Linguistic Laws and Compression in a Comparative Perspective: A Conceptual Review and Phylogenetic Test in Mammals

KANG, TARANDEEP,SINGH

How to cite:

KANG, TARANDEEP,SINGH (2021) *Linguistic Laws and Compression in a Comparative Perspective: A Conceptual Review and Phylogenetic Test in Mammals*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/13923/

Linguistic Laws and Compression in a Comparative Perspective: A Conceptual Review and Phylogenetic Test in Mammals

Tarandeep S. Kang

January 2021

A thesis submitted in fulfilment of the requirements for the degree of Master of Science by Research

in the Psychology Department

Durham University

1 List of Tables

2 Acknowledgements

Science is inherently collaborative, and this work could not have been completed without the help, advice and assistance of many people.

First of all, I owe an enormous amount to my truly amazing supervisory team: Zanna Clay Raphaela Heesen, Jake Dunn and Stuart Semple. Each of them has provided excellent mentorship and feedback throughout every stage of this work. Not least lots of wise advice on the many pitfalls and challenges of doing science. I thank them not only for the scientific advice they have given me, but for the encouragement, and above all the friendship. I couldn't have asked for a better team of supervisors, mentors and colleagues. Thank you all for keeping your sense of humour, and for a relentless attention to detail that has really allowed this work to shine.

I would like to thank all members of the ever-growing C4D Lab as well as members of staff in the Department of Anthropology. Thank you all for providing such wonderful feedback on the many talks and presentations I have given throughout the process of this work. Thanks also, for creating such a wonderfully friendly and supportive work environment, and for the many post-seminar drinks and conversations at White Church.

Research can often be a very isolating experience, but my wonderful friends both at Durham and at home never let me feel this way. Thank you to all members of the Trevelyan College MCR, for such wonderful discussions about work, and also for creating the perfect space to relax after a long day of work. Ed, Fliss, Yasmin, and everybody else, thank you for the

warm welcome right from the start. It is because of all of you that College felt like a home away from home.

My deepest thanks must go my dear friend Gaëlle. I am ever grateful for your long letters and for the fact that you were always willing to check-in and chat. Thank you for the unremitting faith in my abilities, and for all the kindness and love you have shown me. Je t'embrasse fort! I also sincerely thank Maria; you always know just what to say to make me laugh, and never fail to be genuinely interested in my work. I thank you also for being a constant source of wisdom and support to me. I owe and enormous debt of gratitude also to Garance; for reminding me that resting is as important as working, and all of the chats that helped to make lockdown bearable. Merci du fond du cœur! I am grateful also to Giamarco, Nanni and Mara, for the laughs, the support, and for generally being the most wonderful group of friends I could wish for. Thank you also to my best friend from home, David, for always encouraging me to keep going. I love each and every one of you!

I am a member of a tiny minority of postgraduate students with disabilities. In order to study and live away from home, I have depended upon the support of two wonderful carers. Thank you, Natalee and Lindsay, for supporting me throughout this process. Without you, I certainly couldn't have done any of this work!

Finally, I would like to thank my family; my mum, my dad, my brother, and my grandparents for pushing me to do the science I love. Without your love, patience and understanding, none of this would've been possible. This thesis is as much yours as mine.

3 General Abstract

Over the last several decades, the application of “Linguistic Laws” - statistical regularities underlying the structure of language- to studying human languages has exploded. These ideas, adopted from Information Theory, and quantitative linguistics, have been useful in helping to understand the evolution of the underlying structures of communicative systems. Moreover, since the publication of a seminal article in 2010, the field has taken a comparative approach to assess the degree of similarities and differences underlying the organisation of communication systems across the natural world. In this thesis, I begin by surveying the state of the field as it pertains to the study of linguistic laws and compression in nonhuman animal communication systems. I subsequently identify a number of theoretical and methodological gaps in the current literature and suggest ways in which these might be rectified to strengthen conclusions in future and enable the pursuit of novel theoretical questions. In the second chapter, I undertake a phylogenetically controlled analysis, which aims to demonstrate the extent of conformity to Zipf’s Law of Abbreviation in mammalian vocal repertoires. I test each individual repertoire, and then examine the entire collection of repertoires together. I find mixed evidence of conformity to the Law of Abbreviation, and conclude with some implications of this work, and future directions in which it might be extended.

4 Chapter 1: Current Caveats and Future Directions in Compression Research in Animal Communication

4.1 Abstract

For the last few decades, researchers have vigorously extended the application of human linguistic laws to nonhuman animal communication systems. This research has been enormously fruitful in highlighting that statistical regularities can be applied outside of human language and should be considered in a broader evolutionary context. In this review, I will highlight a number of caveats that need to be urgently addressed to aid the interpretation of comparative research in this area. Particularly, I focus on several methodological issues hampering comparative research on linguistic laws relating to analytical procedures and statistical methodology. Research has examined compression (the information-theoretic concept of providing strings as short as possible to encode information) in terms of shorter call duration; however, other features, such as amplitude may also be promising avenues of investigation. Finally, different behavioural contexts require further systematic investigation to gain a deeper understanding of selection pressures. To aid with this, I provide tentative context-based hypotheses, indicating the degree of compression to be found in some key context types. I conclude that although this research area has been highly productive, a number of gaps remain which should be addressed. Over the last few decades this line of research has provided new insights regarding the evolution of underlying structural regularity is in human language as well as no human communication and biological systems more generally. I hope that if the changes I suggest can be carried out, future studies can be strengthened even further.

Introduction

The search for universal principles in language represents a common practice in the field of linguistics. This builds on work of Chomsky, who proposed the theory of Universal Grammar (Chomsky & Smith, 2000) which proposes that "human languages, as superficially diverse as they are, share some fundamental similarities, and that these are attributable to innate principles unique to language: that deep down, there is only one human language" (Dąbrowska, 2015, p. 1). Although the concept of Universal Grammar has subsequently been strongly refuted (Evans & Levinson, 2009; Levinson & Evans, 2010; Tomasello, 2005, 2009, 2010), the search for statistical regularities underlying language and communication remains vigorous. A great deal of evidence for features being common to all languages can be drawn from the study of "linguistic laws" (Altmann & Gerlach, 2016) which - broadly defined - are common statistical patterns underlying language as well as multiple nonhuman animal communication systems. Each different statistical regularity can be considered a separate law.

The present work focusses particularly on two of these statistical laws which have been claimed to be universal properties of language: Zipf's Law of Abbreviation and Menzerath's Law. Zipf's Law of Abbreviation posits a negative relationship between the duration of a word and the frequency of its use (Zipf, 1936, 1949) whereas Menzerath's Law (Altmann, 1980; Menzerath, 1954; Teupenhayn & Altmann, 1984) states that 'the greater the whole, the smaller the parts', such that longer communicative sequences should be composed of shorter elements. Evidence of conformity to Zipf's Law of Abbreviation has so far been found in a variety of human languages, including written, signed, and spoken modalities. For example, there is evidence from dozens of languages that overall, shorter words are used more frequently (Börstell et al., 2016; Casas et al., 2019; Corral & Serra, 2020; Ferrer-i-Cancho &

Hernández-Fernández, 2013; Hernández-Fernández et al., 2016, 2019; King & Wedel, 2020; Mahowald et al., 2018; Piantadosi et al., 2011; Shu et al., 2003; Sigurd et al., 2004; Strauss et al., 2007, pp. 277–294; Teahan et al., 2000; Wang & Chen, 2015), as well as 986 translations of The Universal Declaration of Human Rights (Bentz & Ferrer-i-Cancho, 2015). Similarly, Menzerath's law -,the tendency to find that longer sentences are on average made of shorter subclauses, and longer words are made of shorter syllables,- also appears to be supported in a very wide variety of languages (Altmann, 2014; Araujo et al., 2020; Mačutek et al., 2019; Xu & He, 2020).

Widespread evidence of linguistic laws across human languages has also prompted linguists and psychologists to probe the evolutionary origins of such organisational principles, by making comparisons with communication systems of nonhuman animals (Bickerton, 2003; Hauser et al., 2002; Pinker, 1994). These studies have led to the recent hypothesis that instead of being language-specific, linguistic laws are better conceptualised as biologically-based principles that promote efficiency, and thus influence the design of communication systems across the natural world more generally (Ferrer-i-Cancho, et al., 2013). Ferrer-i-Cancho et al. argue that human languages, as well as other animal communication systems, have evolved under the selection pressure for coding efficiency. Coding efficiency is mathematically related (Ferrer-i-Cancho, 2016; Ferrer-i-Cancho et al., 2020) to the information-theoretic principle (Dębowski & Bentz, 2020) of compression, which posits that systems should be optimised to minimise code lengths. In the context of brevity, i.e. the shortening of the length of elements, these researchers suggest that the information theoretic concept of compression results in brevity, as typically observed in communication systems. Since an initial study of the vocal communication system of macaques (Semple et al., 2010) the vocal repertoires of a great number of species have subsequently been studied, and the majority of results from these studies conform to the patterns predicted by the linguistic laws. A summary of this evidence, comprising 15 studies documenting linguistic laws in animal communication is provided in Table 1. While there remain important and significant differences between animal communication systems and human language, for example the extent of referentiality (Scarantino & Clay, 2015; Scott-Phillips, 2016; Sievers & Gruber, 2016; Townsend & Manser, 2013; Wheeler & Fischer, 2015) or combinatoriality (Bowling & Fitch, 2015; Engesser & Townsend, 2019; Zuidema & de Boer, 2018) this growing body of work suggests that linguistic and non-linguistic communication systems share some fundamental design features relating to signal structure.

Table 1: A Summary of 15 Published Studies Examining Linguistic Laws of Compression in Nonhuman Animal Communication and Behaviour

Study	**Species**	**Behavioural Context** [a]	**Communicative Modality/Behaviour**	**Linguistic Law investigated: Zipf's Law of Abbreviation (ZLA)/ Menzerath's (ML)**
Ferrer-i-Cancho & Lusseau (2009)	Common bottlenose dolphin (*Tursiops truncatus*)	Wave breaking	Movement	ZLA - supported
Ferrer-i-Cancho & Hernández-Fernández (2013)	Common Marmoset (*Callithrix jacchus*), Golden-backed uakari (*Cacajao melanocephalus*) & Common Raven (*Corvus corax*)	Various	Vocal	ZLA- not supported
Semple et al., (2010)	Formosan Macaque (*Macaca cyclopis*)	Various	Vocal	ZLA- supported
Gustisson et al., (2016)	Gelada (*Theropithecus gelada*)	Various	Vocal	ZLA & ML supported
Hailman et al., (1985)	Black-capped chickadee (*Poecile atricapillus*)	'chick-a-dee' calls	Vocal	ZLA supported
Heesen et al., (2019)	Chimpanzee (*Pan troglodytes*)	Play	Gestural	ZLA supported in repertoire subset; ML supported in whole repertoire
Huang et al., (2020)	Cao Vit Gibbon (*Nomascus nasutus*), Western black-crested gibbon (*Nomascus concolor*)	Loud Morning Calls	Vocal	ZLA & ML supported.
Luo et al., (2013)	Black-bearded tomb bat (T*aphozous melanopogon*); Mexican free-tailed bat, (*Tadarida brasiliensis*); Greater horseshoe bat, (*Rhinolophus ferrumequinum*); Least horseshoe bat (*Rhinolophus pusillus*)	Short-Range Social Calls (Distress)	Vocal	ZLA supported in social but not distress contexts
Favaro et al., (2020)	African penguin (*Spheniscus demersus*)	Display Songs	Vocal	ZLA & ML both supported.
Demartsev et al., (2019)	Rock hyrax (*Procavia capensis*).	Long-Range Calls	Vocal	ZLA supported, but for amplitude rather than duration
Fedurek et al., (2017)	Chimpanzee (*Pan troglodytes*)	Pant Hoot	Vocal	ML supported
Safryghin, (2019)	Chimpanzee (*Pan troglodytes*)	Sexual Solicitation	Gestural	ZLA not supported; ML supported
Clink et al., (2020)	Bornean gibbon (*Hylobates muelleri*)	Song Bouts	Vocal	ZLA not supported; ML supported
Clink and Lau (2020)	Tarsier (*Tarsius spectrumgurskyae*), titi monkey (*Plecturocebus cupreus*) and gibbon (*Hylobates funereus*)	Duets	Vocal	ML Supported in only 4/8 call types
Watson et al., (2020)	Mountain gorilla (*Gorilla beringei beringei*)	Close Calls	Vocal	ZLA initially supported. However, once single-unit sequences were removed from analysis, patterns opposing the law were detected

[a] Whether or not the study focused on Compression in a particular behaviour or in a particular behavioural context, and if so, what this was

The evidence reviewed in Table 1 reveals that over the last several decades, there has been considerable interest in studying linguistic laws in animal communication. To date, researchers have mostly focused on exploring linguistic laws in vocal communication, with a heavy focus on primates. Researchers have convincingly demonstrated that patterns supporting linguistic laws can be found outside of human language, something which has given us new insights into the evolutionary origins of language, revealing that compression and brevity most likely share a common origin and evolutionary function across diverse taxa species. However, previous research has also demonstrated, that to some extent ,the degree of conformity to linguistic laws depends upon the size of the acoustic construct being investigated, as well as the context in which it is studied.

Nevertheless, although this research area has been both fruitful and inter-disciplinary, spanning evolutionary biology (Bezerra et al., 2011; Clink et al., 2020; Clink & Lau, 2020; Gustison et al., 2016), linguistics (Altmann, 1980; Menzerath, 1954; Teupenhayn & Altmann, 1984), anthropology (Heesen et al., 2019; Semple et al., In Prep), complexity science (Ferrer-i-Cancho, et al., 2013; Ferrer-i-Cancho et al., 2020; Ferrer-i-Cancho & Hernández-Fernández, 2013; Ferrer-i-Cancho & Lusseau, 2009), and comparative psychology (Watson et al., 2020), there appear to be several methodological issues with the research practices in the field which may hamper progress. Therefore, the aim of this review is to both review current findings of linguistic laws in animal communication as well as to describe some of the apparent methodological issues and some tentative solutions to them. Methodological challenges and new opportunities include considering how to define a call sequence, but also to move beyond the vocal modality when searching for signatures of compression in animal communication. I will also discuss the need for studies of the development of compression in communicative systems, as well as considering how compression may be studied in different structural features of the communicative signal, such as amplitude along with duration. Finally, I will discuss the need to examine individual level and contextual variation in studies of communicative compression. By taking account of these considerations, I hope that future researchers will be able to improve the robustness of their findings and gain a more detailed understanding of the kinds of factors and selection pressures which contribute to the evolution of compression. In general, I hope that by strengthening compression research in this way, researchers will be able to gain new and more detailed insights into the evolution of underlying structures of communication systems both in humans and other animals.

4.2 Defining Communicative Sequences

Studying compression in animal communication involves the study and analysis of communicative sequences. However, particularly in the vocal domain, there remains intense debate over what precisely constitutes a communicative sequence, and how best to analyse them (Bianco et al., 2019; Kershenbaum et al., 2016; Prat, 2019; Schlenker et al., 2016; Valletta et al., 2017). To improve the comparability of research on compression, greater consensus among researchers is needed over the nature of communicative sequences and approaches to analysis. In fact, all studies that I have reviewed here have used a manual approach to call and sequence classification. As well as introducing an element of subjectivity, this approach is often not biologically informed by conducting experiments to ensure such sequences are perceived as by the animals themselves, therefore potentially questioning their ecological validity. To date almost all studies of compression in animal communication systems have made use of standard statistical approaches to quantify similarities and differences between different call types. This includes discriminant function analysis (McGarigal et al., 2000) which uses combinations of acoustic variables to create groups of similar vocalisations (call types) depending on the degree of similarity in the various acoustic parameters. The resulting call types are then typically confirmed as correctly classified using expert judgement, and then classified into sequences based upon existing literature for the species under investigation (Gustison et al., 2016; Semple et al., 2010). However, a notable exception comes from a study of Bornean gibbon song (*Hylobates muelleri*) (Clink et al., 2020) in which machine learning approaches were used in concert with these aforementioned other methods. In this particular case the results are less likely to have been due to bias by human interpretation, and I therefore suggest that such approaches should be more widely adopted in future.

There is also the issue of higher order organisation of call sequences, which is to say that human experimenters may perceive a unit to be part of the sequence, even if this sequence may be instead viewed by the animal as a single unit within a superordinate structure. This type of hierarchical organisation is common in birdsong (Fehér et al., 2009; Hyland Bruno & Tchernichovski, 2019; Kang, 2017; Lipkind et al., 2013) as well as cetacean song (Allen et al., 2019; Cholewiak et al., 2013; Mercado & Handel, 2012). Given recent advances in the understanding of the complexity of animal sequences, we should therefore take a more careful approach when classifying vocal sequences (Kershenbaum et al., 2016). As the study of compression is inherently based upon the study of sequences, it is thus important for researchers to reach greater consensus as well as transparency in how they assign vocal sequences to improve the reliability, generalisability and validity of conclusions drawn. In their review, Kershenbaum et al., (2016) make several important suggestions regarding new approaches to the classification of sequences, in ways that might help to mitigate this human subjectivity. One promising suggestion is the broader use of automated computerised analysis and machine learning, as noted above (Fukuzawa et al., 2020; Noriega et al., 2019; Sainburg et al., 2020; Tchernichovski et al., 2000; Valletta et al., 2017). Although this approach has rarely been taken in a study of compression, it represents a promising new direction that is more data-driven rather than depending on manual human classification. This approach may also increase the generalisability of conclusions, because far larger corpora of data can be analysed far more rapidly than possible using typical manual methods. To some extent this human-based classification relies on subjective judgement which may be open to bias (Fitzpatrick et al., 2009). The more widespread adoption of such or the methodologies reduces this bias and may improve the reliability of classification judgements.

When studying communicative sequences, it is also important that the sequences under investigation are considered in their relevant biological context, such as through the use of playback experiments (Yasukawa & Bonnie, 2017). This approach has been taken with multiple species, to check that differences in sequence structure can be reliably perceived by conspecifics in natural circumstances. For instance, different sequence structures have been shown to reliably elicit different behaviours in the context of predation contexts in Chestnut-crowned babblers (*Pomatostomus ruficeps*) (Engesser et al., 2016, 2017, 2018, 2019); Guereza colobus monkeys (*Colobus guereza*) (Schel et al., 2010) and putty nosed monkeys (*Cercopithecus nictitans*) (Arnold & Zuberbühler, 2006) as well as many other species (Engesser & Townsend, 2019). Making broader use of playback experiments to initially check whether classified sequences can be perceived by conspecifics, may enhance understanding of the production-perception trade-off in sequence production as it relates to compression.

Another alternative approach to identifying communicative sequences involves taking advantage of the biomechanical limits on sound production (Torre et al., 2019). Torre and colleagues investigated compression by examining units known as ‘breath groups’ in human speech. Breath groups are defined as pauses in the production of the vocalisations of any duration required for breathing, or longer. Breath dynamics are well understood in human vocal production (Tsao & Weismer, 1997; Yunusova et al., 2005). Interestingly in a study in which underlying linguistic structure was not available, Torre and colleagues (2019) found stronger evidence of compression when examining human breath groups than when examining phonemes or syllables – suggesting that compression may be acting in other more ecologically and physiologically aspects of the communicative signal. In an ideal world, we would advocate for the broader usage of breath group analysis in not only human language, but also broader comparative studies of animal vocalisation. If this approach were to be adopted, it would enable truly unbiased studies of the evolution of communication and conversion. However, we should acknowledge that in naturalistic recordings of animal vocalisations, such breath groups can often be very difficult to detect.

I hope that some of these alternatives to classification of sequences might enable future research to be conducted with a greater focus on reliability and validity of findings. By automating the process of classification, such as through machine-learning, elements of the subjectivity introduced by human coders can be reduced. If we are to make meaningful progress on deepening our understanding of evolutionary drivers of that have given rise to compression, we must ensure that the sequences we are studying have relevance to our study species under investigation.

4.3 Beyond Signal Duration? Investigating Compression in Call Amplitude

Almost all studies to date examining evidence of compression in animal communication have focussed on signal duration (see Table 1). However, one promising new study (Demartsev et al., 2019) examined compression in the domain of call amplitude. The researchers found that in hyraxes the typical negative relationship between call duration and frequency of usage was not found however there was a negative relationship between call amplitude and frequency of usage, such that low amplitude calls were produced more often than higher amplitude ones. This suggests that for this species' repertoire may have been subject to selection pressure for efficiency regarding aptitude, but not necessarily for duration. In an effort to explain these rather unusual findings, the authors point out that selection for compression does not operate in a vacuum. Indeed, communicative signals often exhibit a balance of features demonstrating an optimum in relation to several selection pressures that have been exerted upon them. In this particular case, it is possible that brief but extremely loud vocalisations -as are common in the hyrax repertoire at short distances- may make them more energetically costly than longer but quieter calls. If we continue with their assumption that efficiency of coding is driven by minimising energetic costs, then it makes sense that in this particular context compression should perhaps act on amplitude more strongly than it does upon duration. The authors point out that such loud calls are often used across long distances and related situations. This would suggest that although there has been previous evidence of compression in long distance vocalisations, this particular category of vocalisations ought to be investigated both in relation to compression for amplitude as well as duration. To gain a deeper insight into influence on other acoustic features, such as amplitude, as well as signal duration. To focus only on duration may potentially limit the picture, because there is likely to be a trade-off between signal duration and amplitude (Nakano & Nagamine, 2019) in many contexts and species. If compression has acted upon amplitude in a similar way as it has upon duration, we would expect that the most common calls in a species' repertoire will also be the quietest. We would expect, in line with Demartsev and colleagues (2019), to find greater evidence of compression acting strongly upon amplitude in contexts which relate to high urgency vocalisations.

Evidence that long-distance calls are typically among the loudest in an animal's vocal repertoire suggests that long-distance calls are adapted to carry over long distances, whereby carrying distance are positively correlated with average signal amplitude (Gustison & Townsend, 2015; Mager & Walcott, 2007; Van Staaden & Römer, 1997). This means that carrying distance may be an important factor to investigate when investigating compression on amplitude. Indeed, both human and nonhuman animals increase the amplitude of their vocalisations as carrying distance increases (Brumm & Slater, 2006; Johnson et al., 1981). The finding that increased amplitude is related to greater carrying distance, would suggest that long-distance calls might be expected to be compressed in terms of duration, as has been supported by previous research (Fedurek et al., 2017; Huang et al., 2020) but, they may not be compressed in terms of amplitude.

Detailed studies of the energetic costs of vocalising have only been conducted in a few animal species including bottlenose dolphins (*Tursiops truncates*) (Noren et al., 2013), zebra finches (*Taeniopygia guttata*) (Zollinger et al., 2011) and humans (Russell et al., 1998). These studies have shown that the energetic cost of vocalisation increases as organisations are lengthened and their amplitude is increased. To enable cross-species comparisons of compression on signal amplitude, detailed studies of energy expenditure and metabolic cost of signalling must first be conducted in a greater variety of species.

To date, only energetic expenditure has been considered as a potential cost when considering the trade-off between compression and redundancy (Ferrer-i-Cancho, et al., 2013; Gustison et al., 2016; Semple et al., 2010). However, eavesdropping by predators (Magrath et al., 2015) may also be another relevant cost related to the selection pressures that shape the degree of compression found in communicative repertoires. Costly effects of eavesdropping by predators may explain the prevalence of acoustic crypsis in multiple species (Dunlop & Noad, 2016; Parks et al., 2019; Schmidt & Belinsky, 2013). When animals use crypsis, communicative signals are usually shorter in duration and have reduced amplitude. Low amplitude signals may function to reduce predation and risk of attracting sexual competitors (Dabelsteen et al., 1998; Nakano et al., 2009; Padilla de la Torre & McElligott, 2017; Reichard & Anderson, 2015; Reichard & Welklin, 2015; Vargas-Castro et al., 2017). A typical example of the use of this strategy comes from a study of southern right whales (*Eubalaena australis*) (Nielsen et al., 2019), in which it was found that mothers with calves produced vocalisations less frequently in a way that is less likely to attract the attention of predators; mainly killer whales (*Orcinus orca*). This is because predators can often eavesdrop on prey vocalisations, to gain access to information about their presence, location, and perhaps even relative abundance. In reducing both the duration and amplitude of vocalisation when predation risk is high, prey species will reduce their relative risk of being eaten, and this is therefore likely to have resulted in selection for brevity as well as low amplitude of signals in contexts where predation risk is especially high. This is particularly the case with alarm calls; these are often produced at high amplitude and they have a putative predator deterrence function (Zuberbühler et al., 1999). In contrast to these high amplitude vocalisations soft contact calls are likely to be relatively cryptic due to the significantly reduced altitude. Animals must clearly balance trade-offs with energetic expenditure when adding redundant information to vocalisations to ensure that they are still relatively easy to disambiguate for receivers. However, well these energetic costs are relatively well understood, there is another kind of cost to vocalisations which contain excessive redundancy. When animals use acoustic crypsis, they tend to reduce the duration of the vocalisations, in effect compressing them, to minimise conspicuousness to predators (Brooker & Wong, 2020; Nielsen et al., 2019; Parks et al., 2019; Ruxton, 2009; Ruxton et al., 2018). There is however a general trade-off here, given that the loudness of alarm calls is also believed to be part of the predator deterrent function. In this way, acoustic crevices may provide another valuable insight to compression researchers; yet once again a single strategy is unlikely to apply to all species. This is why we must once again take full account of habitat ecology and other unique features of the species' ecological niche when studying compression in its communication system. In general, however, the study of acoustic crypsis presents another way of assessing the costs to vocalising individuals, and therefore this is likely to be important. Thus, if compression research investigated other costs to excessive lengthening of vocalisations such as predation in the context of crypsis, we could gain novel insights about selection for efficiency, and the role of other costs besides simply energetic expenditure. To date, researchers have only considered the cost of energetic expenditure regarding signal production, efficiency, and compression. However, based on acoustic crypsis research other factors such as conspicuousness to predators may also drive selection pressures for compression in ecological contexts where predation risk is heightened. Importantly, this selection pressure may act in addition to,- or indeed instead of,- selection for minimisation of energetic costs. Therefore, it is possible for other types of costs besides energy, to potentially act to favour greater compression in animal communication systems.

4.4 Multimodal approaches to understanding compression in animal communication

As is apparent in the literature reviewed here, almost all research on compression in animal communication has thus focused on vocalisations – see Table 1 (except Heesen et al. 2019, Safryghyn, 2019 and Lusseau & Ferrer-i-Cancho, 2009). This is despite evidence that both human language and many other animal communication systems are themselves strong multi-modal (Holler & Levinson, 2019; Kita et al., 2017; Levinson & Holler, 2014; Sekine et al., 2015; So et al., 2009). Understanding the nature of compression within a multi-modal framework is therefore a relevant future direction in compression research.

For instance, in a recent review, it has been hypothesised that gestural and vocal communication may have co-evolved (Levinson & Holler, 2014) and that the interplay between different communicative modalities is complex, and best viewed in a stratified manner. Given that vocal communication appears to interact with other non-verbal behaviours, in recent years we have seen an explosion of interest in the application of multimodal approaches to nonhuman communication systems (Fröhlich & van Schaik, 2018; Higham & Hebets, 2013; Partan & Marler, 1999; Prieur et al., 2020; Slocombe et al., 2011). However, thus far, most studies of animal multimodality have been restricted to the interaction of gestural and vocal domains (Fröhlich & van Schaik, 2018; Genty, 2019; Genty et al., 2014, 2015) despite the fact that animal species are known to communicate in a variety of other modalities, including olfaction, (Bossert & Wilson, 1963), vibration (Markl, 1983), facial expressions and body postures (Chambers & Mogil, 2015; Davila Ross et al., 2008; Hinde & Rowell, 1962; Teufel et al., 2010; Waller et al., 2015). Compression research has been strongly biased on vocal communication (see Table 1) and even those studies which have explored gestures (Heesen et al., 2019) examined them as a single modality. The focus on unimodal signals to date may limit our understanding of the evolution of compression. By taking a multimodal approach we can begin to ask questions about the ways in which multimodal displays might add redundant information and therefore reduce the tendency and need for compression

While there have yet to be studies of compression in multimodal systems, two studies have investigated compression in gestural systems of chimpanzees during play (Heesen et al.,

2019) and sexual solicitation (Safryghin, 2019) contexts. This latter study has shown that the predictions of Zipf's Law of Abbreviation was not necessarily met in sexual solicitation contexts, in and out in an analysis of gestural communication. Results consistent with compression at the higher level of organisation, were found regarding Menzerath's Law consistent with compression at the higher level of organisation. These studies highlight the fact that behavioural contexts and the ecological environment may have a strong influence on the overall degree of compression found in a gestural communication system. Moreover, the degree of compression is not uniform at different levels of organisation. It is notable that Safrgyhn (2019) did not find evidence supporting the flow of brevity in sexual solicitation context and neither did even when the entire repertoire was analysed. However, Heesen et al., (2019) did find evidence of compression in a subset of chimpanzee play gestures. How might we explain this pattern of results when compared with the far larger literature examining compression in vocal communication? It is possible that compression has been more strongly selected for in vocal communication than other modalities. This is because vocal signals are likely to become increasingly compressed as their carrying distance increases (Ferrer-i-Cancho, et al., 2013) and given that gestural and other non-vocal signals typically have shorter broadcast distances, they may have been less strongly influenced by selection pressure for compression. Perhaps, it is not that gestural signals are less compressed overall than their vocal counterparts, but merely that different parts of the repertoire are differentially impacted (Heesen et al., 2019). This type of pattern has been also found in the vocal domain, with the degree of compression being influenced by the size of the construct being analysed. Perhaps it is time therefore to take a more nuanced approach to the study of compression outside of the vocal domain because it is becoming increasingly obvious that the influences of selection pressures for compression may have been different to vocalisations. It would be fruitful for future researchers to explicitly examine the influences of carrying distance in future studies of compression to help reduce the ambiguity among different modalities.

Furthermore, multi-modal signals often contain redundant information provided by the different modalities they are combining, and given that the function of this redundancy can often be to reduce ambiguity (Hebets et al., 2016; Johnstone, 1996), it is possible that compression may be reduced in multimodal signalling contexts; in other words when only one signalling channel is employed, compression should generally be evident as long as it is possible for receivers to disambiguate the signal type. However, where modalities are combined across multiple channels at once, each display may reinforce the others, leading to a degree of redundancy, and appear less compressed. This is because the combination of different modalities might add extra layers of information content to the overall signal. Given that there can be multiple layers of information, with each acting to reinforce the other, there may be less overall pressure for ambiguity minimisation, leading to potentially stronger selection for efficiency. This is because redundancy is typically used to help to disambiguate signals for receivers (Plotkin & Nowak, 2000), and in this hypothetical multimodal signal, each there is no such pressure for disambiguation, given the multiple backup signals.

In sum, despite the multimodal nature of animal communication, researchers have only recently begun to integrate approaches to the study of multiple communicative modalities (Fröhlich et al., 2019). In compression research there has been no such move yet towards multimodal integration, and indeed only two published (Ferrer-i-Cancho & Lusseau, 2009; Heesen et al., 2019) and one unpublished (Safryghin, 2019) study has yet examined evidence of compression outside of the vocal domain. However, it is important to take account of different signal types, from gestures, vocalisations, facial expressions, visual signals, and chemical signals. In the case of several of these modalities we really cannot be certain of the ways in which compression may have acted upon their communicative function. Including multiple communicative modalities may enable a more holistic picture of how compression acts on communication systems.

4.5 Methodological Caveats: Pseudoreplication and Comparing Across Contexts

Many datasets in animal behaviour research contain clustered observations, wherein multiple observations are collected from each individual (Clay & Zuberbühler, 2009, 2011a, 2011b; Townsend et al., 2008). Though not in itself a problem, the failure to acknowledge and correctly model sources of nonindependence can lead to serious issues of pseudoreplication as has been repeatedly highlighted in the field of animal behaviour and communication (Freeberg & Lucas, 2009; Hurlbert, 1984, 2009; Waller et al., 2013). The compression studies discussed above are no exception from this issue. Indeed, several studies have not taken account of subject identities in the statistical analyses (Ferrer-i-Cancho & Hernández-Fernández, 2013; Ferrer-i-Cancho & Lusseau, 2009; Heesen et al., 2019; Semple et al., 2010). However, in others, such as by Gustison et al. (2016) and Watson et al. (2020) individual identity was controlled for as well as behavioural context of production. In these cases, it is possible to examine individual differences as well as contextual variation, and if samples are sufficiently large and diverse, potentially test other hypotheses regarding how compression might develop across the lifespan. Although no such controls were implemented by Ferrer-i-Cancho & Hernández-Fernández (2013) and Semple et al., (2010), it should be noted that both of these studies were analyses of secondary data rather than original sources.

Recent advances in statistical methods, including the use of hierarchical and generalised linear mixed models, mean that clustered datasets from the same individuals can be effectively statistically modelled, with subject identity included as a random factor and potentially interacting with main effects via the inclusion of random slopes. To avoid pseudoreplication, a simulation-based approach has also been advocated (Garamszegi, 2016) which involves repeatedly assigning random subject identities to the samples and then analysing the data using a mixed model with averaged values for each randomly assigned identity. However, a newer simulation (Gratton & Mundry, 2019) suggests that this kind of approach performs just as poorly as does implementing no controls for identity whatsoever. In both cases, type-one error rates are likely to be unacceptably high. Although this applies to much research in animal communication in general, it is likely that taking account of this will be useful for compression research in the future. Specifically, although identifying vocalising individuals in noisy and clustered environments can be challenging, the inclusion of individual identity can bring significant new knowledge and theoretical advances. This is why Gratton and Mundry (2019) suggest that the inclusion of subject identity is of critical importance. Indeed, they advise that if it is not possible, it is preferred that subject identities should be ignored all together, rather than being substituted for by a highly error prone method. The particular advantage of including individual identity in modelling as Gustison and colleagues (Gustison et al., 2016) have done, is that we are able to also investigate Menzerath's Law rather than being constrained to Zipf's Law of Abbreviation. Although no individual-level variation was found in the extent to which call sequences obey the linguistic laws by Gustison et al., (2016) this approach may yield relevant insights if extended to a broader range of species with differing social structures and ecology.

The adoption of more careful statistical modelling strategies might enable us to answer entirely new questions or profound theoretical interest. Although there has been no research examining the ontogeny of compression, accounting for subject identities and age in modelling could enable researchers to detect and model the ontogenetic pathways of communicative compression and its differential usage across time. In particular, it is possible that the degree of compression increases as individuals age. In this way we would be able to extend recent work using a miniature lexicon, which shows that compression images naturally in communication systems, as a function of the degree of experience with the system (Kanwal et al., 2017). Using this approach would enable us to determine whether, and to what extent, the emergence of compression in animal and human communication systems is dependent on developmental experience.

4.6 Context Effects on Compression

A related challenge in animal communication compression research is lack of attention to the potential effects of context on the degree of compression. Several studies have either investigated compression only in one context or taken no account of different contexts from which recordings were taken, in subsequent analyses (Ferrer-i-Cancho & Hernández-Fernández, 2013; Hailman et al., 1985; Semple et al., 2010) (see the Context column of Table 1). This is also likely to influence interpretation of results, because different social and ecological environments are likely to impose different costs on the signaller as well as the receiver. As these costs will vary, it is likely that selective pressure for compression will vary accordingly. It may well be that a given animal species may exhibit differing degrees of compression and its communicative signals, depending on social and ecological context. Primarily these are likely to be related to ecological factors, and the evolutionary urgency of the calling context.

In order to develop the study of compression in animal communication and its variation across behavioural contexts, a number of different contexts could be investigated in future. For example, comparing degrees of compression found in affiliative contexts versus agonistic ones. As discussed previously, the principal aim of compression research appears to be to minimise energy expenditure in communicative act. In this way, play related contexts usually occur when there is an excess of both time and energy (Held & Špinka, 2011) thus compression may be potentially relaxed as there is superfluous energy, and no particular need to conserve it. Vocalisations can also be important in initiating play and given that they can sometimes escalate to violence, they may be important in differentiating exit of play from acts of aggression. In this case there may actually be selection against compression because firstly as discussed above energy is not in short supply, and also ambiguity must be minimised so that violence is not confused for play.

Alarm calls appear to have undergone strong selection to provide reliable information about predator class or location (Zuberbühler, 2009). Their apparent ubiquity makes them another insightful behavioural context in which to study compression. There is also evidence of redundancy in alarm contexts; for example wild chimpanzees (*Pan troglodytes schweinfurthii*) (Crockford et al., 2012) and Thomas langurs (*Presbytis thomasi*) (Wich & de Vries, 2006) have been found to continue informing ignorant group members of danger similar patterns have been observed in langurs (Wich & de Vries, 2006). This argument is advanced by evidence from a simulation study (Nowak & Krakauer, 1999) that showed that information can never be perfectly transmitted in a way that is free of errors; especially in urgent contexts, where survival is at risk, miscommunication carries a very high cost indeed. Clarity is essential to survival (Pinker, 2000). Therefore, in this context redundancy may have some adaptive benefits that compression does not. However, it is possible to also argue for precisely the opposite explanation. This is because, as Zuberbühler (2009) notes, alarm calls also produce costs on the signaller given that they are conspicuous and thus may actually increase the risk of production, rather than deterring the predator. That these two competing explanations appear equally probable at first sight, only underlines the need for empirical testing of the degree of compression evidenced by animal communicative repertoire in this context.

Let us now turn our attention to identify hypotheses regarding the degree of compression we might find in copulatory and sexual contexts. Males (Bateman, 1948; Trivers, 1972) and females (Andersson & Simmons, 2006; Hare & Simmons, 2019; Jennions & Petrie, 2000; Parker, 2006) of many animal species are hypothesised to actively compete for access to the highest quality mating partners. As discussed previously, a high degree of compression can help to avoid eavesdropping by conspecifics and predators; in the mating context it is also likely to be relevant, given that vocalisations can be used in mate attraction (Langmore & Davies, 1997), and may even be inhibited in some species, in a way that is hypothesised to help avoid mate-poaching (Townsend et al., 2008). On the other hand, it may also be beneficial to include redundant information because ambiguity can be costly (Plotkin & Nowak, 2000), especially in evolutionarily urgent contexts. In some species such as the bonobo (*Pan paniscus*) sex does not only serve a reproductive function, but can help with relationship maintenance and formation (Clay et al., 2011; Clay & de Waal, 2015; Clay & Zuberbühler, 2011).

In travel related contexts, animals often produce signals that are intended to be broadcast over long distances which function to locate and recruit group members during travel (Byrne, 1981; Gruber & Zuberbühler, 2013; Schamberg et al., 2016, 2017). Given this, as well as recent evidence that pant hoots, a form of long-distance vocalisation produced by chimpanzees, are aligned with Zipf's Law of Abbreviation (Fedurek et al., 2017), and the fact that they are specifically designed to be broadcast of the long distances, travel related vocalisations should be expected to conform to the Law of Abbreviation and be highly compressed. Specifically, this is because compressed signals tend to have simplified acoustic structures, meaning that they are less likely to be degraded by broadcast distance, and reverberation off solid structures (Ferrer-i-Cancho, et al., 2013).

4.7 Conclusion

In sum, although still in its infancy, the study of compression in nonhuman animal communication systems has already been fruitful. In particular because the research of the last few decades has demonstrated that linguistic laws are not only laws of human language but may well be efficiency principles guiding biological communication systems more broadly. However, despite this progress, significant theoretical and methodological caveats remain. These include more carefully evaluating the nature of communicative sequences to ensure that they are in the ecological context of investigation, and not biased by human intuition. The interdisciplinary field of compression research will also benefit from embracing the study of multimodality as there are likely to be significant differences in the degree of compression found in communication depending upon communicative modalities. The extent of conformity to linguistic laws might also depend upon whether modalities are considered in concert or alone, just that new advances in modelling practices might allow future researchers to thoroughly investigate developmental trajectories of the emergence of compression. It will also be fruitful to begin investigating the developmental pathway of compression in diverse animal species, to look for commonalities and differences of developmental experience, and ecological environment on the degree of compression evidenced in mature individuals. The field will also benefit from reducing the current focus on signal duration to also considering amplitude. This is because, once again the selection pressures on these two facets of communicative signals are likely to have been moderately different. Finally, certain methodological issues, namely pseudoreplication, and a lack of comparison across contexts, remain prevalent in the field, as well as other areas of comparative cognition and animal behaviour. These must be addressed in order to maximise the interpretability of experimental findings. The advances of the last few decades in comparative psychology, and in particular in the application of methodologies from quantitative linguistics, have begun to illustrate commonalities in the structure of human language, and nonhuman animal communication systems. This endeavour has shown that on the structural level, human language and animal communication have much in common, in terms of how they are organised. Linguistic laws were first developed in the context of human language; however recent research has shown that they may in fact be even more general as principles of biological information systems more broadly. By taking account of the recommendations made here, future endeavours to examine the fundamental organisational principles of nonhuman animal communication can provide further novel insights.

Another concluding insight regarding the state of compression research, is that a great deal of the currently available data comes from studies of primates (see Table 1). Although this may be partly explained by their close phylogenetic relatedness to humans, taxonomic biases such as this prevents us attaining a full picture of the evolution of compression and efficiency coding in communication systems. To better understand the selection pressures that have led to evidence of efficiency of coding and brevity across languages and animal communication systems requires a less primate-centric approach. Only by taking such a broader perspective and examining questions of the convergent evolution of compression will we be able to enrichen our understanding of the common selection pressures that have shaped compression more generally. Such tests will also allow us to identify whether or not compression is a universal principle of animal behaviour as has been previously supposed (Ferrer-i-Cancho, et al., 2013). In the next chapter of the thesis I will thus address these issues by presenting a quantitative analysis of mammalian vocal repertoires in an effort to assess conformity to Zipf's Law of Abbreviation beyond the primates. It is hoped that by providing such data others will be encouraged to also investigate linguistic laws, compression and efficiency of coding in a broader range of species and contexts.

We intend to submit the following study for publication in an academic journal. Thus, we have formatted it as we might a typical research paper.

5 Chapter 2:
A Comparative Analysis of Zipf's Law of Abbreviation in Mammal Vocal Repertoires

5.1 Abstract

Many human languages appear to conform to Zipf's Law of Abbreviation, which posits a negative relationship between the duration of words and the frequency of their usage; this is broadly considered to be the result of the process of compression, which is widely discussed in the information theory literature. By comparison, evidence for brevity within animal vocal communication systems remain mixed, partly due to a paucity of research. Thus far, only one broader taxonomic analysis has yet been conducted to investigate evidence for compression in primate vocal repertoires. In the present study, I build upon and extend this analysis by conducting an even broader-scale comparative analysis of the vocal repertoires of 50 diverse mammal species. Mammals are particularly interesting for investigating brevity because they typically have large vocal repertoires; most notably bats. Using data assembled from available published vocal repertoires, I conducted Spearman correlations complemented by a permutation-based procedure to examine the frequency of usage of each call type in the repertoire and its duration. Finally, I ran a phylogenetically-controlled analysis to control for potential effects of phylogenetic relatedness between the species under investigation. Results revealed that six mammal species in our sample showed evidence of compression in their vocal repertoires. Intriguingly, two other species in our sample showed significant positive correlations between call frequency and call duration, suggesting that selection may also favour acoustic redundancy rather than brevity in some cases, potentially when signals need

to be easily disambiguated by receivers. In the remaining N=39 species, I could not detect a significant relationship between call frequency and call duration. However, by using phylogenetically controlled analysis, I found a significant positive relationship between the presence of vocal compression and species phylogenetic relatedness, suggesting that genetic relatedness mediates the tendency to find compression in the vocal repertoires of mammals. Overall, our results suggest that while there may be selection for compression in mammalian vocal signalling in certain contexts and species, the production of signals which could not be easily discriminated might have incurred greater costs than the energetic costs of redundant vocalisations in others, therefore leading to anti-compression. In sum, while finding evidence for brevity in some mammalian vocal repertoires, our results also suggest that selection for brevity might not be the only evolutionarily viable strategy in mammal acoustic communication. In general, I demonstrate that broad-scale phylogenetic analysis of animal vocal repertoire can help to better understand the evolution of efficiency in communication systems.

5.2 Introduction

Beginning in the 1930s, the pioneering work of George Kingsley Zipf (Zipf, 1936, 1949) illustrated compression to be a general design feature of human language by demonstrating a negative relationship between the duration of words and their frequency of use. In subsequent decades, Zipf's Law of Abbreviation (or Brevity; both terms are used interchangeably in the literature), has become one of many such, "linguistic laws" (Chacoma & Zanette, 2020; Corral & García del Muro, 2020; Egghe, 2007). In fact, the study of these patterns is now a cornerstone of quantitative linguistics (Altmann & Gerlach, 2016). Conformity to the law is found in an enormous variety of human languages, and is indeed now believed to be a human language universal (Bentz & Ferrer-i-Cancho, 2015). The concept of brevity is related to the information-theoretic principles of compression and efficiency, which suggest that systems should evolve to be optimised to minimise code length (Ferrer-i-Cancho et al., 2020). The particular function of this is likely to be an increased economisation of energetic expenditure (Ferrer-i-Cancho, et al., 2013; Gustison et al., 2016; Zipf, 1949). In particular this relates to the tendency to assign strings as short as possible when representing information, while also minimising the likelihood of confusion by receivers (Cover & Thomas, 2006). Human languages appear efficient in this regard (Gibson et al., 2019) with evidence of Zipf's Law of Abbreviation so far having been demonstrated in corpora-based analyses of a variety of languages (Börstell et al., 2016; Casas et al., 2019; Ferrer-i-Cancho & Hernández-Fernández, 2013; Hernández-Fernández et al., 2016, 2019; King & Wedel, 2020; Mahowald et al., 2018; Piantadosi et al., 2011; Strauss et al., 2007, pp. 277–294; Tamaoka & Kiyama, 2013; Teahan et al., 2000; Wang & Chen, 2015) as well as 986 individual translations of The Universal Declaration of Human Rights (Bentz & Ferrer-i-Cancho, 2015).

Widespread evidence of linguistic laws across human languages has also prompted linguists and psychologists to probe the evolutionary origins of such organisational principles, by making comparisons with communication systems of nonhuman animals (Bickerton, 2003; Hauser et al., 2002; Pinker, 1994). Although the statistical regularities of Zipf's Law of Abbreviation were initially conceptualised as being human language-specific, subsequent research now suggests that the principles of compression shape the organisation of non-human animal communication (Gustison et al., 2016), as well as behaviour more generally (Ferrer-i-Cancho, et al., 2013) for instance, behavioural displays such as wave-breaking in dolphins can conform to patterns predicted by compression theory (Ferrer-i-Cancho & Lusseau, 2009). In this respect, these principles may in fact be even more general principles of the organisation of information in biological systems, something which has been supported by evidence of compression in the structure of proteins and genes (Caetano-Anollés et al., 2017; Ferrer-i-Cancho & Forns, 2009; Li, 2012; Shahzad et al., 2015). It should be noted, that at this stage I cannot say anything about the evolution of the cognitive capacities for brevity, and that this may indeed only be an emergent structural property of information systems more generally.

Thus far, due to the phylogenetic closeness to humans, research investigating linguistic laws in animal communication has mostly focused on primates (see table 1 of chapter 1 of this Thesis). For instance, a forthcoming study (Semple et al., In Prep) analysed the relationships between the frequency of call usage and call duration in a broad group of 28 primate species. The study found mixed evidence for brevity species in this taxon. This suggests that further broad taxonomic analyses might yield interesting and novel insights However, research into the presence of Zipf's Law of Abbreviation in more diverse mammal species would provide an even better understanding of the breadth of compression in animal communication beyond the primate lineage. Among mammals, bats (Kerth, 2008; Pfalzer & Kusch, 2003) and cetaceans (Janik, 2014; Marino et al., 2007) particularly, have complex social lives, and large vocal repertoires which warrant further investigation. Indeed, some research has already provided evidence for compression in the vocal repertoires of Black-bearded tomb bats (T*aphozous melanopogon*); Mexican free-tailed bats, (*Tadarida brasiliensis*); Greater horseshoe bats, (*Rhinolophus ferrumequinum*); and least horseshoe bats, (*Rhinolophus pusillus*) (Luo et al., 2013). In the current study, I conducted a broader analysis of mammal species more generally. I intend to build upon the study by Semple et al (In Prep) by examining evidence for Zipf's Law of Abbreviation in a broad-scale taxonomic analysis of the vocal repertoires of 50 mammal species using available published data.

As well as examining evidence for brevity within species' repertoires, I also conducted a modified t-test to take account of the phylogenetic relatedness of the species and to demonstrate whether this additional moderator can help to explain the distribution of compression results across the species of our dataset. I expect those species that are more closely related to show greater similarity in their tendency to show compression. This type of taxonomic comparative analysis is more useful than simply focusing on individual species or taxa. Although it has been demonstrated that there can be selection for compression, research has also shown evidence for the opposite, redundancy, which suggests that selection pressure for efficiency can be outweighed by other pressures, such as to minimise the likelihood of ambiguity in communication. This type of cross-species analysis is especially powerful in this regard because although individual species may not yield results appearing compatible with the law, by examining a greater number of species at once, for example at the level of the taxon I unveil the distribution of compression across species more broadly. At this juncture it is not possible to confidently hypothesise as to which species will conform to the law. Given previous data (Semple et al., In Prep) I do however hypothesise that not all species in the group will display evidence of compression.

5.3 Methods

5.3.1 Data Collection

Drawing on a systematic literature search using *Web of Science*, I identified published sources in the literature describing vocal repertoires of mammal species. My primary search term was 'vocal*' in combination with each mammal genus name. I subsequently searched the references within such studies, to identify further potentially relevant studies. I limited my analysis to studies which reported both the frequency of usage of individual call types as well as the average call duration. One additional limiting factor was that the studies must have reported at least five different call types, as it has been shown that at least call types are needed in order for correlation results to be reliable and interpretable (Ferrer-i-Cancho & Hernández-Fernández, 2013). Finally, this sample of five different call types per species must only have consisted of adult calls. The reason for eliminating immature vocalisations is because these were seldomly reported and maturing vocal repertoires may show differing properties of those from adults. Initially, I found 73 repertoire studies of diverse mammals, however following these exclusions my final sample consisted of 48 studies comprising of 50 genera. This dataset results from the fact that some studies covered multiple species, and several species were covered in more than one study. In cases in which more than one repertoire study per species we treated each species repertoire separately. Importantly, although individual studies often differed in their classifications of call types, but the majority of studies made use of discriminant function analysis and cluster analysis to quantify similarities and differences between call types on various acoustic parameters. This type of methodology is commonly used in the study of animal bioacoustics and communication. We also classified whether repertoire studies were conducted with individuals in captivity or the wild.

5.3.2 Analyses

Quantitative analyses were conducted in R Version 4.02 (R Core Team, 2020). The code for all analyses is available in the supplementary material. For each species, I ran a bivariate Spearman Rank correlation between call duration and frequency of use using the correlation package (Makowski et al., 2020) which is part of the ‘easystats’ suite (Lüdecke et al., 2019). For cases in which the analysis indicated ties, I followed up with a nonparametric Kendall

Rank correlation. In both cases I calculated 95% confidence intervals using the Fieller correction (Fieller et al., 1957) which follows current best practice (Bishara & Hittner, 2017). We then used a permutation-based procedure (Ferrer-i-Cancho, et al., 2013) with custom written code (Heesen et al., 2019). This allows us to address criticisms levelled by other researchers That correlation results could be artifacts given the inherent mathematical relationship between d and f (Solé, 2010). The use of this permutation procedure helps us to overcome this. This method is described further detail below.

Following Heesen et al., (2019) and Safryghyn (2019), I first calculated for each species the observed mean code length (L) (the mean duration of each call in the vocal repertoire) following Equation 1, where n is the number of call types in the repertoire, p_i is the normalised probability of occurrence of the i-th most probable call type (calculated as the number of that call type recorded, divided by the total number of all calls recorded) and e_i is the mean duration of that call type.

$$L = \sum_{i=1}^{n} p_i e_i \qquad \text{[Equation 1]}$$

To test for compression and whether Zipf's law is applicable to mammal vocalisations I used a permutation test assessing whether L was significantly small . Following Heesen et al. (2019) and Safryghyn (2019), I created a control distribution of L (L'), which was defined by a permutation function π (i), see Equation 2. In this control distribution, I shuffled all potential values of mean call duration and call frequency and then checked whether my L value indeed remained significantly small within this new control distribution (i.e., an extreme value within the smallest 5% values of L'). I calculated the left p-value to assess whether the repertoire is significantly small by dividing the number of permutations where L’≤L by the number of total permutations (10^5).

$$L' = \sum_{i=1}^{n} p_i e_{\pi[i]} \qquad \text{[Equation 2]}$$

Finally, to control for phylogenetic relatedness, we used data from a mammalian supertree (Bininda-Emonds et al., 2007) to perform a phylogenetic paired t-test (Lindenfors et al., 2010) on the resulting compression values. We used the phyools (Revell, 2012) and ape (Paradis & Schliep, 2019) packages to conduct the test and for all manipulation of phylogenetic trees. This enabled us to detect whether the degree of phylogenetic relatedness between the species related to the degree of compression found in their vocal repertoires. We hypothesise that not all species in our group will display evidence of compression, but that fellow genetic relatedness will be a significant mediator of any effects. The code for all analyses is available in supplementary material.

5.4 Results

5.4.1 Do Mammalian Vocal Repertoires Follow Zipf's Law of Abbreviation?

A set of Spearman Correlations demonstrated that the repertoires of six of the 50 species (those coloured green in Table 2) in the dataset conformed to Zipf’s Law of Abbreviation, whereby mean call type duration was significantly negatively related to frequency of use

(Humpback Whale (*Megaptera novaeangliae*), Yellow squirrel (*Spermophilus fulvus*), Northern treeshrew (*Tupaia belangeri*), Dhole (*Cuon alpinus*), Bearded seal (*Erigignathus barbatus*), and the Natterer's bat (*Myotis nattereri*)) (see Table 2 for a summary of statistical results). I then used a permutation-based test which showed that for 38 of the 50 species, mean call duration, *L*, was significantly smaller than the mean of the permuted distribution of *L'* (Table 2). For all six of the species for which a significantly negative Pearson correlation was found, there was a significant correlation between *D (duration)* and *f (frequency of usage within the repertoire)* after permutation, demonstrating that the main effect was not only an artefact of analysing mean call type duration (Semple et al., 2013). In addition, as well as showing significant negative correlations in six species, I also found evidence of a positive correlation between frequency of use and mean call type duration in four other species (those coloured red in Table 2). However, only two of these (presented in bold type in Table 2) reached statistical significance, both of which I confirmed were not a result of the analysis procedure, as they were also confirmed by the permutation analysis.

Table 2: A Summary of Statistical Tests of Compression in Mammal Vocal Repertoires

Species	**Vocal Repertoire Size**	**N Subjects**	**rs [95% CI][a]**	***p*[b]**	**rt [95% CI][c]**	***p*[d]**	**Study Context[e]**	**Settting[f]**	**Reference**
Humpback Whale (*Megaptera novaeangliae*)	34	Maximum 24 Groups	**r -.55 [-.75, -.26]**	**< .001**			Social	Wild	Dunlop et al., (2007)
Yellow squirrel (*Spermophilus fulvus*)	9	59	**-.83 [-.98, -.05]**	**.042**			Not Reported	Wild	Matrosova et al., (Matrosova et al., 2012)
Northern treeshrew (*Tupaia belangeri*)	9	9	**-.70 [-.93, -.07]**	**.035**	-0.54 [-0.88, 0.20	.046	Not reported	Captive	Zimmermann & Binz (1989)
Dhole (*Cuon alpinus*)	8	9	**-.71 [-.94, 0.00]**	**.05**			Not reported	Captive	Volodin et at al., (2001)
Bearded seal (*Erigignathus barbatus*)	8	Not Reported	**-.79 [-.96, -.18**	**.021**			Not Reported	Wil	Risch et al., (2007)
Natterer's bat (*Myotis nattereri*)	6	Not Reported	**-.90 [-.99, -.09]**	**.037**			Social	Wild	Schmidbauer & Denzinger (2019)
Pale Spear-Nosed Bat (*Phyllostomus discolor*)	13	6	**-.51 [-.83, .06]**	**.156**			Social	Captive	Lattenkamp et al., (2019)
Grey Seal (*Halichoerus grypus*)	9	Not Reported	**.52 [-.22, .88]**	**.154**			Breeding	Wild	Asselin et al., (1993)
European Badger (*Meles meles*)	8	56	**.79 [.18, .96]**	**.021**			Not Reported	Wild	Wong et al., (1999)
Risso's Dolphin (*Grampus griseus*)	11	Not Reported	**.59 [-.02, .88]**	**.057**			Social	Wild	Corkeron & van Parijs (2001)
Sea Otter (*Enhydra lutris*)	13	9	.64 [.14, .88]	.017			Not Reported	Both	McShane et al., (1995)
Florida manatee (*Trichechus manatus*)	5	5-120	.36 [-.77, .94]	.553	.32 [-.79, .94]	.448	Not Reported	Wild	Brady et al., (2020)
Cavy (*Cavia aperea*)	6	66 Adults	-.14 [-.86, 0.76]	.784	-.14 [-.85, .76]	.702	Study Context	Captive	Monticelli & Ades (2013)
Beluga/White Whale (*Delphinapterus leucas*)	16	120	-.2 [-.63, .33]	.92	-.14 [-.60, .38]	.885	Reproductive	Wild	Belikov & Bel'Kovich (2007)
Beluga/White Whale (*Delphinapterus leucas*)	10	Not Reported	-.47 [-.85, .23]	.174			Not Reported	Wild	Karlsen et al., (2002)
Asian parti-colored bat (*Vespertilio sinensis*)	6	15	-.77 [-.97, .11]	.072			Social/Antagonistic	Wild	Luo et al., (2017)
Black rhino (*Diceros bicornis*)	5	13	-.05 [-.89, .87]	.935	t -0.11 [-0.90, 0.86]	0.801	Not Reported	Captive	Budde & Klump (2003)

Species	**Vocal Repertoire Size**	**N Subjects**	**rs [95% CI][a]**	***p*[b]**	**rt [95% CI][c]**	***p*[d]**	**Study Context[e]**	**Settting[f]**	**Reference**
Piebald Shrew (*Diplomesodon pulchellum*)	8	25 Adults	.45 [-.37, .88]	.26			Not Reported	Captive	Volodin et al., (2015)
Silvery mole-rat (*Heliophobius argenteocinereus*)	11	26	r -0.26 [-0.74, 0.40]	.441	-.11 [-.67, .52]	.639	Not Reported	Captive	Knotková et al., (2009)
Degu (*Octodon degus*)	10	17 (11 Adult, 6 Juvenile)	-.1 [-.69, 056]	.776			Not Reported	Captive	Long (2007)
killer whale (*Orcinus orca*)	10	13	-.15 [-.71, .53]	.676			Not Reported	Captive	Dahlheim & Awbrey (1982)
Speckled squirrel (*Spermophilus sus*licus)	8	230 Adult, 82 Juvenile	-0.21 [-.80, .58]	.610			Not Reported	Wild	Matrosova et al., (2012)
feathertail glider (*Acrobates pygmaeus*)	14	30	.22 [-0.35, .67]	.445	.17 [-.40, .64]	.41	Not Reported	Captive	Martin (2019)
Cheetah (*Acinonyx jubatus*)	7	14	.41 [-.49, .89]	.355			Not Reported	Captive	Volodina (2000)
Cheetah (Acinonyx jubatus)	6	13	.09 [-.78, .84]	p 0.872			Not Reported	Captive	Smirnova et al., (2016)
Bearded seal (Erignathus barbatus)	7	Not Reported	r-0.43 [-0.89, 0.48]	p, 0675			Not Reported	Wild	Jones et al., (2014)
bearded seal Erignathus barbatus	8	Not Reported	r -0.19 [-0.79, 0.59]	p, > .999	t -0.07 [-0.74, 0.67]	p, > .999	Not Reported	Wild	Frouin-Mouy et al., (2016)
Ribbon seal Histriophoca fasciata	6	Not Reported	-0.09 [-0.84, 0.78]	0.872			Not Reported	Wild	Jones et al., (2014)
North American River Otter (Lontra canadensis*)*	12	Ten adult otters, and eight pups	-0.12 [-0.65, 0.49]	0.707	t -0.11 [-0.64, 0.50]	0.629	Not Reported	Captive	Almonte (2014)
Hooded Seal Cystophora cristata	8	Not Reported	-0.61 [-0.92, 0.16]	0.108	t-0.47 [-0.88, 0.35]	p 0.105	Breeding	Wild	Ballard & Kovacs (1995)
Crab-Eating Fox Cerdocyon thous	6	8 Adults, 15 Pups	0.09 [-0.78, 0.84]	0.872			Not Reported	Captive	Brady (1981)
Bush Dog Speothos venaticus	7	6 Adults, 2 Pups	r -0.54 [-0.92, 0.36]	p, 0.215			Not Reported	Captive	Brady (1981)
Swift fox (*Vulpes velox*)	19	8 Adults, 19 infants	r -0.11 [-0.54, 0.36]	p, 0.641	t, 0.11 [-0.54, 0.36]	p, 0.527	Not Reported	Captive	Darden & Dabelsteen (2006)
Eastrern quoll (*Dasyurus viverrinus*)	5	15 Adults	r -0.21[-0.92, 0.83]	0.747			Not Reported	Captive	Dorph & McDonald (2017)
Steppe polecat (Mustela eversmannii)	6	6	r 0.18 [-0.55, 0.76]	p, 0.637			Not Reported	Captive	Farley et al., (1987)

Species	**Vocal Repertoire Size**	**N Subjects**	**rs [95% CI][a]**	***p*[b]**	**rt [95% CI][c]**	***p*[d]**	**Study Context[e]**	**Settting[f]**	**Reference**
Yellow mongoose (*Cynictis penicillata*)	9	Not Reported	r 0.18 [-0.55, 0.76]	0.637			Not Reported	Captive	Le Roux et al., (2009)
Giant otter (*Pteronura brasiliensis*)	14	female, and 1 unknown), and 25 cubs (1 male, 1 female, and\ 23 unknown)	-.0.33 [-0.73, 0.24]	0.249	t, -0.22 [-0.67, 0.36]	p, 0.293	Not Reported	Wild	Leuchtenberger et al., (2014)
Giant otter (*Pteronura brasiliensis*)	21	Not Reported	r- 0.04 [-0.46, 0.40]	0.88			Not Reported	Wild & Captive	Mumm & Knönschild (2014)
Weddell seal (Leptonychotes weddellii)	12	Approx 1400 individuals in the region, Approx. 20 female-pup pairs in the vicinity	r- 0.42 [-0.80, 0.21]	0.179			Not Reported	Wild	Pahl et al., (1997)
South-American fur seal (*Arctocephalus australis*)	11	“the N of males within the sampling area ranged from 1-8 (median = 3); females ranged from 8-49 (median = 21	r 0.04 [-0.57, 0.62]	0.908	t 0.04 [-0.57, 0.63]	0.864	Not Reported	Wild	Phillips & Stirling (2001)
Sumatran tiger (*Panthera tigris)*	6	5 (2 Adults & 3 Juveniles)	r, 0.14 [-0.76, 0.787]	p, 0.864			Not Reported	Captive	Rose et al., (2018)
Harbour seal *(Phoca vitulina*)	7	Between 25 and 75 seals hauled out daily. Most were juveniles and adult males (Kovacs et al., 1990)	r 0.14 [-0.68, 0.81]	P 0.760			Not Reported	Wild	Van Parijs & Kovacs (2002)

Species	Vocal Repertoire Size	N Subjects	rs [95% CI][a]	*p*[b]	rt [95% CI][c]	*p*[d]	Study Context[e]	Settting[f]	Reference
Big-footed myotis (Myotis macrodactylus)	8	Not Reported	r -0.14 [-0.77, 0.62]	p 0.736			Social	Wild	Guo et al., (2019)
Bechstein's bat (Myotis bechsteinii)	7	Not Reported	r 0.14 [-0.68, 0.81]	p 0.760			Social		Pfalzer (n.d.); Pfalzer & Kusch (Pfalzer & Kusch, 2003)
Brandt's Bat (*Myotis brandtii*)	6	Not Reported	r 0.2 [-0.73, 0.87]	p 0.700	t 0.14 [-0.76, 0.85]	p 0.702	Social		Pfalzer (n.d.); Pfalzer & Kusch (Pfalzer & Kusch, 2003)
Common pipistrelle (*Pipistrellus pipistrelus*)	5	Not Reported	r 0.3 [-0.79, 0.93]	p 0.624			Social		Pfalzer (n.d.); Pfalzer & Kusch (Pfalzer & Kusch, 2003)
Big brown bat (*Eptesicus fuscus*)	7	38	r -0.04 [-0.77, 0.74]	p 0.939			Social	Individuals born both in captivity and wild	Wright et al., (2013)
Commissaris's long-tongued bat (*Glossophaga commissarisi*)	8	Approx 100	r -0.49 [-0.89, 0.33]	p 0.278	t -0.37 [-0.85, 0.45]	p 0.432	Not Reported	Captive	Knörnschild et al., (2010)
Pallas's long-tongued bat (*Glossophaga soricina*)	15	18	r 0.36 [-0.18, 0.74]	p 0.547	t 0.29 [-0.26, 0.70]	p 0.503	Not Reported	Captive	Knörnschild et al., (2010)
Greater mouse-eared bat *(Myotis myotis)*	5	Approx. 50 Individuals, mostly young	r-0.8 [-0.99, 0.28]	p 0.104	t -0.6 [-0.97, 0.60]	p 0.142	Social	Wild	Pfalzer (n.d.); Pfalzer & Kusch (Pfalzer & Kusch, 2003)
Greater tube-nosed bat (*Murina leucogaster*)	15	14	r .08 [-0.45, 0.57]	p 0.776	t 0.08 [-0.45, 0.57]	p 0.692	Social	Wild-caught but moved to captivity for study	Lin et al., (2015)
Parnell's moustached bat (*Pteronotus parnellii*)	17	Sixty-five (41 males and 24 females) adults	.06 [-.44, .52]	p, 0.829			Communication	Wild-caught but moved to captivity for study	Kanwal et al., (1994)
Little Brown Bat (*Myotis lucifugus*)	12	Between 10,000-20,000 individuals	-.46 [-.82, .15]	p0.131			Social	Wild	Barclay et al., (1979)

Species	Vocal Repertoire Size	N Subjects	rs [95% CI][a]	p[b]	rt [95% CI][c]	p[d]	Study Context[e]	Settting[f]	Reference
Natterer's bat *(Myotis nattereri*)	6	Not Reported	-.6 [-.95, .41]	p 0.416			Social	Wild	Pfalzer (n.d.); Pfalzer & Kusch (Pfalzer & Kusch, 2003)

[a] Spearman's rank correlation coefficient with 95% confidence interval
[b] p value for Spearman's rank correlation coefficient
[c] Kendall's rank correlation coefficient with 95% confidence interval this statistic is not reported for every species, as it was only required for species with data sets containing ties of call usage frequency
[d] p value for Kendall's rank correlation coefficient
[e] Denotes if the repertoire study was conducted in any particular social or behavioural context
[f] Denotes if the repertoire study was conducted with animals housed in the wild or captivity

5.4.2 Is the Degree of Phylogenetic Relatedness Between Species Related to the Degree of Compression Found in their Vocal Repertoires?

Permutation tests revealed that, across the species in our dataset, the observed mean code length was significantly smaller than the mean of the permuted distribution of mean code lengths (Phylogenetic paired t-test: t_{47}=-3.08, p=0.003). This demonstrated there was a significant effect of phylogenetic relatedness on the tendency for compression in the vocal repertoire.

5.5 Discussion

5.5.1 The Extent of Conformity to the Law of Abbreviation in Our Dataset

This study represents the first broad scale comparative analysis investigating evidence of compression across multiple mammalian vocal repertoires. By investigating compression in the vocal repertoires of diverse mammalian species, I found evidence for Zipf's Law of Abbreviation in 6 of the 50 sampled species repertoires. In addition, I also found evidence of vocal redundancy in 2 of the 50 sampled species repertoires in our analysis, as demonstrated by a significant positive correlation between call duration and call frequency. Comparing across species, a phylogenetic analysis revealed that the degree of compression was significantly predicted by phylogenetic relatedness, that is to say species that are more closely related are more likely to show similar levels of compression. Generally, I observed that it is highly sociable species, particularly bats, badgers and squirrels that show strong evidence of compression. Previous research has demonstrated that high sociality in a species is related to larger overall repertoire size (McComb & Semple, 2005). It is possible that this is one of the effects that drives the tendency for highly social species to also demonstrate compression. In light of this, it is likely to be profitable for future researchers to include sociality and group size indices as covariance when examining compression in future. There are likely to be environmental factors that cause highly social group-living species, with generally larger vocal repertoire to produce vocalisations with greater efficiency. What might this be? It is possible that with a greater range of call types in the repertoire, there is less chance of conspecifics confusing individual vocal types for one another. For example, one of the main selection pressures against brevity is the idea of redundancy, which is believed to help receivers disambiguate sounds, as they become less similar after lengthening. It is possible that for species with larger repertoires the selection pressure for the addition of redundant information by lengthening (against compression) is not so strong. This is because with a greater variety of call types, it is easier for conspecifics to disambiguate signal types, even without the addition of redundant information. This is because the individual call types should be sufficiently acoustically distinct, even without the addition of further redundancy. Again, this highlights the need for indices of sociality and overall repertoire size to be included as predictors in future phylogenetically controlled analysis.

Overall, there was mixed evidence regarding optimisation for compression in mammalian vocal repertoires, with some repertoires showing evidence of compression, while others not, and two even having opposite effects. Following a number of studies of individuals species (see Table 1 in Chapter 1 of this thesis) the evidence of this study together with another provided by forthcoming work (Semple et al., In Prep), suggests that this type of broad phylogenetic analysis can provide stronger and more compelling evidence for compression in animal communication systems, when combined with a single species approach, Supporting earlier studies from individual species (IFavaro et al., 2020; Gustison et al., 2016; Huang et al., 2020; Luo et al., 2013), we have taken a broader taxonomic approach to show That in principle memo vocal repertoires can obey the law of brevity, however the evidence was not particularly widespread. In this respect, results support research which suggests that linguistic laws should be viewed in a broader evolutionary context as principles of biological organisation, rather than laws of language or communication specifically (Caetano-Anollés et al., 2017; Ferrer-i-Cancho & Forns, 2009; Li, 2012; Shahzad et al., 2015). Either we cannot say with confidence that compression is a universal Principle of animal behaviour, instead we advocate that researchers should pay greater attention to individual species ecology and environment, as it is most likely to be this way enables it will help us to differentiate the species which show evidence of compression from those which do not. Compression may well be universally selected for, yet as I will discuss below, it is possible that for individual species selection for compression has been outweighed by other evolutionary selection pressures.

Regarding patterns of commonality between species which do and do not demonstrate evidence of brevity, it is at this stage difficult to speculate as to what these may be. To help to determine such influences we recommend their future researchers should focus on deepening our understanding of the behavioural and environmental ecology of the species they are studying. Particularly, a stronger focus might be given to an understanding of the way these factors influence the evolution and development of individual signalling systems. In general, however we have shown their broad multi species comparative analysis such as this and previous work conducted by Semple and his colleagues can potentially be more useful in helping to understand the evolution of compression. This is specifically because we found limited evidence supporting the Law of Abbreviation in individual species the correct ones. However, when we take a broader phylogenetic approach, we find that there does appear to be evidence of selection for compression. This demonstrates that future studies may also benefit from taking such a phylogenetic approach, instead of or indeed in addition to studying single species.

One of the selection pressures typically considered as strongest with regard to selection of vocal compression is that of energy expenditure minimisation (Ferrer-i-Cancho, et al., 2013). Previous studies have shown that foraging vocalisations do not typically incur much of an energetic cost, for bats (Jones, 1999; Speakman & Racey, 1991). Given this evidence that foraging vocalisations do not appear to be energetically costly, it is no surprise that we did not find evidence of brevity in our analysis of vocalisations associated with foraging in big-footed myotis bats (*Myotis macrodactylus*) (Guo et al., 2019). This does not mean that bat vocalisations cannot be compressed in different contexts, for example previous studies have shown that compression is evident in distress related vocalisations in various bat species (Luo et al., 2013) More recent research has also demonstrated that high amplitude organisations are extremely energetically costly for bats (Currie et al., 2020). This would suggest that context in which I amplitude vocalisations are produced would be those in which we might expect to find greater conformity to efficiency of coding and brevity-related principles. This demonstrates that we cannot conclude that a species does not have efficiently coded vocalisations from a single context. Different selection pressures are likely to have differently impacted signal organisation, and multiple contexts must be compared, in order to get a fuller picture of a species' tendency to show compression in its repertoire.

The results show that unexpected patterns of compression emerged for several species; Grey Seals (*Halichoerus grypus*), Risso's Dolphins (*Grampus griseus*) and Sea otters (*Enhydra lutris*) (Asselin et al., 1993; Corkeron & Van Parijs, 2001; McShane et al., 1995) where there was a positive correlation between call duration and frequency of use. These patterns are indicative of redundancy i.e. the opposing concept to efficiency, namely that code lengths encoding information are not as short as they could possibly be (Cover & Thomas, 2006; Ferrer-i-Cancho, et al., 2013). Where there was a positive correlation between call duration and frequency, or permutation analysis showed that repertoire sizes were significantly large, and therefore not likely to be strongly compressed. These findings support conclusions from previous studies with neural networks (Chaabouni et al., 2019), chimpanzee gestural communication (Heesen et al., 2019) and European heraldic symbols (Miton & Morin, 2019) that show that efficiency is not always the default state for communication systems. In line with recent theoretical work (Ferrer-i-Cancho et al., 2020) our results also provide evidence indicative of anti-efficiency coding in animal communication, which was confirmed by permutation analysis. Possible explanations for these effects are discussed in the next section.

5.5.2 Why Might Communicative Redundancy Occur?

Redundancy, such as demonstrated by a positive correlation between call duration and call frequency, is thought to sometimes act in combination with, and in opposition to the pressure for compression. This is because, the pressures for high fidelity information transfer, and reliable disambiguation by receivers may sometimes outweigh the selection for efficiency and compression (Cover & Thomas, 2006; Ferrer-i-Cancho, et al., 2013; Plotkin & Nowak, 2000). It is worth noting that for the grey seal analyses reported in this study the repertoire under consideration relates to underwater vocalisations. These vocalisations are usually designed to carry of a distance, and therefor we might expect similar patterns to emerge here as do with long-distance vocalisations in other species. However, to our knowledge underwater vocalisations have not yet been investigated in the context of compression. It is therefore probable that factors relating to the transmission of sound through water, which is likely to be more difficult than sound transmission through air can encourage selection to favour redundancy over brevity. Underwater vocalisations may also be costly signals in some species (Rogers, 2017) which suggests that they are energetically costly, and cannot be produced indiscriminately. In this case we might expect a high degree of compression to be found in marine species when vocalising underwater. It will be for future researchers to begin to investigate in greater detail the evolutionary selection pressures that might be driving this, however for now, we believe that this pattern of results should help to support the idea that in certain contexts, efficiency is not the only evolutionarily viable communication strategy.

Although further research is required, there may be a number of functions for redundancy in vocal repertoires. Combining the vocalisation types into longer sequences, and lengthening the duration of the sequences, as well as of the component parts is likely to aid intelligibility. This would occur as it would become easier to differentiate potentially similarly structured sounds from each other (Cover & Thomas, 2006; Ferrer-i-Cancho, et al., 2013; Plotkin & Nowak, 2000). It can also help to overcome external factors such as high levels of environmental noise or noise from conspecifics (Bee & Micheyl, 2008; Brumm & Slater, 2006; Hotchkin & Parks, 2013; Tyack, 2008). For some species it is possible that individual vocalisations, if produced at short durations could be difficult for receivers to disambiguate. In an evolutionarily urgent context, such as appropriate responses to group members for community living species with complex and often hierarchical social networks, clarity can often be key. A certain amount of redundancy (Cover & Thomas, 2006, p. 184; Hebets et al., 2016; Rand & Williams, 1970) in the communication system can increase saliency overall, which this may indirectly favour reproductive success (Pinker, 2000). This way, rather than maximising compression, redundancy may also provide a strong selective pressure on the evolution of animal vocal communication. Given previous work showing associations between social group complexity and vocal complexity (Bouchet et al., 2013; McComb & Semple, 2005) it is possible that heightened sociality contributes to driving additional redundancy in vocal communication systems. This complex sociality (and the related addition of redundant information acting against compression) might also explain why evidence for compression in individual primate species was not particularly strong In general we suggest that such variation in species level trends could be integrated into future phylogenetic studies of compression in communication by taking advantage of a statistical approach called phylogenetic regression (Grafen, 1989/1992). In such an analysis, measures of sociality as well as repertoire size and complexity, and other relevant characteristics could be entered as predictors, to better understand their effects.

5.5.3 Statistical and Methodological Considerations

There are also some statistical and methodological issues which also may have influenced the results in this study. Previous research work (Ferrer-i-Cancho & Hernández-Fernández, 2013) has shown that at least five data points are required for the spearman correlation to be reliable. For our study, this resulted in species with smaller repertoires being excluded, thus limiting our small sample sizes and the number of recorded calls. This meant that of from our initial corpus of data (N = 73 species) only fifty were suitable for inclusion. We were also unable to analyse infant and juvenile repertoires, as these are typically smaller (Ames & Vergara, 2020; Kiefer et al., 2006; Knörnschild et al., 2010; Nottebohm & Nottebohm, 1978).

In addition, many of the data collected in these studies was constrained to a single context; this is to say that it is possible that vocalisation times for a specific species may have come only from a mating-related context as in the case of our data for beluga whales and the European badger. Likewise, the only data available for the Asian particoloured bat comes from an antagonistic context. This is important because if repertoire data were collected from multiple different behavioural contexts, we may find that different levels of compression according to context . This follows from previous studies with multiple species including bats (Luo et al., 2013), chimpanzees (Heesen et al., 2019; Safryghin, 2019) and several others (Ferrer-i-Cancho & Hernández-Fernández, 2013) showing that some parts of the repertoire can show evidence of compression while others do not. Only studies investigating the same repertoire is in multiple contexts will allow us access to definitive answers to such questions. As we have discussed above, factors relating to social and ecological contexts are likely to strongly influence selection pressures for brevity in vocal repertoires. It might therefore be that any given species might adhere to compression related predictions. However, from our analysis it is not possible to determine whether these effects are a true representation of the species vocal behaviour, or of selection for compression in a specific context. Future studies should consider the effect of context when examining evidence for vocal compression. Finally, given that in many cases individual level data was not reported, it is possible that some of the studies incurred issues with pseudoreplication. Future work should endeavour to control for subject identity as a random effect in models.

5.5.4 Conclusion

In conclusion, although we have shown that efficiency of coding can be observed in mammal vocal repertoires it may be somewhat rare, being demonstrated in just 12% of cases examined here. The common thread between the species demonstrating evidence of efficiency of coding appears to be that they are all highly social, and tend to live in large groups, for example the cetaceans bats, squirrels, and badgers. The analyses that we have used in the study should be more widely adopted in future in order to gain more detailed understanding of the selection pressures that have acted for and against communicative efficiency. Against a general hypothesis of efficiency in communication, we also found evidence of anti-compression (redundancy) in our analyses, with some species showing a positive correlation between call length and call frequency. Although this has been previously hypothesised, there has not been, to our knowledge, any empirical evidence of this before. Overall, our results add to growing evidence that linguistic laws can be more broadly applied to nonhuman animal communication systems but may operate in combination with other selective forces, including those favouring redundancy. Finally, we hope that broad taxonomic analyses, such as these, can help to tackle the widespread issue of bias against publishing null or negative results. Such results are important for building a picture of how compression and redundancy might evolve in animal communication systems. In the case of Zipf's Law of Abbreviation, it is possible that other studies examining evidence of compression may not have published such results if they failed to find an effect (Csada et al., 1996; Pautasso, 2010; Rosenthal, 1979), something which can potentially undermine the capacity to draw broader phylogenetic conclusions. It is likely that multi species comparisons such as these, which present data in a broader comparative context will be able to help to address potential publication bias.

6 References

Allen, J. A., Garland, E. C., Dunlop, R. A., & Noad, M. J. (2019). Network analysis reveals underlying syntactic features in a vocally learnt mammalian display, humpback whale song. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1917), 20192014. https://doi.org/10.1098/rspb.2019.2014

Altmann, E. G., & Gerlach, M. (2016). Statistical Laws in Linguistics. In M. Degli Esposti, E. G. Altmann, & F. Pachet (Eds.), *Creativity and Universality in Language* (pp. 7–26). Springer International Publishing. https://doi.org/10.1007/978-3-319-24403-7_2

Altmann, G. (1980). Prolegomena to Menzerath's law. *Glotometrika*, *2*, 1–10.

Altmann, G. (2014). Bibliography: Menzerath s Law. *Glottotheory*, *5*(1). https://doi.org/10.1515/glot-2014-0008

Ames, A. E., & Vergara, V. (2020). Trajectories of Vocal Repertoire Development in Beluga (Delphinapterus leucas) Calves: Insights from Studies a Decade Apart. *Aquatic Mammals*, *46*(4), 344–366. https://doi.org/10.1578/AM.46.4.2020.344

Andersson, M., & Simmons, L. W. (2006). Sexual selection and mate choice. *Trends in Ecology & Evolution*, *21*(6), 296–302. https://doi.org/10.1016/j.tree.2006.03.015

Araujo, L. C., Benevides, A., & Pereira, M. (2020). Análise da Lei de Menzerath no Português Brasileiro. *Linguamática*, *12*(1), 31–48. https://doi.org/10.21814/lm.12.1.300

Arnold, K., & Zuberbühler, K. (2006). Semantic combinations in primate calls. *Nature*, *441*(7091), 303–303. https://doi.org/10.1038/441303a

Asselin, S., Hammill, M. O., & Barrette, C. (1993). Underwater vocalizations of ice breeding grey seals. *Canadian Journal of Zoology*, *71*(11), 2211–2219. https://doi.org/10.1139/z93-310

Bateman, A. J. (1948). Intra-sexual selection in Drosophila. *Heredity*, *2*(3), 349–368. https://doi.org/10.1038/hdy.1948.21

Bentz, C., & Ferrer-i-Cancho, R. (2015, October). *Zipf's law of abbreviation as a language universal. Capturing Phylogenetic Algorithms for Linguistics*. Lorentz Center Workshop, Leiden. https://publikationen.uni-tuebingen.de/xmlui/bitstream/handle/10900/68639/Bentz_Ferrer.pdf?sequence=1&isAllowed=y

Bezerra, B. M., Souto, A. S., Radford, A. N., & Jones, G. (2011). Brevity is not always a virtue in primate communication. *Biology Letters*, *7*(1), 23–25. https://doi.org/10.1098/rsbl.2010.0455

Bianco, M. J., Gerstoft, P., Traer, J., Ozanich, E., Roch, M. A., Gannot, S., & Deledalle, C.-A. (2019). Machine learning in acoustics: Theory and applications. *The Journal of the Acoustical Society of America*, *146*(5), 3590–3628. https://doi.org/10.1121/1.5133944

Bickerton, D. (2003). Symbol and Structure: A Comprehensive Framework for Language Evolution. In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution* (pp. 77–93). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199244843.003.0005

Bininda-Emonds, O. R. P., Cardillo, M., Jones, K. E., MacPhee, R. D. E., Beck, R. M. D., Grenyer, R., Price, S. A., Vos, R. A., Gittleman, J. L., & Purvis, A. (2007). The delayed rise of present-day mammals. *Nature*, *446*(7135), 507–512. https://doi.org/10.1038/nature05634

Bishara, A. J., & Hittner, J. B. (2017). Confidence intervals for correlations when data are not normal. *Behavior Research Methods*, *49*(1), 294–309. https://doi.org/10.3758/s13428-016-0702-8

Börstell, C., Hörberg, T., & Östling, R. (2016). Distribution and duration of signs and parts of speech in Swedish Sign Language. *Sign Language & Linguistics*, *19*(2), 143–196. https://doi.org/10.1075/sll.19.2.01bor

Bossert, W. H., & Wilson, E. O. (1963). The analysis of olfactory communication among animals. *Journal of Theoretical Biology*, *5*(3), 443–469. https://doi.org/10.1016/0022-5193(63)90089-4

Bouchet, H., Blois-Heulin, C., & Lemasson, A. (2013). Social complexity parallels vocal complexity: A comparison of three non-human primate species. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00390

Bowling, D. L., & Fitch, W. T. (2015). Do Animal Communication Systems Have Phonemes? *Trends in Cognitive Sciences*, *19*(10), 555–557. https://doi.org/10.1016/j.tics.2015.08.011

Brooker, R. M., & Wong, B. B. M. (2020). Non-visual camouflage. *Current Biology*, *30*(21), R1290–R1292. https://doi.org/10.1016/j.cub.2020.08.001

Brumm, H., & Slater, P. J. B. (2006). Animals can vary signal amplitude with receiver distance: Evidence from zebra finch song. *Animal Behaviour*, *72*(3), 699–705. https://doi.org/10.1016/j.anbehav.2006.01.020

Byrne, R. W. (1981). Distance Vocalisations of Guinea Baboons (Papio Papio) in Senegal: An Analysis of Function. *Behaviour*, *78*(3–4), 283–312. https://doi.org/10.1163/156853981X00365

Caetano-Anollés, G., Minhas, B. F., Aziz, F., Mughal, F., Shahzad, K., Tal, G., Mittenthal, J. E., Caetano-Anollés, D., Koç, I., Nasir, A., Caetano-Anollés, K., & Kim, K. M.

(2017). The Compressed Vocabulary of the Proteins of Archaea. In G. Witzany (Ed.), *Biocommunication of Archaea* (pp. 147–174). Springer International Publishing. https://doi.org/10.1007/978-3-319-65536-9_10

Casas, B., Hernández-Fernández, A., Català, N., Ferrer-i-Cancho, R., & Baixeries, J. (2019). Polysemy and brevity versus frequency in language. *Computer Speech & Language*, *58*, 19–50. https://doi.org/10.1016/j.csl.2019.03.007

Chaabouni, R., Kharitonov, E., Dupoux, E., & Baroni, M. (2019). Anti-efficient encoding in emergent communication. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d\textquotesingle Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 32* (pp. 6293–6303). Curran Associates, Inc. http://papers.nips.cc/paper/8859-anti-efficient-encoding-in-emergent-communication.pdf

Chacoma, A., & Zanette, D. H. (2020). Heaps' Law and Heaps functions in tagged texts: Evidences of their linguistic relevance. *Royal Society Open Science*, *7*(3), 200008. https://doi.org/10.1098/rsos.200008

Chambers, C. T., & Mogil, J. S. (2015). Ontogeny and phylogeny of facial expression of pain: *PAIN*, *156*(5), 798–799. https://doi.org/10.1097/j.pain.0000000000000133

Cholewiak, D. M., Sousa-Lima, R. S., & Cerchio, S. (2013). Humpback whale song hierarchical structure: Historical context and discussion of current classification issues. *Marine Mammal Science*, *29*(3), E312–E332. https://doi.org/10.1111/mms.12005

Chomsky, N., & Smith, N. (2000). *New Horizons in the Study of Language and Mind* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511811937

Clay, Z., & de Waal, F. B. M. (2015). Sex and strife: Post-conflict sexual contacts in bonobos. *Behaviour*, *152*(3–4), 313–334. https://doi.org/10.1163/1568539X-00003155

Clay, Z., Pika, S., Gruber, T., & Zuberbühler, K. (2011). Female bonobos use copulation calls as social signals. *Biology Letters*, *7*(4), 513–516. https://doi.org/10.1098/rsbl.2010.1227

Clay, Z., & Zuberbühler, K. (2011). The Structure of Bonobo Copulation Calls During Reproductive and Non-Reproductive Sex: Structure of Bonobo Copulation Calls. *Ethology*, *117*(12), 1158–1169. https://doi.org/10.1111/j.1439-0310.2011.01975.x

Clink, D. J., Ahmad, A. H., & Klinck, H. (2020). Brevity is not a universal in animal communication: Evidence for compression depends on the unit of analysis in small ape vocalizations. *Royal Society Open Science*, *7*(4), 200151. https://doi.org/10.1098/rsos.200151

Clink, D. J., & Lau, A. R. (2020). Adherence to Menzerath's Law is the exception (not the rule) in three duetting primate species. *Royal Society Open Science*, *7*(11), 201557. https://doi.org/10.1098/rsos.201557

Corkeron, P. J., & Van Parijs, S. M. (2001). Vocalizations of eastern Australian Risso's dolphins, *Grampus griseus*. *Canadian Journal of Zoology*, *79*(1), 160–164. https://doi.org/10.1139/cjz-79-1-160

Corral, Á., & García del Muro, M. (2020). From Boltzmann to Zipf through Shannon and Jaynes. *Entropy*, *22*(2), 179. https://doi.org/10.3390/e22020179

Corral, Á., & Serra, I. (2020). The Brevity Law as a Scaling Law, and a Possible Origin of Zipf's Law for Word Frequencies. *Entropy*, *22*(2), 224. https://doi.org/10.3390/e22020224

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed). Wiley-Interscience.

Crockford, C., Wittig, R. M., Mundry, R., & Zuberbühler, K. (2012). Wild Chimpanzees Inform Ignorant Group Members of Danger. *Current Biology*, *22*(2), 142–146. https://doi.org/10.1016/j.cub.2011.11.053

Csada, R. D., James, P. C., & Espie, R. H. M. (1996). The 'File Drawer Problem' of Non-Significant Results: Does It Apply to Biological Research? *Oikos*, *76*(3), 591. https://doi.org/10.2307/3546355

Currie, S. E., Boonman, A., Troxell, S., Yovel, Y., & Voigt, C. C. (2020). Echolocation at high intensity imposes metabolic costs on flying bats. *Nature Ecology & Evolution*. https://doi.org/10.1038/s41559-020-1249-8

Dabelsteen, T., McGregor, P. K., Lampe, H. M., Langmore, N. E., & Holland, J. (1998). Quiet Song in Song Birds: An Overlooked Phenomenon. *Bioacoustics*, *9*(2), 89–105. https://doi.org/10.1080/09524622.1998.9753385

Dąbrowska, E. (2015). What exactly is Universal Grammar, and has anyone seen it? *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00852

Davila Ross, M., Menzler, S., & Zimmermann, E. (2008). Rapid facial mimicry in orangutan play. *Biology Letters*, *4*(1), 27–30. https://doi.org/10.1098/rsbl.2007.0535

Dębowski, Ł., & Bentz, C. (2020). Information Theory and Language. *Entropy*, *22*(4), 435. https://doi.org/10.3390/e22040435

Demartsev, V., Gordon, N., Barocas, A., Bar-Ziv, E., Ilany, T., Goll, Y., Ilany, A., & Geffen, E. (2019). The "Law of Brevity" in animal communication: Sex-specific signaling optimization is determined by call amplitude rather than duration. *Evolution Letters*. https://doi.org/10.1002/evl3.147

Egghe, L. (2007). Untangling Herdan's law and Heaps' law: Mathematical and informetric arguments. *Journal of the American Society for Information Science and Technology*, *58*(5), 702–709. https://doi.org/10.1002/asi.20524

Engesser, S., Holub, J. L., O'Neill, L. G., Russell, A. F., & Townsend, S. W. (2019). Chestnut-crowned babbler calls are composed of meaningless shared building blocks. *Proceedings of the National Academy of Sciences*, *116*(39), 19579–19584. https://doi.org/10.1073/pnas.1819513116

Engesser, S., Ridley, A. R., Manser, M. B., Manser, A., & Townsend, S. W. (2018). Internal acoustic structuring in pied babbler recruitment cries specifies the form of recruitment. *Behavioral Ecology*, *29*(5), 1021–1030. https://doi.org/10.1093/beheco/ary088

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*(21), 5976–5981. https://doi.org/10.1073/pnas.1600970113

Engesser, S., Ridley, A. R., & Townsend, S. W. (2017). Element repetition rates encode functionally distinct information in pied babbler 'clucks' and 'purrs'. *Animal Cognition*, *20*(5), 953–960. https://doi.org/10.1007/s10071-017-1114-6

Engesser, S., & Townsend, S. W. (2019). Combinatoriality in the vocal systems of nonhuman animals. *Wiley Interdisciplinary Reviews: Cognitive Science*, *10*(4), e1493. https://doi.org/10.1002/wcs.1493

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–448. https://doi.org/10.1017/S0140525X0999094X

Favaro, L., Gamba, M., Cresta, E., Fumagalli, E., Bandoli, F., Pilenga, C., Isaja, V., Mathevon, N., & Reby, D. (2020). Do penguins' vocal sequences conform to linguistic laws? *Biology Letters*, *16*(2), 20190589. https://doi.org/10.1098/rsbl.2019.0589

Fedurek, P., Zuberbühler, K., & Semple, S. (2017). Trade-offs in the production of animal vocal sequences: Insights from the structure of wild chimpanzee pant hoots. *Frontiers in Zoology*, *14*(1), 50. https://doi.org/10.1186/s12983-017-0235-8

Fehér, O., Wang, H., Saar, S., Mitra, P. P., & Tchernichovski, O. (2009). De novo establishment of wild-type song culture in the zebra finch. *Nature*, *459*(7246), 564–568. https://doi.org/10.1038/nature07994

Ferrer-i-Cancho, R. (2016). Compression and the origins of Zipf's law for word frequencies. *Complexity*, *21*(S2), 409–411. https://doi.org/10.1002/cplx.21820

Ferrer-i-Cancho, R., Bentz, C., & Seguin, C. (2020). Optimal Coding and the Origins of Zipfian Laws. *Journal of Quantitative Linguistics*, 1–30. https://doi.org/10.1080/09296174.2020.1778387

Ferrer-i-Cancho, R., & Forns, N. (2009). The self-organization of genomes. *Complexity*, NA-NA. https://doi.org/10.1002/cplx.20296

Ferrer-i-Cancho, R., & Hernández-Fernández, A. (2013). The Failure of the Law of Brevity in Two New World Primates. Statistical Caveats. *Glottotheory*, *4*(1). https://doi.org/10.1524/glot.2013.0004

Ferrer-i-Cancho, R., Hernández-Fernández, A., Lusseau, D., Agoramoorthy, G., Hsu, M. J., & Semple, S. (2013). Compression as a Universal Principle of Animal Behavior. *Cognitive Science*, *37*(8), 1565–1578. https://doi.org/10.1111/cogs.12061

Ferrer-i-Cancho, R., & Lusseau, D. (2009). Efficient coding in dolphin surface behavioral patterns. *Complexity*, *14*(5), 23–25. https://doi.org/10.1002/cplx.20266

Fieller, E. C., Hartley, H. O., & Pearson, E. S. (1957). Tests for Rank Correlation Coefficients. I. *Biometrika*, *44*(3–4), 470–481. https://doi.org/10.1093/biomet/44.3-4.470

Fitzpatrick, M. C., Preisser, E. L., Ellison, A. M., & Elkinton, J. S. (2009). Observer bias and the detection of low-density populations. *Ecological Applications*, *19*(7), 1673–1679. https://doi.org/10.1890/09-0265.1

Freeberg, T. M., & Lucas, J. R. (2009). Pseudoreplication is (still) a problem. *Journal of Comparative Psychology*, *123*(4), 450–451. https://doi.org/10.1037/a0017031

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & van Schaik, C. P. (2019). Multimodal communication and language origins: Integrating gestures and vocalizations: Multimodal communication and language origins. *Biological Reviews*. https://doi.org/10.1111/brv.12535

Fröhlich, M., & van Schaik, C. P. (2018). The function of primate multimodal communication. *Animal Cognition*, *21*(5), 619–629. https://doi.org/10.1007/s10071-018-1197-8

Fukuzawa, Y., Webb, W. H., Pawley, M. D. M., Roper, M. M., Marsland, S., Brunton, D. H., & Gilman, A. (2020). *Koe*: Web-based software to classify acoustic units and analyse sequence structure in animal vocalizations. *Methods in Ecology and Evolution*, *11*(3), 431–441. https://doi.org/10.1111/2041-210X.13336

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336–10341. https://doi.org/10.1073/pnas.1502134112

Garamszegi, L. Z. (2016). A simple statistical guide for the analysis of behaviour when data are constrained due to practical or ethical reasons. *Animal Behaviour*, *120*, 223–234. https://doi.org/10.1016/j.anbehav.2015.11.009

Genty, E. (2019). Vocal–gestural combinations in infant bonobos: New insights into signal functional specificity. *Animal Cognition*, *22*(4), 505–518. https://doi.org/10.1007/s10071-019-01267-0

Genty, E., Clay, Z., Hobaiter, C., & Zuberbühler, K. (2014). Multi-Modal Use of a Socially Directed Call in Bonobos. *PLOS ONE*, *9*(1), e84738. https://doi.org/10.1371/journal.pone.0084738

Genty, E., Neumann, C., & Zuberbühler, K. (2015). Complex patterns of signalling to convey different social goals of sex in bonobos, *Pan paniscus*. *Scientific Reports*, *5*, 16135. https://doi.org/10.1038/srep16135

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How Efficiency Shapes Human Language. *Trends in Cognitive Sciences*, *23*(5), 389–407. https://doi.org/10.1016/j.tics.2019.02.003

Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *326*(1233), 119–157. https://doi.org/10.1098/rstb.1989.0106

Grafen, Alan. (1992). The uniqueness of the phylogenetic regression. *Journal of Theoretical Biology*, *156*(4), 405–423. https://doi.org/10.1016/S0022-5193(05)80635-6

Gratton, P., & Mundry, R. (2019). Accounting for pseudoreplication is not possible when the source of nonindependence is unknown. *Animal Behaviour*, *154*, e1–e5. https://doi.org/10.1016/j.anbehav.2019.05.014

Gruber, T., & Zuberbühler, K. (2013). Vocal Recruitment for Joint Travel in Wild Chimpanzees. *PLoS ONE*, *8*(9), e76073. https://doi.org/10.1371/journal.pone.0076073

Guo, D., Luo, B., Zhang, K., Liu, M., Metzner, W., Liu, Y., & Feng, J. (2019). Social vocalizations of big-footed myotis (*Myotis macrodactylus*) during foraging. *Integrative Zoology*, *14*(5), 446–459. https://doi.org/10.1111/1749-4877.12367

Gustison, M. L., Semple, S., Ferrer-i-Cancho, R., & Bergman, T. J. (2016). Gelada vocal sequences follow Menzerath's linguistic law. *Proceedings of the National Academy of Sciences*, *113*(19), E2750–E2758. https://doi.org/10.1073/pnas.1522072113

Gustison, M. L., & Townsend, S. W. (2015). A survey of the context and structure of high- and low-amplitude calls in mammals. *Animal Behaviour*, *105*, 281–288. https://doi.org/10.1016/j.anbehav.2015.04.021

Hahn, M., Jurafsky, D., & Futrell, R. (2020). Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences*, 201910923. https://doi.org/10.1073/pnas.1910923117

Hailman, J. P., Ficken, M. S., & Ficken, R. W. (1985). The 'chick-a-dee' calls of Parus atricapillus: A recombinant system of animal communication compared with written English. *Semiotica*, *56*(3–4). https://doi.org/10.1515/semi.1985.56.3-4.191

Hare, R. M., & Simmons, L. W. (2019). Sexual selection and its evolutionary consequences in female animals. *Biological Reviews*, *94*(3), 929–956. https://doi.org/10.1111/brv.12484

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Hebets, E. A., Barron, A. B., Balakrishnan, C. N., Hauber, M. E., Mason, P. H., & Hoke, K. L. (2016). A systems approach to animal communication. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1826). https://doi.org/10.1098/rspb.2015.2889

Heesen, R., Hobaiter, C., Ferrer-i-Cancho, R., & Semple, S. (2019). Linguistic laws in chimpanzee gestural communication. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1896), 20182900. https://doi.org/10.1098/rspb.2018.2900

Held, S. D. E., & Špinka, M. (2011). Animal play and animal welfare. *Animal Behaviour*, *81*(5), 891–899. https://doi.org/10.1016/j.anbehav.2011.01.007

Hernández-Fernández, A., Casas, B., Ferrer-i-Cancho, R., & Baixeries, J. (2016). Testing the Robustness of Laws of Polysemy and Brevity Versus Frequency. In P. Král & C. Martín-Vide (Eds.), *Statistical Language and Speech Processing* (Vol. 9918, pp. 19–29). Springer International Publishing. https://doi.org/10.1007/978-3-319-45925-7_2

Hernández-Fernández, A., G. Torre, I., Garrido, J.-M., & Lacasa, L. (2019). Linguistic Laws in Speech: The Case of Catalan and Spanish. *Entropy*, *21*(12), 1153. https://doi.org/10.3390/e21121153

Higham, J. P., & Hebets, E. A. (2013). An introduction to multimodal communication. *Behavioral Ecology and Sociobiology*, *67*(9), 1381–1388. https://doi.org/10.1007/s00265-013-1590-x

Hinde, R. A., & Rowell, T. E. (1962). Communication by Postures and Facial Expressions in the Rhesus Monkey (macaca Mulatta). *Proceedings of the Zoological Society of London*, *138*(1), 1–21. https://doi.org/10.1111/j.1469-7998.1962.tb05684.x

Holler, J., & Levinson, S. C. (2019). Multimodal Language Processing in Human Communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

Huang, M., Ma, H., Ma, C., Garber, P. A., & Fan, P. (2020). Male gibbon loud morning calls conform to Zipf's law of brevity and Menzerath's law: Insights into the origin of human language. *Animal Behaviour*, S000334721930377X. https://doi.org/10.1016/j.anbehav.2019.11.017

Hurlbert, S. H. (1984). Pseudoreplication and the Design of Ecological Field Experiments. *Ecological Monographs*, *54*(2), 187–211. https://doi.org/10.2307/1942661

Hurlbert, S. H. (2009). The ancient black art and transdisciplinary extent of pseudoreplication. *Journal of Comparative Psychology*, *123*(4), 434–443. https://doi.org/10.1037/a0016221

Hyland Bruno, J., & Tchernichovski, O. (2019). Regularities in zebra finch song beyond the repeated motif. *Behavioural Processes*, *163*, 53–59. https://doi.org/10.1016/j.beproc.2017.11.001

Janik, V. M. (2014). Cetacean vocal learning and communication. *Current Opinion in Neurobiology*, *28*, 60–65. https://doi.org/10.1016/j.conb.2014.06.010

Jennions, M. D., & Petrie, M. (2000). Why do females mate multiply? A review of the genetic benefits. *Biological Reviews of the Cambridge Philosophical Society*, *75*(1), 21–64. https://doi.org/10.1017/S0006323199005423

Johnson, C. J., Pick, H. L., Siegel, G. M., Cicciarelli, A. W., & Garber, S. R. (1981). Effects of Interpersonal Distance on Children’s Vocal Intensity. *Child Development*, *52*(2), 721. https://doi.org/10.2307/1129198

Johnstone, R. A. (1996). Multiple displays in animal communication: ‘Backup signals’ and ‘multiple messages’. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1337), 329–338. https://doi.org/10.1098/rstb.1996.0026

Jones, G. (1999). Scaling of echolocation call parameters in bats. *The Journal of Experimental Biology*, *202*(Pt 23), 3359–3367.

Kang, T. S. (2017). *The Impact of Acoustic Feedback and Tutoring on Song Syntax Development in Zebra Finches* [Unpublished MSc Thesis]. University of Warwick.

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's Law of Abbreviation and the Principle of Least Effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, *165*, 45–52. https://doi.org/10.1016/j.cognition.2017.05.001

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. (2016). Acoustic sequences in non-human animals: A tutorial review and prospectus: Acoustic sequences in animals. *Biological Reviews*, *91*(1), 13–52. https://doi.org/10.1111/brv.12160

Kerth, G. (2008). Causes and Consequences of Sociality in Bats. *BioScience*, *58*(8), 737–746. https://doi.org/10.1641/B580810

Kiefer, S., Spiess, A., Kipper, S., Mundry, R., Sommer, C., Hultsch, H., & Todt, D. (2006). First-Year Common Nightingales (Luscinia megarhynchos) Have Smaller Song-Type Repertoire Sizes Than Older Males. *Ethology*, *112*(12), 1217–1224. https://doi.org/10.1111/j.1439-0310.2006.01283.x

King, A., & Wedel, A. (2020). Greater Early Disambiguating Information for Less-Probable Words: The Lexicon Is Shaped by Incremental Processing. *Open Mind*, *4*, 1–12. https://doi.org/10.1162/opmi_a_00030

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, *124*(3), 245–266. https://doi.org/10.1037/rev0000059

Knörnschild, M., Nagy, M., Metz, M., Mayer, F., & von Helversen, O. (2010). Complex vocal imitation during ontogeny in a bat. *Biology Letters*, *6*(2), 156–159. https://doi.org/10.1098/rsbl.2009.0685

Langmore, N. E., & Davies, N. B. (1997). Female dunnocks use vocalizations to compete for males. *Animal Behaviour*, *53*(5), 881–890. https://doi.org/10.1006/anbe.1996.0306

Levinson, S. C., & Evans, N. (2010). Time for a sea-change in linguistics: Response to comments on 'The Myth of Language Universals'. *Lingua*, *120*(12), 2733–2758. https://doi.org/10.1016/j.lingua.2010.08.001

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130302. https://doi.org/10.1098/rstb.2013.0302

Li, W. (2012). Menzerath's law at the gene-exon level in the human genome. *Complexity*, *17*(4), 49–53. https://doi.org/10.1002/cplx.20398

Lindenfors, P., Revell, L. J., & Nunn, C. L. (2010). Sexual dimorphism in primate aerobic capacity: A phylogenetic test: Sexual dimorphism in aerobic capacity. *Journal of Evolutionary Biology*, *23*(6), 1183–1194. https://doi.org/10.1111/j.1420-9101.2010.01983.x

Lipkind, D., Marcus, G. F., Bemis, D. K., 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. https://doi.org/10.1038/nature12173

Lüdecke, D., Waggoner, P., & Makowski, D. (2019). insight: A Unified Interface to Access Information from Model Objects in R. *Journal of Open Source Software*, *4*(38), 1412. https://doi.org/10.21105/joss.01412

Luo, B., Jiang, T., Liu, Y., Wang, J., Lin, A., Wei, X., & Feng, J. (2013). Brevity is prevalent in bat short-range communication. *Journal of Comparative Physiology A*, *199*(4), 325–333. https://doi.org/10.1007/s00359-013-0793-y

Mačutek, J., Chromý, J., & Koščová, M. (2019). Menzerath-Altmann Law and Prothetic /v/ in Spoken Czech. *Journal of Quantitative Linguistics*, *26*(1), 66–80. https://doi.org/10.1080/09296174.2018.1424493

Mager, J. L., & Walcott, C. (2007). Structural and Contextual Characteristics of Territorial “Yodels” Given by Male Common Loons (Gavia immer) in Northern Wisconsin. *The Passenger Pigeoon*, *69*(3), 327–338.

Magrath, R. D., Haff, T. M., Fallow, P. M., & Radford, A. N. (2015). Eavesdropping on heterospecific alarm calls: From mechanisms to consequences: Interspecific eavesdropping. *Biological Reviews*, *90*(2), 560–586. https://doi.org/10.1111/brv.12122

Mahowald, K., Dautriche, I., Gibson, E., & Piantadosi, S. T. (2018). Word Forms Are Structured for Efficient Use. *Cognitive Science*, *42*(8), 3116–3134. https://doi.org/10.1111/cogs.12689

Makowski, D., Ben-Shachar, M., Patil, I., & Lüdecke, D. (2020). Methods and Algorithms for Correlation Analysis in R. *Journal of Open Source Software*, *5*(51), 2306. https://doi.org/10.21105/joss.02306

Marino, L., Connor, R. C., Fordyce, R. E., Herman, L. M., Hof, P. R., Lefebvre, L., Lusseau, D., McCowan, B., Nimchinsky, E. A., Pack, A. A., Rendell, L., Reidenberg, J. S., Reiss, D., Uhen, M. D., Van der Gucht, E., & Whitehead, H. (2007). Cetaceans Have Complex Brains for Complex Cognition. *PLoS Biology*, *5*(5), e139. https://doi.org/10.1371/journal.pbio.0050139

Markl, H. (1983). Vibrational Communication. In F. Huber & H. Markl (Eds.), *Neuroethology and Behavioral Physiology* (pp. 332–353). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-69271-0_24

McComb, K., & Semple, S. (2005). Coevolution of vocal communication and sociality in primates. *Biology Letters*, *1*(4), 381–385. https://doi.org/10.1098/rsbl.2005.0366

McGarigal, K., Stafford, S., & Cushman, S. (2000). Discriminant Analysis. In K. McGarigal, S. Stafford, & S. Cushman, *Multivariate Statistics for Wildlife and Ecology Research* (pp. 129–187). Springer New York. https://doi.org/10.1007/978-1-4612-1288-1_4

McShane, L. J., Estes, J. A., Riedman, M. L., & Staedler, M. M. (1995). Repertoire, Structure, and Individual Variation of Vocalizations in the Sea Otter. *Journal of Mammalogy*, *76*(2), 414–427. https://doi.org/10.2307/1382352

Menzerath, P. (1954). *Die Architektonik des deutschen Wortschatzes*. F. Dümmler.

Mercado, E., & Handel, S. (2012). Understanding the structure of humpback whale songs (L). *The Journal of the Acoustical Society of America*, *132*(5), 2947–2950. https://doi.org/10.1121/1.4757643

Miton, H., & Morin, O. (2019). When iconicity stands in the way of abbreviation: No Zipfian effect for figurative signals. *PLOS ONE*, *14*(8), e0220793. https://doi.org/10.1371/journal.pone.0220793

Nakano, R., Ishikawa, Y., Tatsuki, S., Skals, N., Surlykke, A., & Takanashi, T. (2009). Private ultrasonic whispering in moths. *Communicative & Integrative Biology*, *2*(2), 123–126. https://doi.org/10.4161/cib.7738

Nakano, R., & Nagamine, K. (2019). Loudness–Duration Tradeoff in Ultrasonic Courtship Songs of Moths. *Frontiers in Ecology and Evolution*, *7*, 244. https://doi.org/10.3389/fevo.2019.00244

Nielsen, M. L. K., Bejder, L., Videsen, S. K. A., Christiansen, F., & Madsen, P. T. (2019). Acoustic crypsis in southern right whale mother–calf pairs: Infrequent, low-output calls to avoid predation? *The Journal of Experimental Biology*, *222*(13), jeb190728. https://doi.org/10.1242/jeb.190728

Noren, D. P., Holt, M. M., Dunkin, R. C., & Williams, T. M. (2013). The metabolic cost of communicative sound production in bottlenose dolphins (Tursiops truncatus). *Journal of Experimental Biology*, *216*(9), 1624–1629. https://doi.org/10.1242/jeb.083212

Noriega, F., Montes-Medina, A. C., & Timme, M. (2019). Quantitative analysis of timing in animal vocal sequences. *ArXiv:1902.07650 [q-Bio]*. http://arxiv.org/abs/1902.07650

Nottebohm, F., & Nottebohm, M. E. (1978). Relationship between Song Repertoire and Age in the Canary, Serinus canarius. *Zeitschrift Für Tierpsychologie*, *46*(3), 298–305. https://doi.org/10.1111/j.1439-0310.1978.tb01451.x

Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences*, *96*(14), 8028–8033. https://doi.org/10.1073/pnas.96.14.8028

Padilla de la Torre, M., & McElligott, A. G. (2017). Vocal Communication and the Importance of Mother-Offspring Relations in Cattle. *Animal Behavior and Cognition*, *4*(4), 522–525. https://doi.org/10.26451/abc.04.04.13.2017

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Parker, G. A. (2006). Sexual conflict over mating and fertilization: An overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1466), 235–259. https://doi.org/10.1098/rstb.2005.1785

Parks, S. E., Cusano, D. A., Van Parijs, S. M., & Nowacek, D. P. (2019). North Atlantic right whale (*Eubalaena glacialis*) acoustic behavior on the calving grounds. *The Journal of the Acoustical Society of America*, *146*(1), EL15–EL21. https://doi.org/10.1121/1.5115332

Partan, S., & Marler, P. (1999). BEHAVIOR:Communication Goes Multimodal. *Science*, *283*(5406), 1272–1273. https://doi.org/10.1126/science.283.5406.1272

Pautasso, M. (2010). Worsening file-drawer problem in the abstracts of natural, medical and social science databases. *Scientometrics*, *85*(1), 193–202. https://doi.org/10.1007/s11192-010-0233-5

Pfalzer, G., & Kusch, J. (2003). Structure and variability of bat social calls: Implications for specificity and individual recognition. *Journal of Zoology*, *261*(1), 21–33. https://doi.org/10.1017/S0952836903003935

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529. https://doi.org/10.1073/pnas.1012551108

Pinker, S. (1994). *The language instinct: The new science of language and mind*. Allen Lane, Penguin Press.

Pinker, S. (2000). Survival of the clearest. *Nature*, *404*(6777), 441–442. https://doi.org/10.1038/35006523

Plotkin, J. B., & Nowak, M. A. (2000). Language Evolution and Information Theory. *Journal of Theoretical Biology*, *205*(1), 147–159. https://doi.org/10.1006/jtbi.2000.2053

Prat, Y. (2019). Animals Have No Language, and Humans Are Animals Too. *Perspectives on Psychological Science*, *14*(5), 885–893. https://doi.org/10.1177/1745691619858402

Prieur, J., Barbu, S., Blois-Heulin, C., & Lemasson, A. (2020). The origins of gestures and language: History, current advances and proposed theories. *Biological Reviews*, *95*(3), 531–554. https://doi.org/10.1111/brv.12576

R Core Team. (2020). *R: A Language and Environment for Statistical Computing* (4.0.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rand, A. S., & Williams, E. E. (1970). An Estimation of Redundancy and Information Content of Anole Dewlaps. *The American Naturalist*, *104*(935), 99–103. https://doi.org/10.1086/282643

Reichard, D. G., & Anderson, R. C. (2015). Why signal softly? The structure, function and evolutionary significance of low-amplitude signals. *Animal Behaviour*, *105*, 253–265. https://doi.org/10.1016/j.anbehav.2015.04.017

Reichard, D. G., & Welklin, J. F. (2015). On the existence and potential functions of low-amplitude vocalizations in North American birds. *The Auk*, *132*(1), 156–166. https://doi.org/10.1642/AUK-14-151.1

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things): phytools: R package. *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Rogers, T. L. (2017). Calling underwater is a costly signal: Size-related differences in the call rates of Antarctic leopard seals. *Current Zoology*, *63*(4), 433–443. https://doi.org/10.1093/cz/zox028

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Russell, B. A., Cerny, F. J., & Stathopoulos, E. T. (1998). Effects of Varied Vocal Intensity on Ventilation and Energy Expenditure in Women and Men. *Journal of Speech, Language, and Hearing Research*, *41*(2), 239–248. https://doi.org/10.1044/jslhr.4102.239

Ruxton, G. D. (2009). Non-visual crypsis: A review of the empirical evidence for camouflage to senses other than vision. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1516), 549–557. https://doi.org/10.1098/rstb.2008.0228

Ruxton, G. D., Allen, W. L., Sherratt, T. N., & Speed, M. P. (2018). *Avoiding attack: The evolutionary ecology of crypsis, aposematism, and mimicry* (Second edition). Oxford University Press.

Safryghin, A. (2019). *Language Laws in Chimpanzee Gestures: Insights from Sexual Solicitations* [Unpublished MSc Thesis]. University of Saint Andrews.

Sainburg, T., Thielk, M., & Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLOS Computational Biology*, *16*(10), e1008228. https://doi.org/10.1371/journal.pcbi.1008228

Scarantino, A., & Clay, Z. (2015). Contextually variable signals can be functionally referential. *Animal Behaviour*, *100*, e1–e8. https://doi.org/10.1016/j.anbehav.2014.08.017

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. https://doi.org/10.1016/j.anbehav.2016.10.003

Schamberg, I., Cheney, D. L., Clay, Z., Hohmann, G., & Seyfarth, R. M. (2017). Bonobos use call combinations to facilitate inter-party travel recruitment. *Behavioral Ecology and Sociobiology*, *71*(4), 75. https://doi.org/10.1007/s00265-017-2301-9

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. https://doi.org/10.1016/j.anbehav.2010.07.012

Schlenker, P., Chemla, E., Schel, A. M., Fuller, J., Gautier, J.-P., Kuhn, J., Veselinović, D., Arnold, K., Cäsar, C., Keenan, S., Lemasson, A., Ouattara, K., Ryder, R., & Zuberbühler, K. (2016). Formal monkey linguistics. *Theoretical Linguistics*, *42*(1–2), 1–90. https://doi.org/10.1515/tl-2016-0001

Scott-Phillips, T. C. (2016). Meaning in great ape communication: Summarising the debate. *Animal Cognition*, *19*(1), 233–238. https://doi.org/10.1007/s10071-015-0936-3

Sekine, K., Sowden, H., & Kita, S. (2015). The Development of the Ability to Semantically Integrate Information in Speech and Iconic Gesture in Comprehension. *Cognitive Science*, *39*(8), 1855–1880. https://doi.org/10.1111/cogs.12221

Semple, S., Heesen, R., Ferrer-I-Cancho, R., & Dunn, J. (In Prep). *Efficiency of Coding in Primate Vocal Repertoires*.

Semple, S., Hsu, M. J., & Agoramoorthy, G. (2010). Efficiency of coding in macaque vocal communication. *Biology Letters*, *6*(4), 469–471. https://doi.org/10.1098/rsbl.2009.1062

Semple, S., Hsu, M. J., Agoramoorthy, G., & Ferrer-i-Cancho, R. (2013). The Law of Brevity in Macaque Vocal Communication is not an Artefact of Analysing Mean Call Durations *. *Journal of Quantitative Linguistics*, *20*(3), 209–217. https://doi.org/10.1080/09296174.2013.799917

Shahzad, K., Mittenthal, J. E., & Caetano-Anollés, G. (2015). The organization of domains in proteins obeys Menzerath-Altmann's law of language. *BMC Systems Biology*, *9*(1), 44. https://doi.org/10.1186/s12918-015-0192-9

Shu, H., Chen, X., Anderson, R. C., Wu, N., & Xuan, Y. (2003). Properties of School Chinese: Implications for Learning to Read. *Child Development*, *74*(1), 27–47. https://doi.org/10.1111/1467-8624.00519

Sievers, C., & Gruber, T. (2016). Reference in human and non-human primate communication: What does it take to refer? *Animal Cognition*, *19*(4), 759–768. https://doi.org/10.1007/s10071-016-0974-5

Sigurd, B., Eeg-Olofsson, M., & van Weijer, J. (2004). Word length, sentence length and frequency—Zipf revisited. *Studia Linguistica*, *58*(1), 37–52. https://doi.org/10.1111/j.0039-3193.2004.00109.x

Slocombe, K. E., Waller, B. M., & Liebal, K. (2011). The language void: The need for multimodality in primate communication research. *Animal Behaviour*, *81*(5), 919–924. https://doi.org/10.1016/j.anbehav.2011.02.002

So, W. C., Kita, S., & Goldin-Meadow, S. (2009). Using the Hands to Identify Who Does What to Whom: Gesture and Speech Go Hand-in-Hand. *Cognitive Science*, *33*(1), 115–125. https://doi.org/10.1111/j.1551-6709.2008.01006.x

Solé, R. V. (2010). Genome size, self-organization and DNA's dark matter. *Complexity*, *16*(1), 20–23. https://doi.org/10.1002/cplx.20326

Speakman, J. R., & Racey, P. A. (1991). No cost of echolocation for bats in flight. *Nature*, *350*(6317), 421–423. https://doi.org/10.1038/350421a0

Strauss, U., Grzybek, P., & Altmann, G. (2007). *Word length and word frequency*. Springer.

Tamaoka, K., & Kiyama, S. (2013). The effects of visual complexity for Japanese kanji processing with high and low frequencies. *Reading and Writing*, *26*(2), 205–223. https://doi.org/10.1007/s11145-012-9363-x

Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B., & Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour*, *59*(6), 1167–1176. https://doi.org/10.1006/anbe.1999.1416

Teahan, W. J., Wen, Y., McNab, R., & Witten, I. H. (2000). A Compression-based Algorithm for Chinese Word Segmentation. *Computational Linguistics*, *26*(3), 375–393. https://doi.org/10.1162/089120100561746

Teufel, C., Gutmann, A., Pirow, R., & Fischer, J. (2010). Facial expressions modulate the ontogenetic trajectory of gaze-following among monkeys: Gaze-following in Barbary macaques. *Developmental Science*, *13*(6), 913–922. https://doi.org/10.1111/j.1467-7687.2010.00956.x

Teupenhayn, R., & Altmann, G. (1984). Clause length and Menzerath's law. *Glotometrika*, *6*, 127–138.

Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition* (1. Harvard Univ. Press paperback ed). Harvard Univ. Press.

Tomasello, M. (2009). Universal grammar is dead. *Behavioral and Brain Sciences*, *32*(5), 470–471. https://doi.org/10.1017/S0140525X09990744

Tomasello, M. (2010). *Origins of human communication* (1. MIT paperback ed). MIT Press.

Torre, I. G., Luque, B., Lacasa, L., Kello, C. T., & Hernández-Fernández, A. (2019). On the physical origin of linguistic laws and lognormality in speech. *Royal Society Open Science*, *6*(8), 191023. https://doi.org/10.1098/rsos.191023

Townsend, S. W., Deschner, T., & Zuberbühler, K. (2008). Female Chimpanzees Use Copulation Calls Flexibly to Prevent Social Competition. *PLoS ONE*, *3*(6), e2431. https://doi.org/10.1371/journal.pone.0002431

Townsend, S. W., & Manser, M. B. (2013). Functionally Referential Communication in Mammals: The Past, Present and the Future. *Ethology*, *119*(1), 1–11. https://doi.org/10.1111/eth.12015

Trivers, R. L. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual Selection and the Descent of Man 1871-1971.* (pp. 136–179). Heinemann.

Tsao, Y.-C., & Weismer, G. (1997). Interspeaker Variation in Habitual Speaking Rate: Evidence for a Neuromuscular Component. *Journal of Speech, Language, and Hearing Research*, *40*(4), 858–866. https://doi.org/10.1044/jslhr.4004.858

Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, *124*, 203–220. https://doi.org/10.1016/j.anbehav.2016.12.005

Van Staaden, M., & Römer, H. (1997). Sexual signalling in bladder grasshoppers: Tactical design for maximizing calling range. *The Journal of Experimental Biology*, *200*(Pt 20), 2597–2608.

Vargas-Castro, L. E., Sandoval, L., & Searcy, W. A. (2017). Eavesdropping avoidance and sound propagation: The acoustic structure of soft song. *Animal Behaviour*, *134*, 113–121. https://doi.org/10.1016/j.anbehav.2017.10.008

Waller, B. M., Caeiro, C. C., & Davila-Ross, M. (2015). Orangutans modify facial displays depending on recipient attention. *PeerJ*, *3*, e827. https://doi.org/10.7717/peerj.827

Waller, B. M., Warmelink, L., Liebal, K., Micheletta, J., & Slocombe, K. E. (2013). Pseudoreplication: A widespread problem in primate communication research. *Animal Behaviour*, *86*(2), 483–488. https://doi.org/10.1016/j.anbehav.2013.05.038

Wang, Y., & Chen, X. (2015). Structural complexity of simplified chinese characters. In A. Tuzzi, M. Benešová, & J. Macutek (Eds.), *Recent contributions to quantitative linguistics*. Walter de Gruyter.

Watson, S. K., Heesen, R., Hedwig, D., Robbins, M. M., & Townsend, S. W. (2020). An exploration of Menzerath's law in wild mountain gorilla vocal sequences. *Biology Letters*, *16*(10), 20200380. https://doi.org/10.1098/rsbl.2020.0380

Wheeler, B. C., & Fischer, J. (2015). The blurred boundaries of functional reference: A response to Scarantino & Clay. *Animal Behaviour*, *100*, e9–e13. https://doi.org/10.1016/j.anbehav.2014.11.007

Wich, S. A., & de Vries, H. (2006). Male monkeys remember which group members have given alarm calls. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1587), 735–740. https://doi.org/10.1098/rspb.2005.3320

Xu, L., & He, L. (2020). Is the Menzerath-Altmann Law Specific to Certain Languages in Certain Registers? *Journal of Quantitative Linguistics*, *27*(3), 187–203. https://doi.org/10.1080/09296174.2018.1532158

Yasukawa, K., & Bonnie, K. E. (2017). Observational and experimental methods in comparative psychology. In J. Call, G. M. Burghardt, I. M. Pepperberg, C. T. Snowdon, & T. Zentall (Eds.), *APA handbook of comparative psychology: Basic concepts, methods, neural substrate, and behavior.* (pp. 65–86). American Psychological Association. https://doi.org/10.1037/0000011-004

Yunusova, Y., Weismer, G., Kent, R. D., & Rusche, N. M. (2005). Breath-Group Intelligibility in Dysarthria: Characteristics and Underlying Correlates. *Journal of Speech, Language, and Hearing Research*, *48*(6), 1294–1310. https://doi.org/10.1044/1092-4388(2005/090)

Zipf, G. (1936). *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. George Routledge and Sons Ltd.

Zipf, G. (1949). *Human Behaviour and the Principle of Least Effort*. Addison Wesley.

Zollinger, S. A., Goller, F., & Brumm, H. (2011). Metabolic and Respiratory Costs of Increasing Song Amplitude in Zebra Finches. *PLoS ONE*, *6*(9), e23198. https://doi.org/10.1371/journal.pone.0023198

Zuberbühler, K. (2009). Chapter 8 Survivor Signals. In *Advances in the Study of Behavior* (Vol. 40, pp. 277–322). Elsevier. https://doi.org/10.1016/S0065-3454(09)40008-1

Zuberbühler, K., Jenny, D., & Bshary, R. (1999). The Predator Deterrence Function of Primate Alarm Calls: Predator Deterrence Function of Primate Alarm Calls. *Ethology*, *105*(6), 477–490. https://doi.org/10.1046/j.1439-0310.1999.00396.x

Zuidema, W., & de Boer, B. (2018). The evolution of combinatorial structure in language. *Current Opinion in Behavioral Sciences*, *21*, 138–144. https://doi.org/10.1016/j.cobeha.2018.04.011

7 Supplementary Information (Analysis Code)

7.1 Correlational Analyses

```
library(correlation)


#BATS

#Barbastelle (Barbastella barbastellus)

Barbastelle <- read.csv("~/Desktop/almost final repertoires/bat final/Barbastelle (Barbastella barbastellus).csv")

cor.test(~ Freq + CallDurMs, data = Barbastelle, method = "spearman")

correlation(Barbastelle, method = "spearman",ci = 0.95)

correlation(Barbastelle, method = "kendall",ci = 0.95)



#Bechstein’s bat (Myotis bechsteinii)

Bechstein <- read.csv("~/Desktop/almost final repertoires/bat final/Bechstein’s bat (Myotis bechsteinii).csv", comment.char="#")

cor.test(~ Freq + CallDurMs, data = Bechstein, method = "spearman")

correlation(Bechstein, method = "spearman",ci = 0.95)


#Big Brown Bat
```

```
big.brown.bat <- read.csv("~/Desktop/almost final repertoires/bat final/big brown bat.csv")
cor.test(~ Freq + CallDurMs, data = big.brown.bat, method = "spearman")
correlation(big.brown.bat, method = "spearman",ci = 0.95)

#Brandt's bat (Myotis brandtii)
Brandt <- read.csv("~/Desktop/almost final repertoires/bat final/Brandt's bat (Myotis brandtii) .csv")
cor.test(~ Freq + CallDurMs, data = Brandt, method = "spearman")
correlation(Brandt, method = "spearman",ci = 0.95)
correlation(Brandt, method = "kendall",ci = 0.95)



#Commissaris's long-tongued bat (Glossophaga commissarisi)
Commissaris <- read.csv("~/Desktop/almost final repertoires/bat final/Commissaris's long-tongued bat (Glossophaga commissarisi).csv")
cor.test(~ Freq + CallDurMs, data = Commissaris, method = "spearman")
correlation(Commissaris, method = "spearman",ci = 0.95)
correlation(Commissaris, method = "kendall",ci = 0.95)

#common pipistrelle (Pipistrellus pipistrelus)

common.pipistrelle <- read.csv("~/Desktop/almost final repertoires/bat final/common pipistrelle (Pipistrellus pipistrellus).csv", comment.char="#")
cor.test(~ Freq + CallDurMs, data = common.pipistrelle, method = "spearman")
correlation(common.pipistrelle, method = "spearman",ci = 0.95)
```

```
#Greater mouse-eared bat (Myotis myotis)

Greater.mouse.eared.bat <- read.csv("~/Desktop/almost final repertoires/bat final/Greater
mouse-eared bat (Myotis myotis).csv")
cor.test(~ Freq + CallDurMs, data = Greater.mouse.eared.bat, method = "spearman")
correlation(Greater.mouse.eared.bat, method = "spearman",ci = 0.95)
correlation(Greater.mouse.eared.bat, method = "kendall",ci = 0.95)
#greater tube-nosed bat ( Murina leucogaster)
greater.tube.nosed.bat<- read.csv("~/Desktop/almost final repertoires/bat final/greater tube-
nosed bat ( Murina leucogaster).csv")
cor.test(~ Freq + CallDurMs, data = greater.tube.nosed.bat, method = "spearman")
correlation(greater.tube.nosed.bat, method = "spearman",ci = 0.95)
correlation(greater.tube.nosed.bat, method = "kendall",ci = 0.95)

Leaf.Nosed.Bat <- read.csv("~/Desktop/almost final repertoires/bat final/Leaf-Nosed Bat,
(Carolloa perspicillata).csv")
cor.test(~ Freq + CallDurMs, data = Leaf.Nosed.Bat, method = "spearman")
correlation(greater.tube.nosed.bat, method = "spearman",ci = 0.95)
#Little Brown Bat, (Myotis lucifugus)
Little.Brown.Bat <- read.csv("~/Desktop/almost final repertoires/bat final/Little Brown Bat,
(Myotis lucifugus).csv")
 cor.test(~ Freq + CallDurMs, data = Little.Brown.Bat, method = "spearman")
 correlation(Little.Brown.Bat, method = "spearman",ci = 0.95)
 #Natterer’s bat (Myotis nattereri)
```

```
`Natterer’s.bat` <- read.csv("~/Desktop/almost final repertoires/bat final/Natterer’s bat (Myotis nattereri).csv", comment.char="#")
cor.test(~ f + d, data = `Natterer’s.bat`, method = "spearman")
correlation(`Natterer’s.bat`, method = "spearman",ci = 0.95)

#Natterer’s bat (Myotis nattereri) Schmidbauer & Denzinger
Natterers.bat.Schmidbauer <- read.csv("~/Desktop/almost final repertoires/bat final/Natterer’s bat (Myotis nattereri) Schmidbauer and Denzinger 2019.csv")
cor.test(~ Freq + CallDurMs, data = Natterers.bat.Schmidbauer, method = "spearman")
correlation(Natterers.bat.Schmidbauer, method = "spearman",ci = 0.95)

#Pale Spear-Nosed Bat (Phyllostomus discolor)
Pale.Spear.Nosed.Bat <- read.csv("~/Desktop/almost final repertoires/bat final/Pale Spear-Nosed Bat (Phyllostomus discolor).csv")
cor.test(~ Freq + CallDurMs, data = Pale.Spear.Nosed.Bat, method = "spearman")
correlation(Pale.Spear.Nosed.Bat, method = "spearman",ci = 0.95)
##Pallas's long- (Glossophaga soricina)
Pallas <- read.csv("~/Desktop/almost final repertoires/bat final/Pallas's long-tongued bat (Glossophaga soricina).csv", comment.char="#")
cor.test(~ Freq + CallDurMs, data = Pallas, method = "spearman")
correlation(Pallas, method = "spearman",ci = 0.95)
correlation(Pallas, method = "kendall",ci = 0.95)
```

```
 big.footed.myotis <- read.csv("~/Desktop/almost final repertoires/bat final/big-footed myotis (Myotis macrodactylus) .csv")
 cor.test(~ Freq + CallDurMs, data = big.footed.myotis, method = "spearman")
 correlation(big.footed.myotis, method = "spearman",ci = 0.95)



 Parnells.mustached.bat <- read.csv("~/Desktop/Parnells mustached bat.csv")
 cor.test(~ Freq + CallDurMs, data = Parnells.mustached.bat, method = "spearman")
 correlation(Parnells.mustached.bat, method = "spearman",ci = 0.95)
 correlation(Parnells.mustached.bat, method = "kendall",ci = 0.95)

Asian.parti.colored.bat <- read.csv("~/Desktop/Asian parti-colored bat (Vespertilio sinensis).csv")
 cor.test(~ Freq + CallDurMs, data = Asian.parti.colored.bat, method = "spearman")
 correlation(Asian.parti.colored.bat, method = "spearman",ci = 0.95)



 #CARNIVORES
 North.Atlantic.right.whale.Trygonis <- read.csv("~/Desktop/almost final repertoires/carnivore final/North Atlantic right whale (Eubalaena glacialis) Trygonis .csv")


 Siberian.Ferret <- read.csv("~/Desktop/almost final repertoires/carnivore final/ Siberian Ferret (Mustela eversmanni).csv")
```

```
 Rissos.dolphin <- read.csv("~/Desktop/almost final repertoires/carnivore final/Rissos
dolphin (Grampus griseus).csv", comment.char="#")
 cor.test(~ Freq + CallDurMs, data = Rissos.dolphin, method = "spearman")
 correlation(Rissos.dolphin, method = "spearman",ci = 0.95)
 correlation(Arctic.seal, method = "kendall",ci = 0.95)

 Canadian.harbour.seal <- read.csv("~/Desktop/almost final repertoires/carnivore
final/Canadian harbour seal (Phoca vitulina).csv")
 cor.test(~ Freq + CallDurMs, data = Canadian.harbour.seal, method = "spearman")
 correlation(Canadian.harbour.seal, method = "spearman",ci = 0.95)

 bearded.seal.Risch <- read.csv("~/Desktop/bearded seal Risch.csv")
 cor.test(~ Freq + CallDurMs, data = bearded.seal.Risch, method = "spearman")
 correlation(bearded.seal.Risch, method = "spearman",ci = 0.95)

 bearded.seal.frouin <- read.delim("~/Desktop/almost final repertoires/bearded seal frouin.csv
")
 cor.test(~ f + d, data = bearded.seal.frouin, method = "spearman")
 correlation(bearded.seal.frouin, method = "spearman",ci = 0.95)
 correlation(bearded.seal.frouin, method = "kendall",ci = 0.95)
```

```
 bearded.seal.jones <- read.delim("~/Desktop/ECL and EM Compression formulae/ECL Stuart/bearded seal jones.txt")
 cor.test(~ f + d, data = bearded.seal.jones, method = "spearman")
 correlation(bearded.seal.jones, method = "spearman",ci = 0.95)


 southern.right.whale <- read.csv("~/Desktop/almost final repertoires/carnivore final/southern right whale (Eubalaena australis) .csv", comment.char="#")
 cor.test(~ Freq + CallDurMs, data = southern.right.whale, method = "spearman")
 correlation(southern.right.whale, method = "spearman",ci = 0.95)
 correlation(southern.right.whale, method = "kendall",ci = 0.95)


 Arctic.seal<- read.csv("~/Desktop/almost final repertoires/carnivore final/Arctic seal (Erignathus barbatus).csv")
 cor.test(~ Freq + CallDurMs, data = Arctic.seal, method = "spearman")
 correlation(Arctic.seal, method = "spearman",ci = 0.95)
 correlation(Arctic.seal, method = "kendall",ci = 0.95)

 Australian.fur.seal <- read.csv("~/Desktop/almost final repertoires/carnivore final/Australian fur seals, (Pusillus doriferus).csv")
 cor.test(~ Freq + CallDurMs, data = Australian.fur.seal, method = "spearman")
 correlation(Australian.fur.seal, method = "spearman",ci = 0.95)
```

```
 Bush.Dog <- read.csv("~/Desktop/almost final repertoires/carnivore final/Bush Dog (Speothos venaticus).csv")
 cor.test(~ Freq + CallDurMs, data = Bush.Dog, method = "spearman")
 correlation(Bush.Dog, method = "spearman",ci = 0.95)
 correlation(Bush.Dog, method = "kendall",ci = 0.95)



 cheehtah.volodina <- read.csv("~/Desktop/almost final repertoires/carnivore final/cheehtah volodina.csv")
 cor.test(~ Freq + CallDurMs, data = cheehtah.volodina, method = "spearman")
 correlation(cheehtah.volodina, method = "spearman",ci = 0.95)
 correlation(cheehtah.volodina, method = "kendall",ci = 0.95)

 cheehtah.smirnova <- read.csv("~/Desktop/almost final repertoires/carnivore final/cheehtah smirnova.csv")
 cor.test(~ Freq + CallDurMs, data = cheehtah.smirnova, method = "spearman")
 correlation(cheehtah.smirnova, method = "spearman",ci = 0.95)



 Crab.Eating.Fox <- read.csv("~/Desktop/almost final repertoires/carnivore final/Crab-Eating Fox (Cerdocyon Thous).csv")
 cor.test(~ Freq + CallDurMs, data = Crab.Eating.Fox, method = "spearman")
 correlation(Crab.Eating.Fox, method = "spearman",ci = 0.95)
```

```
 Dhole <- read.csv("~/Desktop/almost final repertoires/carnivore final/Dhole (Cuon alpinus).csv")
 cor.test(~ Freq + CallDurMs, data = Dhole, method = "spearman")
 correlation(Dhole, method = "spearman",ci = 0.95)
 correlation(Dhole, method = "kendall",ci = 0.95)


 Eastern.quoll <- read.csv("~/Desktop/almost final repertoires/carnivore final/Eastern quoll (Dasyurus viverrinus).csv")
 cor.test(~ Freq + CallDurMs, data = Eastern.quoll, method = "spearman")
 correlation(Eastern.quoll, method = "spearman",ci = 0.95)


 European.Badger <- read.csv("~/Desktop/almost final repertoires/carnivore final/European Badger (Meles meles).csv")
 cor.test(~ Freq + CallDurMs, data = European.Badger, method = "spearman")
 correlation(European.Badger, method = "spearman",ci = 0.95)
 correlation(European.Badger, method = "kendall",ci = 0.95)

 giant.otter.leuchctenberger <- read.csv("~/Desktop/almost final repertoires/carnivore final/giant otter (Pteronura brasiliensis) leuchctenberger.csv")
 cor.test(~ Freq + CallDurMs, data = giant.otter.leuchctenberger, method = "spearman")
 correlation(giant.otter.leuchctenberger, method = "spearman",ci = 0.95)
 correlation(giant.otter.leuchctenberger, method = "kendall",ci = 0.95)
```

```
 giant.otter.Mumm.and.Knornschild. <- read.csv("~/Desktop/almost final
repertoires/carnivore final/giant otter (Pteronura brasiliensis) Mumm and Knornschild .csv")
 cor.test(~ Freq + CallDurMs, data = giant.otter.Mumm.and.Knornschild., method =
"spearman")
 correlation(giant.otter.Mumm.and.Knornschild., method = "spearman",ci = 0.95)
 correlation(giant.otter.Mumm.and.Knornschild., method = "kendall",ci = 0.95)


 North.American.River.Otter <- read.csv("~/Desktop/almost final repertoires/carnivore
final/North American River Otter (Lontra canadensis).csv")
 cor.test(~ Freq + CallDurMs, data = North.American.River.Otter, method = "spearman")
 correlation(North.American.River.Otter, method = "spearman",ci = 0.95)
 correlation(North.American.River.Otter, method = "kendall",ci = 0.95)


 Grey.seal <- read.csv("~/Desktop/almost final repertoires/carnivore final/Grey seal
(Halichoerus grypus).csv")
 cor.test(~ Freq + CallDurMs, data = Grey.seal, method = "spearman")
 correlation(Grey.seal, method = "spearman",ci = 0.95)


 Hooded.Seal <- read.csv("~/Desktop/almost final repertoires/carnivore final/Hooded Seal
(Cystophora cristata).csv")
```

```
 cor.test(~ Freq + CallDurMs, data = Hooded.Seal, method = "spearman")
 correlation(Hooded.Seal, method = "spearman",ci = 0.95)
 correlation(Hooded.Seal, method = "kendall",ci = 0.95)

 Raccoon <- read.csv("~/Desktop/almost final repertoires/carnivore final/Raccoon (Procyon lotor).csv")
 cor.test(~ Freq + CallDurMs, data = Raccoon, method = "spearman")
 correlation(Raccoon, method = "spearman",ci = 0.95)
 correlation(Raccoon, method = "kendall",ci = 0.95)

 Ribbon.seal.jones <- read.csv("~/Desktop/almost final repertoires/carnivore final/Ribbon seal (Histriophoca fasciata).csv")
 cor.test(~ Freq + CallDurMs, data = Ribbon.seal.jones, method = "spearman")
 correlation(Ribbon.seal.jones, method = "spearman",ci = 0.95)



 Ringed.Seal <- read.csv("~/Desktop/almost final repertoires/carnivore final/Ringed Seal (Pusa hispida).csv")
  cor.test(~ Freq + CallDurMs, data = Ringed.Seal, method = "spearman")
  correlation(Ringed.Seal, method = "spearman",ci = 0.95)
  correlation(Ringed.Seal, method = "kendall",ci = 0.95)

 sea.otter <- read.csv("~/Desktop/almost final repertoires/carnivore final/sea otter (Enhydra lutris).csv")
 cor.test(~ Freq + CallDurMs, data = sea.otter, method = "spearman")
```

```
 correlation(sea.otter, method = "spearman",ci = 0.95)
 correlation(sea.otter, method = "kendall",ci = 0.95)

 South.American.fur.seal <- read.csv("~/Desktop/almost final repertoires/carnivore final/South American fur seal (Arctocephalus australis).csv")
 cor.test(~ Freq + CallDurMs, data = South.American.fur.seal, method = "spearman")
 correlation(South.American.fur.seal, method = "spearman",ci = 0.95)
 correlation(South.American.fur.seal, method = "kendall",ci = 0.95)



 Sumatran.tiger <- read.csv("~/Desktop/almost final repertoires/carnivore final/Sumatran tiger (Panthera tigris).csv")
 cor.test(~ Freq + CallDurMs, data = Sumatran.tiger, method = "spearman")
 correlation(Sumatran.tiger, method = "spearman",ci = 0.95)




 swift.fox <- read.csv("~/Desktop/almost final repertoires/carnivore final/swift fox (Vulpes velox).csv")
 cor.test(~ Freq + CallDurMs, data = swift.fox, method = "spearman")
 correlation(swift.fox, method = "spearman",ci = 0.95)
 correlation(swift.fox, method = "kendall",ci = 0.95)

 Weddell.Seal <- read.csv("~/Desktop/almost final repertoires/carnivore final/Weddell Seal (Leptonychotes weddellii).csv")
```

```
 cor.test(~ Freq + CallDurMs, data = Weddell.Seal, method = "spearman")
 correlation(Weddell.Seal, method = "spearman",ci = 0.95)
 correlation(Weddell.Seal, method = "kendall",ci = 0.95)


 Yellow.Mongoose <- read.csv("~/Desktop/almost final repertoires/carnivore final/Yellow Mongoose (Cynictis penicillata).csv")
 cor.test(~ Freq + CallDurMs, data = Yellow.Mongoose, method = "spearman")
 correlation(Yellow.Mongoose, method = "spearman",ci = 0.95)

 # OTHER MAMMALS



 Beluga.Belikov <- read.csv("~/Desktop/almost final repertoires/mammal final/Beluga Belikov.csv")
 cor.test(~ Freq + CallDurMs, data = Beluga.Belikov, method = "spearman")
 correlation(Beluga.Belikov, method = "spearman",ci = 0.95)
 correlation(Beluga.Belikov, method = "kendall",ci = 0.95)


 Beluga.karlsen <- read.csv("~/Desktop/almost final repertoires/mammal final/Beluga karlsen.csv")

 cor.test(~ Freq + CallDurMs, data = Beluga.karlsen, method = "spearman")
```

```
 correlation(Beluga.karlsen, method = "spearman",ci = 0.95)
 correlation(Beluga.karlsen, method = "kendall",ci = 0.95)


 Degu <- read.csv("~/Desktop/almost final repertoires/mammal final/Degu (Octodon degus).csv")
 cor.test(~ Freq + CallDurMs, data = Degu, method = "spearman")
 correlation(Degu, method = "spearman",ci = 0.95)
 correlation(Degu, method = "kendall",ci = 0.95)


 Black.rhino <- read.csv("~/Desktop/almost final repertoires/mammal final/Black rhino (Diceros bicornis).csv")
 cor.test(~ Freq + CallDurMs, data = Black.rhino, method = "spearman")
 correlation(Black.rhino, method = "spearman",ci = 0.95)
 correlation(Black.rhino, method = "kendall",ci = 0.95)

 Humpback.whale <- read.csv("~/Desktop/almost final repertoires/mammal final/Humpback whale (Megaptera novaeangliae).csv")
 cor.test(~ Freq + CallDurMs, data = Humpback.whale, method = "spearman")
 correlation(Humpback.whale, method = "spearman",ci = 0.95)
 correlation(Humpback.whale, method = "kendall",ci = 0.95)
```

```
 Killer.whale <- read.csv("~/Desktop/almost final repertoires/mammal final/Killer whale
(Orcinus orca).csv")
 cor.test(~ Freq + CallDurMs, data = Killer.whale, method = "spearman")
 correlation(Killer.whale, method = "spearman",ci = 0.95)
 correlation(Killer.whale, method = "kendall",ci = 0.95)
```

```
 piebald.shrew <- read.csv("~/Desktop/almost final repertoires/mammal final/piebald shrew
Diplomesodon pulchellum).csv")
 cor.test(~ Freq + CallDurMs, data =piebald.shrew, method = "spearman")
 correlation(piebald.shrew, method = "spearman",ci = 0.95)
```

```
 silvery.mole.rat <- read.csv("~/Desktop/almost final repertoires/mammal final/silvery mole-
rat (Heliophobius argenteocinereus).csv")
 cor.test(~ Freq + CallDurMs, data = silvery.mole.rat, method = "spearman")
 correlation(silvery.mole.rat, method = "spearman",ci = 0.95)
 correlation(silvery.mole.rat, method = "kendall",ci = 0.95)
```

```
 European.squirrel <- read.csv("~/Desktop/almost final repertoires/mammal
final/Spermophilus  citellus (European  squirrel).csv")
```

```
cor.test(~ Freq + CallDurMs, data = European.squirrel, method = "spearman")
correlation(European.squirrel, method = "spearman",ci = 0.95)
```

```
 Yellow.squirrel <- read.csv("~/Desktop/almost final repertoires/mammal final/Spermophilus fulvus (Yellow  squirrel).csv")
 cor.test(~ Freq + CallDurMs, data = Yellow.squirrel, method = "spearman")
 correlation(Yellow.squirrel, method = "spearman",ci = 0.95)
```

```
 Speckled.squirrel <- read.csv("~/Desktop/almost final repertoires/mammal final/Spermophilus  suslicus (Speckled squirrel).csv")
 cor.test(~ Freq + CallDurMs, data = Speckled.squirrel, method = "spearman")
 correlation(Speckled.squirrel, method = "spearman",ci = 0.95)
```

```
 Spotted.paca <- read.csv("~/Desktop/almost final repertoires/mammal final/Spotted paca (Cuniculus paca).csv")
 cor.test(~ Freq + CallDurMs, data = Spotted.paca, method = "spearman")
 correlation(Spotted.paca, method = "spearman",ci = 0.95)
 correlation(Spotted.paca, method = "kendall",ci = 0.95)
```

```
 Wild.cavy <- read.csv("~/Desktop/almost final repertoires/mammal final/Wild cavy (Cavia aperea).csv")
 cor.test(~ Freq + CallDurMs, data = Wild.cavy, method = "spearman")
 correlation(Wild.cavy, method = "spearman",ci = 0.95)
 correlation(Wild.cavy, method = "kendall",ci = 0.95)



 Florida.manatee <- read.csv("~/Desktop/almost final repertoires/mammal final/Florida manatee (Trichechus manatus latirostris).csv")
 cor.test(~ Freq + CallDurMs, data = Florida.manatee, method = "spearman")
 correlation(Florida.manatee, method = "spearman",ci = 0.95)
 correlation(Florida.manatee, method = "kendall",ci = 0.95)



 feathertail.glider <- read.csv("~/Desktop/almost final repertoires/mammal final/feathertail glider (Acrobates pygmaeus).csv")
 cor.test(~ Freq + CallDurMs, data = feathertail.glider, method = "spearman")
 correlation(feathertail.glider, method = "spearman",ci = 0.95)
 correlation(feathertail.glider, method = "kendall",ci = 0.95)



 Tree.Shrew <- read.csv("~/Desktop/almost final repertoires/mammal final/Tree Shrew.csv")
```

```
 cor.test(~ Freq + CallDurMs, data = Tree.Shrew, method = "spearman")
 correlation(Tree.Shrew, method = "spearman",ci = 0.95)
 correlation(Tree.Shrew, method = "kendall",ci = 0.95)
```

7.2 Permutation Analyses

```
data1 <- read.table ("Murina leucogaster.txt", header=T)
reps <- 100000
results <- rep(0, reps)
x <- c(data1$p)
y <- c(data1$d)
L <- sum(x*y)
print (c("real L is", L))
sortvector <- 1:length(x)
for (i in 1:reps){
  sortvector <- sample(sortvector, replace = F)
  xtemp <- x[sortvector]
  L_temp <- sum(xtemp *y)
  results[i] <- L_temp
}
hist(results)
print(c("mean", mean(results)))
is_small <- sum(results < L)
print(c("P of being so small is estimated as ", is_small/reps))
```

7.3 Phylogenenetic Analyses

```
setwd("~/Documents/phylogenetics")



library(ape)
library(phytools)
library(phylotools)
library(phangorn)
library(geiger)



# read supertree containing all mammmal species source: https://doi.org/10.1111/j.1461-0248.2009.01307.x
(trees <- read.nexus("mammalsupertree.tre"))

(tree <- trees$mammalST_MSW05_bestDates)   # Selecting first tree from multi-tree phylo object.
```

```
  plot(tree)


  # list species to be retained on tree
  prunespecies <- c("Trichechus_manatus","Cavia_aperea","Delphinapterus_leucas","Diceros_bicornis","Diplomesodon_pulchellum","Heliophobius_argenteocinereus","Megaptera_novaeangliae","Octodon_degus","Orcinus_orca","Spermophilus_fulvus",

"Spermophilus_suslicus","Tupaia_belangeri","Acrobates_pygmaeus","Acinonyx_jubatus","Cuon_alpinus","Pusa_hispida","Histriophoca_fasciata","Lontra_canadensis","Halichoerus_grypus","Cystophora_cristata",

"Cerdocyon_thous","Speothos_venaticus","Vulpes_velox","Dasyurus_viverrinus","Mustela_eversmanii","Cynictis_penicillata","Pteronura_brasiliensis","Enhydra_lutris","Leptonychotes_weddellii","Arctocephalus_australis",

"Panthera_tigris","Meles_meles","Grampus_griseus","Phoca_vitulina","Myotis_macrodactylus","Barbastella_barbastellus","Myotis_bechsteinii","Myotis_brandtii","Pipistrellus_pipistrellus","Vespertilio_sinensis",

"Eptesicus_fuscus","Glossophaga_commissarisi","Glossophaga_soricina","Myotis_myotis","Murina_leucogaster","Pteronotus_parnellii","Myotis_lucifugus","Myotis_nattereri","Phyllostomus_discolor","Erignathus_barbatus")
```

```
  (pr.tree <- keep.tip(tree, prunespecies))

  plot(pr.tree)

data <- read.csv("summary results for phylo avg-1.csv") # Read data file containing
compression values
data

##Automate the check between the tree and data using name.check##
name.check(pr.tree,data, data.names=data$Species)

#******tell R that in the dataframe the column values representing data match to the species
names in the Species column*****#

rownames(data) <- data$Species

#check normality
qqnorm(data$mean_ECL)
qqnorm(data$ECL_value)
shapiro.test(data$mean_ECL)
shapiro.test(data$ECL_value)
# Normality checks using above Shapiro test and plots fail therfore apply  log transformation

#****** Add those logged variables into your data frame*******#
```

```
data$logECL_value <- log(data$ECL_value+1)
data$logmean_ECL <- log(data$mean_ECL+1)
qqnorm(data$logmean_ECL)
qqnorm(data$logECL_value)
shapiro.test(data$logmean_ECL)
shapiro.test(data$logECL_value)
# Appears more normal, proceed with phylogenetic paired t-test

#******check logged data appear in dataframe*******#
data

# Run T-Test
data_t_test <- data[, c("logmean_ECL", "logECL_value")]
data_t_test
phyl.pairedttest(pr.tree, data_t_test)
```