Hernández-Fernández, A.; González, I. Compression principle and Zipf's Law of brevity in infochemical communication. "Biology Letters", Juliol 2022, vol. 18, núm. 7, p. 1-7. DOI10.1098/rsbl.2022.0162

Compression principle and Zipf's law of brevity in infochemical communication

Antoni Hernández-Fernández*[a] and Iván G. Torre*†[b,c]

[a]Complexity and Quantitative Linguistics Lab, Institut de Ciències de l'Educació, Universitat Politècnica de Catalunya, Av. Doctor Marañón 44-50, 08028 Barcelona, Catalonia, Spain.

[b]Language and Speech Laboratory, Universidad del País Vasco, Justo Vélez de Elorriaga Kalea, 1, 01006, Vitoria, Spain

[c]Departamento de Matemática Aplicada, Universidad Politécnica de Madrid, Avda. Puerta de Hierro, 2-4, 28040 Madrid, Spain

June 20, 2022

Abstract

Compression has been presented as a general principle of animal communication. Zipf's law of brevity is a manifestation of this postulate and can be generalised as the tendency of more frequent communicative elements to be shorter. Previous works supported this claim, showing evidence of Zipf's law of brevity in animal acoustical communication and human language. However, a significant part of the communicative effort in biological systems is carried out in other transmission channels, such as those based on infochemicals. To fill this gap, we seek, for the first time, shreds of evidence of this principle in infochemical communication by analysing the statistical tendency of more frequent infochemicals to be chemically shorter and lighter. We analyse data from the largest and most comprehensive open-access infochemical database known as Pherobase, recovering Zipf's law of brevity in interspecific communication (allelochemicals) but not in intraspecific communication (pheromones). Moreover, these results are robust even when addressing different magnitudes of study or mathematical approaches. Therefore, different dynamics from the compression principle would dominate intraspecific chemical communication, defying the universality of Zipf's law of brevity. To conclude, we discuss the exception found for pheromones in the light of other potential communicative paradigms such as pressures on successful communication or the Handicap principle.

Keywords: compression, infochemicals, information theory, language, communication.

1 Introduction

Chemical communication plays a paramount role in the complex dynamic of ecological systems and has governed life from the very beginning, from DNA replication to the internal communication mechanisms of living beings [1]. It is crucial for understanding the behaviour of organisms [2] and has shaped ecological communities [3, 4]. One form of chemical communication is infochemicals, substances emitted outside of the body and carrying information to interact with other individuals from the same or different species. Infochemicals play a relevant role in complex ecosystems and have recently been the subject of intense research in ecology [5, 6], evolution [7] and environmental

*equal contribution
†corresponding author

[8] or agricultural sciences [9], among other scientific fields.

Infochemicals have evolved under communication networks' selective pressures [10], from non-communicative compounds with a phylogenetic pattern [11, 12], to aggregates specialised in efficiently accomplishing the transmission of biological information. Each infochemical has been classically assorted considering five adaptative functions depicted in Figure 1: pheromones, involved in intraspecific communication; attractants, that cause aggregation of individuals; allomones, when they benefit the sender; kairomones when the beneficial is the receiver; and finally synomones, benefiting both signaller and receiver in mutualistic interactions. Definitely, in a relevant distinction here, attractants, allomones, kairomones and synomones are allelochemicals that mediate interspecies communications, while pheromones intercede in intraspecific communication [13].

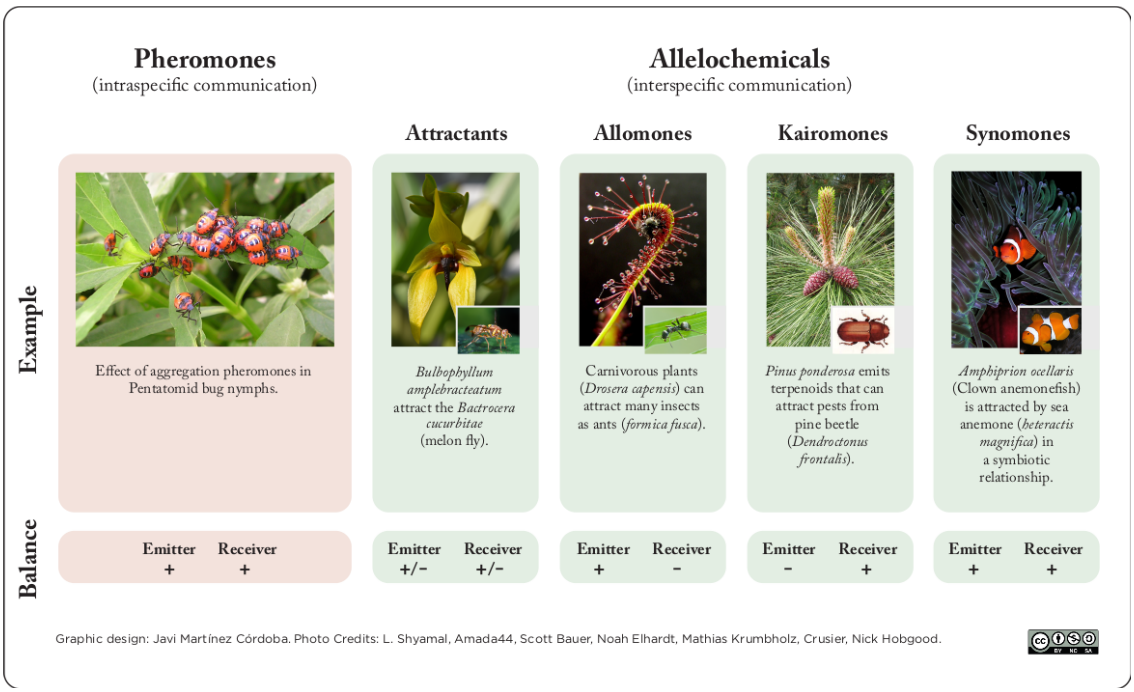

Figure 1: **Infochemical communication is classified by its adaptive function.** Pheromones mediate in intraspecific communication, being beneficial to individuals of the species, while allelochemicals intercede in interspecific communication. Allelochemicals are classified into four subcategories: allomones, which benefit the emitter but not the receiver; kairomones, profiting the receiver but not the emitter; synomones, which satisfy both sender and receiver in symbiotic interactions; and attractants which, depending on the case, can benefit or harm either the sender or the receiver.

Historically, researchers have established parallels with Linguistics intending to improve the understanding of chemical communication. One paradigmatic case is the description of DNA structure [14, 15] where similarities between the genetic code and the verbal language have been validated [16]. Language patterns can be theorized and quantitatively measured with the use of linguistic laws, statistical regularities shared across human communication [17, 18], and also discovered in biological systems [19] as diverse as the vocalizations [20, 21, 22, 23, 24] and gestures of other primates [25], genomics [26, 27], proteomics [28, 29, 30, 31] and chemical communication systems [32].

One important linguistic law is Zipf's law of brevity [33, 34], or just Brevity law: the tendency for the most frequent communicative elements to be shorter or briefer [18]. This law is a consequence of the so-called principle of compression [35] which in standard information theory consists of assigning codes as short as possible to the more frequently occurring symbols and longer codewords to the symbols that infrequently occur [36]. In this theoretical framework, optimal non-singular coding predicts that in language, the length of a word should grow approximately as the logarithm of its frequency of occurrence [36]. Thus, if f is the frequency of a communicative element and ℓ is its length, Zipf's law of brevity can be formulated as [18]:

$$f \sim \exp(-\lambda\ell), \quad \lambda > 0 \tag{1}$$

where λ is a fitting parameter.

In chemical communication, Zipf's law of brevity predicts that the more frequent a chemical compound is, the lighter and shorter. The size of infochemicals can be measured by considering their molecular weight (MW) or by counting their number of carbons (NC), which, in general, correlate together [37]. Therefore, it would be expected that the most frequent infochemicals have low MW and are composed of fewer carbons. Many species share the same chemical substances in their communication systems [4, 38, 39], so we can define the degree of use [73] of an infochemical as the number of species that interacts with it. Therefore, from an ecological perspective, the infochemical degree of use would be equivalent to the total number of species that are sensitive to it.

For the first time, we empirically explore and discuss whether Zipf's law of brevity and compression principle hold in the infochemical communication system that governs much of the communication between insects, plants, and other organisms. To address this critical issue, here we conduct a systematic exploration of Pherobase, the largest available infochemical database [39]. We will show that Zipf's law of brevity is fulfilled for allelochemicals –interspecific communication– but not for pheromones –intraspecific communication–, suggesting that chemical communication between individuals of the same species is strongly influenced by other factors beyond the principle of compression.

2 Materials and methods

Information was fetched from Pherobase, the largest and most comprehensive open-access infochemical database [39]. Pherobase contains elaborate and verified infochemical information, making it a very reliable and valuable resource widely used by the scientific community [40, 41, 42]. With 36,133 recorded interactions for 1,393 different infochemicals at the time of this study [72] – see Table 1–, each infochemical is classified with details on what its adaptive function is, depending on the species interacting with it [39]. Among them, the statistical preponderance of insects (34.55%) and arachnids (7.61%) is remarkable (see section 1 of the Supplementary Information for more details on the most represented taxonomies). The same infochemical (type) can be used by several species in the same or different ecosystem with different functions. Therefore, the aggregation of sub-types (e.g. Attractants, Allomones) is not necessarily equal to the total number of types of the aggregation (e.g. Allelochemicals) –see r in the third column of Table 1–. For all infochemicals, it is obtained the molecular weight (MW), chemical composition, number of carbons (NC), and the number of species (degree of use) that interact with the compound, in each specific adaptive function – see Figure 1–. The group of Synonomes was excluded from the research since only three different items were reported.

Firstly, it is explored the probability distribution of infochemical sizes considering MW and NC as study units. As it will be shown the probability distributions are subexponential heavy-tailed [43], so lognormal (LND) $p(x,\mu,\sigma)=\frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(lnx-\mu)^2}{2\sigma^2}}$, gamma $p(x,\kappa,\theta)=\frac{1}{\Gamma(\kappa)\theta^{\kappa}}x^{\kappa-1}e^{-\frac{x}{\theta}}$ – and Weibull $p(x,k,\lambda)=\frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}e^{(x/\lambda)^k}$ candidate distributions are fitted to data by maximum likelihood estimation method [44]. Kolmogorov-Smirnov test is carried out for goodness of fit and mean loglikelihood is used for model selection. Then, following [45], one-sided Kendall correlation test [46] is used to assess if there is a negative monotonic relationship between the use degree of infochemicals and the compound size, considering MW and NC. Additionally, Spearman

	N	r	$H(r)$	degree-MW		degree-NC	
				τ	$p-value$	τ	$p-value$
All infochemicals	36 133	1 393	9.0	-0.11	$p < 10^{-4}$	-0.09	$p < 10^{-4}$
Pheromones	18 973	1 178	8.9	-0.01	$p = 0.36$	0.01	$p = 0.74$
Alelochemicals	17 063	829	8.1	-0.22	$p < 10^{-4}$	-0.19	$p < 10^{-4}$
Attractants	10 761	531	7.4	-0.14	$p < 10^{-4}$	-0.13	$p < 10^{-4}$
Allomones	5 654	500	7.3	-0.21	$p < 10^{-4}$	-0.21	$p < 10^{-4}$
Kairomones	642	201	7	-0.22	$p < 10^{-4}$	-0.15	$p < 10^{-2}$

Table 1: **Parameters and correlations across infochemical adaptive function**. Total number of interactions (N), number of different infochemicals (r), entropy of the rank empirical distribution function ($H(r)$) and one sided Kendall correlation (τ) between degree of use and infochemical size both both in continuous magnitudes –Molecular Weight (MW)– and in discrete units –Number of Carbons (NC)–.

and Pearson correlation tests are also provided in section 2 of the Supplementary Information (SI).

We also explore the level of information contained on the probability distributions of ranks by analyzing the Shannon entropy $H(r) = -\sum_{i=1}^{n} P(r_i)\log_2 P(r_i)$ [47], where $P(r_i)$ is the probability degree associated of the infochemical with rank r, after sorting infochemical type by use degree in descending order and being $r = 1$ the most frequent one [33, 34]. Alternative entropy estimators are provided and discussed in section 7 of SI [48]. Finally, the mean square error method is used to fit equation 1 to assess the relationship between infochemical size and degree.

3 Results

3.1 Degree and allelochemical size are negatively correlated

We start by analysing the Pherobase dataset to address if there is a negative correlation between the size of infochemicals and their degree of use. The size is measured using two different magnitudes: MW and NC. The first can be approximated as a continuous magnitude, while the last can be understood as a discrete unit. Given that choosing the magnitude of the study is still an open problem and that both alternatives are related to the energy used in communication, this approach will allow us to address recent discussions in the study of linguistic laws [18, 49, 50].

For a given adaptive function –e.g. Pheromones– and following [45], we compute a one-sided Kendall correlation test (τ) to explore manifestations of the compression principle (Spearman and Pearson are also considered in section 2 of SI). Our results are summarised in Table 1, revealing a strong negative correlation between infochemical size and degree of use when all infochemicals are studied together. Different behaviours are observed when the subtypes of infochemicals are assessed. Interestingly, Zipf's law of brevity is fulfilled in interspecies communication –allelochemicals–, while not in Pheromones. Moreover, allelochemicals are also studied with more granularity, finding likewise strong negative correlation regardless of their adaptive function. These results are robust for both units of study: MW and NC. Additionally, the Shannon entropy of rank probability distribution has been estimated for each adaptive group, uncovering that pheromones are more entropic than allelochemicals subgroups.

3.2 Infochemical size is lognormally distributed when using MW

Following previous studies in human communication [18, 49], we explore the marginal distribution of infochemical size considering MW and NC as units of study. To begin with, using the method of maximum-likelihood estimation (MLE), we have fitted the data to three possible theoretical

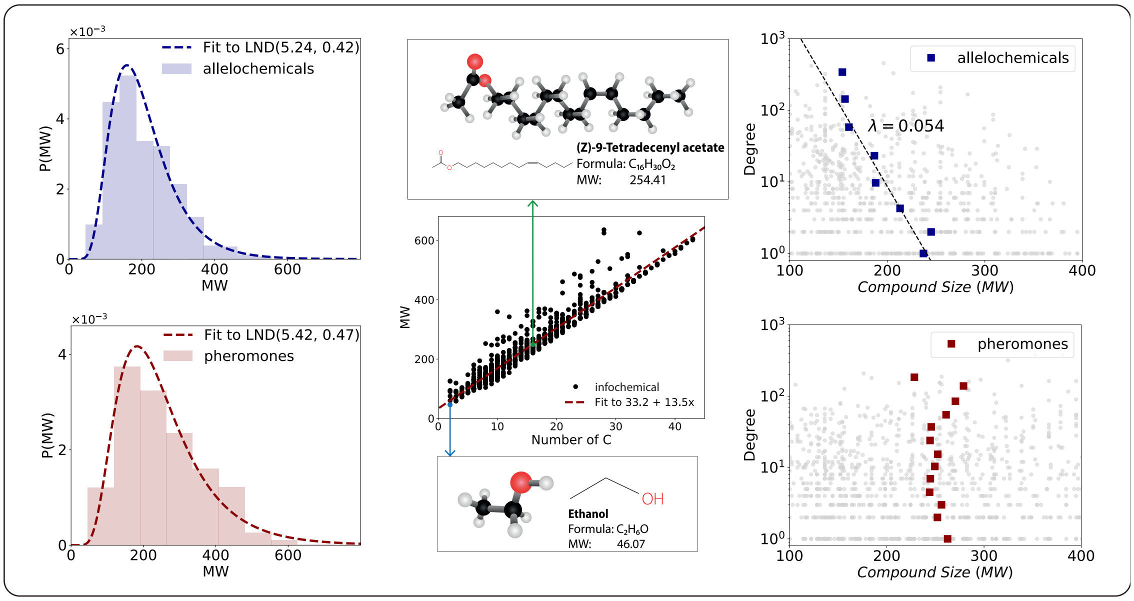

Figure 2: **Brevity law in infochemical communication**. First two left panels show the molecular weight distribution of allelochemicals (blue bars) and pheromones (red bars) with fitting to best LND distribution (dotted line). The central panel illustrates the almost linear relationship between the Number of Carbons (NC) and Molecular Weight (MW), with two illustrative examples of infochemicals, including the most frequent one: Ethanol. Finally, the right panels reveal that Brevity law holds for interspecies communication –allelochemicals– while not for intraspecific communication –pheromones–.

distributions: LND, gamma and Weibull (we use Kolmogorov–Smirnov distance ks for goodness of fits, and mean loglikelihood $\langle \mathscr{L} \rangle$ for model selection, see table 2 and, additionally, tables 4 and 5 of SI). Jointly considering $\langle \mathscr{L} \rangle$ and ks, MW distribution is best explained by LNDs, in line with results reported in [18] where continuous physical magnitudes were used. Left panels of Figure 2 show estimated MW distribution for allelochemicals (blue) and for pheromones (red). Colour bars correspond to empirical binned data, and dotted lines are the MLE fit to LND. Nonetheless, the gamma probability distribution is the most plausible when considering NC –discrete magnitudes– for addressing the probability size distribution of infochemicals. These results agree with previously reported investigations where the authors examined word length distribution in texts with discrete units of study [49].

3.3 MW and NC are strongly correlated

The appropriate choice of study units to explore compression principle and Brevity law remains an open question that should be carefully considered. Nevertheless, the chosen magnitudes should contain information on the energy and communication effort intended to be minimized. Here, we have considered both a discrete unit of study –NC– and a continuous magnitude –MW–, and we explore their relationship. In the central panel of Figure 2 each infochemical is represented with a black circle. Dotted red line represents the linear regression between the MW and NC of each infochemical. It has been found that MW and NC are strongly correlated, so both magnitudes should be appropriate for exploring manifestations of the compression principle.

	Type	LND				gamma				Weibull	
		$\langle \mathscr{L} \rangle$	D_{ks}	μ	σ	$\langle \mathscr{L} \rangle$	D_{ks}	κ	θ	$\langle \mathscr{L} \rangle$	D_{ks}
MW	All infochemicals	-6.05	0.05	5.4	0.46	-6.06	0.06	4.89	0.02	-6.10	0.06
	Pheromones	-6.09	0.05	5.43	0.47	-6.09	0.06	4.84	0.019	-6.13	0.05
	Allelochemicals	-5.78	0.06	5.24	0.42	-5.79	0.08	6.01	0.029	-5.84	0.08
	Attractans	-5.69	0.07	5.20	0.39	-5.69	0.09	6.74	0.034	-5.73	0.1
	Allomones	-5.78	0.08	5.22	0.42	-5.80	0.09	5.67	0.028	-5.86	0.11
	Kairomones	-5.61	0.15	5.03	0.43	-5.66	0.18	5.22	0.031	-5.77	0.19
NC	All infochemicals	-3.44	0.06	2.61	0.55	-3.42	0.08	3.71	0.23	-3.44	0.08
	Pheromones	-3.49	0.07	2.66	0.56	-3.47	0.08	3.71	0.22	-3.48	0.08
	Allelochemicals	-3.14	0.08	2.43	0.49	-3.13	0.09	4.54	0.35	-3.16	0.1
	Attractans	-3.11	0.1	2.41	0.48	-3.08	0.1	4.83	0.39	-3.09	0.12
	Allomones	-3.11	0.09	2.39	0.49	-3.11	0.11	4.41	0.35	-3.15	0.12
	Kairomones	-3.00	0.16	2.23	0.52	-3.01	0.19	3.92	0.37	-3.06	0.21

Table 2: **Estimated parameters of infochemical size distribution**. Probability distribution candidates are fitted using MLE, model selection is based on maximizing mean log-likelihood $\langle \mathscr{L} \rangle$ and goodness of fit is based on Kolmogorov-Smirnov distance D_{ks}, where lower values mean that data are more plausible to follow the distribution. The estimated parameters for LND(μ, σ) and gamma(κ, θ) are provided.

3.4 Infochemical degree of use follows Brevity law –except for pheromones–

To round off, we finally consider Zipf's law of brevity in detail. In the right panels of Figure 2, we scatter plot MW versus degree of use of infochemicals with grey circular dots. Then, we apply a logarithmic binning over the degree axis (blue for allelochemicals and red for pheromones) in order to counterbalance low sampling effects [18]. Other infochemical subgroups are also considered in Figure 3 of SI. Data is then fitted to equation 1 and represented with a red dotted line showing an excellent agreement with the square colour bins used as visual information. Pheromones are not fitted to equation 1 because in section 3.1, it was previously shown that no correlation was held.

4 Discussion

We have validated with unprecedented detail, by studying more than 36,000 infochemical interactions, that Zipf's law of brevity operates at interspecies infochemical communication. Those results are in line with precedent studies in language, and other communication systems [18, 20, 21, 22, 51, 52, 53], suggesting the universality of this linguistic law and the principle of compression, from which it derives. Nevertheless, we have found no evidence that Zipf's law of brevity holds in intraspecific chemical communication –pheromones–.

Previous work pointed out that communication efficiency is equivalent to minimising the length l of a communicative element [27, 35, 54]. In contrast, unambiguity and maximum accuracy in communication is analogous to a non-singular scheme of communication, where a unique element is assigned to each meaning [36, 45]. Pheromones may be statistically uniformly distributed so that even if there were optimal coding [36, 45], it would not manifest itself in the form of Zipf's law of brevity. The failure to obtain a significant p-value for the pheromones (see Table 1) does not mean that Zipf's law of brevity is not operating but raises the fact that other communicative principles are present [35, 55].

In some noisy or adverse domains, selection pressures have maximized transmission through redundancy in opposite to efficiency or energy saving, as it happens in acoustic long-distance calls [50, 55, 56]. This could be happening in the case of intraspecific pheromone communication, and it is consistent with the unrevealed Shannon entropy, significantly higher when comparing pheromones to other allelochemicals [36, 57].

The Handicap principle goes even further regarding the energy effort spent on communication [58]. It suggests that some individuals may expend even more effort than necessary as long as they demonstrate that others can not afford it. The so-called handicap hypothesis (see [59] about terminology) has evolved considerably from the original works [60, 61], early criticized [62, 63], and discussed for its lack of evidence [59]. The fact that pheromones do not comply with Zipf's law of brevity and their higher entropy may be an indirect evidence of the Handicap principle.

Pheromones have an intraspecific communication function, including a sexual role, but these uses are not granular detailed in the Pherobase dataset. Nevertheless, concerning sex pheromones, they can attract both the attention of predators and potential mates [60, 65, 66]. Natural selection may favour the evolution of chemical compounds whose reproductive benefits outweigh the adverse effects on survival so that the extra cost of increasing the chemical length of a substance may be offset by making it harder to be detected by predators [59, 66, 65]. For this reason, the size increase of pheromones could be a consequence of evolutionary survival pressures. However, the factors affecting signalling during communication are manifold and can significantly skew the direction of evolutionary pressures (see [66] for a review).

On the other hand, considering that pheromones are issued in parallel to other communication channels [4], the communicative complexity [38] and the redundancy of pheromone signals [64] may also explain the non-compliance with Zipf's law of brevity. Honeybee queens emit multiple redundant signals to ensure the message is correctly perceived, avoiding possible signal detection and identification errors[64]. In contrast, the parallel study of ubiquitous substances in many species and ecosystems may be essential to understanding interkingdom communication (e.g. melatonin case in [67]).

Moreover, at the outset, we have explored the influence of the units of study to measure the manifestation of Zipf's law of brevity in chemical communication. We have obtained a LND when analysing the infochemicals size distribution when measuring it with MW –a continuous magnitude– but when considering the NC –a discrete unit–, the gamma distribution is more plausible (Table 2). Previous works suggest that the functional form of Zipf's law of brevity depends crucially on the length distribution function of the communicative elements, and they obtained a gamma distribution for text characters –a discrete magnitude– [49]. On the other hand, in a previous study of Zipf's law of brevity in speech, considering for length a continuous unit of study, LND was more plausible [18]. Our findings are consistent with both results, and further work is needed to elucidate the influence of study units on signal size distributions in communication systems.

Interestingly, it has been shown that both magnitudes of study are strongly correlated, finding a linear correlation between the NC and MW. This is explained by the tendency of infochemicals to be carbon-chain compounds [32, 37]. This result would suggest both magnitudes are equivalent, but in general, it is theoretically preferable to deal with continuous physical quantities as close as possible to the observable reality [18].

Finally, the conclusions reached by this work are shaped by the limitations of the Pherobase dataset itself. This dataset is based on scientific evidence, so many infochemicals are still not reported; therefore, the degree of use of specific infochemicals may be underestimated. Besides, some taxonomic groups prevail in Pherobase, as reported in Section 1 of the SI. This fact, however, does not detract from the validity of the overall study carried out here, although a detailed analysis of the different families, classes and orders should be further addressed in the future.

Future investigations should consider a detailed, granular exploration of infochemical communication within ecosystems to determine the dynamic of ecological interactions that it is happening in each case. Although this study has considered infochemical communication from a global perspective, beyond being a limitation, it could be instrumental in exobiology. Exobiology is a discipline where very diverse approaches do necessarily coexist in the search for life outside the Earth [68, 69]. We suggest that the infochemicals distribution on Earth can be a comparative indicator to detect life on other planets, and the presence of specific carbon chains an indicative of chemical communication [70, 71].

Data accessibility

Pherobase is a freely accessible database. Curated data used in this study containing infochemical information is now freely available in Dryad Digital Repository https://doi.org/10.5061/dryad.905qfttnx [72]. All the necessary scripts and coding used during this research for processing data, computing and figure generation are freely available in https://github.com/ivangtorre/compression-and-brevity-law-in-infochemical-communication . We have used Python 3.8 and R 4.1 with fitdistrplus library for the analysis. Other libraries such as Numpy 1.22, Pandas 1.0 or Matplotlib 3.2 are also used.

Acknowledgements

We gratefully acknowledge Ramon Ferrer-i-Cancho for his seminal contributions and Nicolás Hernández for valuable theoretical discussions. We want to thank Javier Martínez Córdoba for his collaboration in graphic design and all the people who have contributed to Pherobase.

References

[1] W. C. Agosta. *Chemical communication: the language of pheromones.* Scientific American Library, New York, USA, 1992.

[2] Dustin J Penn. Chemical communication. *Chemical ecology: from genes to ecosytem. Dicke, M. and Takken, W.(eds), Wageningen, Springer*, -:9–18, 2006.

[3] Mark E Hay. Marine chemical ecology: chemical signals and cues structure marine populations, communities, and ecosystems. *Annual review of marine science*, 1:193—212, 2009.

[4] Tristram D Wyatt. *Pheromones and animal behaviour: communication by smell and taste.* Cambridge university press, 2003.

[5] Louise EM Vet and Marcel Dicke. Ecology of infochemical use by natural enemies in a tritrophic context. *Annual review of entomology*, 37(1):141–172, 1992.

[6] MW Sabelis, A Janssen, and MR Kant. The enemy of my enemy is my ally. *Science*, 291(5511):2104–2105, 2001.

[7] Caroline Müller, Barbara A Caspers, Jürgen Gadau, and Sylvia Kaiser. The power of infochemicals in mediating individualized niches. *Trends in Ecology & Evolution*, 35(11):981–989, 2020.

[8] Matthew S Savoca, Martha E Wohlfeil, Susan E Ebeler, and Gabrielle A Nevitt. Marine plastic debris emits a keystone infochemical for olfactory foraging seabirds. *Science advances*, 2(11):e1600395, 2016.

[9] Haibo Zhou, Longsheng Chen, Yong Liu, Julian Chen, and Frédéric Francis. Use of slow-release plant infochemicals to control aphids: a first investigation in a belgian wheat field. *Scientific Reports*, 6(1):1–8, 2016.

[10] Brian Wisenden & Norman Stacey. Fish semiochemicals and the evolution of communication networks. *Animal Communication Networks*, -:540–567, 01 2005.

[11] Sandra Steiger, Thomas Schmitt, and H Martin Schaefer. The origin and dynamic evolution of chemical information transfer. *Proceedings of the Royal Society B: Biological Sciences*, 278(1708):970–979, 2010.

[12] Matthew RE Symonds and Mark A Elgar. The mode of pheromone evolution: evidence from bark beetles. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1541):839–846, 2004.

[13] Donald A Nordlund and WJ Lewis. Terminology of chemical releasing stimuli in intraspecific and interspecific interactions. *Journal of Chemical Ecology*, 2(2):211–220, 1976.

[14] V. Brendel and H.G. Busse. Genome structure described by formal languages. *Nucleic Acids Research*, 12(5):2561–2568, 03 1984.

[15] David B Searls. The linguistics of dna. *American Scientist*, 80(6):579–591, 1992.

[16] Gemma Bel-Enguix and M.D. Jiménez-López. Genetic code and verbal language. In *Biology, Computation and Linguistics*, pages 85–103. IOS Press, Amsterdam, 2011.

[17] Eduardo G Altmann and Martin Gerlach. Statistical laws in linguistics. In *Creativity and Universality in Language*, pages 7–26. Springer, 2016.

[18] Iván G. Torre, Bartolo Luque, Lucas Lacasa, Christopher T. Kello, and Antoni Hernández-Fernández. On the physical origin of linguistic laws and lognormality in speech. *Royal Society open science*, 6(8):191023, 2019.

[19] Stuart Semple, Ramon Ferrer-i Cancho, and Morgan L Gustison. Linguistic laws in biology. *Trends in Ecology & Evolution*, 2021.

[20] Stuart Semple, Minna J Hsu, and Govindasamy Agoramoorthy. Efficiency of coding in macaque vocal communication. *Biology letters*, 6(4):469–471, 2010.

[21] Stuart Semple, Minna J Hsu, Govindasamy Agoramoorthy, and Ramon Ferrer-i Cancho. 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, 2013.

[22] Morgan L. Gustison, Stuart Semple, Ramon Ferrer-i Cancho, and Thore J. Bergman. Gelada vocal sequences follow menzerath's linguistic law. *Proceedings of the National Academy of Sciences*, -, 2016.

[23] Stuart K Watson, Raphaela Heesen, Daniela Hedwig, Martha M Robbins, and Simon W Townsend. An exploration of menzerath's law in wild mountain gorilla vocal sequences. *Biology Letters*, 16(10):20200380, 2020.

[24] Tarandeep Kang. *Linguistic Laws and Compression in a Comparative Perspective: A Conceptual Review and Phylogenetic Test in Mammals*. PhD thesis, Durham University, 2021.

[25] Raphaela Heesen, Catherine Hobaiter, Ramon Ferrer i Cancho, and Stuart Semple. Linguistic laws in chimpanzee gestural communication. *Proceedings of the Royal Society B: Biological Sciences*, 286(1896):20182900, 2019.

[26] Wentian Li. Menzerath's law at the gene-exon level in the human genome. *Complexity*, 17(4):49–53, 2012.

[27] Ramon Ferrer-i Cancho, Núria Forns, Antoni Hernández-Fernández, Gemma Bel-enguix, and Jaume Baixeries. The challenges of statistical patterns of language: The case of menzerath's law in genomes. *Complexity*, 18(3):11–17, 2013.

[28] Sertac Eroglu. Language-like behavior of protein length distribution in proteomes. *Complexity*, 20(2):12–21, 2014.

[29] Khuram Shahzad, Jay E Mittenthal, and Gustavo Caetano-Anollés. The organization of domains in proteins obeys menzerath-altmann's law of language. *BMC systems biology*, 9(1):1–13, 2015.

[30] Gustavo Caetano-Anollés, Bushra F Minhas, Fayez Aziz, Fizza Mughal, Khuram Shahzad, Guy Tal, Jay E Mittenthal, Derek Caetano-Anollés, Ibrahim Koç, Arshan Nasir, et al. The compressed vocabulary of the proteins of archaea. In *Biocommunication of Archaea*, pages 147–174. Springer, 2017.

[31] Vladimír Matlach, Daniel Dostál, and Marian Novotnỳ. Secondary structures of proteins follow menzerath–altmann law. *International Journal of Molecular Sciences*, 23(3):1569, 2022.

[32] Antoni Hernández-Fernández and Ramon Ferrer-i-Cancho. The infochemical core. *Journal of Quantitative Linguistics*, 23(2):133–153, 2016.

[33] George Kingsley Zipf. *The psycho-biology of language: An introd. to dynamic philology.* Mifflin, 1935.

[34] George Kingsley Zipf. *Human behaviour and the principle of least-effort.* Addison-Wesley, 1949.

[35] R. Ferrer-i Cancho, A. Hernandez Fernandez, D. Lusseau, G. Agoramoorthy, M. Hsu, and S. Semple. Compression as a universal principle of animal behavior. *Cognitive science*, 37(8):1565–1578, Jan 2013.

[36] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, New York, NY, USA, 2006.

[37] Edward O Wilson and William H Bossert. Chemical communication among animals. *Recent progress in hormone research*, 19:673, 1963.

[38] Tristram D Wyatt. Pheromones and signature mixtures: defining species-wide signals and variable cues for identity in both invertebrates and vertebrates. *Journal of Comparative Physiology A*, 196(10):685–700, 2010.

[39] Ashraf El-Sayed. The pherobase: Database of insect pheromones and semiochemicals. url-http://www.pherobase.com, 2021. Accessed and downloaded 2021.

[40] Genevieve M Tauxe, Dyan MacWilliam, Sean Michael Boyle, Tom Guda, and Anandasankar Ray. Targeting a dual detector of skin and co2 to modify mosquito host seeking. *Cell*, 155(6):1365–1379, 2013.

[41] Ingmar Weiss, Thomas Rössler, John Hofferberth, Michael Brummer, Joachim Ruther, and Johannes Stökl. A nonspecific defensive compound evolves into a competition avoidance cue and a female sex pheromone. *Nature communications*, 4(1):1–11, 2013.

[42] Bao-Jian Ding, Per Hofvander, Hong-Lei Wang, Timothy P Durrett, Sten Stymne, and Christer Löfstedt. A plant factory for moth pheromone production. *Nature communications*, 5(1):1–7, 2014.

[43] Ivan Voitalov, Pim van der Hoorn, Remco van der Hofstad, and Dmitri Krioukov. Scale-free networks well done. *Physical Review Research*, 1(3):033034, 2019.

[44] Scott R Eliason. *Maximum likelihood estimation: Logic and practice*, volume 96. Sage Publications, 1993.

[45] Ramon Ferrer-i Cancho, Christian Bentz, and Caio Seguin. Optimal coding and the origins of zipfian laws. *Journal of Quantitative Linguistics*, 0:1–30, 2020.

[46] Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

[47] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[48] Peter Hall and Sally C Morton. On the estimation of entropy. *Annals of the Institute of Statistical Mathematics*, 45(1):69–88, 1993.

[49] Álvaro Corral and Isabel Serra. The brevity law as a scaling law, and a possible origin of zipf's law for word frequencies. *Entropy*, 22(2):224, 2020.

[50] Dena J Clink, Abdul Hamid Ahmad, and Holger Klinck. 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, 2020.

[51] Bo Luo, Tinglei Jiang, Ying Liu, Jing Wang, Aiqing Lin, Xuewen Wei, and Jiang Feng. Brevity is prevalent in bat short-range communication. *Journal of Comparative Physiology A*, 199(4):325–333, 2013.

[52] Mingpan Huang, Haigang Ma, Changyong Ma, Paul A Garber, and Pengfei Fan. 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, 2020.

[53] Daria Valente, Chiara De Gregorio, Livio Favaro, Olivier Friard, Longondraza Miaretsoa, Teresa Raimondi, Jonah Ratsimbazafy, Valeria Torti, Anna Zanoli, and Cristina Giacoma. Linguistic laws of brevity: conformity in indri indri. *Animal Cognition*, 2021:1–10, 2021.

[54] Ramon Ferrer-i Cancho. Compression and the origins of zipf's law for word frequencies. *Complexity*, 21(S2):409–411, 2016.

[55] R. Ferrer-i Cancho and A. Hernandez Fernandez. The failure of the law of brevity in two new world primates: statistical caveats. *Glottotheory*, 4(1):45–55, Mar 2013.

[56] BM Bezerra, AS Souto, AN Radford, and G Jones. Brevity is not always a virtue in primate communication. *Biology Letters*, 7:23 – 25, 2011.

[57] Thierry Lafouge, Abdellatif Agouzal, and Genevieve Lallich. The deconstruction of a text: the permanence of the generalized zipf's law—the inter-textual relationship between entropy and effort amount. *Scientometrics*, 104(1):193–217, 2015.

[58] Amotz Zahavi. The fallacy of conventional signalling. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 340(1292):227–230, 1993.

[59] Dustin J Penn and Szabolcs Számadó. The handicap principle: how an erroneous hypothesis became a scientific principle. *Biological Reviews*, 95(1):267–290, 2020.

[60] Amotz Zahavi. Mate selection—a selection for a handicap. *Journal of theoretical Biology*, 53(1):205–214, 1975.

[61] Amotz Zahavi. The cost of honesty (further remarks on the handicap principle). *Journal of theoretical Biology*, 67(3):603–605, 1977.

[62] John Maynard Smith. Sexual selection and the handicap principle. *Journal of Theoretical Biology*, 57(1):239–242, 1976.

[63] John Krebs. Sexual selection and the handicap principle. *Nature*, 261(5557):192–192, 1976.

[64] Sarah A. Princen, Ricardo Caliari Oliveira, Ulrich R. Ernst, Jocelyn G. Millar, Jelle S. van Zweden, and Tom Wenseleers. Honeybees possess a structurally diverse and functionally redundant set of queen pheromones. *Proceedings of the Royal Society B: Biological Sciences*, 286(1905):20190517, 2019.

[65] Amotz Zahavi and Avishag Zahavi, The handicap principle: A missing piece of Darwin's puzzle. Oxford: Oxford University Press, 1999.

[66] John A. Endler, Some general comments on the evolution and design of animal communication systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 340(1292), 215-225, 1993.

[67] Susan J. Murch and Lauren A.E. Erland, A Systematic Review of Melatonin in Plants: An Example of Evolution of Literature. *Frontiers in Plant Science*, 12, fpls.2021.683047, 2021.

[68] Vera M Kolb. *Handbook of Astrobiology.* CRC Press, 2018.

[69] Andrea Butturini, Daniel García-Castellanos, Carme Jordi, and Ignasi Ribas. *(In)habitabilidad planetaria: Fundamentos de astrogeobiología.* Marcombo, 2020.

[70] Eric Herbst and Robin T. Garrod. Synthetic approaches to complex organic molecules in the cold interstellar medium. *Frontiers in Astronomy and Space Sciences*, 8, 2022.

[71] Jonathan Tennyson and Sergey N. Yurchenko. High accuracy molecular line lists for studies of exoplanets and other hot atmospheres. *Frontiers in Astronomy and Space Sciences*, 8, 2022.

[72] Antoni Hernandez-Fernandez and Ivan G. Torre. Data from: Compression principle and Zipf's law of brevity in infochemical communication *Dryad Digital Repository*, https://doi.org/10.5061/dryad.905qfttnx

[73] Newman, M. E. The structure and function of complex networks *SIAM review*, 45(2), 167-256. 2003.