LINGUISTIC LAWS IN PRIMATE VOCAL COMMUNICATION

Steven Moran[*1,2], Quentin Gallot[1], Auriane Le Floch[1,3], Marco Maiolini[1], Adrian Soldati[1,4], and Klaus Zuberbühler[1,4]

[*]Corresponding Author: steven.moran@unine.ch
[1]Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland
[2]Department of Anthropology, University of Miami, Coral Gables, USA
[3]Institut des Sciences Cognitives Marc Jeannerod, CNRS, Lyon, France
[4]School of Psychology and Neuroscience, University of St Andrews, St Andrews, UK

Researchers have increasingly focused on the commonalities between animal communication and language with the aim to identify what makes language unique and to shed light on how it evolved. One area of recent study is whether linguistic laws also characterize vocal communication in other animals. Given their relevance for the evolution of language, here we provide a comprehensive review of the presence or absence of linguistic laws reported in studies on primate vocal communication. We find that primate vocal production follows Menzerath's law, while adherence to the other laws is mixed. Moreover, we raise three important points to consider when studying linguistic laws: the role of sexual dimorphism in vocal production, the criteria used to define a vocal sequence, and the choice of vocal units for analysis. Thus, this review provides a road map for future studies investigating linguistic laws in primates and is aimed at making results more comparable across species and signals.

1. Introduction

In ancient history, there was a great deal of interest in the communication systems of animals (Fögen, 2014). In contrast, in the last two centuries linguists and biologists have increasingly focused on the commonalities between animal communication systems and language. The aim is to identify what makes language unique and to shed light on its origin and how it evolved.

The best known account is Hockett's (1959, 1960) universal design features of human language, which among others include: semanticity, discreteness, and arbitrariness. That is, vocalizations (or signs) bear meaning, are made up of discrete elements (e.g., consonants and vowels in words), and those elements are typically arbitrary in terms of their mapping between their form and meaning. For instance, there is nothing iconic or onomatopoeic about the speech sounds in the various words for *Felis catus*, e.g., 'cat', 'chat' and 'Katze'.

Hockett used the design features to distinguish between animal communication systems and language, noting that most of them (originally 9 out of 13) are attested in other species. Displacement (the ability to communicate things in remote space or time), productivity (the ability to communicate information that has never been said before and yet still be understood), and duality of patterning[1] (meaningful messages are comprised of combinations of meaningless elements, e.g., ‘cat’ vs ‘act’ vs ‘tack’) were deemed to be unique to language.[2]

Ever since Hockett proposed the design features, they have been used to juxtapose human language with communication systems of other animals for descriptive purposes, even though the study of language evolution itself has evolved in the last sixty years. Their use is limited as an evolutionary theoretical framework (Wacewicz & Żywiczyński, 2015). Furthermore, the increased access to large datasets and machine learning has given rise to the application of quantitative and information-theoretic approaches in biological systems, including those that characterize language and that have long been known in linguistics.

One area of study to emerge in the last two decades by biologists is the exploration of linguistic laws in biological systems, including at the molecular and organismal levels, and in studies on species abundance (see Semple, Cancho, and Gustison (2021) and references inter alia). As the authors point out, at the level of organisms much research focuses on animal communication, specifically the vocal repertoires and vocalizations of birds and of marine and terrestrial mammals. As reported in these studies, current findings are mixed as to whether various properties of vocal repertoires adhere to statistical regularities that appear in language.

Here we provide a comprehensive review of the current state of the art of linguistic laws reported in primate vocal communication. We note in which species, signals, and contexts that there is support for the presence or absence of different linguistic laws. Our work builds on that of Semple et al. (2021) and we provide an online and openly accessible table of the current reported findings. These results shed light on, and provide a road map for, future studies on primates and other animal species to investigate communication systems, linguistic laws, and language evolution.

2. Linguistic laws

Linguistic laws are mathematically formulated statistical regularities discovered by linguists studying properties of (written) language. These patterns appear at various linguistic levels (e.g., sounds in words, words in text) and their existence is often attributed to universal principles of efficiency in human language (cf. Levshina and Moran (2021)). There are four well-studied linguistic laws.

The most well-known is colloquially referred to as Zipf’s law or Zipf’s principle of least effort (Zipf, 1936, 1949). It states that the frequency distribution of

[1] Aka, double articulation (Martinet, 1949).

[2] Two recent takes aiming to elucidate the unique aspects of language in the animal kingdom, or the lack thereof, are Van Schaik (2016) and Golston (2018). Van Schaik notes that questions, negation, and the use of phonemes and recursion (i.e., productivity) are notably absent in the extant primates. Golston argues that grammaticalized features of languages, such as number, person, and gender, originate in cognitive capacities of other animals, and as such, have a very deep evolutionary past.

words in a language is inversely proportional to their frequency rank. For example, in a corpus of English the highest frequency words include functional words, e.g., ‘the’, ‘to’, ‘of’, which among other things, encode grammatical structure. Zipf theorized that this distribution of words is due to the principle of least effort, i.e., speakers minimize their effort when transmitting messages (hence why the most frequent words in a language tend to also be the shortest). Work by Yu, Xu, and Liu (2018) shows that Zipf’s law holds across a sample of 50 languages. However, why the Zipfian distribution – which is one of several power law distributions – occurs is still not well understood. Power law distributions are found in numerous biological, physical, and human-made phenomena, and they emerge naturally from latent variables (Aitchison, Corradi, & Latham, 2016) or from almost nothing (Piantadosi, 2014).

Zipf’s law of abbreviation (Zipf, 1936, 1949), also known as the law of brevity, states that more frequently used words will tend to be shorter than less frequently used words within a language. This law in particular captures an aspect of the efficiency of language use, e.g., word length optimization (Kanwal, Smith, Culbertson, & Kirby, 2017), and is seen for example in lexical reductions due to clipping (e.g., Auto $<$ Automobil in German) or in the preference to use shorter synonyms. Shorter words require less biomechanical effort, e.g., car instead of automobile (Zipf, 1936, 33).

Menzerath-Altmann’s law, aka Menzerath’s law, states that the size of constituents within a linguistic construction decrease in size as constituents get longer, and vice versa (Menzerath, 1954; Altmann, 1980). For example, longer words tend to have shorter syllables and shorter morphemes, and longer sentences tend to have shorter clauses. Menzerath-Altmann’s law appears in many multi-level systems, including proteins, genes, genomes, and genetics (Nikolaou, 2014; Ferrer-I-Cancho & Forns, 2010; Shahzad, Mittenthal, & Caetano-Anollés, 2015; Sun & Caetano-Anollés, 2021).

Lastly, Herdan-Heaps’ law (aka as Heaps’ law (Heaps, 1978) in linguistics or Herdan’s law (Herdan, 1960, 1964) in information retrieval (Egghe, 2007)) is an empirical formulation of the number of distinct words within a document, or set of documents (treated as one text), as a function the text’s length. Herdan-Heaps’ law states that as the size of a text increases, the number of discoverable unique words (i.e., word types) decreases as a function of the text’s length. For example, Kornai (2002) verified Herdan-Heaps’ law in a corpus of 50 million words, i.e., the distribution the type-token ratio of words did not flatten, indicating that adding more text would lead to more unique words. Brants and Franz (2006) created the first one trillion word corpus of English by aggregating internet webpages and they showed that there were nearly 14 million word types in the corpus – again with no indication that all word types had been discovered.[3]

[3] Note that word types included *everything* in the corpus. For example, according to Google’s

Of the four linguistic laws, to the best of our knowledge only Herdan-Heaps' law has not yet been studied in animal communication systems. The other laws, however, have been claimed to exist, or not exist, in various signals in various species. In the next section, we describe the materials and methods used in our review of linguistic laws reported in primate vocal communication, before turning to the results and discussion.

3. Materials and methods

We have identified 19 published studies on linguistic laws and primate vocalizations. These include species from different taxa, such as lemurs (*Indri indri*), tarsiers (*Tarsius spectrumhurskyae*), Old World monkeys (*Macaca cyclopis*, *Theropithecus gelada*), New World monkeys (*Saimiri sciureus*, *Callithrix jacchus*, *Cacajao melanocephalus*, *Plecturocebus cupreus*), lesser apes (*Hylobates muelleri*, *Nomascus nasutus*, *Nomascus concolor*), and great apes (*Pan troglodytes schweinfurthii*, *Gorilla beringei beringei*).

For our review of the existing literature, we identified and extracted information for the following variables (together with the bibliographic metadata): species, linguistic law investigated, evidence (i.e., was the linguistic law supported or not by the data), dependent variable (e.g., units, sequences, bouts), signal investigated (e.g., pant hoots, wobbles), signal modality (e.g., wobble cycles, vocal sequences), definition and criteria of the signal (as per the authors' description), the context of production, sample size, number of subjects, sex, age, social system, size of repertoire reported, and any pertinent or interesting comments or findings.[4]

4. Results

The oldest paper that we identified is by McCowan, Doyle, and Hanser (2002), who use comparative measures from information theory to compare the development of vocal repertoires of bottle nose dolphins (*Tursiops truncatus*), squirrel monkeys (*Saimiri sciureus*), and humans (*Homo sapiens sapiens*). The authors did not test for a particular linguistic law in the sequences of vocalizations. Instead, they used the Zipf coefficient to approximate diversity and repertoire complexities and how they change during development from infant to adult. Their findings suggest that in species that are capable of vocal learning, repertoire structure diversity decreases and becomes more organized into adulthood, i.e., less entropic. Thus, they report that chuck calls of squirrel monkeys exhibit similar developmental patterns as in human language acquisition.

spell correction system, there are nearly 600 different spellings for the query 'britney spears' (`http://archive.google.com/jobs/britney.html`).

[4] `https://docs.google.com/spreadsheets/d/1aOvPk7hZSHDaXnzTS00LkqQ3vQTw_Bmxoi5Csx7kQQA/edit?usp=sharing`

Other authors use a similar approach to McCowan et al. (2002) in that they investigate power law coefficients in relation to Shannon entropy (Kershenbaum et al., 2021), study developmental trends from infancy to adulthood (Gultekin, Hildebrand, Hammerschmidt, & Hage, 2021), or compare one or more linguistic laws across different species, e.g., macaques (*Macaca cyclopis*), marmosets (*Callithrix jacchus*), and uakaris (*Cacajao melanocephalus*) (Bezerra, Souto, Radford, & Jones, 2011; Ferrer-I-Cancho & Hernández-Fernández, 2013; Ferrer-I-Cancho et al., 2013; Kershenbaum et al., 2021).

The majority of the current published research, however, investigates a linguistic law within a single species' repertoire, including chimpanzees (Fedurek, Zuberbühler, & Semple, 2017), geladas (Gustison, Semple, Cancho, & Bergman, 2016; Gustison & Bergman, 2017; Gustison, Tinsley Johnson, Beehner, & Bergman, 2019), gibbons (Clink, Ahmad, & Klinck, 2020; Clink, Tasirin, & Klinck, 2020; Huang, Ma, Ma, Garber, & Fan, 2020), gorrilas (Watson, Heesen, Hedwig, Robbins, & Townsend, 2020), indris (Valente et al., 2021; Zanoli et al., 2020), and macaques (Semple, Hsu, & Agoramoorthy, 2010; Semple, Hsu, Agoramoorthy, & Cancho, 2013). Follow-up studies on the same species, but perhaps on different aspects of vocalizations, include (some of) the same authors, e.g., the work on Menzerath's law in geladas (Gustison et al., 2016; Gustison & Bergman, 2017; Gustison et al., 2019) or Zipf's law of abbreviation in macaques, marmosets, and uakaris (Semple et al., 2010, 2013; Ferrer-I-Cancho et al., 2013; Ferrer-I-Cancho & Hernández-Fernández, 2013).

In terms of the representation of linguistic laws, we found that six publications investigate Menzerath's law (Clink & Lau, 2020; Fedurek et al., 2017; Gustison et al., 2016; Watson et al., 2020; Gustison et al., 2019; Gustison & Bergman, 2017) and six investigate Zipf's law of abbreviation (Gultekin et al., 2021; Semple et al., 2010; Bezerra et al., 2011; Ferrer-I-Cancho et al., 2013; Ferrer-I-Cancho & Hernández-Fernández, 2013; Semple et al., 2013). A further three investigate both Menzerath's law and Zipf's law of abbreviation (Clink et al., 2020; Huang et al., 2020; Valente et al., 2021).

Regarding Menzerath's law in primate vocal repertoires, there is overwhelming support reported, but with some caveats. Gustison et al. (2016) find that adult male gelada vocalizations follow Menzarath's law, i.e., longer sequences are composed of calls of shorter duration. Gustison and Bergman (2017) report that wobble cycle duration was shorter when the number of wobbles or lip smacks was greater. The authors also report that Menzerath's law was identified separately in both inhaled and exhaled wobbles and suggest therefore that the compression of vocal signals by geladas operates at multiple levels (as Menzerath's law does at different linguistic levels in language).

Menzerath's law is also found in four of eight calls by tarsiers, titi monkeys (*Plecturocebus cupreus*), and male Bornean gibbons (*Hylobates muelleri*) (Clink & Lau, 2020). These findings raise several important points for consideration when studying linguistic laws in animal communication systems (see also Semple et al. (2021)). First, sexual dimorphism may play an important part in the evolutionary development of vocal repertoires (e.g., in orangutans and gorillas, who differ greatly in size by sex). Additionally, sexes may differ with regard to the context of production, as well as a call's acoustic features. Thus, ideally the vocalizations of both sexes need to be tested to prevent bias, e.g., Odom, Hall, Riebel, Omland, and Langmore (2014). Second, at what level or which part of the vocal production is analyzed (e.g., duration of notes, proportion of call types, single vs multi-unit sequences) and how a sequence is defined, i.e., which criteria are used to separate the units from each other (e.g., length of silence gaps). Third, a linguistic law may be found in certain call types or certain levels of analysis, but not others. For example, Watson et al. (2020) investigated close-calls (double-grunt, single-grunt, grumble, hum, mixed) of mountain gorillas in sequences made up of more than one unit. The authors initially found positive evidence for Menzerath's law, but then report that the relationship was due to the difference between single and multi-unit sequences (leaving single units out of the analysis resulted in longer sequences being typically composed of longer units). Hence, Watson et al. (2020) report that close calls by mountain gorillas only partially adhere to Menzerath's law.

Other than mountain gorillas, the only non-human great ape vocalizations that have so far been studied with regard to linguistic laws are chimpanzee pant hoots. Fedurek et al. (2017) find support for Menzerath's law in the number and duration of calls within the pant hoot and also for entire vocal sequences. They also report that these findings hold between the duration of adjacent phases in the pant hoot. Pant hoots were investigated in the context of feeding and traveling and only pant hoots produced by males and that contained the climax phase were included in their investigation.

The results from studies of various species and whether their vocalizations adhere to Zipf's law of abbreviation (and Menzerath's law) are more mixed. Bezerra et al. (2011) investigate 12 call types of marmosets (excluding predator-specific alarm calls) and 7 from uakaris produced by both male and female adults across all contexts. They find no support for Zipf's law of abbreviation (see also Ferrer-I-Cancho and Hernández-Fernández (2013)). Likewise, Gultekin et al. (2021) find no support for Zipf's law of abbreviation in marmosets throughout their development.

However, Semple et al. (2010) examine the full repertoire of macaque vocalizations (*Macaca cyclopis*), in all contexts, from all age ranges and sexes. They report support for Zipf's law of abbreviation and show that more frequent vocalizations are shorter in duration. In a follow up study, Semple et al. (2013) report that the support for Zipf's law of abbreviation that they found is not an artefact of their previous analysis of mean call duration.

Support for Zipf's law of abbreviation and Menzerath's law was recently reported by Valente et al. (2021) for indris' songs. They investigated songs produced in the context of territory defense and long distance communication by males and females (sub-adults and adults). Although not an investigation of any particular linguistic law in indris, we note the work of Zanoli et al. (2020), who undertake a Levenshtein distance analysis[5] of adult male and female indri songs. They report that the songs of female indris are less stereotyped that those of males. This work again highlights the importance of evaluating both sexes and the opportunity to apply computational linguistic approaches and measures outside of linguistic laws.

Lastly, we note two studies on whether gibbons' vocalizations adhere to Zipf's law of abbreviation and Menzerath's law. Clink et al. (2020) find no support for Zipf's law of abbreviation, but they report strong support for Menzerath's law in Bornean gibbons. Huang et al. (2020) report that both laws are confirmed in male gibbon calls. These two studies again highlight the importance of which species and which calls are taken under consideration and how bouts and sequences are calculated.

Clink et al. (2020) study solo singing bouts in multi-phase vocal sequences by male Bornean gibbons. These lengthy singing bouts are comprised of a discrete number of note types and are composed of a large repertoire of phrases. Sequences of notes are calculated by two-second breaks or more. Clink et al. (2020) report no negative relationship between duration of a signal and the frequency of its use (Zipf's law). The authors also report a strong negative correlation between the number of notes in a phase and the notes' mean duration. They conclude that individual variation produces strong individual signatures.

In contrast, the study by Huang et al. (2020) focuses on western black-crested gibbons and Cao-vit gibbons. The authors study the loud morning calls of sub-adult and adult males in various contexts (resource/territorial defense, mate defense, mate attraction, group cohesion). They define a song bout as all notes in a song with silence periods of less than ten minutes, where a note is a single continuous sound produced either through inhalation or exhalation. Huang et al. (2020) report that the most common notes of the male gibbons are shortest in duration (in accordance with Zipf's law) and that longer sequences consist on average of shorter notes (i.e., Menzerath's law).

5. Discussion

We conducted a comprehensive review of the current published findings regarding the presence or absence of so-called linguistic laws in the vocal communication systems of primates. The exploration of linguistic laws outside of the human

[5] Levenshtein distance is a measure of the difference between two sequences. Although its roots are in mathematics, it is often used in computer science, information retrieval, and linguistics, to calculate the pairwise similarity between two strings of texts for applications such as spell checking and fuzzy matching, e.g., 'britney spears' vs 'brittany spears' has a Levenshtein distance of 3.

language is a relatively new area of research, which aims to shed light on the similarities and differences with regard to the statistical regularities present in animal communication systems and in biological systems more broadly (Semple et al., 2021)

Here we focus on reported findings of linguistic laws in the vocalization systems of primates because of its potential link to the origin and evolution of spoken language. However, we note that this review is restricted because it does not include studies on gestures, facial expressions, or multi-modal systems (cf. Liebal, Slocombe, and Waller (2022)). Although such work is beyond the scope of this paper, we have nevertheless begun to collect this information for future research (e.g., Hobaiter and Byrne (2014), Heesen, Hobaiter, Cancho, and Semple (2019)), as we plan to integrate more animal species and communicative signals for meta-analysis (Gurevitch, Koricheva, Nakagawa, & Stewart, 2018).

What our review of the existing literature says about the vocal communication systems of primates and whether or not they adhere to linguistic laws is mixed. We find that studies of primate vocalizations and Menzerath's law are in general supportive. However, investigations of Zipf's law of abbreviation in primates are more mixed, i.e., different studies support, or do not support, brevity in primate vocalization systems. Nevertheless, we find that the studies surveyed so far highlight many areas for further research. For example, our findings confirm that within the research of linguistic laws in primate vocalizations, most studies are descriptive in nature. Therefore, we encourage future researchers to consider the four levels of analysis raised by Semple et al. (2021) – i.e., exploratory and descriptive statistics, descriptive mathematical modeling, inferential and predictive mathematical modeling, and general theory – and what questions they aim to address. Nevertheless, descriptive studies are critical for developing future theories about how animal communication systems and languages have evolved independently after the split from a common ancestor.

Finally, while we acknowledge the disparate nature of the studies reviewed here, there are issues and limitations regarding the comparability of species, vocalization types (e.g., duration, number of calls per phase), choice of dependent variables, sexual differences, context of call production, age range, among others. To address these issues, we provide a list in the supplementary materials in which we highlight differences in authors' definitions of units and sequences, the number of subjects in their sample, sample size of vocalizations, and discrepancies in reported repertoire sizes. Moreover, we note that how to test for linguistic laws is still an area of debate even within linguistics, e.g., Zipf's distribution (Piantadosi, 2014). Thus, we hope that studies on animal communication will strengthen this research avenue and ultimately help to elucidate how language evolved in our species.

6. Supplementary Materials

A full table of results is available at: `https://docs.google.com/spreadsheets/d/1aOvPk7hZSHDaXnzTS00LkqQ3vQTw_Bmxoi5Csx7kQQA/edit?usp=sharing`.

Contributions

SM designed research. SM, QG, ALF, MM, & AS identified variables, collected data, and performed analysis. SM wrote the paper. All authors read, contributed to, and approved the paper. KZ & SM acquired funding.

Acknowledgements

SM & MM were funded by the Swiss National Science Foundation (grant #186841). ALF & QG were funded by the University of Neuchâtel. KZ, ALF & QG were funded by the Swiss National Science Foundation (grant #185324).

References

Aitchison, L., Corradi, N., & Latham, P. E. (2016). Zipf's law arises naturally when there are underlying, unobserved variables. *PLOS Computational Biology*, *12*(12), e1005110.

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

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.

Brants, T., & Franz, A. (2006). Web 1T 5-gram Version 1 (2006). *Linguistic Data Consortium, Philadelphia*.

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.

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.

Clink, D. J., Tasirin, J. S., & Klinck, H. (2020). Vocal individuality and rhythm in male and female duet contributions of a nonhuman primate. *Current Zoology*, *66*(2), 173–186.

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.

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.

Ferrer-I-Cancho, R., & Forns, N. (2010). The self-organization of genomes. *Complexity*, *15*(5), 34–36.

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), 45–55.

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.

Fögen, T. (2014). Animal communication. In G. L. Campbell (Ed.), *The Oxford Handbook of Animals in Classical Thought and Life* (pp. 216–232). Oxford University Press.

Golston, C. (2018). Phi-features in animal cognition. *Biolinguistics*, *12*, 055–098.

Gultekin, Y. B., Hildebrand, D. G. C., Hammerschmidt, K., & Hage, S. R. (2021). High plasticity in marmoset monkey vocal development from infancy to adulthood. *Science Advances*, *7*(27), eabf2938.

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, *555*(7695), 175–182.

Gustison, M. L., & Bergman, T. (2017). Divergent acoustic properties of gelada and baboon vocalizations and their implications for the evolution of human speech. *Journal of Language Evolution*, *2*(1), 20–36.

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

Gustison, M. L., Tinsley Johnson, E., Beehner, J. C., & Bergman, T. J. (2019). The social functions of complex vocal sequences in wild geladas. *Behavioral Ecology and Sociobiology*, *73*(1), 14.

Heaps, H. S. (1978). *Information retrieval, computational and theoretical aspects.* Academic Press.

Heesen, R., Hobaiter, C., Cancho, R. Ferrer-i, & Semple, S. (2019). Linguistic laws in chimpanzee gestural communication. *Proceedings of the Royal Society B*, *286*(1896), 20182900.

Herdan, G. (1960). *Type-token mathematics.* Mouton.

Herdan, G. (1964). *Quantitative linguistics.* Butterworth.

Hobaiter, C., & Byrne, R. W. (2014). The meanings of chimpanzee gestures. *Current Biology*, *24*(14), 1596–1600.

Hockett, C. F. (1959). Animal "languages" and human language. *Human Biology*, *31*(1), 32–39.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*, 89–96.

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*, *160*, 145-155.

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.

Kershenbaum, A., Demartsev, V., Gammon, D. E., Geffen, E., Gustison, M. L., Ilany, A., & Lameira, A. R. (2021). Shannon entropy as a robust estimator of Zipf's law in animal vocal communication repertoires. *Methods in Ecology and Evolution*, *12*(3), 553–564.

Kornai, A. (2002). How many words are there? *Glottometrics*, *4*, 61–86.

Levshina, N., & Moran, S. (2021). Efficiency in human languages: Corpus evidence for universal principles. *Linguistics Vanguard*, *7*(S3).

Liebal, K., Slocombe, K. E., & Waller, B. M. (2022). The language void 10 years on: multimodal primate communication research is still uncommon. *Ethology Ecology & Evolution*, *34*(3), 274–287.

Martinet, A. (1949). La double articulation linguistique. *Travaux du Cercle linguistique de Copenhague*, *5*, 30–37.

McCowan, B., Doyle, L. R., & Hanser, S. F. (2002). Using information theory to assess the diversity, complexity, and development of communicative repertoires. *Journal of Comparative Psychology*, *116*(2), 166–172.

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

Nikolaou, C. (2014). Menzerath–Altmann law in mammalian exons reflects the dynamics of gene structure evolution. *Computational Biology and Chemistry*, *53*, 134–143.

Odom, K. J., Hall, M. L., Riebel, K., Omland, K. E., & Langmore, N. E. (2014). Female song is widespread and ancestral in songbirds. *Nature Communications*, *5*(1), 1–6.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin and Review*, *21*(5), 1112–1130.

Semple, S., Cancho, R. Ferrer-i, & Gustison, M. L. (2021). Linguistic laws in biology. *Trends in Ecology & Evolution*, *37*(1), 53–66.

Semple, S., Hsu, M., Agoramoorthy, G., & Cancho, R. Ferrer-i. (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.

Semple, S., Hsu, M. J., & Agoramoorthy, G. (2010). Efficiency of coding in macaque vocal communication. *Biology Letters*, *6*(4), 469–471.

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), 1–13.

Sun, F., & Caetano-Anollés, G. (2021). Menzerath–Altmann's law of syntax in RNA accretion history. *Life*, *11*(6), 489.

Valente, D., De Gregorio, C., Favaro, L., Friard, O., Miaretsoa, L., Raimondi, T., Ratsimbazafy, J., Torti, V., Zanoli, A., Giacoma, C., & Gamba, M. (2021).

Linguistic laws of brevity: conformity in *Indri indri*. *Animal Cognition*, *24*(4), 897–906.

Van Schaik, C. (2016). *The primate origins of human nature.* John Wiley & Sons, Ltd.

Wacewicz, S., & Żywiczyński, P. (2015). Language evolution: Why Hockett's design features are a non-starter. *Biosemiotics*, *8*(1), 29–46.

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.

Yu, S., Xu, C., & Liu, H. (2018). Zipf's law in 50 languages: its structural pattern, linguistic interpretation, and cognitive motivation. *arXiv preprint arXiv:1807.01855*.

Zanoli, A., De Gregorio, C., Valente, D., Torti, V., Bonadonna, G., Randrianarison, R. M., Giacoma, C., & Gamba, M. (2020). Sexually dimorphic phrase organization in the song of the indris (*Indri indri*). *American Journal of Primatology*, *82*(6), e23132.

Zipf, G. K. (1936). *The psychobiology of language.* London: Routledge.

Zipf, G. K. (1949). *Human behavior and the principle of least effort.* New York: Addison-Wesley.