: "…natural selection can only favour the emergence of syntax if the number of required signals exceeds a threshold value" (Nowak, Plotkin, & Jansen, 2000). Although intuitively appealing, the hypothesis is difficult to test because it presumes species-specific thresholds, but there is no theory as to how they could be determined. Nevertheless the hypothesis predicts that, in closely related species, syntax is only present in species that have reached the threshold, i.e. the ones with larger repertoires.

A typology of syntax in animal communication

There is a long ethological tradition of studying the surface features of animal communication, i.e., the way species combine elements of their signal repertoires into sequences. Pioneering were studies of birdsong that have revealed, for example, that syntax plays a role in social interactions (Marler & Peters, 1988). Birdsong functions to attract mates and keep rivals away and, as such, mainly contains information about caller identity (Catchpole & Slater, 1995). More recently, research on bird syntax has shifted towards the question of how meaning is conveyed by combinations of signals that carry their own meaning, with relevant work on babblers (Engesser, Ridley, & Townsend, 2016) and Japanese tits ((Engesser et al., 2016; Suzuki, Wheatcroft, & Griesser, 2016). For mammals, studies exist on rock hyraxes (Kershenbaum, Ilany, Blaustein, & Geffen, 2012) and various primates (Crockford & Boesch, 2005; Hedwig, Mundry, Robbins, & Boesch, 2015; Schamberg, Cheney, Clay, Hohmann, & Seyfarth, 2016; Zuberbuhler, 2018).

In primates, early studies reported syntactic structures for example in Cebus and squirrel monkey calls (Newman, Katzlieblich, Talmageriggs, & Symmes, 1978; Robinson, 1984). More recently, combinatorial calling has been found in various primate alarm and contact calls (alarms: Diana monkeys (Stephan & Zuberbuhler, 2008), Campbell's monkeys (Lemasson, Ouattara, Bouchet, & Zuberbuehler, 2010; Ouattara, Lemasson, & Zuberbuhler, 2009a; Zuberbühler, 2002), King Colobus (Schel, Tranquilli, & Zuberbuhler, 2009); contact calls: Diana monkeys, Campbell's monkeys (Candiotti, Zuberbühler, & Lemasson, 2012; Coye, Ouattara, Arlet, Lemasson, & Zuberbuhler, 2018; Coye, Ouattara, Zuberbuhler, & Lemasson, 2015; Coye, Zuberbuhler, & Lemasson, 2016)).

Putty-nosed monkeys have been particularly well studied, with males producing two alarm call types, pyows and hacks, either singly or in combination. Series of pyows are given mainly to terrestrial disturbances (e.g. leopards) and series of hacks to aerial dangers (e.g. crowned eagles) (Arnold & Zuberbuhler, 2006). In addition, males can combine both calls into 'pyow-hack' sequences, which carry a different meaning (travel) unlinked to the meanings of the component calls (Arnold & Zuberbühler, 2006). The composed meaning appears to reside in the pyow-hack transition, regardless of the number of pyows and hacks (Arnold & Zuberbühler, 2012). Listeners respond to the combinatorial features, by perceiving the combination as a meaningful unit, which is different from its components (Arnold & Zuberbühler, 2008).

Although these bird and monkey studies are relevant, an evolutionarily informed theory of syntax necessarily requires data from our closest relatives, the apes. Gibbon song has long been of interest (Demars & Goustard, 1972), although structural changes were not usually linked with changes in meaning. More recently, it was found that white-handed gibbon songs were produced both as duets and to predators;

snakes, clouded leopards and humans. Crucially, predator and non-predator songs were assembled in different ways, albeit from the same basic song units, with syntactic differences particularly visible during the early parts of a song, with indications that recipients discriminated between the different song types (Clarke, Reichard, & Zuberbuhler, 2006). For chimpanzees, an early study found also evidence for regular use of call combinations but no clear semantic effects (Crockford & Boesch, 2005). Some progress has been made with a study on the syntax of pant hoot call utterances. Using machine learning and automated feature extraction, the study produced evidence for encoding of age, rank, identity and context, across the four phases (fig. 1), (Fedurek, Zuberbühler, & Dahl, 2016). The introduction and build-up phases are low amplitude signals and contained mainly caller identity information, suggesting they are directed at nearby individuals. The climax phase, in contrast, is acoustically conspicuous and high-amplitude and contained information on both identity and social status (low vs. high rank), presumably targeting faraway group members and neighbouring groups. This is a relevant finding because, in chimpanzees, decisions about whether to engage in intergroup conflict are largely based on attending to neighbouring pant hoot vocalisations (Herbinger, Papworth, Boesch, & Zuberbuehler, 2009; Wilson, Hauser, & Wrangham, 2001; Wilson et al., 2014; Wilson et al., 2012; Wilson, Hauser, & Wrangham, 2007). Finally, pant hoots usually end with low-amplitude, let-down units, which inform nearby group members about the caller's forthcoming behavioural intentions (feeding vs. travelling). Callers can omit one or several of the four phases, allowing them to target specific audiences with specific information (fig. 1).

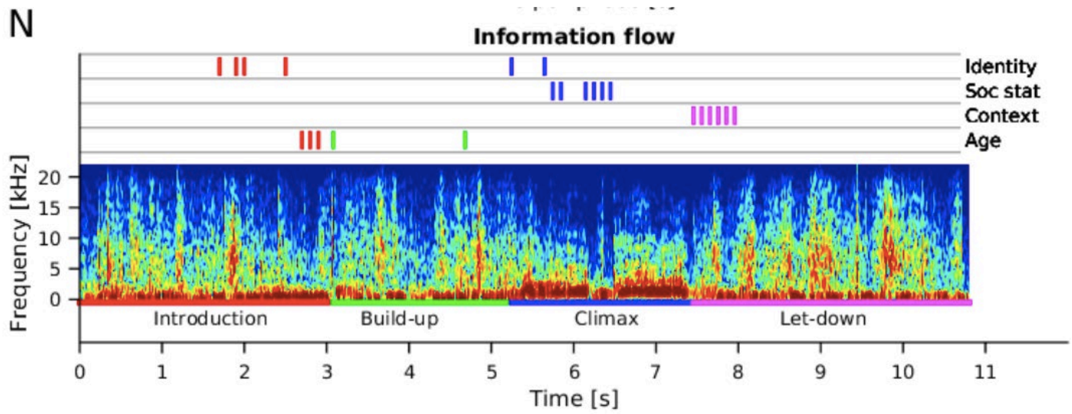

Figure 1. Four phases of chimpanzee pant hoot vocalisations: introduction, build-up, climax, let-down. Adults usually produce pant hoot phases in this order, although one or several phases can be omitted. Different phases convey different sets of information, as indicated by the top information flow panel (reprinted from (Fedurek, Zuberbühler, & Dahl, 2016) licensed under a Creative Commons Attribution 4.0 International License).

Another line of research in great apes is on the combined use of different modalities. Here, call-gesture combinations were generally rare and only used in social interactions of very positive or very negative connotations. Gestures were often added when vocal utterances failed to achieve a desired social goal, an expression of underlying persistence (Hobaiter, Byrne, & Zuberbuhler, 2017). For bonobos, (Schamberg et al., 2016) demonstrated natural call combinations in a wild population, whereas for captive groups there was evidence for call/gesture combinations, notably to disambiguate meaning (Genty, Clay, Hobaiter, & Zuberbühler, 2014). These studies demonstrate that call/call and call/gesture combinations exist and as such provide the groundwork for further research on the evolution of signal combinations in our closest living relative. Also relevant is the finding that when encountering food, bonobos produced sequences of call types that depended on the perceived quality of food (Clay & Zuberbuhler, 2009). Playback experiments confirmed that listeners were able to attend to the different sequences and make predictions about what type of food the caller has found (Clay & Zuberbühler, 2011). For gorillas, finally, grunts, grumbles and hums can be given singly or in combinations (Harcourt & Stewart, 1996; Harcourt, Stewart, & Hauser, 1993; Salmi, Hammerschmidt, & Doran-Sheehy, 2013; Seyfarth, Cheney, Harcourt, & Stewart, 1994; Stewart & Harcourt, 1994). When produced in isolation, grunts were given by individuals resting in close proximity of each other, whereas grumbles were given during foraging. When produced in combination, grumbles appeared to lose their foraging meaning, suggesting that call combinations have less to do with augmenting semantics but to mark social roles during communicative interactions (Hedwig et al., 2015).

The studies reviewed so far have revealed a bewildering range of combinatorial structures that can be grouped as follows (Zuberbuhler, 2018). First, in *merged units* callers combine vocal structures, mainly to convey identity and event information (Diana monkeys (Candiotti et al., 2012; Coye et al., 2016); Campbell's monkeys (Candiotti et al., 2012; Coye et al., 2018)). Related to this is *suffixation,* as found in male Campbell's monkey alarm calls. Here, callers add acoustically invariable 'oo'-units to three distinct alarm calls, to indicate that danger is non-urgent (Ouattara, Lemasson, & Zuberbuehler, 2009b). At the call sequence level, there are examples of *permutations*, i.e., ordered call deliveries, as found in alarm calling of male Campbell's and, as discussed before, male putty-nosed monkeys (Arnold & Zuberbühler, 2006; Ouattara, Lemasson, & Zuberbuehler, 2009a). Another line of inquiry has been on New World monkeys in Brazil. Black-fronted titi monkeys produce alarm call sequences to refer to both predator type and location (Cäsar, Zuberbühler, Young, & Byrne, 2013), although meaning is encoded by a *probabilistic (stochastic*) rather than a categorical mechanism (Berthet et al submitted). Finally, there is evidence for meaning being conveyed by utterances of varying lengths, further assembled into more complex sequences, a *numeric* and seemingly *hierarchical* structure (Diana monkeys (Zuberbühler, 2000), Campbell's monkeys (Lemasson et al., 2010), black-and-white Colobus monkeys (Schel, Candiotti, & Zuberbühler, 2010).

Syntax in deep structure

The studies reviewed so far have revealed little about any underlying cognitive processes and it is even possible that, what appears as syntax, is not linked to any interesting mental processing. For example, syntactic regularities in signal sequences could emerge merely by accident due to physiological constraints (e.g. structural changes due to increasing exhaustion), semantic constraints (e.g. responding to X may warrant some call types but not others), pragmatic constraints (e.g. more urgent calls may be produced before less urgent calls), or on-going changes in the environmental context triggering calls (P Schlenker and E Chemla, pers. comm.).

In humans, however, syntax is tightly linked with how events in the external world are perceived, structured and mentally represented. In particular, humans have a natural propensity to decompose events into actors, actions, and patients to the effect that there is a curious correspondence between the components of natural events and the grammatical functions of language. Sentences are structured in that they contain agents (doer, cause, experiencer), actions (what), patients, targets or beneficiaries (to whom) who experience the action or state of affair (e.g. "the eagle attacked the monkey"). Arguments usually have additional components, such as the manner (how) by which an action is carried out or the instrument used (with what) for this purpose.

Additionally, arguments can contain information about location (where), origin or direction (from - to where), or time (when) an action was, is or will be carried out (e.g. "the eagle attacked the monkey *from above*".

Another useful description of events is in terms of predication *("the eagle attacked…"),* modification (*"the large eagle…"*) and coordination *("eagles and leopards")* (Townsend, Engesser, Stoll, Zuberbuhler, & Bickel, 2018). Languages have means to express these event features in ways to make them evident to listeners, usually with specific syntactic functions. For example, to syntactically distinguish an agent from a patient, some languages use phonological case marking while others use word order.

The hypothesis here is that, during human evolution, these event-bound cognitive universals (agents, patients, actions, manners, etc.) have become externalised and assimilated into the communication system. This hypothesis is supported by work on Nicaraguan sign language, which has shown that deaf children will gradually and without specific tutoring develop syntactic structures in spontaneous sign language that enables them to encode the core components of an event, rather than referring to entire events holistically (Senghas, Kita, & Ozyurek, 2004). Modern humans, in other words, have a natural propensity to mark the key components of external events with (arbitrary) syntactic features.

How do animals perceive natural external events? There is evidence from artificial language studies that marine mammals can be trained to discriminate agents from patients (Herman, Richards, & Wolz, 1984). In natural communication, a study on chimpanzee vocal behaviour has found acoustic differences in screams given in different social roles, i.e. when the caller was the actor or the patient in an aggressive act, which was discriminated by others (Slocombe, Kaller, Call, & Zuberbuehler, 2010; Slocombe & Zuberbuhler, 2005).

Human event perception, however, is vastly more complex than marking agents and patients. Complex event perception is likely to have evolved first, possibly due to increasingly complex social systems and associated brain enlargements. Syntax, in this view, is a mere by-product of perceiving external events in decomposed ways and of the ability to mark these components with communication signals. The human road to syntax may have built on this predisposition, completed with the advent of unprecedented vocal control, allowing event perception to become linguistically encoded with grammatical functions.

Current issues

Syntax without precursors

A particularly contentious on-going debate is whether animal data can contribute in meaningful ways to questions about syntax evolution in humans. One argument is that studies of animal communication are irrelevant, because the only relevant property of human syntax is its generative, hierarchical nature, for which there is no evidence in animal communication. Cognitively, the argument goes, this is achieved by a single mental operation, merge, which takes two syntactic elements and assembles them to form a set (Bolhuis, Tattersall, Chomsky, & Berwick, 2014)(Bolhuis, 2017; Bolhuis, Beckers, Huybregts, Berwick, & Everaert, 2018; Bolhuis et al., 2014; Townsend et al., 2018).

The ‘merge’ view of language, however, is not universally accepted, even amongst linguists. For example, much of ordinary language use is based on accessing prefabricated phrases from a vast memory stock. Although the retrieved utterances may be analysed in terms of syntactic structure, language users simply retrieve them ‘wholesale’ to fit into appropriate slots (Townsend et al., 2018). Prefabricated expressions account for up to half of all phrases used in conversations (Van Lancker-Sidtis & Rallon, 2004), suggesting that evolutionary investigations of syntax should also focus on non-generative, non-hierarchical combinatorial systems, as frequently seen in animal communication.

Varieties of merge

An evolutionarily more fruitful proposal has been to distinguish between different levels of ‘merge’, with increasing generative capacity (Rizzi, 2016). According to this, 0-merge systems operate only with individual items from the lexicon. In fact, this has been the default view of animal communication for decades, i.e., that animal signals function as holistic units without any recourse to combinations (Hauser, 2000). 1-merge systems, next, have combinatorial properties insofar as they allow for the formation of two-unit expressions, although the system then stops, with no recursive procedures (i.e., word-word merges). Following this are 2-merge systems that allow for recursion insofar as merged expressions (e.g. word-word or ‘phrase’) can enter new merges, with its own components, but this requires more memory capacity. Thus, 2-merge systems can potentially generate an unlimited set of expressions (word-phrase merges) and are thus truly generative. Finally, 3-merge systems are characterised by the ability to merge already merged expressions (phrase-phrase merges), which requires further memory capacity. Sentence formation in human language requires a 3-merge system, as subjects and predicates consist of merged expressions. Current evidence suggests that animal calling goes beyond 0-merge systems but stops at 1-merge systems, without any recursive applications.

Compositionality

Are humans thus unique in having higher-level hierarchical syntax to generate meaning? Most definitions of human language require compositionality, that is, that simple expressions are used to build more complex expressions, whose meaning is determined by the meanings of the constituent simple expressions *and* the rule that combines them. The meaning of the whole is determined by the meaning of its parts and how they are put together, the principle of compositionality.

In several theory papers, primate call systems have been analysed in such ways, which has led to the conclusion that some systems, particularly Campbell's monkey call suffixation and putty-nosed monkey call permutations, have weak *compositional* properties (fig. 2), a claim with implications for evolutionary theories of language (Schlenker et al., 2014; Schlenker, Chemla, Arnold, & Zuberbuhler, 2016; Schlenker, Chemla, Casar, Ryder, & Zuberbuhler, 2017; Schlenker, Chemla, et al., 2016a, 2016b; Schlenker, Chemla, & Zuberbuhler, 2016).

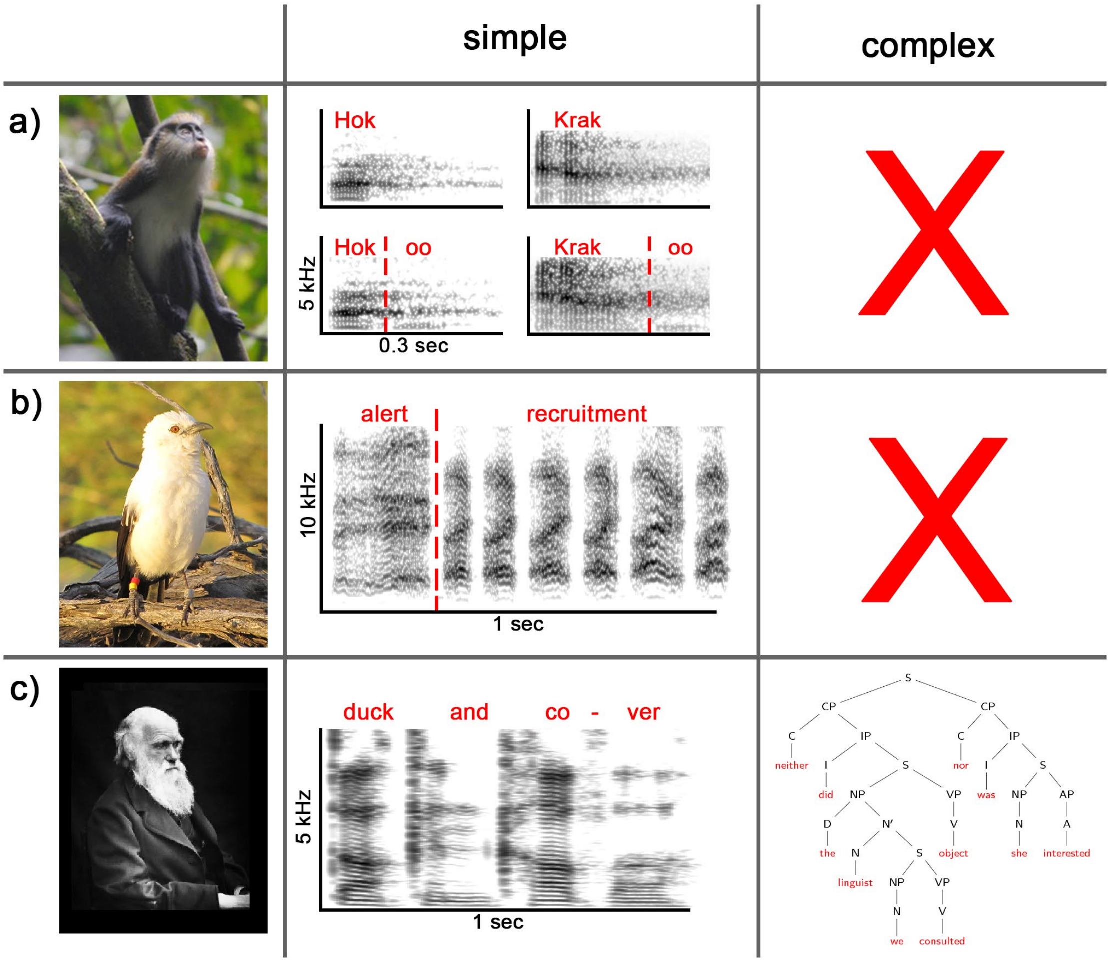

Figure 2. Examples of simple and complex compositionality in animals and humans. a) Male Campbell's monkeys produce 'krak' alarms (to leopards) and 'hok' alarms (to eagles), but both calls can also be merged with an '-oo' suffix to generate 'krak-oo' (to a range of disturbances) and 'hok-oo' (to non-ground disturbances) (Ouattara, Lemasson, & Zuberbuhler, 2009b). In playback experiments, suffixation has shown to be meaningful to listeners, suggesting that it is an evolved communication function (Coye et al., 2015). This system may qualify as limited compositionality, as the meanings of krak-oo and hok-oo are directly derived from the meanings of krak/hok plus the meaning of—oo. b) Compositionality in birds: Pied babblers produce 'alert' calls in response to unexpected but low-urgency threats and 'recruitment' calls when recruiting conspecifics to new foraging sites (Engesser, 2016; Engesser et al., 2016).

When encountering a terrestrial threat that requires recruiting group members (in the form of mobbing), pied babblers combine the two calls into a larger structure, and playback experiments have indicated that receivers process the call combination compositionally by linking the meaning of the independent parts. c) Compositionality in humans: humans are capable of producing both simple, non-hierarchical compositions (e.g., 'Duck and cover!') and complex hierarchical compositions and dependencies. Photo in panel A credited to Erin Kane. Photo in panel B credited to Sabrina Engesser. A, adjective; AP, adjective phrase; C, conjunction; CP, conjunction phrase; D, determiner; I, Inflection-bearing element; IP, inflectional phrase; N, (pro-)noun; NP, noun phrase; S, sentence; V, verb; VP, verb phrase (reprinted from (Townsend et al., 2018) under the Creative Commons Attribution license).

Conclusion

Animal communication research has long worked under the assumption that animal calls are structurally simple, holistic signals that develop under strong genetic control (Snowdon et al., 1992; Snowdon & Hausberger, 1997). As call producers, animals (including primates) were thought to be cognitively unengaged, merely responding with acoustically invariable signals to evolutionarily urgent situations in more or less automated ways (Tomasello, 2008). This point has also been made for great apes, despite the fact that chimpanzees and bonobos have excelled in terms of social cognition and visually based communication (Call & Tomasello, 2008, 2007). More recently, the stance has come under scrutiny, due to a range of empirical developments. First, although primates do not imitate sounds, they have considerable degrees of control over their vocal output, which enables them to refrain from signalling and to modify parts of their vocal repertoire in communicatively functional ways (Lameira, Maddieson, & Zuberbuehler, 2014). Moreover, it has become clear that primate vocal behaviour goes beyond producing single calls to single events, with a steady stream of studies reporting various forms of signal combinations, sometimes even in compositional ways (Zuberbuhler, 2018). As a consequence, research on animal syntax is currently amongst the most productive areas in animal behaviour research, with results being debated across disciplines (Bolhuis et al., 2018; Schlenker, Chemla, et al., 2016b).

However, the currently available data do not yet give rise to an empirically informed evolutionary theory of human syntax. Instead, the current literature provides a bewildering diversity of combinatorial systems in animal communication, with no clear evolutionary trends or obvious phylogenetic patterns (Zuberbuhler, 2018). Equally, there is no conceptual agreement in how to integrate the different phenomena into a coherent evolutionary theory of syntax (Kershenbaum et al., 2014; Zuberbuhler, 2018).

Human syntax is the result of mental processes but this is rarely addressed by animal communication studies. Testing animals with artificial grammars has produced interesting findings, revealing something about the limits of computational capacities, but results are difficult to interpret because stimulus sequences are devoid of semantic relations. Yet "…what distinguishes true language from just collections of uttered words is that the semantic relations among the words are conveyed by syntactic and morphological structure" (Jackendoff, 2007).

Future research should focus on how animals, and especially non-human primates, naturally discriminate and mentally represent natural events and whether these representation correspond to the main grammatical functions of human language (actors, patients, descriptions of objects etc.). Data on whether animals perceive events as functionally structured is likely to produce important progress and lead to a better understanding of the evolutionary road to syntax.

References

Arnold, K., & Zuberbuhler, K. (2006). The alarm-calling system of adult male putty-nosed monkeys, Cercopithecus nictitans martini. *Animal Behaviour, 72*, 643-653.

Arnold, K., & Zuberbühler, K. (2006). Semantic combinations in primate calls. *Nature, 441*(7091), 303-303.

Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology, 18*(5), R202-R203. doi: 10.1016/j.cub.2008.01.040

Arnold, K., & Zuberbühler, K. (2012). Call combinations in monkeys: Compositional or idiomatic expressions? *Brain and Language, 120*(3), 303-309. doi: 10.1016/j.bandl.2011.10.001

Bolhuis, J. J. (2017). Making sense of language in the light of evolution. *Mind & Language, 32*(5), 591-596. doi: 10.1111/mila.12161

Bolhuis, J. J., Beckers, G. J. L., Huybregts, M. A. C., Berwick, R. C., & Everaert, M. B. H. (2018). Meaningful syntactic structure in songbird vocalizations? *Plos Biology, 16*(6). doi: 10.1371/journal.pbio.2005157

Bolhuis, J. J., Tattersall, I., Chomsky, N., & Berwick, R. C. (2014). How Could Language Have Evolved? *Plos Biology, 12*(8). doi: 10.1371/journal.pbio.1001934

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12*(5), 187-192.

Call, J., & Tomasello, M. (Eds.). (2007). *The gestural communication of apes and monkeys*: Taylor & Francis Lea.

Candiotti, A., Zuberbühler, K., & Lemasson, A. (2012). Context-related call combinations in female Diana monkeys. *Animal Cognition, 15*(3), 327-339. doi: 10.1007/s10071-011-0456-8

Cäsar, C., Zuberbühler, K., Young, R. J., & Byrne, R. W. (2013). Titi monkey call sequences vary with predator location and type. *Biology Letters, 9*(5). doi: 10.1098/rsbl.2013.0535

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

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory, 2*, 113-124. doi: doi:10.1109/TIT.1956.1056813

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Clarke, E., Reichard, U. H., & Zuberbuhler, K. (2006). The Syntax and Meaning of Wild Gibbon Songs. *Plos One, 1*(1). doi: 10.1371/journal.pone.0000073

Clay, Z., & Zuberbuhler, K. (2009). Food-associated calling sequences in bonobos. *Animal Behaviour, 77*(6), 1387-1396. doi: Doi 10.1016/J.Anbehav.2009.02.016

Clay, Z., & Zuberbühler, K. (2011). Bonobos extract meaning from call sequences. *Plos One, 6*(4). doi: 10.1371/journal.pone.0018786

Coye, C., Ouattara, K., Arlet, M. E., Lemasson, A., & Zuberbuhler, K. (2018). Flexible use of simple and combined calls in female Campbell's monkeys. *Animal Behaviour, 141*, 171-181. doi: 10.1016/j.anbehav.2018.05.01

Coye, C., Ouattara, K., Zuberbuhler, K., & Lemasson, A. (2015). Suffixation influences receivers' behaviour in non-human primates. *Proceedings of the Royal Society B-Biological Sciences, 282*(1807). doi: 10.1098/rspb.2015.0265

Coye, C., Zuberbuhler, K., & Lemasson, A. (2016). Morphologically structured vocalizations in female Diana monkeys. *Animal Behaviour, 115*, 97-105. doi: 10.1016/j.anbehav.2016.03.010

Crockford, C., & Boesch, C. (2005). Call combinations in wild chimpanzees. *Behaviour, 142*, 397-421.

Demars, C., & Goustard, M. (1972). STRUCTURE AND DEVELOPMENTAL PATTERN OF SONORAL EMISSIONS OF HYLOBATES HYLOBATES-CONCOLOR. *Bulletin Biologique de la France et de la Belgique, 106*(3), 177-191.

Engesser, S. (2016). *Vocal combinatoins in the Southern Pied Babbler (Turdoides bicolor) and the Chestnut-Crowned Babbler (Pomatostomus ruficeps).* PhD, Zurich, Zurich.

Engesser, S., Ridley, A., & Townsend, S. (2016). Meaningful call combinations and compositional processing in the southern pied babbler. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1600970113

Fedurek, P., Zuberbühler, K., & Dahl, C. D. (2016). Sequential information in a great ape utterance. *Scientific Reports, 6*, 38226.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science, 303*(5656), 377-380.

Friederici, A. D. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends in Cognitive Sciences, 8*(6), 245-247. doi: 10.1016/j.tics.2004.04.013

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

Genty, E., Clay, Z., Hobaiter, C., & Zuberbühler, K. (2014). Multi-modal use of a socially directed call in bonobos. *Plos One, 9*(1). doi: DOI 10.1371/journal.pone.0084738

Harcourt, A. H., & Stewart, K. J. (1996). Function and meaning of wild gorilla 'close' calls .2. Correlations with rank and relatedness. *Behaviour, 133*, 827-845.

Harcourt, A. H., Stewart, K. J., & Hauser, M. (1993). Functions of wild gorilla "close" calls: I. Repertoire context, and interspecific comparison. *Behaviour, 124*, 89-122.

Hauser, M. D. (2000). A primate dictionary? Decoding the function and meaning of another species' vocalizations. *Cognitive Science, 24*(3), 445-475.

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*(2), 423-435. doi: 10.1007/s10071-014-0812-6

Herbinger, I., Papworth, S., Boesch, C., & Zuberbuehler, K. (2009). Vocal, gestural and locomotor responses of wild chimpanzees to familiar and unfamiliar intruders: a playback study. *Animal Behaviour, 78*(6), 1389-1396. doi: 10.1016/j.anbehav.2009.09.010

Herman, L. M., Richards, D. G., & Wolz, J. P. (1984). Comprehension of sentences by bottlenosed dolphins. *Cognition, 16*(2), 129-219.

Hobaiter, C., Byrne, R. W., & Zuberbuhler, K. (2017). Wild chimpanzees' use of single and combined vocal and gestural signals. *Behavioral Ecology and Sociobiology, 71*(6). doi: 10.1007/s00265-017-2325-1

Hochmann, J. R., Azadpour, M., & Mehler, J. (2008). Do humans really learn A(n) B-n artificial grammars from exemplars? *Cognitive Science, 32*(6), 1021-1036. doi: 10.1080/03640210801897849

Jackendoff, R. (2007). *Language, Consciousness, Culture*. Cambridge MA: MIT Press.

Kershenbaum, A., Bowles, A. E., Freeberg, T. M., Jin, D. Z. Z., Lameira, A. R., & Bohn, K. (2014). Animal vocal sequences: not the Markov chains we thought they were. *Proceedings of the Royal Society B-Biological Sciences, 281*(1792). doi: 10.1098/rspb.2014.1370

Kershenbaum, A., Ilany, A., Blaustein, L., & Geffen, E. (2012). Syntactic structure and geographical dialects in the songs of male rock hyraxes. *Proceedings of the Royal Society B-Biological Sciences, 279*(1740), 2974-2981. doi: 10.1098/rspb.2012.0322

Lameira, A. R., Maddieson, I., & Zuberbuehler, K. (2014). Primate feedstock for the evolution of consonants. *Trends in Cognitive Sciences, 18*(2), 60-62. doi: 10.1016/j.tics.2013.10.013

Lemasson, A., Ouattara, K., Bouchet, H., & Zuberbuehler, K. (2010). Speed of call delivery is related to context and caller identity in Campbell's monkey males. *Naturwissenschaften, 97*(11), 1023-1027. doi: 10.1007/s00114-010-0715-6

Marler, P., & Peters, S. (1988). The role of song phonology and syntax in vocal learning preferences in the song sparrow, Melospiza melodia. *Ethology, 77*, 125-149.

Newman, J. D., Katzlieblich, A., Talmageriggs, G., & Symmes, D. (1978). SYLLABLE CLASSIFICATION AND SEQUENCING IN TWITTER CALLS OF SQUIRREL-MONKEYS (SAIMIRI-SCIUREUS). *Zeitschrift Fur Tierpsychologie-Journal of Comparative Ethology, 47*(1), 77-88.

Nowak, M. A., Plotkin, J. B., & Jansen, V. A. A. (2000). The evolution of syntactic communication. *Nature, 404*(6777), 495-498. doi: 10.1038/35006635

Ouattara, K., Lemasson, A., & Zuberbuehler, K. (2009a). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences of the United States of America, 106*(51), 22026-22031. doi: 10.1073/pnas.0908118106

Ouattara, K., Lemasson, A., & Zuberbuehler, K. (2009b). Campbell's Monkeys Use Affixation to Alter Call Meaning. *Plos One, 4*(11). doi: 10.1371/journal.pone.0007808

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

Ouattara, K., Lemasson, A., & Zuberbuhler, K. (2009b). Campbell's Monkeys Use Affixation to Alter Call Meaning. *Plos One, 4*(11).

Rizzi, L. (2016). Monkey morpho-syntax and merge-based systems. *Theoretical Linguistics, 42*(1-2), 139-145. doi: 10.1515/tl-2016-0006

Robinson, J. G. (1984). Syntactic structures in the vocalizations of wedge-capped capuchin monkeys, Cebus olivaceus. *Behaviour, 90*, 46-79.

Salmi, R., Hammerschmidt, K., & Doran-Sheehy, D. M. (2013). Western Gorilla Vocal Repertoire and Contextual Use of Vocalizations. *Ethology, 119*(10), 831-847. doi: 10.1111/eth.12122

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. doi: 10.1016/j.anbehav.2016.10.003

Schel, A. M., Candiotti, A., & Zuberbühler, K. (2010). Predator-deterring alarm call sequences in Guereza colobus monkeys are meaningful to conspecifics. *Animal Behaviour, 80*(5), 799-808. doi: 10.1016/j.anbehav.2010.07.012

Schel, A. M., Tranquilli, S., & Zuberbuhler, K. (2009). The Alarm Call System of Two Species of Black-and-White Colobus Monkeys (Colobus polykomos and Colobus guereza). *Journal of Comparative Psychology, 123*(2), 136-150.

Schlenker, P., Chemla, E., Arnold, K., Lemasson, A., Ouattara, K., Keenan, S., . . . Zuberbuhler, K. (2014). Monkey semantics: two 'dialects' of Campbell's monkey alarm calls. *Linguistics and Philosophy, 37*(6), 439-501. doi: 10.1007/s10988-014-9155-7

Schlenker, P., Chemla, E., Arnold, K., & Zuberbuhler, K. (2016). Pyow-hack revisited: Two analyses of Putty-nosed monkey alarm calls. *Lingua, 171*, 1-23. doi: 10.1016/j.lingua.2015.10.002

Schlenker, P., Chemla, E., Casar, C., Ryder, R., & Zuberbuhler, K. (2017). Titi semantics: Context and meaning in Titi monkey call sequences. *Natural Language & Linguistic Theory, 35*(1), 271-298. doi: 10.1007/s11049-016-9337-9

Schlenker, P., Chemla, E., Schel, A. M., Fuller, J., Gautier, J. P., Kuhn, J., . . . Zuberbuhler, K. (2016a). Formal monkey linguistics. *Theoretical Linguistics, 42*(1-2), 1-90. doi: 10.1515/tl-2016-0001

Schlenker, P., Chemla, E., Schel, A. M., Fuller, J., Gautier, J. P., Kuhn, J., . . . Zuberbuhler, K. (2016b). Formal monkey linguistics: The debate. *Theoretical Linguistics, 42*(1-2), 173-201. doi: 10.1515/tl-2016-0010

Schlenker, P., Chemla, E., & Zuberbuhler, K. (2016). What Do Monkey Calls Mean? *Trends in Cognitive Sciences, 20*(12), 894-904. doi: 10.1016/j.tics.2016.10.004

Senghas, A., Kita, S., & Ozyurek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science, 305*(5691), 1779-1782.

Seyfarth, R. M., Cheney, D. L., Harcourt, A. H., & Stewart, K. J. (1994). The acoustic features of gorilla double grunts and their relation to behavior. *Am. j. Primatol., 33*(1), 31-50.

Slocombe, K. E., Kaller, T., Call, J., & Zuberbuehler, K. (2010). Chimpanzees Extract Social Information from Agonistic Screams. *Plos One, 5*(7). doi: 10.1371/journal.pone.0011473

Slocombe, K. E., & Zuberbuhler, K. (2005). Agonistic screams in wild chimpanzees (Pan troglodytes schweinfurthii) vary as a function of social role. *Journal of Comparative Psychology, 119*(1), 67-77.

Snowdon, C. T., Elowson, A. M., Nishida, T., McGrew, W. C., Marler, P., Pickford, M., & de Waal, F. B. M. (Eds.). (1992). *Ontogeny of primate vocal communication*.

Snowdon, C. T., & Hausberger, M. (Eds.). (1997). *Social influences on vocal development*. Cambridge: Cambridge University Press.

Stephan, C., & Zuberbuhler, K. (2008). Predation increases acoustic complexity in primate alarm calls. *Biology Letters, 4*(6), 641-644. doi: Doi 10.1098/Rsbl.2008.0488

Stewart, K. J., & Harcourt, A. H. (1994). Gorillas' vocalizations during rest periods: signals of impending departure? *Behaviour, 130*, 29-40.
Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nature Communications, 7*. doi: 10.1038/ncomms10986
Tomasello, M. (2008). *Origins of human communication*. Cambridge MA: MIT Press.
Townsend, S. W., Engesser, S., Stoll, S., Zuberbuhler, K., & Bickel, B. (2018). Compositionality in animals and humans. *Plos Biology, 16*(8). doi: 10.1371/journal.pbio.2006425
van Heijningen, C. A. A., de Visser, J., Zuidema, W., & ten Cate, C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Sciences of the United States of America, 106*(48), 20538-20543. doi: 10.1073/pnas.0908113106
Van Lancker-Sidtis, D., & Rallon, G. (2004). Tracking the incidence of formulaic expressions in everyday speech: methods for classification and verification. *Language & Communication, 24*(3), 207-240. doi: 10.1016/j.langcom.2004.02.003
Wilson, B., Smith, K., & Petkov, C. I. (2015). Mixed-complexity artificial grammar learning in humans and macaque monkeys: evaluating learning strategies. *European Journal of Neuroscience, 41*(5), 568-578. doi: 10.1111/ejn.12834
Wilson, M., Hauser, M., & Wrangham, R. (2001). Does participation in intergroup conflict depend on numerical assessment, range location, or rank for wild chimpanzees? *Animal Behaviour, 61*(6), 1203-1216.
Wilson, M. L., Boesch, C., Fruth, B., Furuichi, T., Gilby, I. C., Hashimoto, C., . . . Wrangham, R. W. (2014). Lethal aggression in Pan is better explained by adaptive strategies than human impacts. *Nature, 513*(7518), 414-+.
Wilson, M. L., Boesch, C., Furuichi, T., Gilby, I. C., Hashimoto, C., Hohmann, G., . . . Wrangham, R. W. (2012). Rates of lethal aggression in chimpanzees depend on the number of adult males rather than measures of human disturbance. *American Journal of Physical Anthropology, 147*, 305-305.
Wilson, M. L., Hauser, M. D., & Wrangham, R. W. (2007). Chimpanzees (Pan troglodytes) modify grouping and vocal behaviour in response to location-specific risk. *Behaviour, 144*, 1621-1653.
Zuberbuhler, K. (2018). Combinatorial capacities in primates. *Current Opinion in Behavioral Sciences, 21*, 161-169. doi: 10.1016/j.cobeha.2018.03.015
Zuberbühler, K. (2000). Referential labelling in Diana monkeys. *Animal Behaviour, 59*(5), 917-927.
Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour, 63*, 293-299.